Casos prácticos de calidad en la producción cartográfica

Francisco Javier Ariza López
José Luis García Balboa
Raúl Amor Pulido

Universidad de Jaén

Para mi Vega,

Con cariño.

Raúl

Abril 2005

Casos Prácticos de Calidad en la Producción Cartográfica

UNIVERSIDAD DE JAÉN

Casos Prácticos de Calidad en la Producción Cartográfica

Francisco Javier Ariza López
José Luis García Balboa
Raúl Amor Pulido

Casos prácticos de calidad en la producción cartográfica / Francisco Javier Ariza López, José Luis García Balboa, Raúl Amor Pulido. — Jaén : Universidad de Jaén, Servicio de Publicaciones e Intercambio Científico, [2004]
ISBN 84-8439-239-2
1. Cartografía 2. Calidad 3. Control I. García Balboa, José Luis, coaut. II. Amor Pulido, Raúl, coaut. III. Universidad de Jaén. Servicio de Publicaciones e Intercambio Científico, ed. IV. Título
528.9
658.562

© Francisco Javier Ariza López y José Luis García Balboa (2004)
© Raúl Amor Pulido, del anexo estadístico (2004)
© UNIVERSIDAD DE JAÉN
Primera edición, Diciembre de 2004

Depósito Legal: J - 600 - 2004

I.S.B.N.: 84-8439-239-2

Difusión: Publicaciones de la Universidad de Jaén
Vicerrectorado de Extensión Universitaria
Campus Las Lagunillas, s/n - 23071 JAÉN
Tel.: 953 212 355 - Fax: 953 212 235

Impreso por: Gráficas "LA PAZ" de Torredonjimeno, S. L.
Avda. de Jaén, s/n - 23650 TORREDONJIMENO (Jaén)
Teléfono 953 571 087 - Fax 953 571 207

A nuestras familias

ÍNDICE GENERAL

ÍNDICE GENERAL 9

ÍNDICE DE FIGURAS 15

ÍNDICE DE TABLAS 19

PRÓLOGO 25

PRESENTACIÓN 27

AGRADECIMIENTOS 31

BLOQUE 1: MEJORA DE LA CALIDAD 33

CASO 1: Elaboración de un Procedimiento Específico 37

CASO 2: Utilización del diagrama de Pareto 40

CASO 3: Utilización del diagrama de Ishikawa 46

CASO 4: Análisis Modal de Fallos y Efectos 50

CASO 5: Utilización del Despliegue de la Función de Calidad 56

BLOQUE 2: TAMAÑOS MUESTRALES 63

CASO 6: Determinación del tamaño muestral en la verificación de la clasificación de imágenes de satélite 66

CASO 7: Muestreo de los trabajos de revisión catastral de titularidad 70

CASO 8: Planes de muestreo para la aceptación de lotes (I) 74

CASO 9: Planes de muestreo para la aceptación de lotes (II) 80

BLOQUE 3: CONTROL DE PROCESOS 87

CASO 10: Análisis de varianza 90

CASO 11: Análisis cluster (agrupamiento) 93

CASO 12: Análisis de series temporales en la digitalización gráfica manual 97

CASO 13: Empleo de cuantiles en la determinación del tamaño de celda 102

CASO 14: Evaluación de la calidad geométrica de un escáner 107

CASO 15: Control de una impresión sobre papel 112

BLOQUE 4: CONTROL DE CALIDAD POSICIONAL 119

CASO 16: Estudio teórico de la calidad posicional de una cartografía 123

CASO 17: Aplicación del estándar NMAS 126

CASO 18: Aplicación del estándar EMAS 131

CASO 19: Aplicación del estándar NSSDA 136

CASO 20: Control de una rectificación 139

BLOQUE 5: CONTROL DE CALIDAD TEMÁTICA 145

CASO 21: Parámetros de exactitud temática en la clasificación 148

CASO 22: Normalización de la matriz de confusión y cálculo de pseudoceros 158

CASO 23: Comparación de dos matrices de confusión 162

CASO 24: Exactitud temática tras la generalización de la leyenda 164

CASO 25: Estimación de incertidumbre en el modelado cartográfico 168

BLOQUE 6: SIMULACIÓN 173

CASO 26: Aproximación a la banda de error 175

CASO 27: Simulación en el modelado de propagación del error 179

CASO 28: Simulación para la comparación de matrices de confusión 189

BLOQUE 7: CONTROL DE OTRAS COMPONENTES DE UNA BDG 193

CASO 29: Control de la compleción y de la calidad temática 195

CASO 30: Control de la coherencia lógica 203

CASO 31: Elaboración de metadatos 210

ANEXO: ESTADÍSTICA 215

1. INTRODUCCIÓN 215

2. ESTADÍSTICA DESCRIPTIVA 216
2.1. Distribución de frecuencias 217
2.1.1. Distribución de frecuencias unidimensional 217
2.1.2. Distribución de frecuencias bidimensional 217
2.2. Resumen numérico de los datos muestrales 219
2.2.1. Medidas de posición 220
2.2.2. Medidas de dispersión 221
2.3. Tipificación de una variable 223
2.4. Regresión y correlación simple 224
2.4.1. Regresión 225
2.4.2. Correlación 229
2.4.3. Predicción 233
2.4.4. Modelo lineal general de regresión 233
2.4.4. Diagnóstico de influencia 236

3. PROBABILIDAD 239
3.1. Variables aleatorias 240
3.1.1. Variables discretas 241
3.1.2. Variables continuas 242
3.2. Características de una distribución de probabilidad 243
3.2.1. Esperanza matemática 243
3.2.2. Varianza 244
3.2.3. Mediana 244
3.2.4. Moda 244
3.3. Distribuciones de probabilidad más importantes 244
3.3.1. Distribución Uniforme discreta sobre n puntos 245
3.3.2. Distribución Binomial 245
3.3.3. Distribución de Poisson 247
3.3.4. Distribución Hipergeométrica 248
3.3.5. La distribución Uniforme 249
3.3.6. La distribución Normal 250
3.3.7. Distribuciones asociadas a la distribución Normal 252
3.3.8. Distribuciones multidimensionales 255
3.4. Teorema Central del Límite 257

4. INFERENCIA ESTADÍSTICA 260
4.1. Inferencia paramétrica 262
4.1.1. Contrastes de hipótesis paramétricos y estimación por intervalos 263
4.1.2. Bondad de ajuste 269
4.1.3. Pruebas de homogeneidad de varianzas 272
4.2. Inferencia no paramétrica 273
4.2.1. Test de aleatoriedad (test de rachas de Wald-Wolfowitz) 274
4.2.2. Prueba de la $\chi 2$ (tablas de contingencia) 275
4.2.3. Prueba de Kolmogorov-Smirnov 276
4.3. Análisis de la varianza 277
4.3.1. Análisis de la varianza para un factor 278
4.3.2. Análisis de la varianza para dos factores 280
4.3.3. Test de Kruskal-Wallis 284
4.4. Eliminación de outliers 285

5. ESTADÍSTICA CIRCULAR 286
5.1. Principales distribuciones circulares 286
5.1.1. Uniforme circular 287
5.1.2. Distribuciones envueltas 288
5.1.3. Distribución Normal Circular WN(μ, ρ) 288

6. ANÁLISIS CLUSTER 289
6.1. Análisis Cluster no jerárquico 290
6.2. Análisis Cluster jerárquico 291

7. SIMULACIÓN 293
7.1. Generación de números aleatorios 293
7.1.1. Generación de números pseudo-aleatorios 294
7.1.2. Generación de valores de variables aleatorias discretas 295
7.1.3. Generación de valores de variables aleatorias continuas 295
7.2. Método bootstrap para estimar errores cuadráticos medios 298

8. MUESTREO 300
8.1. Muestreo con probabilidades iguales 301
8.1.1. Muestreo aleatorio simple 301
8.1.2. Muestreo con reemplazamiento 302
8.2. Muestreo estratificado 302
8.3. Muestreo por conglomerados 304
8.3.1. Muestreo por conglomerados en una ETAPA 305
8.3.2. Muestreo bietápico o por conglomerados con submuestreo 309
8.4. Planes de muestreo 312
8.4.1. Planes de aceptación por muestreo para atributos 315

9. ANÁLISIS UNIVARIANTE DE SERIES TEMPORALES 319
9.1. Procesos autorregresivos (AR) 321
9.2. Procesos de media móvil (MA) 323
9.3. Procesos mixtos autorregresivos de medias móviles (ARMA) 324
9.4. Procesos ARIMA 325
9.5. Método de Box-Jenkins para el análisis de Series Temporales 325
9.5.1. Identificación del modelo ARIMA 326
9.5.2. La estimación del modelo ARMA 329

9.5.3. Diagnosis del modelo 331

APÉNDICES

Apéndice A: TABLAS ESTADÍSTICAS 333

Apéndice B: GLOSARIO 373

Apéndice C: REFERENCIAS 383

ÍNDICE DE FIGURAS

Figura 1.1 - Ciclo PDCA 35
Figura 1.2 - Ejemplos de cabeceras y pies de formatos de un manual de calidad 39
Figura 1.3 - Diagrama de Pareto 41
Figura 1.4 - Diagrama de Ishikawa para la generalización de una serie con problemas 47
Figura 1.5 - Esquema general del proceso AMFE 51
Figura 1.6 - Ejemplo de interactuación de causas para generar modos de fallo: a) independencia, b) concurrencia, c) encadenamiento y d) ejemplo de modelo complejo 52
Figura 1.7 - Complejidad de la combinación de causas, modos y efectos 52
Figura 1.8 - La Casa de la Calidad con las distintas matrices que la forman 56
Figura 1.9 - Encadenamiento de las casas de la calidad 58
Figura 1.10 - Ejemplo de aplicación de DFC al caso de un mapa con problemas. ... 59

Figura 2.1 - Ejemplo de certificación catastral descriptiva y gráfica de una parcela urbana 70
Figura 2.2 - Comportamiento de la curva característica frente: a) número de aceptación Ac, b) tamaños de muestra n, c) tamaño del lote N 75

Figura 3.1 - Propietario de un proceso 88
Figura 3.2 - Esquema de un análisis cluster 93
Figura 3.3 - Gráfico de dispersión de los datos 94
Figura 3.4 - Dendograma 94
Figura 3.5 - Obtención de la distancia perpendicular entre la línea original y la digitalizada 99
Figura 3.6 - Modelo AR(1) con $c = 0$ y $\phi = 0,6$, para un error aleatorio de 0,5 mm al 95 % de confianza 100
Figura 3.7 - Modelo AR(1) con $c = 0$ y $\phi = 0,9$, para un error aleatorio de 0,5 mm al 95 % de confianza 100

Figura 3.8 - Fichero vectorial de la cobertura de usos del suelo de Andalucía 102
Figura 3.9 - Error relativo cometido con cada percentil en función del tamaño de celda 105
Figura 3.10 - Ejemplo de trazado de los vectores diferencia tras la transformación afín 109
Figura 3.11 - Patrón de resoluciones USAF y sección de su digitalización con escáner 110

Figura 4.1 - Distribución Normal unidimensional 120
Figura 4.2 - Distribución Normal circular 120

Figura 5.1 - Localización de la escena dentro de la provincia de Jaén 153
Figura 5.2 - Composición RGB de la escena 155
Figura 5.3 - Muestro aleatorio simple de la imagen Landsat-TM (385 muestras) .. 156
Figura 5.4 - Aumento del porcentaje de acuerdo por agregación de clases 164

Figura 6.1 - Banda de error de una poligonal: a) valores de los errores en los vértices; b) comportamiento de la banda de error 175
Figura 6.2 - Rasterización acumulativa de un segmento. 2000 iteraciones 177
Figura 6.3 - Rasterización acumulativa de un polígono (cuadrado). 1000 iteraciones 177
Figura 6.4 - Curvado de la rasterización del segmento 177
Figura 6.5 - Curvado de la rasterización del cuadrado 177
Figura 6.6 - Vista tridimensional del curvado de probabilidades para el caso del segmento 177
Figura 6.7 - Vista tridimensional del curvado de probabilidades para el caso del cuadrado 177
Figura 6.8 - Modelo de Erosión Potencial y de simulación del proceso 183
Figura 6.9 - a) Máscara de la zona de trabajo, b) Factor C, c) Factor K, d) Factor LS, e) Factor R, f) MDT de la zona 184
Figura 6.10 - Resultados de la simulación: a) Erosión media según la EUPS para 10 iteraciones clasificada en el intervalo [0, 65], b) Probabilidades, en tantos por uno, del riesgo medio de erosión mayor de 10 Ton/ha/año para 10 iteraciones 187

Figura 7.1 - Relaciones entre el proceso de abstracción y la compleción 196
Figura 7.2 - Esquema con las acepciones de compleción 196
Figura 7.3 - Universo de discurso 201
Figura 7.4 - Base de datos geográfica 202
Figura 7.5 - Ejemplo de inconsistencia topológica 205

Anexo: Estadística

Figura A.1 - Tipificación de una distribución N(5;2,5) 224
Figura A.2 - Funciones de probabilidad y distribución de las caras de un dado. 245
Figura A.3 - Funciones de probabilidad y de distribución de una Binomial (8,0.4) 246
Figura A.4 - Funciones de probabilidad y distribución de Poisson ($\lambda = 10$) 248
Figura A.5 - Función de probabilidad y distribución de Hipergeométrica (N = 1000, n = 25, k = 400) 249
Figura A.6 - Función de densidad y de distribución de una distribución U(5,10) .. 250
Figura A.7 - a) Función de densidad de la distribución N(0,1); b) Función de distribución de la N(0,1) 251
Figura A.8 - Función de densidad de la distribución N(6,2) 252
Figura A.9 - Función de densidad y distribución de la distribución χ (10) 253
Figura A.10 - Función de densidad de la distribución t(10) 254
Figura A.11 - Función de densidad y de distribución de la F(10,10) 254
Figura A.12 - Función de densidad y de distribución de la M(10;0,3;0,7) 256
Figura A.13 - Función de densidad y de distribución de la Normal bivariante con $\mu = (0, 0)$, $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_{12} = \sigma_{21} = 1/3^{1/2}$ 257
Figura A.14 - Ejemplo de Dendograma 292

ÍNDICE DE TABLAS

Tabla 1.1 - Herramientas y técnicas para la mejora de la calidad 34
Tabla 1.2 - Relación entre las etapas del ciclo PDCA, los pasos del análisis y las herramientas a aplicar 35
Tabla 1.3 - Problemas de retraso en un suministro 41
Tabla 1.4 - Total de casos de cada tipo de problema detectado en el lote 43
Tabla 1.5 - Distribución de los problemas detectados en el lote 43
Tabla 1.6 - Factores equivalentes a las cinco emes en la digitalización manual con tableta 48
Tabla 1.7 - Valoración de los factores O, S y D en el Producto 53
Tabla 1.8 - Valoración de los factores O, S y D en el Proceso 53
Tabla 1.9 - Ejemplo de tabla resumen de un análisis AMFE 55
Tabla 1.10 - Principales características del mapa de carreteras 58
Tabla 1.11 - Voz del cliente 61

Tabla 2.1 - Cociente multinomial/binomial entre los tamaños de muestreo 68
Tabla 2.2 - Ejemplo de tabla para la presentación de las estimaciones de tamaño de muestreo 69
Tabla 2.3 - Criterios de paso entre niveles inspección de la Norma UNE 66020-1 . 77
Tabla 2.4 - Códigos de tamaño de muestra según la UNE 66020-1 77
Tabla 2.5 - Plan de muestreo simple para inspección normal 78
Tabla 2.6 - Condiciones contractuales de diferentes procedimientos de muestreo ... 79
Tabla 2.7 - Secuencia de lotes y número de rechazos encontrados 79
Tabla 2.8 - Ejemplo de Ficha de control de cada marco de muestreo 82
Tabla 2.9 - Ejemplo de Informe sobre los resultados de inspección sobre un Lote .. 82

Tabla 3.1 - Residuos (mm) del ajuste de cada hoja de cada muestra 92
Tabla 3.2 - Residuos (mm) del ajuste mínimo cuadrático de una serie de 30 muestras 92
Tabla 3.3 - Serie procedente de la digitalización del operador A 101
Tabla 3.4 - Serie procedente de la digitalización del operador B 101
Tabla 3.5 - Listado de los 10 elementos de menor tamaño superficial 103
Tabla 3.6 - Tamaños de celdilla estimados 104
Tabla 3.7 - Resumen estadístico de errores relativos en pruebas de rasterización .. 105
Tabla 3.8 - Tamaño del fichero con cada tamaño de celda 105

Tabla 3.9 - Características del escáner 110
Tabla 3.10 - Coordenadas calibradas e imagen de los puntos de control de una placa utilizada para calibrar un escáner 111
Tabla 3.11 - Niveles de calidad del producto establecidos en el QATAP 113
Tabla 3.12 - Atributos de impresión y de acabado establecidos en el QATAP 114
Tabla 3.13 - Atributo de impresión "error en la colocación de una imagen" 115
Tabla 3.14 - Atributo de acabado "tapas dañadas" 115

Tabla 4.1 - Componentes de error en distintas fase de la producción cartográfica . 120
Tabla 4.2 - Valores y desviaciones de las distintas variables y factores de escala .. 125
Tabla 4.3 - Conjunto de puntos a analizar (ejemplo) 128
Tabla 4.4 - Diferencias entre coordenadas-mapa y coordenadas-terreno (en metros) para los datos de ejemplo mostrados en la Tabla 4.3 128
Tabla 4.5 - Coordenadas mapa y terreno del conjunto de puntos a analizar 129
Tabla 4.6 - Diferencias entre coordenadas-mapa y coordenadas-terreno (en metros) (ejemplo) 133
Tabla 4.7 Coordenadas de un conjunto de puntos en un espacio origen y en un espacio destino 143

Tabla 5.1 - Matriz de confusión de la clasificación de una imagen 145
Tabla 5.2 - Notación seguida 148
Tabla 5.3 - Ejemplo de matriz de confusión 149
Tabla 5.4 - Parámetros de calidad temática 149
Tabla 5.5 - Clases de la imagen 150
Tabla 5.6 - Matriz de confusión de la clasificación de una imagen 150
Tabla 5.7 - Datos de la imagen 153
Tabla 5.8 - Descripción de la zona 154
Tabla 5.9 - Descripción de la imagen 154
Tabla 5.10 - Clases de la imagen 156
Tabla 5.11 - Matriz de confusión 157
Tabla 5.12 - Ejemplo de matriz de confusión a normalizar 159
Tabla 5.13 - Ejemplo de primer paso de la primera iteración (por columnas) 159
Tabla 5.14 - Ejemplo de segundo paso de la primera iteración (por filas) 159
Tabla 5.15 - Normalización matriz de confusión (Tabla 5.6) (100 iteraciones) 159
Tabla 5.16 - Resultado de la aplicación de pseudoceros según el método de Feinberg a la matriz de la Tabla 5.6 160
Tabla 5.17 - Matriz de confusión A (385 muestras) 163
Tabla 5.18 - Matriz de confusión B (788 muestras) 163
Tabla 5.19 - Leyenda del Corinne Landcover a cinco niveles 165

Tabla 5.20 - Leyenda del Corinne Landcover a tres niveles 165
Tabla 5.21 - Leyenda 166
Tabla 5.22 - Matriz de confusión de la imagen 167
Tabla 5.23 - Modelo de Capacidad de Pasto del Suelo (CPS) 168
Tabla 5.24 - Modelo para el tipo de drenaje (TD) 169
Tabla 5.25 - Modelo para la capacidad de sustento (CS) 169
Tabla 5.26 - Modelo de tipo de suelo (TS) en función de su composición 169
Tabla 5.27 - Modelo para la capacidad de retención de agua (CR) 169
Tabla 5.28 - Datos para el cálculo del modelo de capacidad de pasto de un suelo . 171

Tabla 6.1 - Código del programa de simulación para Arc/Info 185
Tabla 6.2 - Valores de C y P para los cultivos 187
Tabla 6.3 - Principales parámetros del Suelo 187
Tabla 6.4 - Desviaciones 187
Tabla 6.5 - Ejemplo de matriz de confusión 190
Tabla 6.6 - Matriz de confusión de la Tabla 6.5 en forma de lista. 190

Tabla 7.1 - Proceso de abstracción: desde el mundo real al terreno nominal 195
Tabla 7.2 - Ejemplo de diferentes problemas compleción de objetos o atributos ... 197
Tabla 7.3 - Informe de compleción consistente en una serie temporal de observaciones sobre plantaciones de avellano 198
Tabla 7.4 - Matriz de confusión entre entidades 199
Tabla 7.5 - Matriz de confusión entre categorías atributo de una entidad (árbol) ... 199
Tabla 7.6 - Especificaciones de producto 200
Tabla 7.7 - Indicadores y medidas de la coherencia de una BDG 203
Tabla 7.8 - Posibles errores en la adherencia a un formato 206
Tabla 7.9 - Control de la cartografía en la empresa GrafCan 206
Tabla 7.10 - Especificaciones del producto 207
Tabla 7.11 - Base de datos 207
Tabla 7.12 - Componentes del informe de metainformación en la Norma UNE 148001 211
Tabla 7.13 - Ejemplo de informe de calidad en MIGRA (UNE 148001) 212

Anexo: Estadística

Tabla A.1 - Tabla de contingencia o distribución bidimensional 218
Tabla A.2 - Distribuciones marginales 219

Tabla A.3 - Distribuciones condicionadas 219

Tabla A.4 - Funciones de decisión sobre la media 265
Tabla A.5 - Funciones de decisión para la diferencia de medias 266
Tabla A.6 - Tabla de contingencia 275
Tabla A.7 - Disposición de las observaciones 278
Tabla A.8 - Análisis de varianza para un solo factor 279
Tabla A.9 - Disposición de las muestras en el modelo equilibrado sin interacción 280
Tabla A.10 - Análisis de varianza para dos factores sin interacción 282
Tabla A.11 - Análisis de varianza para dos factores con interacción 283
Tabla A.12 - Intervalos de confianza y tamaños para totales, medias y proporciones en muestreos sin reemplazamiento 301
Tabla A.13 - Intervalos de confianza y tamaños para totales, medias y proporciones en muestreos con reemplazamiento 302
Tabla A.14 - Intervalos de confianza y tamaños para totales, medias y proporciones en muestreos estratificados 304
Tabla A.15 - Intervalos de confianza para un Muestreo por conglomerados del mismo tamaño, con probabilidades iguales y sin reemplazamiento 306
Tabla A.16 - Intervalos de confianza para un Muestreo por conglomerados de distinto tamaño, con probabilidades iguales y sin reemplazamiento 307
Tabla A.17 - Intervalos de confianza para un Muestreo por conglomerados combinado con estratificación 308
Tabla A.18 - Intervalos de confianza para un Muestreo por conglomerados combinado con estratificación 309
Tabla A.19 - Intervalos de confianza para un Muestreo bietápico por conglomerados con submuestreo con probabilidades iguales en cada etapa y tamaños de conglomerados iguales 310
Tabla A.20 - Intervalos de confianza para un Muestreo bietápico por conglomerados con submuestreo con probabilidades iguales en cada etapa y tamaños de conglomerados desiguales 311
Tabla A.21 - Intervalos de confianza para un Muestreo bietápico con probabilidades desiguales en la primera etapa y con probabilidades iguales en la segunda 312
Tabla A.22. Comparación de niveles de inspección 316
Tabla A.23. Criterios de pasos entre niveles de inspección 317
Tabla A.24 - Resumen de las principales características de las funciones de autocorrelación y autocorrelación de los modelos AR, MA y ARMA. 328

APÉNDICE A

Tabla 1: Función de distribución Binomial 334
Tabla 2: Función de distribución de Poisson 338
Tabla 3: Función de distribución y de probabilidad Hipergeométrica 342
Tabla 4: Función de distribución Normal (0, 1) 345
Tabla 5: Función de distribución de la Normal circular 346
Tabla 6: Función de distribución $\chi 2$ 348
Tabla 7: Función de distribución t-Student 349
Tabla 8: Función de distribución F de Snedecor 350
Tabla 9: Prueba de Kolmogorov-Smirnov sobre Bondad de Ajuste 354
Tabla 10: Prueba Lilliefors sobre Normalidad 355
Tabla 11: Prueba de homogeneidad de varianzas de Cochran 356
Tabla 12: Test de Rachas 357
Tabla 13: Prueba de Kolmogorov-Smirnov (muestras del mismo tamaño) 359
Tabla 14: Prueba de Kolmogorov-Smirnov (muestras de distinto tamaño) 360
Tabla 15: Distribución acumulativa de Poisson (probabilidad × 1000 de c defectos cuando el número esperado es de c′) 361
Tabla 16: Valores medios para obtener las probabilidades acumulativas de Poisson deseadas 363
Tabla 17: Muestreo por atributos. Letras de código para tamaño de muestra (UNE 66020, ISO 2859, MIL-STD-105E) 364
Tabla 18: Muestreo por atributos. Tabla maestra para inspección normal. Muestreo simple (UNE 66020, ISO 2859, MIL-STD-105E) 365
Tabla 19: Muestreo por atributos. Tabla maestra para inspección rigurosa. Muestreo simple (UNE 66020, ISO 2859, MIL-STD-105E) 366
Tabla 20: Muestreo por atributos. Tabla maestra para inspección reducida. Muestreo simple (UNE 66020, ISO 2859, MIL-STD-105E) 367
Tabla 21: Sistema Philips de muestreo estándar 368
Tabla 22: Método Dodge-Romig. Tolerancia de lote con muestreo simple 369
Tabla 23: Método Dodge-Romig. Tolerancia de lote con muestreo doble 370
Tabla 24: Método Dodge-Romig. LCMS en muestreo simple 371
Tabla 25: Método Dodge-Romig. LCMS en muestreo doble 372

PRÓLOGO

La cartografía matemática y temática continúa aportando los datos necesarios para configurar el modelo de territorio que permite diseñar actuaciones en el amplio campo de las ingenierías. Los mapas y los planos numéricos y, en un sentido más amplio, las bases de datos geográficos (BDG) y los sistemas de información geográfica, aportan a los técnicos un conocimiento espacial, que es fundamental para trabajar. La cualidad más valorada del producto cartográfico, además de su precisión espacial, que depende de todo el proceso de producción cartográfica, es su configuración temática, lo que determina la calidad final. Y precisamente de calidad global del producto cartográfico trata esta monografía dirigida y coordinada por Francisco Javier Ariza López.

El Dr. Ariza pone a nuestro servicio una nueva publicación. Desde el curso 1994/95 en que comenzó su actividad docente en la Escuela Politécnica Superior de la Universidad de Jaén en la disciplina de Producción Cartográfica, ha dedicado gran parte de su tiempo a investigar la calidad de la cartografía. Su preparación, capacidad de trabajo e ilusión, le han llevado a crear un grupo de investigación relevante y a dar a conocer las nuevas tendencias sobre el control de procesos y productos cartográficos, como ya quedó de manifiesto en su anterior trabajo: Calidad en la Producción Cartográfica (Ed. Ra-Ma, 2002).

Esta obra incluye una treintena de casos prácticos estructurados en siete bloques relativos a: la mejora de la calidad, los tamaños muestrales, el control de procesos, el control posicional, el control de la componente temática, el uso de la simulación y otras componentes de la calidad de las BDG; por lo que presenta un abanico muy adecuado a las necesidades actuales de control en las instituciones cartográficas. La estadística y las normas son pieza clave en este manual, lo cual le da un marcado carácter ingenieril. Otro valor añadido de esta obra es el Anexo estadístico que la acompaña y, fundamentalmente, el disponer de un espacio web complementario para la descarga de ficheros e informaciones.

Por último, sólo me queda felicitar a los autores y pedir a Pachi Ariza que siga en su línea de trabajo, y que sus investigaciones en el marco de nuevos métodos para controlar la calidad posicional en cartografía, basadas en la obtención y explotación de curvas operativas, y en el uso de elementos lineales, sean un éxito, y que pronto podamos disfrutar con un nuevo trabajo.

Rafael Ferrer Torío.
Santander, 1 de octubre de 2004.

PRESENTACIÓN

Existen numerosos manuales dedicados a la calidad en un sentido general y de gestión (por ejemplo, Feigenbaum, 1990; Sebastián y col., 1998; Cuatrecasas, 2001), o centrados en aspectos más específicos como puede ser su componente estadística y de control (Hansen y Ghare, 1990; Besterfield, 1995), los procesos de mejora (Vilar, 1997; Galoway, 1998; Domingo y Arranz, 2000), o la implantación de Sistemas de Gestión de la Calidad (Cuatrecasas, 1999), etc. Sin embargo, dentro de un campo más propio de la Ingeniería Cartográfica, y de la Geomática[1], aunque actualmente se da un verdadero auge en los procesos de normalización[2] y de aplicación de los principios de la calidad en instituciones y empresas, los tratados que abordan estos temas son muy escasos y se encuentran en lengua inglesa (Veregin, 1989; Guptil y Morrison, Eds., 1995; Wenzhong, Fisher y Goodchild, Eds., 2002). En castellano las posibilidades son aún más reducidas, pudiendo destacar manual *Calidad en la Producción Cartográfica* (Ariza, 2002) que, centrado en aspectos teóricos, fue publicado por el autor principal de esta obra.

Pensamos que el cocimiento teórico se completa y culmina cuando realmente puede ser llevado a la práctica y que, por ello, son de especial valor aquellos manuales que conjugan las explicaciones teóricas con casos prácticos aplicados. Este es el objetivo básico de esta obra. De manera aplicada se muestran una selección de casos con los que estudiantes y técnicos del sector cartográfico podrán desarrollar unas capacidades concretas, muy aplicadas y orientadas a un quehacer diario de aplicación de la calidad a la cartografía.

---

[1] Ciencia y tecnología que trata de la obtención, análisis, distribución y uso de geoinformación. La Geomática incluye un amplio rango de disciplinas que unidas crean un detallada y comprensible representación del mundo físico. Geomática es un término genérico que engloba las disciplinas de topografía (incluyendo los sistemas de posicionamiento global "GPS"); cartografía Fotogrametría, radargrametría, técnicas cartográficas, AM/FM y producción de cartas de navegación); teledetección (adquisición de datos y aplicación); así como la creación y mantenimiento de los SIG. También puede ser considerada para incluir la componente espacial de los procesos relativos a la Tierra (p.e. meteorología), características sociales (p.e. datos censales) y actividad económica (p.e. transporte de mercancías). (Definición adaptada de *Canadian Geospatial Data Infrastructure*, http: //www.geoconnections.org/ CGDI.cfm/fuseaction/ faq_cgdiGeoConnections.welcome/gcs.cfm.

[2] En la actualidad existe una gran actividad en este campo normativo. A nivel nacional los procesos normativos en el ámbito de la información geográfica los realiza el Comité Técnico 148 de AENOR (Asociación Española de Normalización, www.aenor.es). En Europa esta actividad la desarrolla el Comité Técnico 287 de CEN (Comité Europeo de Normalización www.cenorm.be). Con ámbito internacional es el Comité Técnico 211 de ISO (*International Standard Organization*, www.iso.ch). Junto a estas organizaciones oficiales existen otras, por ejemplo, el *Open GIS Consortium* (OGC, www.opengis.org) que también desarrollan una intensa actividad de normalización en el ámbito geomático.

Dado que entendemos la Producción Cartográfica desde una perspectiva propia de la ingeniería de procesos y productos, nos preocupa el rigor en la toma de decisiones y por ello, en todos los ejemplos propuestos en este manual hay siempre un trasfondo estadístico, y en otros casos la aplicación de alguna norma.

Se trata de un manual de práctico y por ello hay que llegar a él con ciertos conocimientos teóricos. Su aprovechamiento requiere conocimientos previos sobre: calidad, calidad en cartografía, procesos cartográficos, Sistemas de Información Geográfica y estadística. La obra anteriormente referida (*Calidad en la Producción Cartográfica*, Ariza, 2002) es el complemento teórico apropiado a este manual, y una fuente de referencias de valor para ampliar los conocimientos específicos y generales sobre calidad..

Para facilitar la componente estadística, que soporta muchos de los ejemplos que se incluyen, se ha incorporado un completo anexo dedicado a presentar los fundamentos básicos de las herramientas estadísticas que se han de aplicar. Como complemento a este anexo se acompaña el Apéndice A, que presenta veinticuatro tablas estadísticas cuyo uso puede ser necesario para el desarrollo de los casos prácticos.

Para facilitar el manejo de los principales conceptos y fuentes bibliográficas, en el Apéndice B se presenta un glosario con las definiciones más importantes, y en el Apéndice C se incluyen las referencias bibliográficas empleadas.

Creemos que este manual presenta una perspectiva general de la aplicación práctica de la calidad en el moderno sector cartográfico. Como es lógico esta perspectiva es limitada pues solamente se presentan una treintena de casos, no obstante consideramos que son suficientes para adentrar y familiarizar en la aplicación práctica de los aspectos y campos más importantes de la calidad[3]. Dejamos en manos del lector la opción de profundizar en los mismos mediante el uso de las referencias que se incluyen, o de cualquier otra fuente; en este sentido el uso de Internet puede dar resultados sorprendentemente satisfactorios.

No existe una estructura fija en la presentación de los casos prácticos. En algunos se incluye una presentación teórica, en otros se plantea un ejemplo o directamente se procede a la formulación del trabajo a realizar. Todo ello ha dependido básicamente del nivel de complejidad y de la complementariedad con los contenidos teóricos desarrollados en la obra *Calidad en la Producción Cartográfica* (Ariza, 2002). Además, pensamos que los ejercicios que se incluyen han de considerarse como problemas abiertos, importando más la reflexión, análisis e imbricación en los procesos, que el puro resultado final. Por otra parte, todos no poseen el mismo grado de dificultad. Unos

[3] En relación al posible uso docente de este material, consideramos que el número de ejercicios que se incluye es demasiado extenso para una asignatura con una carga de créditos reducida o mediana; por ello sugerimos que el profesor elija los casos que considere más convenientes.

pueden resolverse de manera individual y otros han de hacerse en grupo, para desarrollar las habilidades que se pretenden. En algunos basta con aplicar una ecuación o extraer un valor de una tabla; en otros casos será necesario proceder a algún tipo de cálculo con un conjunto de datos, y en algunos otros será imprescindible llegar a entrar en las arenas de la programación. En relación a eso último, pensamos que ello no debe ser limitación para el desarrollo de los ejercicios puesto que un técnico capacitado del ámbito cartográfico, debe tener desarrolladas y puestas en práctica sus habilidades de programador. Además, la realización de los propios programas es siempre una forma muy eficaz de aprendizaje. Por otra parte, conviene destacar que muchos de los programas pueden plasmarse como macros de una hoja de cálculo, alternativa que simplifica mucho su programación.

Como complemento de valor añadido a la obra, se ha puesto a disposición de los lectores el espacio WEB:

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

en el que puede encontrar los datos de ejemplo[4], en formato digital, para facilitar el desarrollo de los ejercicios, así como algunos programas realizados por los autores para resolver los casos en que se necesitan; pero también se incluye abundante información complementaria y enlaces a otras direcciones de interés. El objetivo de este espacio es establecer un conjunto de recursos temáticos relativos a la calidad en cartografía que favorezcan su aplicación práctica.

Respecto a la organización del manual, éste se divide en siete bloques o capítulos que cubren aspectos de la calidad desde distintas perspectivas, si bien algunas veces, confluyen aunque con matizaciones distintas. La estructura de cada bloque consiste en una breve introducción general a la temática del bloque (por ejemplo, calidad temática, calidad posicional, etc.), una presentación de los casos a desarrollar y las principales fuentes bibliográficas que soportan el capítulo. En cuanto al número de ejercicios por capítulo o bloque, éste no es fijo, hay varios bloques con tres casos propuestos y otros con cuatro, cinco y hasta seis. Los bloques considerados son los siguientes:

— Bloque 1: Mejora de la calidad. Propone cinco casos en el ámbito de los Sistemas de Gestión de la Calidad con clara orientación a la mejora de productos y/o procesos.

— Bloque 2: Tamaños muestrales. Siendo el muestreo estadístico una de las bases del control de calidad en procesos y de la aceptación de productos, se ha considerado oportuno incluir aquí cuatro ejercicios que, mediante la aplicación de la estadística y de normas, abren las perspectivas de cómo afrontar problemas de muestreo en cartografía.

[4] Consideramos que el mayor provecho de un ejercicio práctico se obtiene cuando los datos, condicionantes e hipótesis hacen referencia a las circunstancias particulares de cada cual. Por ello animamos a los lectores a aplicar los ejemplos directamente sobre sus propios datos y circunstancias.

— Bloque 3: Control de procesos. Con seis casos propuestos, este bloque presenta y propone ejemplos de uso de herramientas estadísticas avanzadas para analizar procesos cartográficos.

— Bloque 4: Control de la calidad posicional. Se presenta aquí tanto la aplicación de métodos matemáticos como estadísticos para el control de la componente posicional. En total son cinco ejercicios, de los que tres proponen la aplicación de test de control posicional (NMAS, EMAS y NSSDA), y los otros dos utilizan métodos matemáticos y procesos estadísticos más complejos para tener un mayor control del error.

— Bloque 5: Control de la calidad temática. Tomando como base la matriz de confusión se presentan cuatro ejercicios orientados a comparar índices globales y las propias matrices. Se incluye un quinto caso dedicado al control de un modelo cartográfico de carácter temático que es un ejemplo muy revelador de lo que pueden ser los niveles de calidad temática de este tipo de cartografías.

— Bloque 6: Simulación. Posiblemente este sea el bloque cuyo desarrollo requiera más esfuerzo por parte del lector, debido a la necesidad de entrar en unos niveles de programación más avanzados que los que se necesitan para resolver cualquier otra propuesta de este manual. Son tres ejercicios cuya temática enlaza con otros bloques, pero que requieren de la simulación a muy distintos niveles: simulación en el espacio, simulación dentro de una herramienta SIG y simulación numérica. Entendemos que estas tres opciones de simulación son las básicas que se han de desarrollar en el ámbito de la cartografía.

— Bloque 7: Control de bases de datos geográficas. Forman este bloque un total de tres casos que se centran en lo que denominamos "otras componentes de la calidad".

Respecto al orden de presentación de los bloques, no es más que una de las muchas opciones posibles con coherencia. Este orden de presentación no debe, en ningún caso, limitar el uso del manual, pues son pocas los ejercicios que requieren de otro anterior para su culminación.

Finalmente, esperamos que el trabajo aquí recogido sea de valor para los lectores del mismo, permitiéndoles alcanzar mayores niveles de profundización y conocimiento en los procesos de control y mejora, y que todo ello sirva para elevar la calidad de nuestros productos y procesos, y por ende de nuestra industria cartográfica.

Jaén, octubre de 2004.

AGRADECIMIENTOS

Deseamos dejar constancia de nuestro agradecimiento a todas aquellas personas, instituciones y empresas que, de una manera directa o indirecta, nos han ayudado en la realización de esta obra y, esencialmente, al Servicio de Publicaciones de la Universidad de Jaén por considerarla para su publicación.

Especial agradecimiento a Don Rafael Ferrer Torío (Dpto. de Ingeniería Geográfica y Técnicas de Expresión Gráfica de la Universidad de Cantabria) por su amable prólogo y numerosas sugerencias sobre el texto, y a los alumnos de la Asignatura Producción Cartográfica, de la Titulación de Ingeniería en Geodesia y Cartografía de la Universidad de Jaén, por sus numerosas críticas y propuestas sobre los primeros borradores de este material.

Dado que toda obra es manifiestamente mejorable, huelga decir que será bien venida toda advertencia sobre los fallos y errores que pueda contener el texto, pero también cualquier sugerencia, crítica y propuesta que se nos desee hacer sobre sus contenidos y sus posibles actualizaciones.

Jaén, octubre de 2004.
Los autores.

BLOQUE 1

MEJORA DE LA CALIDAD

La mejora de la calidad es la parte de la gestión de la calidad orientada a aumentar la capacidad de cumplir con los requisitos de la calidad (ISO 9000: 2000). Dicho de otra forma, los gestores de la calidad no pretenden sólo, y exclusivamente, cumplir con los requisitos que, en un momento dado ha de poseer un producto determinado, sino el poder avanzar en un proceso de mejora sostenida. Se pretende que el producto o servicio que se oferta, y por ende la propia empresa, mejoren con el tiempo para que puedan destacar sobre sus competidores.

La trilogía de Juran establece un paradigma básico de la mejora continua:

— La planificación, que permite utilizar conocimientos ya adquiridos para controlar los sucesos del futuro.

— El control, que basado en herramientas estadísticas, tomará como estándar el resultado de la planificación.

— La mejora, que como actividad sistemática, científica y planificada ha de servir para corregir deficiencias. Se considera que para ser efectiva debe ser continua, lo que supone potenciar su organización.

La calidad se podrá mejorar cuando se disponga de información adecuada, y ésta se analice convenientemente para obtener resultados de valor para la empresa y su Sistema de Gestión de la Calidad (SGC). Las fuentes de información para esta mejora podrán ser muy diversas: gráficos de control, quejas de los clientes, sugerencias de los empleados, tormenta de ideas, círculos de calidad, etc. No obstante, existe un conjunto de herramientas[1], técnicas y métodos consolidado a nivel de gestión (Tabla 1.1), para

1 La diferencia entre herramienta y técnica estriba, fundamentalmente, en que las primeras son métodos más sencillos y practicables por todas las personas de una organización, cuya aplicación generalizada está enfocada a mejorar la efectividad en la dinámica de gestión de cada una de éstas personas. Las técnicas

el trabajo individual o en equipos; algunos de estos elementos se van a presentar en este bloque de casos.

Tabla 1.1 - Herramientas y técnicas para la mejora de la calidad

Herramientas	**Técnicas**
Herramientas básicas (o herramientas de Ishikawa): Diagrama de Pareto. Diagrama Causa-Efecto (Ishikawa). Gráficos de control (Shewhart). Diagrama de dispersión. Histograma. Estratificación. Hoja de verificación. Herramientas avanzadas: Diagrama de afinidad. Diagrama de relaciones. Matrices de priorización. Diagrama matricial. Diagrama de árbol. Diagrama de flechas (PERT). Diagrama de proceso de decisión.	Análisis del Modo Fallo, Efectos y Criticidad (AMFEC, o FMECA). Despliegue de la Función de Calidad (DFC). Diseño de experimentos. Tormenta de ideas.

Cada una de estas herramientas tiene un carácter y finalidad diferente, con posibilidad de funcionar de forma interrelacionada, y a diferentes niveles de gestión. Algunas de ellas requieren trabajo en equipo y ser rediseñadas, o modificadas, para ajustarse mejor a las necesidades de cada organización.

Desde una perspectiva más abstracta, la mejora no es mas que un proceso de resolución de problemas, es decir, la aplicación del ciclo de resolución de problemas. En calidad este ciclo se denomina Ciclo PDCA (Planear, Realizar, Comprobar, Actuar; en inglés *PDCA*: *Plan, Do, Check, Act*) (Figura 1.1). La mejora continua será pues una aplicación de dicho ciclo en espiral.

La Tabla 1.2 presenta las relaciones entre las distintas etapas del ciclo PDCA y las herramientas básicas y avanzadas que se pueden aplicar.

son herramientas más complejas, que requieren expertos en su utilización, y cuyo objetivo también es la mejora. En algunos casos la diferencia también puede considerarse en función del nivel en que se utilizan: estratégico (qué hacer), táctico (cómo hacerlo) y operacional (hacerlo).

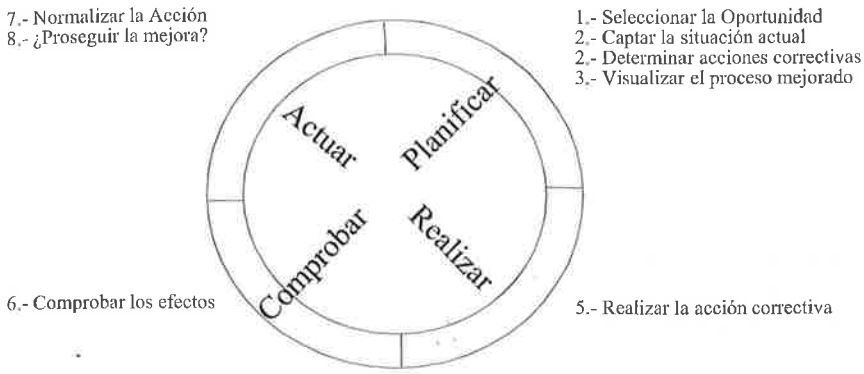


Figura 1.1 - Ciclo PDCA

Tabla 1.2 - Relación entre las etapas del ciclo PDCA, los pasos del análisis y las herramientas a aplicar

PDCA	7 PASOS	7 HERRAMIENTAS BÁSICAS	7 HERRAMIENTAS DE GESTIÓN
PLAN	Seleccionar un tema Recoger y analizar datos Analizar causas	Hojas de chequeo, gráficos, diagrama de Pareto, diagrama de flujo del proceso, histograma, diagrama de dispersión y diagrama de causa - efecto	Diagrama de afinidad Diagrama de relaciones Diagrama matricial
DO	Planificar e implantar la solución		Diagrama de árbol Diagrama matricial Diagrama de flechas Diagrama PDPC
CHECK	Evaluar efectos	Hojas de chequeo, gráficos, diagrama de Pareto, histograma, diagrama de dispersión y diagrama de causa - efecto y gráfico de control	
ACT	Normalizar la solución Reflexión en el proceso y próximo problema		Diagrama de flechas Diagrama PDPC Diagrama de afinidad
Aporta repetición	Aporta pasos	Aporta herramientas	Aporta herramientas

Fuente: FFII (1999)

Los contenidos de este bloque son los siguientes:

— Caso 1: Elaboración de un Procedimiento Específico. Pretendemos resaltar la importancia del Sistema Documental dentro de un Sistema de Gestión de Calidad. El trabajo se centra en la elaboración de un Procedimiento Específico, documento que se sitúa en uno de los niveles más concretos de dicho Sistema.

— Caso 2: Utilización del diagrama de Pareto. Se pretende dar a conocer y aprender a utilizar esta herramienta, a la vez simple y efectiva. Se propone un

ejercicio para enfocar los esfuerzos de mejora en la producción de una serie cartográfica.

— Caso 3: Utilización del diagrama de Ishikawa. En este caso se propone el empleo de esta herramienta para analizar y conocer los problemas de una digitalización manual mediante tablero.

— Caso 4: Análisis Modal, de Fallos y Efectos. Se presenta someramente la herramienta AMFE y un ejemplo de su uso. Se propone una aplicación libre de la misma en el ámbito de la ingeniería cartográfica.

— Caso 5: Utilización del Despliegue de la Función de Calidad. Aquí se muestra una herramienta de mayor nivel de gestión que permite el desarrollo de ideas para mejorar productos, en función de las cualidades de los productos competidores y de la voz, u opinión, de los clientes. En este caso se plantea utilizarla en una tarea de mejora de un mapa turístico.

Para la aplicación concreta de las herramientas y técnicas de mejora que se van a exponer no hay que reseñar ningún soporte estadístico concreto. No obstante, los resultados de los análisis estadísticos pueden, y deben, ser fundamento de muchas de las decisiones y orientaciones que se tomen en la aplicación de las herramientas de mejora.

En relación con las herramientas que aquí se presentan, Ariza (2002), Cuatrecasas (2001), Sebastián, Bargueño y Novo (1998) y Vilar (1997) son referencias adecuadas para ampliar conocimientos sobre algunas de ellas. En Internet se pueden encontrar también numerosos recursos dedicados a estas herramientas y ejemplos de aplicación, por lo general en ámbitos no cartográficos.

1.1. CASO 1

Elaboración de un Procedimiento Específico

SISTEMA DOCUMENTAL DE UN SISTEMA DE GESTIÓN DE LA CALIDAD

En los sistemas de gestión de la calidad todo ha de quedar perfectamente descrito para que en cualquier momento se sepa qué hacer, y cómo hacerlo; por ello ha de existir un sistema documental (manual de calidad, planes, registros, etc.) bien estructurado y actualizado. De forma bastante coloquial pero ajustada, podemos decir que el SGC consiste en "decir lo que se va a hacer, y hacer lo que se ha dicho", de ello la importancia y existencia de un sistema documental, debiendo quedar todo reflejado en documentos escritos y registros. La documentación permite la comunicación del propósito y la coherencia de la acción (ISO 9000:2000).

LOS PROCEDIMIENTOS EN LA GESTIÓN DE LA CALIDAD

Una pieza clave del sistema documental de gestión de la calidad son los procedimientos, que pueden ser de dos tipos:

- Procedimientos Generales (PG). Uno por cada proceso descrito en el manual de calidad. Éstos son una descripción de una forma global, sin llegar a definir tareas concretas. Los contenidos son: objeto, alcance, responsabilidades, definiciones, ejecución, referencias, registros, anexos.

- Procedimientos Específicos (PE). Han de contener información detallada de cómo se desarrollan todas y cada una de las actividades productivas, complementarias y auxiliares, que se hayan considerado; llegando a instrucciones de trabajo y especificaciones pormenorizadas (por ejemplo,. el desarrollo de un test de control posicional de una BDG). Por tanto, son la expresión escrita del Know-how de la empresa.

Para que un procedimiento sea efectivo debe informar sobre alguno, o la totalidad, de los siguientes aspectos:

- Descripción de las acciones y decisiones.

- Información sobre los equipos mínimos necesarios.

- Criterios para la toma de decisiones.
- Criterios para seleccionar alternativas.
- Condiciones iniciales para comenzar el proceso.
- Condiciones finales del proceso.
- Cualquier precaución necesaria para evitar incidentes.
- Controles necesarios para verificar la calidad.
- Cualquier formulario que se deba rellenar.
- Referencias cruzadas con otros documentos.
- Origen de la información.
- Diagrama de flujo.

Los procedimientos suelen tener la siguiente estructura (Senlle, Vilar, 1996):

- Objetivo: Presenta los objetivos que se pretenden lograr.
- Alcance: Campo o área de aplicación y limitaciones de uso.
- Responsabilidades: establece las responsabilidades de cada actividad descrita en el procedimiento.
- Definiciones: aclara conceptos para evitar toda ambigüedad.
- Ejecución: descripción cronológica de la operativa (qué hacer, cuándo, cómo, quién,...).
- Referencias: citas a otros documentos o normas aplicables, internas o externas, y que no se incluyan en los anejos.
- Anejos: se adjunta los formularios, fichas, especificaciones, dibujos, etc. necesarios.
- Registros: Registros y formatos que se utilizan al ejecutar el procedimiento, dónde, cuándo y quién se encarga de ellos.

Los procedimientos y manuales de calidad suelen editarse con un formato normalizado de tal manera que presentan todos la misma estructura. Es común incluir unas cajas en cabeza y pié de página, que con formato variable, presentan la información más relevante sobre el procedimiento (fecha de revisión, nivel de difusión, página, persona que lo ha elaborado, revisado y aprobado, etc.). A continuación, en la Figura 1.2 se presentan algunos ejemplos.

Empresa	Manual de Calidad	Capítulo	
	Apartado del manual	Fecha	
		Rev.:	Pág.:

Empresa	Manual de Procedimientos	Nombre del Documento	
	Nivel de difusión	Fecha Creación: Fecha Revisión:	
		Rev.:	Pág.:

Redacción	Revisión	Aprobación
Firma: Fecha:	Firma: Fecha:	Firma: Fecha:

Figura 1.2. Ejemplos de cabeceras y pies de formatos de un manual de calidad

TRABAJO A REALIZAR

Desarrollar un procedimiento que describa cómo ha de hacerse la toma de datos en campo mediante el método de radiación para el levantamiento de coordenadas X, Y, Z de los puntos capturados, indicando objetivo, alcance, responsabilidades, definiciones, ejecución (incluyendo diagrama de flujo), así como los anejos y registros necesarios.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

1.2. CASO 2

Utilización del diagrama de Pareto

EL DIAGRAMA DE PARETO

El diagrama de Pareto es un método gráfico de análisis, similar al diagrama de barras o frecuencias, que permite visualizar ordenadamente, según su importancia, los factores que afectan a un proceso o una situación. Los diagramas de Pareto sirven para observar el comportamiento propio de muchos procesos naturales, sociales y productivos, en los que existen unos pocos factores vitales (lo poco que es vital), de gran importancia para el resultado del proceso, mientras que otros muchos factores (lo mucho que es trivial o útil) poseen una influencia muy reducida.

Según el principio de Pareto, de manera general, el 80% de los problemas están causados por el 20% de las causas. Por tanto, las acciones de mejora o corrección deberán realizarse sobre los factores que presenten mayor influencia.

El procedimiento de construcción del diagrama es sencillo. Las diferentes etapas pueden resumirse en (Sebastián, Bargueño y Novo, 1999):

1. Elegir las causas a analizar.

2. Elegir la unidad de medida para el análisis.

3. Determinar el intervalo de tiempo en el que los datos van a analizarse.

4. Colocar sobre el eje horizontal, y de izquierda a derecha, las causas en orden decreciente de la unidad de medida.

5. Trazar dos ejes verticales, uno a cada lado del eje horizontal. En éste último se reflejarán las causas y en los ejes verticales el porcentaje de casos de cada causa y el porcentaje acumulado.

6. Representar el diagrama construyendo para cada causa una barra de altura proporcional al porcentaje de casos.

7. Construir la línea de frecuencia acumulativa a partir del porcentaje de casos acumulado.

8. Completar el gráfico con las informaciones necesarias para comprender el análisis (título, período, responsable de los datos, etc.).

EJEMPLO DE APLICACIÓN

La Figura 1.3 presenta un ejemplo de diagrama de Pareto aplicado al caso de las informaciones incluidas en la Tabla 1.3, que se corresponden con datos relativos a un suministro que posee problemas de retraso, y que ha generado quejas de un cliente. Para mayor simplicidad, se supondrá que cada retraso está originado por un único motivo.

Tabla 1.3 - Problemas de retraso en un suministro

Motivo	Casos	% del total	% acumulado
Problemas de mala calidad	45	0,45	0,45
Problemas en el proceso productivo	35	0,35	0,80
Errores en la planificación de la producción	22	0,22	1,02
Errores en la documentación de la expedición	13	0,13	1,15
Problemas del transporte	12	0,12	1,27
Vacaciones	7	0,07	1,34
Otros motivos esporádicos	6	0,06	1,40
Falta de material	5	0,05	1,45
Falta de personal en la producción	3	0,03	1,48
Falta de personal en el control de la calidad	2	0,02	1,50

Nota: El total de suministros al año es de 10000.

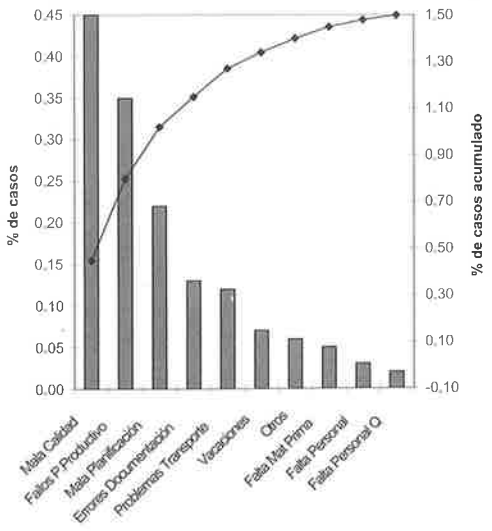


Figura 1.3 - Diagrama de Pareto

Supóngase que la actuación sobre cualquier motivo del retraso implica una misma tasa de éxito[2] (por ejemplo el 80%). Partiendo de un número de retrasos totales del 1,5% (150 sobre 10000), actuando sobre distintas causas se conseguirá disminuir este valor:

— Si se actúa sobre la primera causa (problemas de mala calidad, 45 retrasos):

$$(150 - (0{,}8 \cdot 45)) \cdot 100/10000 = 1{,}14\% \text{ de retrasos totales}$$

— Si se interviene sobre la quinta causa (problemas del transporte, 12 retrasos):

$$(150 - (0{,}8 \cdot 12)) \cdot 100/10000 = 1{,}40\% \text{ de retrasos totales}$$

Y por tanto, de forma conjunta:

$$(150 - (0{,}8 \cdot (45+12))) \cdot 100/10000 = 1{,}04\% \text{ de retrasos totales}$$

También se pueden buscar combinaciones de causas tal que, al actuar sobre las mismas, se pueda alcanzar un objetivo (por ejemplo el 0.5% de retrasos):

— Caso 1, Se interviene sobre: Mala calidad (45 retrasos), problemas en el proceso productivo (35 retrasos) y errores en la planificación (22 retrasos).

Se supone una efectividad del 98% sobre cualquier motivo de retraso.

Los retrasos disminuyen: $(150 - (0{,}98 \cdot (45+35+22))) \cdot 100/10000 = 0{,}5\%$

— Caso 2, Se actúa sobre: Mala calidad (45 retrasos), problemas en el proceso productivo (35 retrasos), errores en la documentación (13 retrasos) y problemas en el transporte (12 retrasos) (total de 105 retrasos).

Se supone una efectividad del 95,2% sobre cualquier motivo de retraso.

Comprobando: $(150 - (0{,}952 \cdot (45+35+13+12)) \cdot 100/10000 = 0{,}5\%$

Así, en el Caso 2, dado que se va a intervenir sobre un mayor número causas, la efectividad puede ser menor. Evidentemente, se debe ajustar el método con la realidad del proceso, considerando distintas efectividades para cada causa. También deberían distinguirse los costes de actuación sobre cada una, y tener este factor en consideración a la hora de decidir cómo actuar en la mejora.

2 La tasa de éxito ha de entenderse como el rendimiento del proceso de aplicación de la mejora. Es decir, dada una causa *C* que genera un problema que se cuantifica en un número *D* de defecto en la producción, si se implementa adecuadamente la mejora existirá una reducción observable del número de defectos en un porcentaje equivalente a la tasa de éxito.

TRABAJO A REALIZAR

Supóngase que un organismo público ha contratado los servicios de una empresa de ingeniería cartográfica para la elaboración de una serie cartográfica[3] E10k formada por 1000 hojas. Considerando que la entrega se iba a realizar en cuatro lotes de 250 hojas cada uno, la empresa decidió aplicar sobre el primer lote la herramienta "diagrama de Pareto" con la información relativa a los problemas encontrados en el control de calidad interno. En la Tabla 1.4 se presenta el tipo y número de problemas encontrados, y en la Tabla 1.5 la distribución de estos problemas en el lote, ya que en una misma hoja puede encontrarse más de un problema. Esta última tabla tiene gran interés práctico si se conoce la correlación entre los distintos factores. Si la correlación es negativa, puede ocurrir que, al actuar sobre uno de ellos en el sentido de la mejora otro factor empeore.

Tabla 1.4 - Total de casos de cada tipo de problema detectado en el lote

Tipo de problema	Casos
1 - No pasa el control de calidad posicional en Y.	81
2 - Problemas de cruces entre curvas de nivel.	51
3 - Errores en la rotulación de los cultivos.	31
4 - Errores en la ubicación de topónimos.	24
5 - Ausencia de topónimos.	19
6 - No pasa el control de calidad posicional en X.	13
7 - Errores en los rótulos de las vías de comunicación.	8
8 - Problemas en el título y el número de hoja.	7
9 - Problemas de rotulación de la cuadrícula de coordenadas.	5
10 - Errores en la rotulación de municipios.	4
11 - Desplazamiento en la ubicación de la leyenda.	2

Tabla 1.5 - Distribución de los problemas detectados en el lote

Nº de mapas	Tipo de problema encontrado										
	1	2	3	4	5	6	7	8	9	10	11
38	x										
17		x									
10			x								
3				x							
6					x						
5						x					
4							x				
5								x			
2									x		
2										x	

3 La expresión de la escala se basa en la recomendación del Consejo Superior Geográfico (CSG, 1992) para referirlas. La regla consiste en utilizar la letra "E" seguida del denominador de la escala expresado en miles: por ejemplo, una escala 1:1.000 se denotará como E1 y una escala 1:50.000 como E50. Para mayor claridad, y en línea con las normas de otros países, nosotros añadimos una "K" al final. Es decir, para los ejemplos anteriores: E1K y E50K

Tabla 1.5 - Distribución de los problemas detectados en el lote

Nº de mapas	Tipo de problema encontrado										
	1	2	3	4	5	6	7	8	9	10	11
1											x
18	x	x									
7	x		x								
6	x	x	x								
6				x	x						
5	x	x		x							
3	x	x	x	x							
3	x			x	x	x					
2	x					x					
2	x		x	x	x						
2			x				x		x		
1	x	x		x	x	x		x		x	
1	x	x	x	x	x	x	x	x	x		x
1	x					x	x			x	
Total de casos	81	51	31	24	19	13	8	7	5	4	2

Analizando el diagrama de Pareto se podría mejorar el proceso de producción de los tres lotes restantes, detectando cuáles son las acciones de mejora más importantes que deben efectuarse. Así, el trabajo a realizar en este caso consiste en:

— Elaborar y analizar un diagrama de Pareto a la vista de los datos que se presentan en la Tabla 1.4. Con ello se conseguiría tener una visión inicial del tipo y cantidad de problemas que presenta el lote.

— Empleando la Tabla 1.5, calcular cuál es el número inicial de mapas incorrectos, y determinar cuál sería el porcentaje de hojas que todavía tienen algún defecto si se interviene sucesivamente para atajar cada problema: inicialmente sobre las causas del problema 1, después sobre las causas que originan los problemas 1º y 2º, y así sucesivamente. Considérese, para cada una de las intervenciones, diferentes tasas de éxito, por ejemplo 75%, 90%, 95%.

— Finalmente, elaborar una gráfica que relacione el número de mapas que todavía presentan deficiencias con los problemas sobre los que se ha actuado, y con la tasa de éxito supuesta. Se analizará y se sacarán conclusiones a la vista del diagrama de Pareto y de la referida gráfica.

Un aspecto a destacar es el hecho de haber supuesto que los problemas detectados son independientes, es decir, que al actuar sobre uno de ellos, para reducirlo, no muestra correlación, positiva o negativa, con otro cualquiera. Esto no tiene por qué ser así en la realidad, ya que, por ejemplo, si se actúa para reducir el número de mapas que no pasan el control de calidad posicional en la componente Y, esto influirá, casi con toda probabilidad, en el mismo problema detectado para la componente X, originando una correlación positiva entre ambos.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

1.3. CASO 3

Utilización del diagrama de Ishikawa

EL DIAGRAMA DE CAUSA EFECTO (DIAGRAMA DE ISHIKAWA)

El diagrama de causa efecto, propuesto y empleado por Ishikawa, es una herramienta de análisis que permite visualizar cómodamente, de forma sencilla y sinóptica, las distintas causas y subcausas que originan un efecto concreto. También se denomina diagrama de raspa de pez por la forma que adopta: las causas principales se derivan directamente del espinazo, mientras las causas secundarias se derivan de éstas últimas, y así sucesivamente, por lo que tienen un sentido claramente jerárquico.

El diagrama de Ishikawa no sólo se emplea para observar un efecto negativo que pueda producirse e investigar las causas que han podido producirlo, sino también para mejorar una cierta característica de calidad mediante el conocimiento de las causas que pueden provocarlo. el diagrama de causa efecto es ilustrativo, sirviendo para que todos los implicados conozcan, en detalle, el proceso con que trabajan, visualizando con claridad las relaciones entre los efectos y sus causas. Permite guiar las discusiones, exponer con claridad los orígenes de un problema, y encontrar rápidamente las causas asignables cuando el proceso se aparta de su funcionamiento habitual. Esta herramienta también es adecuada tras un análisis de Pareto para profundizar sobre las causas de un factor sobre el que se quiere actuar.

Según lo comentado, las principales aplicaciones del diagrama causa efecto son (Sebastián, Bargueño y Novo, 1999):

- Mejora de procesos, uso más eficiente de recursos y reducción de costes.
- Eliminación de las causas que originan productos defectuosos para así atender las demandas de los clientes.
- Establecimiento de procedimientos normalizados.
- Formación y educación del personal, entrenándolos en toma de decisiones y en actividades de grupo.
- Establecimiento de bancos de datos que permitan, en otro momento, observar y analizar lo sucedido.

El análisis de dispersiones (para analizar la relación existente entre dos clases de datos), la enumeración de causas, y la clasificación de la operaciones, son los métodos más comunes para generar este diagrama. Un planteamiento clásico consiste en analizar la repercusión de las cinco emes (5M) sobre el efecto, éstas son: material, mano de obra, medio, máquina y método.

La Figura 1.4 presenta un ejemplo de aplicación de un diagrama de Ishikawa para el estudio de las causas que originaron el fracaso en el proceso de generalización de una serie cartográfica bajo unas circunstancias concretas de personal, programas, información y métodos de trabajo.

Figura 1.4 - Diagrama de Ishikawa para la generalización de una serie con problemas

Las fases a seguir en la realización de un diagrama causa efecto son las siguientes:

— Definir claramente el efecto o característica de calidad a analizar. En el diagrama se traza horizontalmente una flecha gruesa y larga que representa el proceso y a la derecha se escribe el efecto o característica de calidad.

— Identificar algunas categorías de causas, esto es, grandes áreas que sirvan para desarrollar el análisis en detalle, según lo comentado anteriormente se podrían utilizar las 5M si se consideran adecuadas. Cada una de estas categorías de causas etiquetarán cada una de las flechas secundarias que confluyen en la línea horizontal.

— Para cada una de estas divisiones escribir las causas, de nivel inferior, que influyan sobre las mismas. Este proceso se puede repetir hasta que se alcancen

las causas últimas en cada rama del diagrama. Lo que se tendrá finalmente es un diagrama con todas las causas posibles. Puede resultar muy adecuado realizar previamente una tormenta de ideas de las causas potenciales del problema. Hay que indicar que se puede colocar una posible causa en más de una categoría.

— Seguidamente es conveniente seleccionar e identificar las causas que son susceptibles de tener una mayor influencia sobre el efecto, para a continuación preguntarse el porqué de las mismas, intentando hallar la raíz del problema.

Una vez completado el diagrama hay que evaluarlo y determinar las causas más significativas que han podido contribuir al efecto analizado. La identificación de esas causas necesita una valoración crítica de todas ellas haciendo incluso uso de datos históricos, si se tienen, para analizar lo sucedido en casos precedentes.

La decisión de cuáles son las causas más significativas puede hacerse por unanimidad entre los miembros del grupo de trabajo, o incluso por votación, si aquella no existe. Estas causas (4 ó 5 suelen ser suficientes) se jerarquizan en orden de importancia, para después ser examinadas, contrastando su veracidad y definiendo las acciones necesarias para eliminarlas o reducir su influencia en el efecto.

TRABAJO A REALIZAR

Se aplicará el diagrama de Ishikawa para analizar los diferentes problemas que pueden aparecer en las tareas de digitalización manual de una cartografía mediante tablero. Concretamente, el efecto a analizar es la "mala calidad posicional de la cartografía digitalizada". Se deberán analizar las componentes de material, mano de obra, medio, máquina y método, cuya identificación en esta ocasión es muy clara, como se recoge en la Tabla 1.6.

Tabla 1.6 - Factores equivalentes a las cinco emes en la digitalización manual con tableta

Componente	Equivalente en la digitalización manual con tableta
Material	Cartografía a digitalizar
Mano de obra	Operador
Medio	Entorno de trabajo
Máquina	Tableta digitalizadora y programa
Método	Método de digitalización

Las cinco componentes de la Tabla 1.6 constituirán el primer nivel en la jerarquía del diagrama, debiéndose analizar cada uno de ellos e ir estableciendo sucesivamente niveles inferiores. Finalmente hay que evaluar el diagrama elaborado, detectando cuáles han de ser las causas más significativas del problema analizado y las acciones que son recomendables para reducirlas o eliminarlas.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

1.4. CASO 4

Análisis Modal de Fallos y Efectos

EL MÉTODO AMFE

Al igual que otras muchas técnicas que se aplican en el ámbito de la calidad, el Análisis Modal de Fallos y Efectos (AMFE)[4] es una herramienta que tiene sus orígenes en la industria aerospacial, y que luego ha sido adoptada por otros sectores, como el automovilístico. En la actualidad se requiere su uso en algunas especificaciones técnicas correspondientes a licitaciones de cartografía por entidades financieras internacionales.

El Análisis Modal de Fallos y Efectos (AMFE) es una metodología que permite analizar la calidad, seguridad y/o fiabilidad del funcionamiento de un sistema. Dado que el análisis ha de llevar a la mejora, se puede considerar una herramienta de mejora por los caminos de la prevención gracias a cierta capacidad de predicción que incorpora. Esta herramienta puede aplicarse a los procesos, medios de producción y productos, y en este último caso, tanto al diseño como a su mejora.

El objetivo original de la herramienta es limitar al máximo la posibilidad de fallos en un sistema o producto, para ello se analizan los fallos existentes, los modos de fallo y sus efectos. Todo ello servirá para priorizar las actuaciones y, a partir de ello, las causas que los originan, aspecto que permitirá desarrollar mejoras gracias que aseguren que se reducen estas causas. La Figura 1.5 presenta un esquema general del proceso AMFE que se describe a continuación.

Para la realización de AMFE es necesario un gran conocimiento en la materia y por ello suelen intervenir equipos pluridisciplinares donde participan miembros de los distintos departamentos de la organización que están implicados en el proceso o producto.

La base para el análisis debe ser un diagrama que descomponga, al nivel adecuado, cada uno de los elementos y relaciones de interés. Junto con el diagrama se podrán utilizar otras herramientas básicas para profundizar sobre cada uno de los aspectos de interés. Todo ello ha de llevar a:

4 Esta técnica también se conoce como FMECA de Failure Mode, Effect and Critical Analysis, o por su traducción al castellano como AMFEC (Análisis del Modo de Fallos, sus Efectos y Criticidad).

- Explicitar los posibles fallos de un producto o proceso, llevando a la estimación de su probabilidad de ocurrencia, probabilidad de no detección y la severidad o gravedad de sus consecuencias.

- Prever los medios y acciones que permitan detectar los fallos, disminuir los riesgos y hasta cambiar el método de producción.

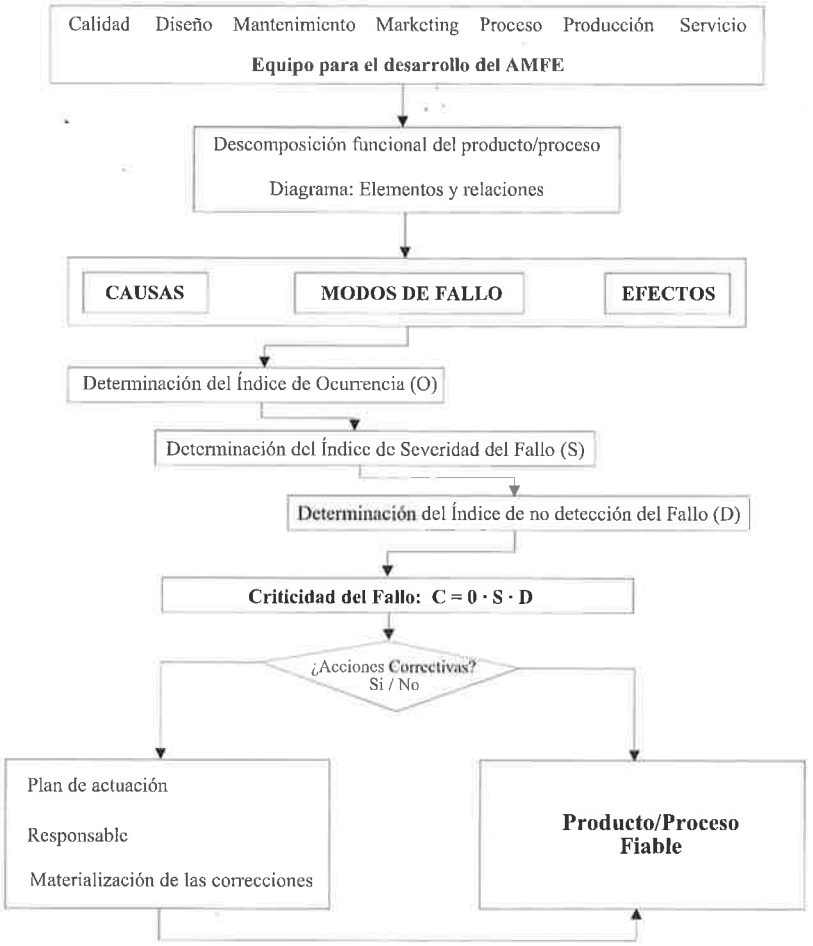


Figura 1.5 - Esquema general del proceso AMFE

Un concepto básico de esta metodología es el de modo de fallo. Por modo de fallo se entiende cada una de las maneras en que un producto (pieza elemental o compleja) o sistema puede fallar potencialmente respecto a las especificaciones que se le han establecido.

Indudablemente, dado que los fallos pueden deberse a causas únicas o a combinaciones, concatenaciones de causas, etc., todas estas posibilidades han de

estudiarse. Para ello se utilizan diagramas donde se presentan las causas y las relaciones lógicas que las relacionan para generar un modo de fallo. La Figura 1.6 presenta unos esquemas básicos de interactuación.

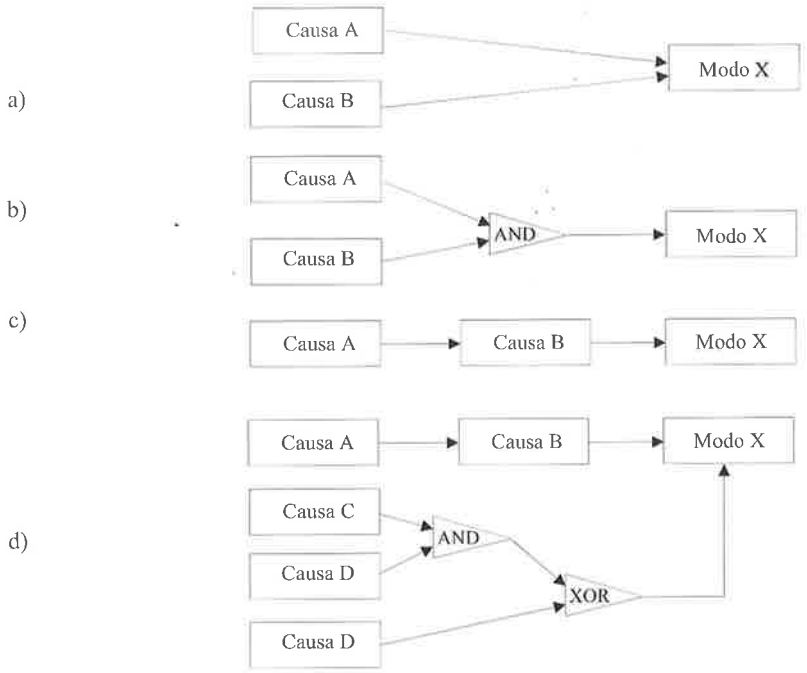


Figura 1.6 - Ejemplo de interactuación de causas para generar modos de fallo: a) independencia, b) concurrencia, c) encadenamiento y d) ejemplo de modelo complejo

Sin embargo, ha de entenderse que cada causa o conjunto de causas no origina un único modo de fallo y que un único modo de fallo no origina un único efecto. Por el contrario, lo que existe es un entrelazado grande de todos ellos, como pone de manifiesto la Figura 1.7.

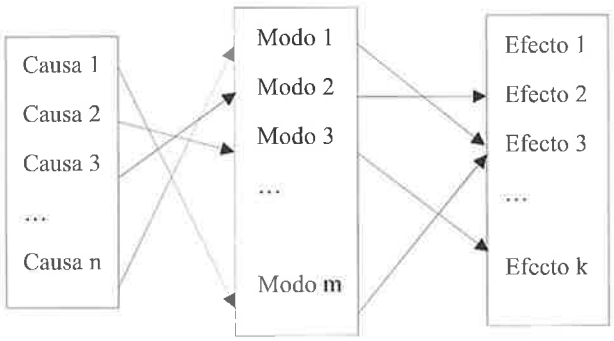


Figura 1.7 - Complejidad de la combinación de causas, modos y efectos

El aspecto más interesante de esta herramienta es que para cada modo de fallo se calcula un índice de criticidad, según:

$$C = O \times S \times D$$

donde:

C Índice de criticidad del fallo.
O Índice de ocurrencia del fallo.
S Índice de severidad del fallo.
D Índice de no detección del fallo.

Tabla 1.7 - Valoración de los factores *O*, *S* y *D* en el Producto

Factor (*O*) Probabilidad de Ocurrencia del fallo			**Factor (*S*) Severidad o gravedad del fallo**		**Factor (*D*) No Detección del defecto**	
O	Frecuencia del fallo		*S*	Severidad del fallo	*D*	Probabilidad no detección
1	0: 0,01%	Muy escasa	1	Fallo de difícil detección por el cliente	1	0,5%
2	0,01:0,02%	Escasa	2:4	Fallo detectado por el cliente, le molesta pero no afecta a la funcionalidad	2	5:10%
3	0,02:0,05%	Escasa	5:7	Fallo que detecta el cliente, le molesta y afecta a la funcionalidad	3	10:20%
4	0,05:0,10%	Media	8:9	Fallo que incomoda mucho al cliente, produce devolución y una costosa reparación	4	20:30%
5	0,11:0,20%	Media	10	Fallo que pone en peligro la seguridad del cliente	5	30:40%
6	0,21:0,50%	Media			6	40:50%
7	0,51:1,0%	Frecuente			7	50:60%
8	1,01:2,0%	Frecuente			8	60:70%
9	2,01:10,0%	Frecuente			9	70:80%
10	>10%	Elevada			10	80:100%

Tabla 1.8 - Valoración de los factores O, S y D en el Proceso

Factor (*O*) Probabilidad de Ocurrencia del fallo			**Factor (*S*) Severidad o gravedad del fallo**		**Factor (*D*) No Detección del defecto**	
O	Frecuencia del fallo		*S*	Severidad del fallo	*D*	Severidad del Fallo
1	Cp > 1,33 0:0,01%	Esporádico	1	Fallo sin influencia en las siguientes operaciones	1	Ninguna influencia en el proceso
2	1,33 > Cp >= 1 0,01:0,1%	Muy poco frecuente	2	Fallo con una pequeña influencia en las operaciones siguientes, sin llegar a parar el proceso productivo.	2	Alta posibilidad de detección
3	1>Cp>=0,83 0,1:0,5%	Poco frecuente	3	Fallo con perturbación de la producción, es decir, produce parada y rechazos.	3	Poca posibilidad de detección por parte del operario, pero si le produce descontente
4	Cp < 0,83 0,5:2,0%	Frecuente	4	Fallo que origina el paro de la producción y grandes rechazos.	4	El fallo es de difícil detección
5	Proceso fuera de control 2:10%	Muy frecuente	5	Fallo que produce los efectos del punto anterior e incide en la seguridad de las operaciones siguientes.	5	No existe, prácticamente, posibilidad de detección

Nota: Cp <> Capacidad del proceso. Este Cp viene dado como Cp = (Ts -Ti)/(6s), donde Ts es la tolerancia superior, Ti es la tolerancia inferior, s es la desviación típica o estándar.

Este índice *C* permite tener una valoración global de la criticidad y, a partir de ella, decidir si se debe actuar o, si se tienen diversos modos de fallo, en qué orden o prioridad.

Los valores de los índices *O, S* y *D* están tabulados tanto para productos y procesos, por lo que se hace necesario disponer de datos históricos y estadísticas adecuadas para saber el nivel real de cada uno de ellos. Las Tablas 1.7 y 1.8 presentan unas escalas de valoración para dichos índices.

Según se deduce de las Tablas 1.7 y 1.8, el índice de criticidad *C* puede variar en los siguientes intervalos:

$$Producto: C \in [1, 1000]$$
$$Procesos: C \in [1, 125]$$

Como orientación se aconseja desarrollar acciones cuando (AEC, 1990):

$$Producto: C >= 60$$
$$Proceso: C >= 20$$

O cuando, si bien los valores de *C* no sean altos, si se cumplen estas condiciones:

$$Producto: O >= 5 \ Y \ D >= 5$$
$$Proceso: O >= 3 \ Y \ D >= 3$$

Finalmente, el AMFE se plasma en un cuadro como el que se presenta en la Tabla 1.9. Este cuadro posee un encabezado en el que se especifica si se trata de un proceso o producto y se añade la información adecuada para tener bien referido el análisis. La parte baja de la tabla resumen se conforma por tres partes o grupos de columnas:

— Estado actual (columnas de la 1ª a la 6ª): Se identifica la operación, el componente o pieza, así como el modo de fallo, el efecto que supone dicho fallo, los controles existentes y las estimaciones numéricas de los índices *O, S, D* y *C* para la situación actual.

— Acciones (columnas 7ª y 8ª): Se especifica la acción o acciones a desarrollar para cada modo de fallo, y el (los) responsable(s).

— Resultados (columnas 9ª y 10ª): Se incluyen las acciones implantadas y los resultados alcanzados sobre los índices *O, S, D* y *C*.

La Tabla 1.9 presenta un ejemplo de todo lo anterior para el caso de un producto cartográfico impreso en una empresa con notables problemas en su taller de artes gráficas.

Tabla 1.9 - Ejemplo de tabla resumen de un análisis AMFE

Producto	X	Proceso	X	Hoja *1 de 1*
Producto: ___Mapa impreso ___ Función: ___ ========== _____ Especificación: __Pliego PPT 10/04/99____		Proceso: ___Impresión y doblado_ Operación: B&N, Color y doblado_		Animador: _López López, José _____ Fecha realización: __19/10/2002_____ Fecha revisión: _____

Estado Actual										Acciones		Resultados				
1	2	3	4	5	6					7	8	9	10			
Nº Op	Componente Conjunto Pieza	Modo Fallo	Efecto Fallo	Causa Fallo	Condiciones Actuales					Acciones Correctoras	Responsabilidad	Acciones Implantadas	Resultado			
					Control	O	S	D	C				O	S	D	C
	Impresión colores	Falta legibilidad	Rechazo de productos	Registro colores	Control en la recepción	4	3	3	2 4	Revisión sistema obtención fotolitos y marcas de registro Ajuste máquina offset Incrementar controles impresión	Resp. fotomecánica Resp. Taller impresión	La recomendada	1	3	2	6
	Fallos impresión texto B&N	Falta legibilidad	Rechazo de productos	Problemas entintado	Control en la recepción	4	3	2	2 4	Limpieza máquina offset	Resp. Taller impresión	Se aplica lo recomendado	1	3	2	6
	Doblado	Mal doblado	Rechazo de productos	Problemas doblado	Puesta a punto 1 vez al mes	3	3	2	1 8	Ajuste velocidad y tensión	Resp. Talleres	Se aplica lo recomendado	1	3	2	6

TRABAJO A REALIZAR

Se buscarán en Internet varios ejemplos de aplicación del AMFE a productos y procesos o servicios. Posteriormente, tras estudiar y analizar los ejemplos, se formará un grupo de trabajo se decidirá sobre un producto, servicio o proceso cartográfico al que aplicar la metodología AMFE.

Si no se tienen registros o controles estadísticos para determinar las condiciones actuales, y las previsibles en el futuro, se realizará una encuesta a expertos en la materia.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

1.5. CASO 5

Utilización del Despliegue de la Función de Calidad

EL DESPLIEGUE DE LA FUNCIÓN DE CALIDAD (DFC)

Esta metodología de Despliegue de la Función de Calidad (DFC) (*Quality Function Deployment, QFD*) también se conoce con el nombre de la Casa de la Calidad (*the house of quality*). El DFC se cimenta en la firme creencia de que los productos deben ser diseñados para reflejar los gustos y los deseos de los clientes (Hauser y Clausing, 1989). Se trata de un mapa conceptual, con él un equipo interdepartamental de la empresa puede exponer, y organizar, las pruebas y datos que necesita para establecer los objetivos de diseño o de desarrollo de un producto. Es un método o rutina para la planificación, comunicación y colaboración entre los encargados del marketing, los ingenieros de diseño, los responsables de fabricación, etc., orientada al desarrollo estructurado de productos, que permite especificar e incluir claramente las necesidades y deseos del cliente en los productos.

La presentación gráfica del DFC es la que da nombre de casa dado que se basa en distintas matrices que se presentan como rectángulos y un tejado a dos aguas que se corresponde con otra matriz del proceso (Figura 1.8).

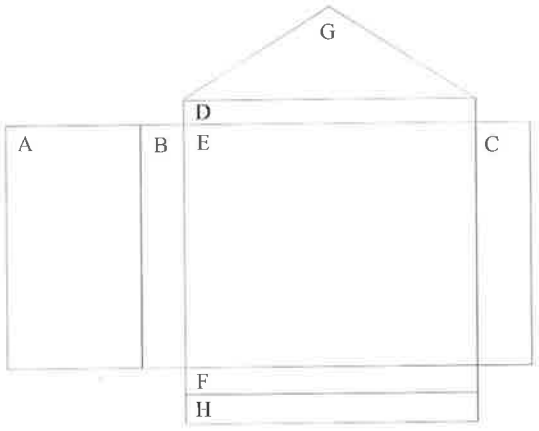


Figura 1.8 - La Casa de la Calidad con las distintas matrices que la forman

Como punto de partida se toma la voz del cliente (qué quiere el cliente), que se plasma o despliega en unas exigencias que se denominan atributos del cliente (AC). Los atributos del cliente se agrupan (matriz A) y ponderan en función de la importancia (matriz B) y se compara la situación actual de nuestro producto y los competidores (matriz C de percepciones del cliente) asignando valores que mostrarán claramente las oportunidades de mejora y las ventajas propias frente a los

competidores. Se trata por tanto, de una fase de estudio de mercado en la que se ha de conseguir, al menos, la siguiente información:

- Identificación de los clientes y de sus necesidades, tanto explícitas como latentes. No todos los clientes son usuarios finales, ya que los ACs pueden incluir las exigencias de una norma o estándar, las exigencias de los vendedores, de los transportistas, e incluso de un departamento que es suministrado por otro departamento de una misma empresa.

- Conocimiento y análisis de las quejas y sugerencias de los clientes para, posteriormente, reflejarlo en la mejora de las calidades más importantes.

- Análisis de la competencia, interesa saber qué organiza y decide en relación al mismo producto o similar, cómo lo consigue y con qué recursos cuenta.

Los AC se trasladan a características de la ingeniería (CI) lo que supone la respuesta técnica al problema (matriz D), es decir, así se expresan los cómos que resolverán los qués a partir de un despliegue funcional de los elementos técnicos. Cada CI puede afectar a uno o más AC, presentándose con signos + y - las necesidades de aumentar o disminuir el CI. Los CI marcados deben describir el producto en términos mensurables y deberán afectar directamente a las percepciones del cliente.

Mediante técnicas de consenso se rellena la matriz de relaciones (matriz E), indicando con una simbología cuánto afecta cada CI a cada AC. En la parte F de la matriz técnica se establecen ahora las medidas objetivas para cada uno de los CI anteriores tanto para nuestro producto como para el de la competencia (*benchmarking*). La cumbre de la casa (matriz G) es una matriz en forma de A que ayuda a los ingenieros a establecer las correlaciones técnicas entre los distintos CI, facilitando la creatividad de la ingeniería, y la posibilidad de cambios. En esta matriz las relaciones también se pueden expresar de forma simbólica.

En la matriz H, al pie de la casa, se incluye una valoración de la dificultad técnica de cada CI, de su importancia, de su coste estimado, así como los objetivos marcados para las medidas objetivo de nuestros CI.

Finalmente, cabe destacar, que no existen reglas rígidas ni en las simbologías ni configuraciones, etc., y que lo que hay que intentar es hacerla lo más útil posible. Igualmente, se debe apuntar la posibilidad de escalonar las casas de tal manera que los cómos de una fase lleven a los qués de la siguiente, con lo que finalmente se consigue que la voz del cliente alcance todo el proceso (Figura 1.9).

En suma, la principal ventaja de la casa de la calidad es "la calidad en casa", que hace que la gente piense en la dirección adecuada y que piense conjuntamente con los demás (Hauser y Clausing, 1989).

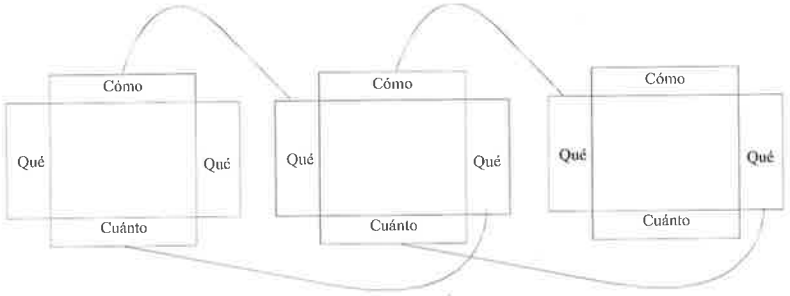


Figura 1.9 - Encadenamiento de las casas de la calidad

EJEMPLO DE APLICACIÓN

Supóngase que se tiene un producto que se corresponde con el Mapa de Carreteras a escala E400K cuyas características principales son las que aparecen en la Tabla 1.10.

Tabla 1.10 - Principales características del mapa de carreteras
Soporte papel. Papel de 70 gr/m^2. Tamaño desplegado 85 × 140 cm. Tirada a color por cuatricromía. Actualización carreteras a 1992. Actualización estaciones de servicio a 1995. No aparecen las propiedades de las áreas de servicio en la simbología (cafetería, hotel, taller, etc.). Fondo con fisiografía expresada mediante tintas hipsométricas.

Llegado el momento de volver a hacer una nueva versión del mismo se desea obtener un producto más competitivo que el anterior pues se han detectado fallos de muy diversa índole. Respecto a dos aspectos concretos, la voz del cliente dice:

— En cuanto al plegado:

Plegado complicado.
Muchos pliegues.
Se rompe por los pliegues con mucha facilidad.

— En cuanto a la semiología gráfica:

Excesiva densidad de topónimos.
Los colores de las carreteras no coinciden con los de las señales de tráfico.

Estos elementos conforman los atributos del cliente (AC) de la matriz A de la casa de la calidad de la Figura 1.10.

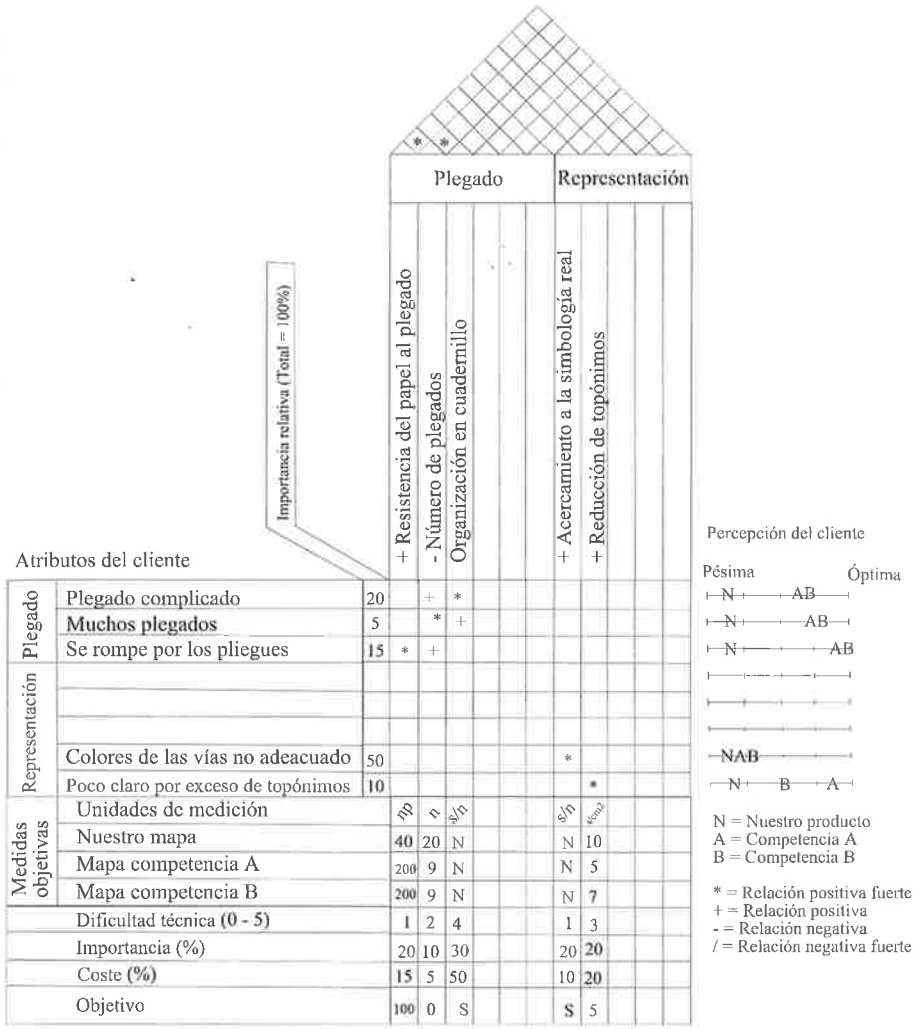


Figura 1.10 - Ejemplo de aplicación de DFC al caso de un mapa con problemas.

En primer lugar se procede a valorar la importancia relativa de cada uno de los AC, de tal manera que el total sume un 100%. Como se deriva de los pesos asignados, el atributo más apreciado es el cambio de la simbología tradicional a nuevos colores que reflejen la asignación que las señales de carretera hacen de las autovías, carreteras nacionales, autonómicas, etc.

Para cada uno de los AC se realiza un mapa perceptual que nos permite hacer la comparación con la competencia (elementos de la matriz C en la Figura 1.6). Según se

observa, el producto tiene claras posibilidades de mejora en todos los aspectos, pues siempre se encuentra peor valorado que la competencia.

La siguiente pregunta que podemos hacernos es cómo modificar el producto. Para lo cual los ingenieros especialistas en la materia nos indican, entre otras, las siguientes opciones, que constituirán las características de ingeniería (CI):

— Para el plegado:
  Resistencia al plegado.
  Número de pliegues.
  Organización como un cuaderno (encuadernación en espiral, etc.).

— Para la semiología:
  Acercamiento a la simbología del modelo real (las carreteras).
  Reducción de topónimos.

Estas CI se presentan en la Figura 1.10 en la parte correspondiente a la matriz D con la simbología + / – apropiada, según el caso de necesitar incrementar o disminuir las características de ingeniería. Ya podemos establecer la matriz de relaciones (E) utilizando también una simbología adecuada (*, +, –, /) para marcar el tipo de relación entre las CI y los AC, tras lo cual se procede a desarrollar la matriz F de medidas objetivas con sus unidades, donde se comparará nuestro producto con el de los competidores. Como medidas en nuestro caso se han adoptado el número de plegados *np* que aguanta el papel, el número de pliegues *n* que hay que dar al mapa para cerrarlo, la encuadernación en espiral, que se considera como una variable binaria (*Si/No*), de igual forma que la aproximación a la simbología del modelo real, y la densidad de topónimos se mide en número de textos por cm^2 (t/cm^2) del mapa.

Pasamos ahora al característico tejado (matriz G) en el que se marca el grado de relación entre las soluciones técnicas propuestas, utilizando para ello una simbología que, en este caso, coincide con la usada en la matriz E. Finalmente, en la matriz H se incorpora una valoración de la dificultad técnica de introducir cada CI, y una valoración porcentual de la importancia y del coste de cada CI, así como el valor objetivo propuesto para nuestro producto. Como se puede observar, las mayores dificultades técnicas están en los cambios que supone la organización en cuadernillo y en la generalización de la toponimia, operaciones que también son las de mayor coste junto con el incremento que se derivará de la utilización de un papel de mayores prestaciones.

TRABAJO A REALIZAR

Supóngase que un estudio de ingeniería cartográfica pretende producir una nueva edición de un mapa turístico de un parque natural. Sin embargo, no sólo se pretende actualizar la información, sino también mejorar el producto a partir de las

consultas realizadas a una muestra heterogénea de usuarios de la edición anterior (mapa N), y de otros dos mapas de la competencia (mapas A y B), para así dotarlo en el futuro de una mayor competitividad en el mercado.

Tabla 1.11 - Voz del cliente

Atributos del cliente	Mapa N	Mapa A	Mapa B
Semiología gráfica			
Escala del mapa	2 E100k	4 E50k	3 E66,5k
Trazado de la escala gráfica	4 8 cm	3 5 cm	3 4 cm
Falta de claridad, debido a la información altimétrica del fondo	4 coropletas	3 coropletas	2 isohipsas
Grosor de las líneas de las carreteras	3 1,5 mm	5 0,5 mm	4 0.75 mm
Apreciación del límite del parque	1	1	1
Información gráfica representada			
Red Compleción de las vías de comunicación	4	2	2
Indicación del norte	No falta	No falta	Sí falta
Información kilométrica de las principales vías de comunicación	Sí falta	Sí falta	Sí falta
Información altimétrica	2 coropletas	3 coropletas y picos	4 isohipsas, collados y picos
Información de interés turístico (campings, vistas, áreas recreativas, hoteles, restaurantes...)	3	2	5
Leyenda			
La información gráfica no es fiel a la leyenda	5	5	2
Categorías de vías de comunicación insuficientes	1 2 categorías	2 3 categorías	4 4 categorías
Categorías de hidrografía insuficientes	2 2 categorías	2 2 categorías	2 2 categorías
Actualización (0 = desconocida)	3 6 años	0 ?	2 10 años
Contenido de información escrita (fauna, flora, historia, fiestas, artesanía, gastronomía...)	No	3 fauna, flora, artesanía	3 historia, fiestas, gastronomía
Presencia de otros mapas complementarios (situación, accesos, de tiempo, callejeros...)	No	2 situación	5 situación, acceso, tiempo, callejeros
Información en varios idiomas	No	No	No
Papel			
Poco resistente a la humedad	4	5	4
Se rasga y dobla	2	3	1
Se rompe por los pliegues	1	2	1
Manejabilidad			
Tamaño grande	4 formato A1	2 formato A0	2 formato A0
La leyenda no se encuentra fácilmente	5	3	2
Plegado complicado	4 8 pliegues	2 16 pliegues	2 16 pliegues
Precio			
Caro	4 4€	3 5€	1 9€

Las quejas de los usuarios respecto a los mapas N, A y B han sido estructuradas en varios apartados y valoradas entre 1 y 5 (1 = pésimo, 2 = malo, 3= regular, 4 = bueno, 5 = óptimo). La Tabla 1.11 muestra el resultado de este estudio, para cada mapa, y para cada ítem bajo cuestión presenta la valoración dada por los usuarios consultados en una encuesta. En cada una de las celdas, junto a la valoración dada, también se incluye alguna palabra o comentario breve explicativo.

Por tanto, dada la diversidad de las modificaciones que se prevén necesarias, se pide emplear el Despliegue de Función de Calidad para conocer cuáles son las ventajas y carencias que presenta el mapa N frente a la competencia y establecer cuáles serían las actuaciones a realizar para lograr su mejora. Asimismo, hágase un análisis crítico del método y de las conclusiones que se obtienen en este caso particular.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

BLOQUE 2

TAMAÑOS MUESTRALES

Para la toma de decisiones siempre es conveniente tener un conocimiento completo de la situación. Muchas veces esto no es posible, en este caso se procura tener un conocimiento que, aunque aproximado, tenga base estadística.

Un ejemplo claro son los sondeos electorales que utilizan los partidos políticos para la elaboración de sus estrategias. El conocimiento completo del apoyo que dará el electorado no es posible hasta después del acto electoral; por ello que se recurre a la realización de encuestas o muestreos previos.

En el ámbito de la producción industrial, y de la producción cartográfica en particular, ocurre lo mismo. Por ejemplo, en la aceptación de una edición de un producto como un Atlas impreso por un subcontratista, lo ideal sería conocer qué defectos tiene, si son o no aceptables todos y cada uno de los ejemplares que forman la edición. Sin embargo, si la edición es de un número elevado, la revisión completa de todos los ejemplares supondría un gran coste económico y temporal.

En otros casos, la forma de saber si un elemento de una serie o lote cumple o no con las especificaciones es sometiéndolo a un ensayo destructivo. Aquí, resulta absurdo comprobar todos los elementos pues significa destruirlos. Para dar soluciones a estos y otros muchos problemas, dentro de la Estadística existe una rama que se denomina Teoría de Muestras. El objeto de ésta no es otro que determinar el número de elementos que han de tomarse de una población para conformar una muestra o subconjunto representativo de la misma. Las propiedades de interés se determinarán en la muestra y se extenderán al completo de la población asumiendo un riesgo acotado estadísticamente.

La gran ventaja de esta forma de actuar es el ahorro económico que se obtiene al reducir los trabajos de comprobación a un número más reducido de elementos. Sin embargo, debido a que se trabaja con un planteamiento estadístico, existen riesgos.

Estos riesgos se denominan: a) riesgo del productor (probabilidad de que se le rechace un trabajo bueno) y b) riesgo del cliente o usuario (probabilidad de aceptar como bueno un trabajo que realmente es malo). A estos riesgos también se les conoce por riesgos de primera y segunda especie, respectivamente; presentes en todo test de hipótesis en estadística.

La determinación de las presiones de muestreo o tamaños muestrales se realiza bajo ciertas hipótesis estadísticas sobre la población origen, basándose en las propiedades que se desean determinar, y la precisión y riesgo que se asumen. En todo ello hay una componente económica importante y con doble vertiente: el coste de los trabajos de muestreo propiamente dichos, y el coste de una mala decisión de aceptación/rechazo.

En este bloque de casos se presentarán, a modo de ejemplo, tres casos centrados en temas cartográficos:

— Caso 6: Determinación del tamaño muestral en la verificación de la clasificación de imágenes de satélite. Se trata de determinar el número apropiado de unidades de muestreo (ya sean celdillas, agrupaciones de celdillas, polígonos o agrupaciones de polígonos) que dan significación estadística al trabajo de verificación de una clasificación o control temático de una cartografía. Para ello se consideraran las distribuciones binomial y multinomial. Un objetivo será analizar la diferencia entre los tamaños sugeridos por ambas distribuciones, modificando el nivel de confianza y el error aceptable.

— Caso 7: Muestreo de los trabajos de revisión catastral de titularidad. Aquí se presenta otro ejemplo de la necesidad de establecer un muestreo en el control de los trabajos. Además, se introduce el problema derivado del reducido tamaño de la población estadística, que obliga a introducir varias etapas en el control.

— Caso 8: Planes de muestreo para la aceptación de lotes (I). Se muestra un uso sencillo de la ISO 2859 (UNE 66020, MIL STD 105E), que se ha venido utilizando para la determinación del tamaño de la muestra en la aceptación de lotes por atributos. Así, con este caso se pretende resaltar la necesidad de efectuar una inspección de los productos recibidos de otras empresas o de departamentos de una misma empresa. El trabajo propuesto se centra en la aceptación de un lote de cartografía en función de posibles defectos en el estado del papel.

— Caso 9: Planes de muestreo para la aceptación de lotes (II). Se presenta un caso algo más complejo que el analizado en el Caso 8, pues en esta ocasión se trabaja directamente con las instancias que aparecen en una BDG.

Para facilitar el seguimiento de estos casos prácticos, en el Anexo de Estadística se pueden revisar los conocimientos relativos a las distribuciones de probabilidad que se emplearán en ellas (binomial, multinomial, normal). La obra de Congalton y Green (1998) y las normas UNE 66020-1 (ISO 2859-1) e ISO 19114 son elementos base de consulta de este bloque.

2.1. CASO 6

Determinación del tamaño muestral en la verificación de la clasificación de imágenes de satélite

ESTIMACIÓN DE TAMAÑOS MUESTRALES EN TELEDETECCIÓN

Como regla general, en los trabajos de Teledetección se recomienda tomar al menos cincuenta muestras por cada clase, y tal que el tamaño sea acorde con la resolución del sensor (Curran, 1985; Hay 1979; Stehman, 1996). No obstante, el número idóneo de muestras se puede establecer con un planteamiento estadístico, existiendo dos corrientes: por un lado la más extendida, que asume el comportamiento binomial aplicado a poblaciones finitas, y por otro aquella que considera un comportamiento multinomial.

Comportamiento binomial de la clasificación

Con el modelo binomial se considera la clasificación como un suceso en el que sólo caben dos posibilidades, la asignación correcta o incorrecta (David, 1986). Puede calcularse a partir de la función de densidad binomial por una aproximación a la normal (en caso de que el tamaño muestral n fuera menor de 60 muestras, se ha de aproximar a una distribución t de Student). En este supuesto no es necesario conocer el número de clases de la clasificación. Su uso queda justificado por proponer un número menor de muestras que bajo el supuesto de comportamiento multinomial, y por su facilidad de cálculo (Belart, 1999). En la Ec. 2.1 se muestra la expresión para calcular el tamaño (n) de la muestra según una distribución binomial:

$$n = \frac{z^2_{\alpha/2}\, pq}{\varepsilon^2} \qquad (2.1)$$

donde:

- $z_{\alpha/2}$ Valor de la función de distribución normal con nivel de confianza α ($\alpha/2$ en cada cola).
- p Probabilidad de asignación correcta y q la probabilidad de asignación incorrecta, cumpliéndose que $1 = p + q$, *y por ello* $q = 1 - p$).
- ε Nivel de error aceptable, o precisión del muestreo.

Como en la práctica se desconocen los valores de p, q, se puede suponer el caso más desfavorable, es decir $p = q = 0,5$, con lo que se tiene la expresión de la Ec. 2.2.

$$n = \frac{z^2_{\alpha/2}}{4\varepsilon^2} \qquad (2.2)$$

Por ejemplo, si se parte de un nivel de confianza del 95%, es decir, $\alpha = 0,05$ y como error se admite un ± 5%, aplicando la Ec. 2.2 se obtiene:

$$n = \frac{1,96^2}{4 \cdot 0,05^2} = 384,16 \cong 385 \ \ muestras$$

Algunos autores aseguran que este tipo de distribución no es válido para elaborar la matriz de error, puesto que en la matriz de error caben más de dos posibilidades, ya que intervienen *m* categorías y para una categoría dada hay sólo una respuesta correcta y (*m-1*) incorrectas.

Comportamiento multinomial de la clasificación

Por otro lado la corriente avalada por Tortora (1978), y utilizada por Congalton y Green (1998) en Teledetección, asocia la clasificación de la imagen a una distribución multinomial. Ésta es una distribución de probabilidad discreta, en la que un atributo posee mas de dos posibilidades, que es lo que realmente sucede en una matriz de error, dado que puede existir confusión entre clases. Por tanto, la distribución binomial con dos clases de atributos es un caso especial de la multinomial (Sokal y Rohlf, 1979).

La expresión simplificada que proporciona el número de muestras según la distribución multinomial, suponiendo que el nivel de error aceptable es igual para todas las clases, se muestra en la Ec. 2.3

$$n = \frac{\chi^2_{1,\alpha/k}\, pq}{\varepsilon^2} \qquad (2.3)$$

donde:

$\chi^2_{1,\alpha/k}$	Percentil[1] correspondiente a α/k de una distribución Chi cuadrado con 1 grado de libertad, con el número de clases k y el nivel de significación α..
p	Probabilidad de asignación correcta y q la probabilidad de asignación incorrecta en una clase determinada.
ε	Nivel de error aceptable o precisión del muestreo.

Como se puede observar, la aproximación multinomial y la binomial tienen expresiones similares, con la diferencia de la utilización de la distribución Chi cuadrado

[1] Debe advertirse que según el criterio de asignación del percentil a la cola o al grueso de la distribución, y en función de la tabla de la χ^2 que se utilice, el percentil puede denotarse como α/k o como $1-\alpha/k$.

(multinomial) frente a la Normal (binomial). Como en la práctica se desconocen los valores de p, q, se supone el caso más desfavorable, es decir $p = q = 0,5$, con lo que se tiene la expresión de la Ec. 2.4. Así, la formulación multinomial supone la necesidad de muestreos sensiblemente mayores (>30%) respecto a la binomial, como se muestra en el resumen numérico recogido en la Tabla 2.1.

$$n = \frac{\chi^2_{1,\alpha/k}}{4\varepsilon^2} \qquad (2.4)$$

Tabla 2.1 - Cociente multinomial/binomial entre los tamaños de muestreo

α	Número de clases							
	3	4	5	6	7	8	9	10
0,010	1,30	1,38	1,44	1,48	1,54	1,56	1,60	1,63
0,025	1,39	1,49	1,57	1,63	1,69	1,74	1,78	1,82
0,050	1,49	1,62	1,73	1,81	1,89	1,94	2,00	2,05
0,075	1,58	1,74	1,87	1,97	2,05	2,13	2,20	2,26
0,100	1,68	1,86	2,00	2,12	2,22	2,31	2,38	2,45

Por ejemplo, si partimos de un nivel de confianza del 95%, es decir, $\alpha = 0,05$, como error ε se admite un ± 5%, y el número de clases es 10, aplicando la Ec. 2.4 se obtiene:

$$n = \frac{7,87943}{4 \cdot 0,05^2} = 787,94 \cong 788 \ \ muestras$$

TRABAJO A REALIZAR

Supuesta la necesidad de efectuar un trabajo de verificación de una cartografía de usos del suelo con 10 clases, realizada bien por fotointerpretación o por clasificación de una imagen de satélite, se pretende conocer cuál es el tamaño muestral adecuado para construir la matriz de confusión. Dado que el coste del muestreo aumenta con el número de muestras, se quiere estudiar su evolución en función del porcentaje de error aceptable y del nivel de confianza.

Para ello se pide hallar el tamaño muestral según las distribuciones binomial y multinomial, combinando diferentes valores de error aceptable y de nivel de confianza, presentando los resultados en una tabla, tomando como ejemplo la Tabla 2.2.

Tabla 2.2 - Ejemplo de tabla para la presentación de las estimaciones de tamaño de muestreo

	error aceptable (ε)					
	1%		2%		...	
α	Binomial	Multinomial	Binomial	Multinomial	Binomial	Multinomial
1%						
5%						
...						

Los resultados obtenidos han de ser sometidos a un análisis exhaustivo para poder extraer unas conclusiones. Para ello se sugiere comenzar por elaborar varias gráficas que relacionen las cuatro variables (tamaño muestral, error, nivel de confianza y distribución), para observar cuáles son los valores máximos y mínimos obtenidos, y las tendencias al modificar el error o el nivel de confianza, y demás factores intervinientes.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

2.2. CASO 7

Muestreo de los trabajos de revisión catastral de titularidad

NECESIDAD DEL MUESTREO DE LOS TRABAJOS CATASTRALES

En los últimos años, los trabajos destinados a la revisión del Catastro son encargados a empresas de servicios, por lo que es imprescindible proceder a realizar un control de calidad que permita valorarlos y contrastarlos con la información aportada por la realidad, y así decidir racionalmente si procede continuar o no con las contratas.

En un término municipal pueden existir miles de parcelas, por lo que no se controla toda la información, sino que se aplica una técnica de muestreo, con el objetivo doble de obtener en el control una precisión elevada pero al menor coste posible. Los errores existentes pueden ser los derivados de la titularidad de la parcela o de su descripción (ver ejemplo de certificación catastral de urbana en la Figura 2.1). Los primeros tienen un coste menor, puesto que se verifican por correo o telefónicamente, mientras los segundos implican el trabajo de campo de los Técnicos.

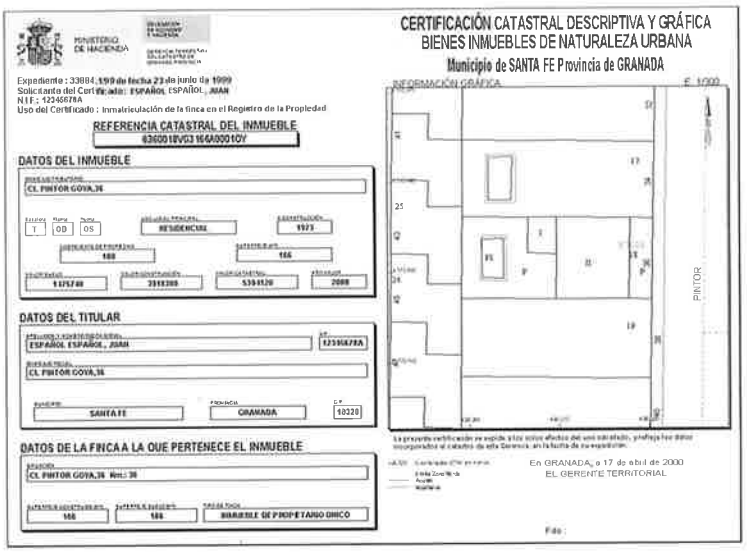


Figura 2.1 - Ejemplo de certificación catastral descriptiva y gráfica de una parcela urbana

MÉTODO PARA LA REVISIÓN CATASTRAL DE TITULARIDAD

Estadísticamente, cada término municipal o polígono catastral, según se considere, constituye una población diferente, que deberá muestrearse por separado mediante un muestreo aleatorio simple. Los tamaños poblacionales que se manejan son lo suficientemente pequeños como para que tenga interés realizar la corrección de los tamaños muestrales por población finita, por lo que la expresión que proporciona el tamaño muestral n es la derivada de una distribución binomial, aplicando la referida corrección (Ec. 2.5):

$$n = \frac{m}{1 + \frac{m}{N}} \qquad m = \frac{z^2_{\alpha/2}\, pq}{\varepsilon^2} \tag{2.5}$$

donde:

- $z_{\alpha/2}$ Valor de la distribución normal con nivel de significación deseado α ($\alpha/2$ en cada cola).
- p Probabilidad de asignación correcta y q la probabilidad de asignación incorrecta.
- ε Nivel de error aceptable o precisión del muestreo.
- m Tamaño de la muestra, antes de la corrección por población finita.
- N Tamaño de la población (número de parcelas del término municipal).

En la Ec. 2.5, N es conocido y α suele ser de 0,05. Sin embargo, los valores de p y ε deben ser fijados siguiendo algún criterio. Normalmente se suele utilizar un valor pequeño para ε, por ejemplo del ± 2,5% y el peor caso para p, es decir $p = q = 0,5$. Estos valores son adecuados para poblaciones muy grandes, pero actuando de esta forma pueden aparecer problemas en términos municipales con un número de parcelas poco elevado, ya que la potencia del muestreo es tanto mayor cuanto mayor es la población. Por ejemplo (Coduras y Costa, 1990), para un polígono con 500 parcelas, un nivel de significación del 95%, un nivel de error del ± 2,5% y una probabilidad de asignación correcta de 0,5 se obtiene $n = 377$ muestras, es decir, el 75% de la población.

Por ese motivo, en casos como éste, es adecuada la utilización de un muestreo piloto previo al muestreo, tal que ayude a concretar los valores del diseño de un segundo muestreo, ya definitivo. Es decir, se parte de un muestreo con más de una etapa, tal que en la fase piloto se trata de averiguar si el porcentaje de errores correspondientes a un término municipal, o polígono catastral concreto, supera o no a un valor crítico β. Los sucesivos pasos del método polietápico pueden ser los siguientes:

— Obtener el tamaño de muestra piloto a partir de la suposición de un nivel bajo de precisión (por ejemplo, $\varepsilon_0 = \pm 5\%$) y un valor moderado de p (por ejemplo p_0

$= 0,9$). Así se obtendrá un valor n_1 de parcelas. Por tanto el número de propietarios a analizar es $k_1 = n_1 / h$, siendo h el promedio de parcelas por propietario.

— Tras contrastar la titularidad de las n_{r1} parcelas reales verificadas (que seguramente no coincida con n_1 porque cada propietario tendrá un número de parcelas distinto de h) de los k_1 propietarios elegidos al azar, se obtiene el valor p_1 y q_1 ($q_1 = 1 - p_1$). Conocido este valor, se despeja ε_1 de la Ec. 2.5.

— Entonces pueden tener lugar tres situaciones diferentes:

Si $(\beta > q_1 + \varepsilon_1)$, entonces el trabajo de la empresa es satisfactorio.

Si $(\beta < q_1 - \varepsilon_1)$, el trabajo de la empresa no es satisfactorio.

Si $(q_1 - \varepsilon_1 < \beta < q_1 + \varepsilon_1)$, entonces deberemos precisar en mayor medida nuestra estimación de q, estrechando el intervalo, es decir, reduciendo ε_1 hasta ε_2, tal que ε_2 haga salir a β del intervalo. Se calculará el nuevo tamaño de la muestra, n_2, que será mayor que n_1, e implicará la necesidad de verificar algunos propietarios más, volviendo a repetir el proceso.

TRABAJO A REALIZAR

Se quiere verificar el trabajo entregado por una empresa para un polígono catastral de 684 parcelas, que se sabe pertenecen a 237 propietarios. Para ello se efectuará un muestreo mediante el método expuesto. Se pide determinar si el trabajo de la empresa es válido según las siguientes hipótesis y datos:

— Los valores supuestos, en primera instancia, para p y ε son: $p_0 = 0,9$ y $\varepsilon_0 = 5\%$.

— El valor crítico β se fijará en el 3%. El nivel de confianza será del 95%.

— Como resultados de un primer muestreo telefónico considerar los siguientes resultados: $q_1 = 6\%$, $n_{r1} = 122$.

— Si hace falta un segundo muestreo, considerar que $q_2 = 7\%$, y que de la casualidad de que $n_{r2} = n_2$.

Además, compárese el tamaño final de la muestra de parcelas verificadas con el tamaño que vendría dado por el caso en el que ε se considera del ± 2,5% y el peor caso para p, es decir $p = q = 0,5$.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

2.3. CASO 8

Planes de muestreo para la aceptación de lotes (I)

LA INSPECCIÓN POR PLANES DE MUESTREO

La inspección en la recepción es un control básico para que no se transfiera la mala calidad entre empresas o departamentos. Las inspecciones de recepción de productos se pueden realizar tanto sobre productos procedentes de proveedores externos como de departamentos de la propia empresa (idea del cliente interno).

El control de la recepción de productos se realiza para asegurar que aquellos poseen la calidad suficiente. De esta forma, mediante el análisis de las propiedades en un número de elementos, se decide sobre la aceptación o no del conjunto de artículos que conforman el lote. La base que se considera para decidir la aceptación o no de un lote de tamaño N_{lote} son las propiedades determinadas sobre n artículos inspeccionados, tal que $n < N_{lote}$. Estas propiedades pueden ser del tipo cuantitativo (variables) o cualitativo (atributos). Dado que la aceptación o rechazo se realiza sobre un muestreo estadístico, existe riesgo, tanto para el proveedor como para el cliente.

Los planes de muestreo lote a lote consisten en estimar la calidad de un lote a partir de la inspección de una muestra obtenida de él. El plan se define en función de la magnitud del lote[1] y la calidad considerada. Los factores son (AEC, 1990): el tamaño de la muestra n, el número máximo de unidades defectuosas[2] que se admiten (número de aceptación, Ac), y en otras ocasiones el número de rechazo Re, y finalmente el método de muestreo. El tamaño de la muestra n dependerá del máximo porcentaje de defectuosos admisible, de los riesgos que se toleran, de la garantía que ofrece el suministrador y del tamaño del lote, entre otros.

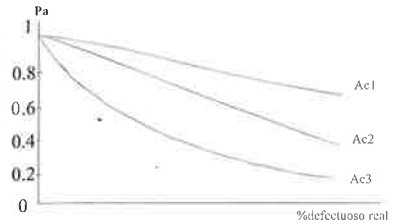
La probabilidad de aceptación Pa de un lote que contiene un determinado porcentaje de unidades defectuosas, y al que se le aplica un plan de muestreo definido por $\{n, Ac\}$, es el valor que se refiere, en término medio, a cuántas de cada 100 veces que se realizaran muestreos del tipo $\{n, Ac\}$ resultaría aceptado el lote.

Mediante un planteamiento analítico (ver Hansen, 1990) en el que se parte del tamaño del lote y del nivel de defectuoso real existente, se puede obtener la curva

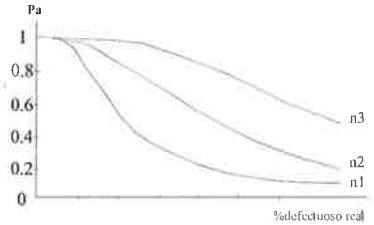
1 Al lote se le imponen fundamentalmente dos características básicas, por un lado que sea homogéneo y, por otro, que tenga un tamaño adecuado dado que las muestras a tomar en lotes grandes son, porcentualmente, más pequeñas que para lotes de menor tamaño.

2 Un defecto es cualquier disconformidad de una unidad de producto frente a los requisitos o exigencias establecidos.

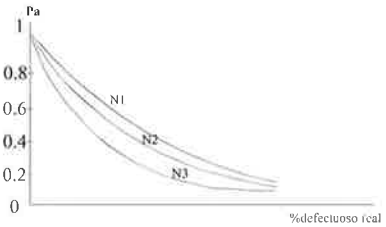
característica u operativa de un plan de muestreo sobre dicho lote. Esta curva indica la probabilidad de aceptación, *Pa*, del lote en función del tamaño de muestra, el *Ac* y porcentaje de defectuosos que realmente tiene. La eficacia de cualquier plan de muestreo se define mediante esta curva cuya forma puede variar como se aprecia en la Figura 2.2 para el caso de distintos valores de aceptación, de tamaño de muestra y de población. La curva tiene menos capacidad discriminante cuanto más plana es.



a) Ac1 > Ac2 > Ac3



b) n1 > n2 > n3



c) N1 < N2 < N3

Figura 2.2 - Comportamiento de la curva característica frente: a) número de aceptación *Ac*, b) tamaños de muestra *n*, c) tamaño del lote N_{lote}

Evidentemente, el porcentaje real de defectos que tiene un lote es desconocido cuando llega para su inspección; no obstante, la curva característica nos indica, como ya se ha apuntado, la probabilidad de aceptarlo, *Pa*, suponiendo que realmente tuviera un porcentaje determinado de defectos. Igualmente, dicha curva da información sobre la probabilidad de rechazo, *Pre*, o número de veces, en termino medio, que se rechazará un lote con un determinado porcentaje defectuoso, por cada cien veces que se presente a inspección. Indudablemente se cumple que $Pa + Pre = 1$.

Dado que cada lote presentará un número de defectuosos distinto y que, por lo general, se pueden recibir numerosos lotes de un mismo suministrador, se puede definir el Nivel de Calidad Aceptable (*NCA*) de un producto como el máximo número de defectuosos que admite el cliente como promedio de los porcentajes de defectuosos que aparecen en los lotes. Esto supone una cierta calidad media de los lotes por lo que se pueden aceptar lotes con porcentajes superiores al valor estipulado siempre y cuando se compensen con otros de menor porcentaje de defectuosos. Como es lógico pensar, el proveedor tendrá que producir con un nivel de defectos menor o, a lo sumo, igual al *NCA* para que sus lotes sean aceptados con alta probabilidad.

Dado que el *NCA* es un valor promedio, al objeto de protegernos respecto a porcentajes de defectuosos elevados, se establece el concepto de calidad límite *CL*. Con él nos referimos a un porcentaje máximo de defectuoso que se puede admitir en un lote aislado, de tal manera que no se admitan lotes con calidad inferior al mismo.

La curva operativa permite determinar el riesgo de primera especie (α) o del productor (proveedor o vendedor) y el riesgo de segunda especie (β) o del usuario (cliente o comprador). El primero hace referencia a la probabilidad de rechazar en el muestreo un lote que debería aceptarse ya que realmente tienen un porcentaje de defectuoso menor o igual al *NCA* pero que en el muestreo ha dado valores altos de defectos. Esta probabilidad se suele mantener en niveles del 5%. Por tanto, a partir de la abscisa con valor *NCA*, se obtiene el punto sobre la curva con coordenadas (*NCA*, $1-\alpha$) que indicará:

- Probabilidad aceptación = $1 - \alpha$
- Probabilidad de rechazo = α

El riesgo del usuario se corresponde con la probabilidad de aceptar lotes con un porcentaje de defectuosos igual a *CL*, este valor se suele tomar cercano al 10%. En este caso el punto de la curva es el (*CL*, β):

- Probabilidad aceptación = β

PROCEDIMIENTO DE MUESTREO (UNE 66020-1)

La UNE 66020-1:2001 (ISO 2859-1:1999) está basada en la Norma MIL STD 105 que, en sus inicios, fue desarrollada por el ejército de los Estados Unidos de Norte América durante la Segunda Guerra Mundial y ha sido ampliamente aplicada en el ámbito industrial desde ese momento. Es una norma bastante amplia y que se basa en el uso del *NCA*.

En la norma se establecen diversos niveles de uso o aplicación, pero lo normal es utilizar los denominados niveles de inspección generales, dejando los especiales para casos muy excepcionales (p.e. falta de recursos). Dentro del nivel general existen tres posibilidades codificadas como I, II y III. El normal o aconsejado para un producto nuevo es el II, recurriendo al nivel III, o riguroso, cuando hay evidencias de falta de calidad. El nivel I también se denomina nivel reducido de inspección, y se aplica, como se ha comentado, cuando hay evidencias del buen comportamiento del suministro. Los criterios de paso entre niveles se recogen en la Tabla 2.3.

Tabla 2.3 - Criterios de paso entre niveles de inspección de la Norma UNE 66020-1		
de	**a**	**Criterio**
I	II	1 rechazo
II	III	2 de 5 consecutivos rechazados
III	II	5 consecutivos sin rechazo
II	I	10 consecutivos sin rechazo

El uso de la norma se realiza mediante tablas. En el caso del muestreo simple se utilizan dos, una primera tabla (Tabla 2.4) que, en función de los niveles de inspección y del tamaño del lote da el código de entrada en la segunda tabla (Tabla 2.5), y esta última en la que con la clave anterior se obtienen el tamaño de la muestra y los valores de los números de aceptación (máximo número de defectos admitidos en la muestra) y el de rechazo (número de defectos que si se alcanza suspende la inspección y lleva a rechazar el lote), para cada *NCA*.

Por ejemplo, para un muestreo propio de un nivel de inspección normal (II) sobre un lote de 250 unidades, de la Tabla 2.4 se deduce que el código asignado es el G. Entrando con esta clave, la Tabla 2.5 indica que se debe tomar una muestra con 32 elementos y, dependiendo del *NCA* que se desee, dicha tabla indica los valores de aceptación y rechazo (por ejemplo, para *NCA = 1,5*, se obtienen los valores de criterio de aceptación = 1, y de rechazo = 2).

Tabla 2.4 - Códigos de tamaño de muestra según la UNE 66020-1								
Tamaño		**Niveles de inspección especiales**				**Niveles de inspección generales**		
		S1	**S2**	**S3**	**S4**	**I**	**II**	**III**
2	8	A	A	A	A	A	A	B
9	15	A	A	A	A	A	B	C
16	25	A	A	B	B	B	C	D
26	50	A	B	B	C	C	D	E
51	90	B	B	C	C	C	E	F
91	150	B	B	C	D	D	F	G
151	280	B	C	D	E	E	G	H
281	500	B	C	D	E	F	H	J
501	1200	C	C	E	F	G	J	K
1201	3200	C	D	E	G	H	K	L
3201	10000	C	D	F	G	J	L	M
10001	35000	C	D	F	H	K	M	N
35001	150000	D	E	G	J	L	N	P
150001	500000	D	E	G	J	M	P	Q
más de	500001	D	E	E	J	N	Q	R

Tabla 2.5 - Plan de muestreo simple para inspección normal

Código	Tamaño muestra	Niveles de calidad aceptable (NCA)									
		0,15	0,25	0,4	0,65	1,0	1,5	2,5	4,0	6,5	10
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓
B	3	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓
C	5	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2
D	8	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3
E	13	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4
F	20	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6
G	32	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8
H	50	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11
J	80	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15
K	125	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22
L	200	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑
M	315	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑
N	500	2 3	3 4	5 6	7 8	10 11	14 15	21 2	⇑	⇑	⇑
P	800	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑
Q	1250	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑
R	2000	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑

⇓ = Utilizar el primer plan de muestreo situado debajo de las flechas. Si el tamaño de la muestra es igual o superior al tamaño del lote, efectuar una inspección al 100%.

⇑ = Utilizar el primer plan de muestreo situado encima de las flechas.

Ac = Número de aceptación.

Re = Número de rechazo.

TRABAJO A REALIZAR

Supóngase que se ha recibido, de imprenta, un lote de cartografía (correspondiente a un callejero de una ciudad) en soporte papel de 7500 unidades (N = 7500). A la hora de controlar la recepción del producto, se decide atender a diferentes factores del estado del papel: a) correcto cortado y plegado de la hoja, b) inexistencia de roturas o rasgaduras, c) inexistencia de arrugas y pliegues.

Se pide calcular el tamaño de la muestra (n) y el número de aceptación (Ac) según el procedimiento de muestreo UNE 66020. Las condiciones contractuales figuran en la Tabla 2.6. En ella se ha de observar que no se establece el mismo nivel de exigencia para los distintos defectos.

Tabla 2.6 - Condiciones contractuales de los diferentes procedimientos de muestreo		
DEFECTOS	**Procedimiento UNE 66020**	
	Nivel de inspección	**NCA**
Deficiencias en el plegado y cortado	II	0,25%
Papel rajado o rasgado	II	0,40%
Papel arrugado o con pliegues.	II	0,65%

Como segundo ejercicio práctico se sugiere completar las columnas 4ª a 8ª de Tabla 2.7, al considerar que se aplica la UNE 66020 sobre los datos que se presentan en las columnas 2ª y 3ª. En este caso el NCA establecido es del 1% para todos los lotes. El nivel de inspección de partida es I, pero éste puede verse modificado según la aceptación o rechazo de un cierto número de lotes previos, de acuerdo a la Tabla 2.3.

Tabla 2.7.- Secuencia de lotes y número de rechazos encontrados							
1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª
Lote	**Tamaño del lote**	**No conformidades**	**Nivel de Inspección**	**Código de tamaño**	**Tamaño de la muestra**	**Nº de rechazo**	**¿Lote rechazado?**
1	1000	2	I				
2	750	3					
3	600	2					
4	400	4					
5	500	5					
6	250	3					
7	250	1					
8	200	1					
9	100	0					
10	100	1					
11	100	0					
12	100	0					

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

2.4. CASO 9

Planes de muestreo para la aceptación de lotes (II)

INSPECCIÓN DE UNA BDG POR PLANES DE MUESTREO

Este caso práctico muestra un ejemplo de aplicación de los planes de muestreo al caso del control de calidad de los propios elementos de una BDG. Los fundamentos son pues los mismos que los del caso práctico anterior, la cual conviene desarrollar previamente a ésta. Por otro lado, el Caso 29, dedicado al control de la compleción y de la calidad temática, muestra también un contenido muy relacionado con el de ésta.

Como hipótesis inicial se va a considerar que la BDG de interés es una serie cartográfica digital, que se produce por un mismo subcontratista, y en un intervalo de tiempo acotado en el que no cambian los métodos de ejecución.

La primera cuestión a aclarar es qué se considera como población y, por tanto su tamaño. Se podría considerar que si la serie cartográfica está formada por hojas, cada una de ellas es un posible elemento de control y que, por tanto, el tamaño de la población y lote es el número de hojas de la serie. No obstante, este razonamiento directo presenta algunos inconvenientes:

- Se tendrán lotes relativamente pequeños, lo que lleva a una mayor necesidad de muestreo porcentual.
- Los elementos de control serán extensiones geográficas muy grandes, con los problemas económicos que ello conlleva.

Una alternativa usual y propuesta por varias fuentes (Newby, 1992, ISO 19114) es considerar que el lote está formado por todos los cuadrados de 1 × 1 Km. que abarca el área geográfica de interés. De esta forma aumenta el número de elementos a considerar, reduciéndose el muestreo porcentual, tal que se concretan marcos de control de un tamaño mucho más manejables y cuya distribución en el espacio muestrea, por lo general, con mayor eficacia las posibles variaciones del territorio.

La idea base es pues determinar el número de elementos que formarán la muestra (marcos de control). Esto se podrá hacer de una manera totalmente aleatoria, o con cierta estratificación, según el caso. La correlación de los defectos también debe guiar el diseño de este muestreo, de tal forma que si los elementos están muy correlados interesa separar los marcos en el muestreo, hacer más pequeños los marcos

de control, e incluso tratar elementos correlados como un único elemento, pues todo ello permite hacer el muestreo más eficiente (ISO 19114).

Una vez determinados el número y posición de los marcos de control, se va a campo y se procede a la contabilización de los defectos de la cartografía respecto al terreno (todos los marcos de muestreo se verificarán en campo). Las no conformidades serán de dos tipos: omisiones y comisiones. En el primer caso los elementos que existen en el terreno no han sido capturados y en el segundo, elementos que aparecen en la BDG realmente no existen en el terreno. Las pruebas para validar la exactitud temática se pueden realizar sobre este mismo material. En este caso se inspeccionarán los atributos de los elementos. Como regla general podrán existir omisiones de atributos, pues está presente el objeto pero no su atributo, pero también se darán casos de errores en los atributos soportados por los objetos de interés.

En las estadísticas e índices, por ejemplo, el porcentaje de errores de omisión/comisión, los valores que deberían tomarse como base de los cálculos son siempre los relativos al mundo real, pero estos no se conocen, por ello se aproximan por los correspondientes a los existentes en la BDG.

Otra cuestión de especial cuidado es el de los elementos que están en los bordes de los marcos de muestreo.

No obstante, no todos los fallos deben pesar por igual. Se debe pues establecer una ponderación que considere la importancia de los mismos (por ejemplo, críticos, mayores y menores) y NCA distintos para cada uno de ellos. Otra posibilidad es ponderarlos en función de la abundancia de los mismos, por ejemplo, peso 1 para aquellos elementos muy numerosos que no se van a controlar, o a los que no se les ha asignado NCA; un peso de 2 ó 3 puede ser adecuado para aquellos cuyos NCA = 4 ó 15, y pesos de 3 para aquellos que son realmente escasos en el lote (ISO, 19114).

MUESTREO EN LOS MARCOS DE CONTROL

Cada uno de los marcos de control deberá ser identificado y visitado en campo. El resultado de la inspección en campo debe quedar recogido en un formulario o ficha con unos contenidos similares a los que se muestran en la Tabla 2.8. Cada producto, en función de sus especificaciones, incluirá mayor o menor número de tipos de elementos.

Para el correcto desarrollo se debe ir al campo con una copia de la BDG y que se corresponda con el área propia del marco de control, y con los datos cuyo cálculo sea directo sobre dicha muestra.

Tabla 2.8.- Ejemplo de Ficha de control de cada marco de muestreo

BDG:	MAPA TOPOGRÁFICO DE ANDALUCÍA
MARCO:	MTA10-MC1048-1
FECHA:	Abril 2003
TECNICO	López López, José

Tipo de elemento	Número de elementos	Compleción Omisión/Comisión Nº Errores	Exactitud Temática Nº Errores
Vía comunicación TIPO 1	4		
Vía comunicación TIPO 2	6	1 / -	
Vía comunicación TIPO 3	14	2 / 1	
Ferrocarril	4		1
Línea Ala Tensión 10-50 KV	3		
Línea Alta Tensión > 50 KV	5		
Transformador	1		
Construcción	30	- / 3	2
Aprisco	12	1 / 1	
...			

INFORME DE CALIDAD SOBRE LA INSPECCIÓN DEL LOTE

Cuando se ha realizado la inspección de cada uno de los marcos de control se procede a hacer un informe general sobre todo el lote y a determinar si éste pasa o no los NCA establecidos. La Tabla 2.9 presenta un ejemplo de este tipo de informe.

Tabla 2.9.- Ejemplo de Informe sobre los resultados de la inspección sobre un Lote

BDG:	MAPA TOPOGRÁFICO DE ANDALUCÍA
MARCO:	MTA10-MC1048-1, MTA10-MC1046-1, MTA10-MC1047-1, MTA10-MC1047-2,
FECHA:	Abril 2003
TECNICO:	López López, José

1ª Tipo de Elemento	2ª Atributo	3ª Tamaño Lote	4ª Tamaño muestra	5ª Tamaño a nivel de inspección 1	6ª Nivel de Inspección	7ª NCA	8ª Valor de aceptación	9ª Valor de Aceptación para NCA 1*	10ª Compleción (omisiones comisiones) Nº Errores	11ª Corrección clasificación
Vía comunicación		4712	184	80	I	4	10	7	2	-
	Clases 1, 2 y 3					4	10	7	2	
	Matricula					1	0	0		
	Estado					1	0	0		
	Gálibo mínimo					1	0	0		
	Peso máximo					0	184	80		
Ferrocarril		1880	52	50	I	15	14	14	1	1
	Clase					15	14	14	2	
	Matrícula					0	52	50		
	Electrificación					0	52	50		
Línea alta tensión		542	18	32	S2	4	1	3		
	Clases									
Transformador										
Construcción		6447	222	80	I	4	14	7	4	4
	Uso					0	222	80	2	
	Estado					0	222	80		
Aprisco		379	10	20	S3	4	1	2		
	Estado					0	10	20		

Como se observa en dicha Tabla 2.9, una vez se ha realizado el trabajo de campo, sobre los marcos de control, y dado que se dispone de toda la BDG como producto digital, se puede contabilizar el número total de instancias o elementos que forman la población del lote (columna 3ª), y puesto que se ha extraído una muestra a partir de los marcos de control, se puede establecer el tamaño real de la muestra para cada elemento que, como es lógico pensar, en el caso más general será distinto para todos ellos (ver columna 4ª).

La columna 5ª va mostrando los tamaños de muestra que indica la Norma UNE 66020 (ISO 2859) para un nivel general de inspección de Tipo I. Pero como puede ocurrir que el tamaño real de la muestra (columna 4ª) sea inferior o superior al que marca la columna 5ª, esto supone un nivel de inspección efectivo distinto para cada tipo de elemento. Este nivel de inspección efectivo es el que se muestra en la columna 6ª. Ejemplo de esto último es el caso de los apriscos. En los marcos de control sólo se han encontrado 10 elementos de este tipo, mientras que por el tamaño de lote (379), con un nivel general de inspección de Tipo I, se deberían muestrear 20 elementos. Dado que sólo se dispone de 10 de ellos, esto equivale a un código de tamaño de muestra D en la tabla que la norma establece para los planes de muestreo (ver Tabla 2.5, Caso nº 8), por lo que entrando en la tabla de códigos para tamaño de muestra (ver Tabla 2.4, Caso nº 8), para el tamaño de lote de 379, se obtiene que el nivel de inspección real que se está aplicando es el especial denominado S3.

La columna 7ª marca el NCA que se desea para cada tipo de elemento, y la columna 8ª el valor de aceptación para el número de elementos de la muestra. La columna 9ª indica el valor de aceptación para la NCA dada si el tamaño fuera el que se indica en la columna 5ª.

Las dos últimas columnas muestran el total de errores.

Indudablemente, a la Tabla 2.9 anterior se le puede añadir una nueva columna que indique si pasa o no cada uno de los controles realizados sobre cada tipo de elemento.

AGREGACIÓN DE LOS RESULTADOS DE LA CALIDAD

Una vez realizado el control sobre la muestra, se puede presentar un resultado único que informe sobre la calidad final del producto, esto es lo que se denomina ADQR (*Agregated data quality result*). Hay diversas formas de obtener valores para un ADQR en los que se pueden combinar resultados procedentes de diversos alcances, elementos, subelementos, etc. Las forma más comunes son:

— % de conforme/no conforme: En este caso se consideran resultados booleanos, la calidad agregada se determina de la siguiente forma:

$$ADQR = V_1 \times V_2 \times V_3 \times ..... V_n$$

donde:

n	Número de medidas.
Vi	variable booleana Vi.

Está claro que si se obtiene un $ADQR = 1$, es por que todos los elementos bajo consideración cumplen. Sin $ADQR = 0$, será porque, al menos, uno de ellos no cumple. Este método es muy exigente y no informa de la magnitud ni dónde se produce la no conformidad.

— Conforme/no conforme promediado: Al igual que en el caso anterior se consideran variables booleanas que indican si hay conformidad o no de un control para un elemento determinado, pero además se utiliza un conjunto de pesos, que suman la unidad, y cuyos valores están en el intervalo $[0, 1]$, la formulación es pues:

$$ADQR = w_1 V_1 + w_2 V_2 + w_3 V_3 + ... + w_n V_n$$

donde:

n	Número de medidas.
Wi	Peso de la variable Vi.
Vi	variable booleana Vi.

El peso que se da a cada variable debe ajustarse, no existiendo normas al respecto. La ventaja de la medida que se obtiene es que es directamente interpretable como cercanía al cumplimiento total, pero no da un valor que marque dónde se producen las conformidades o no conformidades.

— Valores Máximo/Mínimo: Se toma una de las dos alternativas siguientes:

$$ADQR = MAX\ (V_i;\ i = 1, ...n\ )$$
$$ADQR = MIN\ (V_i;\ i = 1, ...n\ )$$

donde:

n	Número de medidas.
Vi	variable booleana Vi.

Se obtiene una magnitud que puede ser interpretada como la distancia o proximidad de que el producto cumpla con las especificaciones. Si se informa adecuadamente, se puede saber en dónde ocurre la mayor/menor disconformidad. No obstante se pierde información sobre resultados del control sobre otros elementos.

— Subconjunto de medidas: En el caso de tomar un número de medidas que excede al que realmente se necesita para cubrir las especificaciones o propósito

del producto, se puede tomar un subconjunto de estas y aplicarles cualquier técnica de las comentadas, o cualquier otra.

TRABAJO A REALIZAR

Para mayor comodidad, el trabajo a realizar se va a realizar sobre dos BDG. La primera de ellas se corresponde con lo que podemos considerar es la verdad nominal y, por tanto, hace las veces de lo que existe realmente en el terreno. La segunda se corresponde con lo que sería un trabajo cartográfico que hay que controlar y cuya ejecución se debería haber realizado según ciertas especificaciones.

El trabajo a desarrollar consiste pues en definir un plan de muestreo completo, establecer los marcos de control, preparar la ficha de control, la tabla resumen y un índice general que indique el nivel de calidad alcanzado en los trabajos.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar el material necesario para este caso.

BLOQUE 3

CONTROL DE PROCESOS

Un proceso es la transformación de unas entradas en unas salidas, bienes o servicios, con una calidad técnica, todo ello mediante la aplicación de un conjunto de elementos materiales (materias primas, equipos, entorno de trabajo), inmateriales (métodos de trabajo, organización, etc.) y humanos.

La determinación de lo que es un proceso puede hacerse a distintos niveles, por ejemplo, un proceso puede ser la restitución fotogramétrica o, dentro de ella, la orientación interna. El determinar qué es un proceso para nosotros dependerá de los intereses, la perspectiva de gestión y de lo crítica que sea la actividad.

En los inicios de la producción industrial los productos sólo se controlaban al final de todo el proceso, no obstante, en la actualidad esto ha cambiado bastante. La idea es que cuanto antes se detecten los productos no conformes tanto menor coste tienen las no calidades para la empresa. Por ello se cuida el aseguramiento y los controles a los niveles más bajos posibles, utilizando la información generada en el control para reconducir la actividad a niveles de calidad adecuados. Bajo esta idea el papel de propietario del proceso (Figura 3.1) es fundamental dado que será él quien, teniendo en cuenta las medidas objetivas realizadas sobre los resultados del proceso, los objetivos del propio proceso, y lo establecido en las normas, procedimientos y demás especificaciones, adopte las acciones correspondientes.

Un proceso queda especificado cuando existe una descripción detallada del mismo, cuando se identifican de sus factores críticos y su comportamiento estadístico.

En este bloque se presentarán algunas herramientas que permiten conocer y caracterizar ciertos procesos. Existen otras muchas herramientas que podrían ser aplicadas, por lo que los casos propuestos no deben ser entendidos como únicos casos de posible aplicación; todo lo contrario, son ejemplos particulares de un amplio abanico de posibilidades:

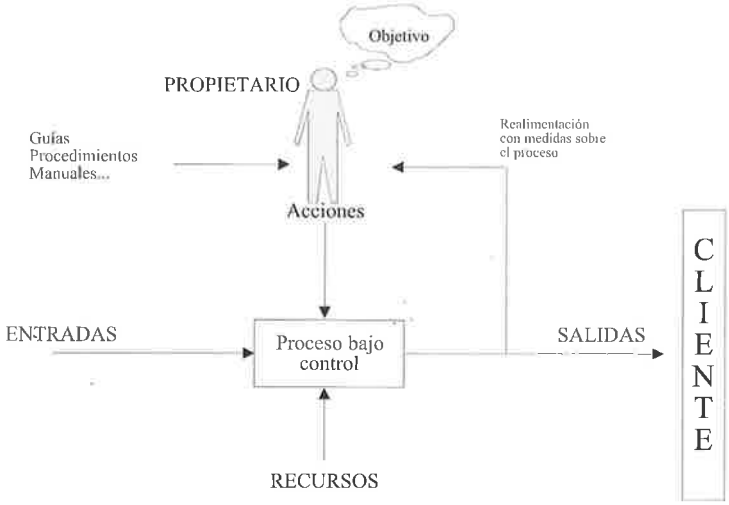


Figura 3.1 - Propietario de un proceso

— Caso 10: Análisis de varianza. Presenta la utilidad de esta herramienta en el control de los procesos cartográficos. Se aplica sobre varias series de datos, sometidas a tratamientos diferentes para verificar si existen, o no, diferencias significativas entre ellas, evidenciando que no ha existido homogeneidad en el proceso productivo.

— Caso 11: Análisis cluster (agrupamiento). Muestra una herramienta estadística que permite separar un conjunto de datos en varios grupos. Este proceso es de interés cuando se sospecha que no hay homogeneidad en alguna de las variables que se pretenden controlar en un lote de un producto.

— Caso 12: Análisis de series temporales en la digitalización gráfica manual. Se incluye un ejemplo de comparación de dos conjuntos de datos procedentes de la digitalización gráfica manual, mediante tablero. Este análisis se efectúa en cuanto a su fidelidad respecto a la información original. Para ello se empleará un modelo autorregresivo AR(1), además de parámetros estadísticos simples, como la media, desviación típica, etc.

— Caso 13: Empleo de cuantiles en la determinación del tamaño de celda. Presenta un método, de base estadística, para la elección del tamaño de celda en la rasterización, a partir de la acotación del error que se desea cometer.

— Caso 14: Evaluación de la calidad geométrica de un escáner. Guía de manera práctica cómo evaluar, antes de su empleo, la calidad geométrica que proporciona un dispositivo de este tipo. Para este control se presenta un proceso

sencillo que se puede aplicar sobre cualquier tipo de escáner y que permite realizar un rápido análisis.

— Caso 15: Control de una impresión sobre papel. Partiendo de la necesidad de controlar los trabajos de reproducción en el ámbito cartográfico, se presenta un ejemplo real de un programa de aseguramiento de la calidad en la impresión.

Son numerosos los manuales estadísticos que abordan estos temas, existiendo incluso referencias especializadas para algunas de estas herramientas. Pensando en una mayor comodidad del lector, se ha incluido un Anexo de Estadística en el que es posible revisar los aspectos más básicos relativos a todas estas herramientas.

3.1. CASO 10

Análisis de varianza

EL ANÁLISIS DE VARIANZA

El análisis de varianza (ANOVA, de "*ANalysis Of VAriance*") permite determinar la existencia de comportamientos con diferencias estadísticas significativas debido a la presencia de diversos factores. Su uso es muy importante en el control y mejora de la calidad de los procesos pues permite discriminar la bondad de unos procedimientos, materiales, operarios, etc. frente a otros.

Por ejemplo, se puede verificar algún aspecto (estabilidad dimensional del papel, sistematismos en la planimetría, etc.) sobre varios grupos de muestras de cartografía impresa sobre papel, cada una procedente de un lote diferente, aunque de la misma empresa. Así, se podría verificar si la producción ha sido homogénea, o bien, si ha habido diferencias en el proceso productivo que conducen a la existencia de lotes con mayor o menor calidad en algunos aspectos. Mediante el diseño de experimentos, el ANOVA también permite comprobar si la presencia de un determinado factor va a influir o no en el resultado de un proceso o experimento. Por ejemplo, se puede comprobar si la temperatura o la humedad influyen en la deformación del soporte, si el cambio de operador influye más que otros factores, etc.

Como condiciones previas para su aplicación, todos los grupos han de cumplir las hipótesis de aleatoriedad, normalidad y homogeneidad de sus varianzas; por ello, con anterioridad al propio análisis de varianza, se hace necesario comprobar que éstas se cumplen. En caso de que alguna no se cumpliera, se optaría por emplear el test no paramétrico de Kruskal-Wallis. Éste test no presupone ninguna distribución en los datos, pero, por este mismo motivo, no es tan potente como el ANOVA. Al igual que al análisis ANOVA, el test de Kruskal-Wallis contrasta si las medianas de los grupos son iguales.

El análisis de varianza para un sólo factor se establece como un test de medias múltiples, fraccionando una muestra de datos en tres ó más grupos para contrastar la hipótesis nula de que las medias de las poblaciones de los grupos son iguales. Para ello se fundamenta en descomponer la varianza en dos componentes: la varianza dentro de los grupos y entre los grupos. Si la varianza entre los grupos es significantemente mayor que dentro de los grupos, esto indica que la muestra procede de diferentes poblaciones.

Por tanto, las hipótesis a contrastar son las siguientes:

H_0: Todas las medias son iguales. $\mu_1 = \mu_2 = ... = \mu_n$
H_1: Al menos una media es diferente. $\mu_i \neq \mu_j = ... = \mu_n$

Primeramente hay que calcular la suma de los cuadrados entre grupos (SC_E, Ec. 3.1), dentro de los grupos (SC_D, Ec. 3.2) y total (SC_T, Ec. 3.3), comprobando que se ha de cumplir que $SC_T = SC_D + SC_E$:

$$SC_E = \sum_{j=1}^{k} \left[n_j \left(\overline{X}_j - \overline{\overline{X}} \right)^2 \right] \tag{3.1}$$

$$SC_D = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)^2 = SC_T - SC_E \tag{3.2}$$

$$SC_T = \sum_{i=1}^{N} \left(X_i - \overline{\overline{X}} \right)^2 = \sum_{i=1}^{N} X_i^2 - N\overline{\overline{X}}^2 \tag{3.3}$$

donde:

N Número total de casos.
n_j Número de casos de cada grupo.
k Número de grupos.

A continuación se calculan la varianza estimada entre grupos ($\hat{\sigma}_E^2$) y dentro de los grupos ($\hat{\sigma}_D^2$). Para ello hay que dividir SC_E y SC_D por los grados de libertad, que serán $(k\text{-}1)$ y $(N\text{-}k)$, respectivamente.

Por último, habrá que efectuar el cociente $F = \hat{\sigma}_E^2 / \hat{\sigma}_D^2$, que será el estadístico a contrastar con la distribución $F_{k-1,N-k,\alpha}$ de Snedecor, siendo α el nivel de significación. Si F es menor que $F_{k-1,N-k,\alpha}$ entonces H_0 se acepta.

Si se rechaza la hipótesis nula para ese nivel de significación α, se puede estudiar entre qué grupos hay diferencias. No habrá diferencias significativas entre dos tratamientos genéricos p y q si se cumple la Ec. 3.4.

$$F_{pq} = \frac{\left(\overline{X}_p - \overline{X}_q\right)^2}{\hat{\sigma}_D^2 \cdot \dfrac{n_p + n_q}{n_p \cdot n_q}} < F_{k-1,N-k,\alpha} \cdot (k-1) \tag{3.4}$$

TRABAJO A REALIZAR

Se han recibido tres lotes (A, B, C) de un mapa impreso por parte de una empresa de ingeniería cartográfica. La empresa receptora quiere verificar si la estabilidad dimensional del papel empleado es la misma o bien existe algún lote con papel de diferente calidad.

La Tabla 3.1 presenta tres muestras de diez unidades de cada lote con los valores residuales de un ajuste por mínimos cuadrados de una transformación afín con las cuatro esquinas del mapa.

Se pide saber si existen o no diferencias significativas entre los distintos lotes, efectuando para ello un análisis de varianza de un sólo factor, con un nivel de significación del 5%. En caso de existir diferencias significativas, se pretende saber entre qué lotes existen diferencias.

Se conoce que cada grupo ha mostrado un comportamiento aleatorio y de distribución Normal, y no se ha rechazado la hipótesis de homogeneidad de varianzas entre los distintos grupos (nivel de significación del 5%).

Tabla 3.1 - Residuos (mm) del ajuste de cada hoja de cada muestra

A	B	C
0,085	0,121	0,234
0,096	0,138	0,240
0,104	0,160	0,276
0,125	0,176	0,294
0,142	0,185	0,303
0,153	0,201	0,318
0,165	0,206	0,330
0,178	0,208	0,346
0,185	0,222	0,363
0,203	0,241	0,393

Además, se pide plantear tres casos en los que se considere que podría tener interés el análisis de varianza dentro de los procesos cartográficos, topográficos, geodésicos, etc., explicando para cada uno de ellos la manera de encarar el análisis.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

3.2. CASO 11

Análisis cluster (agrupamiento)

EL ANÁLISIS CLUSTER

El análisis cluster permite generar agrupaciones o subconjuntos de un conjunto de datos. Esto es posible en el caso de que no exista homogeneidad entre ellos debido a la presencia de algún factor segregador, es decir, que los grupos existan en la propia estructura de los datos de partida. El funcionamiento de esta técnica es justamente el contrario al ANOVA. En el caso del análisis de varianza se tienen grupos ya formados y se pretende comprobar si estadísticamente se comportan igual, en el análisis cluster se generan automáticamente agrupaciones o subgrupos de datos cuyos comportamientos estadísticos son más cercanos entre sí. La importancia de esta técnica en la cartografía actual es grande, de hecho, la creación automática de agrupamientos es la base de los procesos de clasificación automatizada de imágenes en Teledetección. En el ámbito de la calidad de los procesos, esta herramienta es muy útil ya que permite segregar grupos por detección de anomalías debido a errores o perturbaciones involuntarias, o bien a alteraciones deliberadas en los mismos. La Figura 3.2 esquematiza el funcionamiento y resultados que se obtienen de esta herramienta.

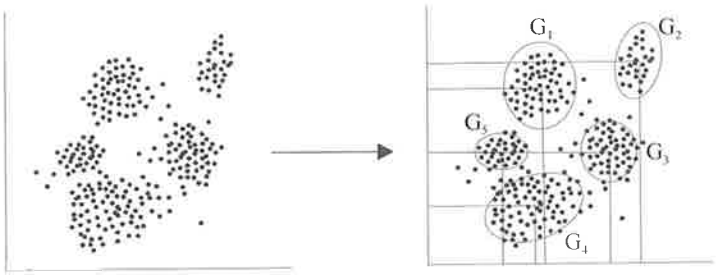


Figura 3.2 - Esquema de un análisis cluster

Por ejemplo, si se tiene una serie de medidas de una misma distancia, tomadas por un mismo operador, y no se tiene la certeza de la utilización de un único tipo de aparato (por ejemplo un distanciómetro) de unas determinadas características (definidas en un pliego de condiciones), se puede recurrir al análisis cluster para comprobar si se pueden efectuar agrupaciones con los datos de la serie. La existencia de más de un grupo podría evidenciar la utilización de instrumentos de diferentes características, de metodologías distintas, o incluso de condiciones de trabajo distintas. Así, por ejemplo, según las agrupaciones que se muestran sobre un único eje en la Figura 3.3, cabría la posibilidad de existencia de dos grupos.

Figura 3.3 - Gráfico de dispersión de los datos

En el análisis cluster no jerarquizado, es decir, aquel en el que cada grupo no depende de otro en una jerarquía, una clasificación correcta debe ser aquella en la que la dispersión dentro de cada grupo sea la menor posible (criterio de la varianza), lo que conlleva una varianza residual (suma de varianzas dentro de cada grupo) mínima. Los métodos jerárquicos se emplean para detectar la estructura de los datos. El análisis comienza considerando tantos grupos como individuos se estudian, con lo cual su varianza sería nula, y sucesivamente se van agrupando en función de la similitud existente entre ellos, aumentando progresivamente la varianza de los nuevos grupos. Si en un paso se agrupan dos o varios cluster, quedan agrupados para el resto del proceso. El último agrupamiento comprendería todos los datos y su varianza sería máxima. Todo el proceso se puede representar mediante un diagrama en forma de árbol invertido (dendograma), como el que muestra la Figura 3.4. Este dendograma ha sido calculado buscando el vecino más próximo con la distancia euclídea, y en el eje X se representa el orden que ocupa cada dato, de menor a mayor valor de la variable.

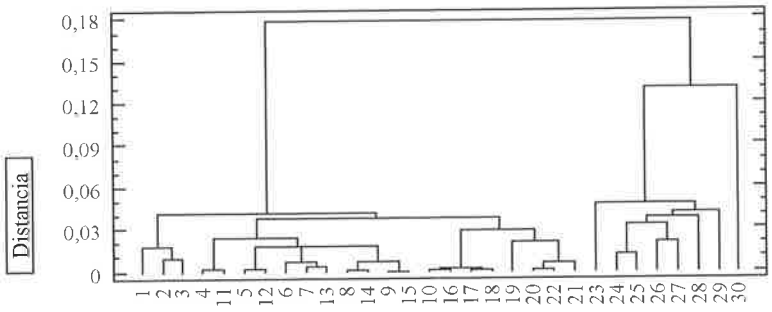


Figura 3.4 - Dendograma

Así, para establecer el número de agrupamientos a realizar se puede partir de un análisis visual del gráfico de dispersión unidimensional de los datos (Figura 3.3) y del dendograma (Figura 3.4).

Dentro de los métodos de agrupamiento no jerarquizados, se encuentra el método de las k-Medias de MacQueen (1967). El término hace referencia al proceso de asignar cada dato a aquel cluster (de los k cluster existentes) con centroide (media) más próximo. Así, el algoritmo de MacQueen para clasificar m datos en k clusters sigue los siguientes pasos:

— Tomar los primeros k datos del conjunto a clasificar y considerarlos como clusters formados por un único dato.

— Asignar cada uno de los restantes m-k datos a aquel cluster con centroide más cercano. A lo largo de este proceso, y tras cada asignación, hay que recalcular el centroide del cluster en cuestión.

— Después de que todos los datos hayan sido asignados en el paso anterior, y se conozcan los centroides de los k clusters, se hará otra pasada en la que se asignarán definitivamente los m datos al cluster cuyo centroide sea el más cercano. Finalmente se recalculan los centroides de los cluster tras la asignación.

El método de las k-Medias tiene la ventaja de ser rápido, además la solución se ve poco afectada por la existencia de casos atípicos. Por otro lado, los inconvenientes son principalmente: a) la dependencia del orden preliminar en que están dispuestos los datos, sobre todo en aquellos conjuntos de datos donde los clusters no estén muy separados, y b) la toma de decisión sobre qué número de grupos que se debe establecer. En relación a esto último, la solución que suelen dar los expertos es utilizar un método jerárquico para decidir el número de grupos y después un algoritmo no jerárquico, como es el de las k-Medias, para realizar el agrupamiento.

Se puede mejorar este método empleando el algoritmo de las k-Medias convergente, que sigue las siguientes etapas:

— Partir de una separación inicial de los datos en varios clusters. Esto se puede realizar, por ejemplo, con los pasos 1 y 2 del método de MacQueen ordinario.

— En una nueva pasada, asignar cada dato al cluster con centroide más próximo y, si este dato había sido clasificado previamente en otro cluster, asignarlo actualizando los centroides de los dos cluster, el que gana y el que pierde un elemento.

— Repetir el paso 2 hasta que se alcance la convergencia, es decir, hasta que en una pasada ningún dato sea cambiado de cluster.

TRABAJO A REALIZAR

Supóngase que el mapa impreso al que se refiere el Caso 10 (Análisis de varianza) se ha recibido en un único lote, y que también se quiere verificar la homogeneidad de la estabilidad dimensional del papel. Para ello se decide aplicar el análisis cluster a una muestra de hojas extraídas del lote.

Para éste caso, la Tabla 3.2 presenta los valores residuales de un ajuste por mínimos cuadrados de una transformación afín con las cuatro esquinas del mapa

obtenidos de 30 muestras del lote. Supóngase que la aleatoriedad de los datos ya ha sido contrastada mediante un contraste de rachas.

Tabla 3.2 - Residuos (mm) del ajuste mínimo cuadrático de una serie de 30 muestras		
0,085	0,393	0,104
0,138	0,142	0,240
0,276	0,201	0,222
0,294	0,346	0,121
0,160	0,096	0,178
0,185	0,330	0,234
0,206	0,185	0,208
0,363	0,303	0,153
0,125	0,318	0,165
0,176	0,203	0,241

Se pide determinar, por análisis cluster, si se puede considerar como una muestra homogénea o si, por el contrario, se puede considerar la existencia de una estructura en grupos, lo que manifestaría la existencia de diferencias significativas en la estabilidad dimensional en el lote. El método a emplear será el de las k-Medias estándar.

Por otro lado, la empresa de Cartografía también quiere servirse de esta herramienta y necesita conocer posibles aplicaciones. Indicar y explicar tres ejemplos de posible uso de esta herramienta de análisis (señalando el objetivo, metodología, material necesario, aplicabilidad, etc.)

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

3.3. CASO 12

Análisis de series temporales en la digitalización gráfica manual

EL ANÁLISIS UNIVARIANTE DE SERIES TEMPORALES

El análisis de series temporales es una herramienta estadística de uso práctico y cotidiano en ámbitos como la climatología, econometría y seguimiento de mercados, medicina, etc. No obstante, estas técnicas también pueden llevarse al campo de la producción y control en las industrias cartográficas para el estudio de la evolución temporal de índices. En algunos casos, previa transformación, también pueden servir para estudiar propiedades métricas de las BDG.

Este tipo de análisis permite construir modelos explicativos de la estructura y evolución de una variable observada a lo largo del tiempo. Una hipótesis básica es la existencia de autocorrelación en los datos (Pindyck y Rubenfeld, 1980). Los datos están disponibles en intervalos regulares de tiempo y se desea utilizar la posible inercia en el comportamiento de la serie para prever su evolución futura.

Se dice que un proceso estocástico es estacionario, en sentido débil, si existen y son estables la media, la varianza y las covarianzas. Una clase importante de procesos estacionarios son los autorregresivos (AR) de orden p, que resultan de imponer una dependencia lineal entre las variables del proceso. Es decir, cada observación se forma como promedio ponderado de las p observaciones precedentes.

También existen los llamados procesos de media móvil (MA) de orden q, donde cada observación se genera mediante una media ponderada de las perturbaciones aleatorias dentro de un retardo de q períodos.

Los procesos AR y MA son complementarios, pues los AR imponen restricciones sobre las pautas de decrecimiento de los coeficientes y los MA suponen un número de términos finito. Los procesos ARMA (p,q) intentan combinar estas ventajas y permiten representar utilizando pocos parámetros procesos cuyos q primeros coeficientes son cualesquiera, y los siguientes decrecen según leyes simples.

La gran mayoría de los procesos que se observan no son estacionarios, y su nivel medio varía con el tiempo. Pero es frecuente que el proceso se convierta en estacionario al diferenciarlo:

— Se llama proceso en primera diferencia del $\{z_t\}$ a un nuevo proceso $\{w_t\}$ obtenido como $w_t = z_t - z_{t-1}$.

— Se llama proceso en segunda diferencia del $\{z_t\}$ a un nuevo proceso obtenido como $w_t - w_{t-1}$.

— Un proceso es integrado de orden h cuando al diferenciarlo h veces se convierte en estacionario.

En el caso de procesos no estacionarios, por diferenciaciones sucesivas se pueden convertir en estacionarios. En estos casos se define el modelo ARIMA(p,d,q), donde d es el número de diferencias a considerar. Si el proceso ya es estacionario, se toma $d = 0$, con lo que se tiene el modelo ARMA(p,q).

APLICACIÓN DEL ARMA A LA DIGITALIZACIÓN GRÁFICA MANUAL

La hipótesis de autocorrelación de los datos no es lejana a la realidad si se piensa cómo sucede el proceso de digitalización gráfica manual. El operador de digitalización mueve continuamente el cursor intentando ajustarse a la línea, cosa que no puede conseguir perfectamente, por lo que en un proceso de retroalimentación y autoajuste, la línea digitalizada corta continuamente la línea modelo. Además, un operario entrenado, no debe tener sesgo en su trabajo, por lo que los errores deberán compensarse y distribuirse según una función de distribución normal de media cero.

De esta forma, es perfectamente factible el considerar que una serie de puntos digitalizados en modo de flujo continuo, en el que se almacena un punto cada cierto intervalo de tiempo, se pueda analizar como una serie temporal estacionaria, es decir, estabilizada en el tiempo sobre un valor constante. Al suponer que los errores se distribuyen aleatoriamente según una normal, la serie queda caracterizada por tres parámetros: media, varianza y autocorrelación.

Keefer y col. (1991) proponen utilizar un modelo autorregresivo de medias móviles ARMA (1,0) (p = 1, q = 0, o simplemente AR(1)). Así, cada digitalización está correlada con la inmediatamente anterior, sumándole además un ruido aleatorio independiente (Ec. 3.5).

$$z_t = c + \phi z_{t-1} + a_t \qquad (3.6)$$

donde:

z_t	Valor de la serie temporal en el instante t, tal que $1 \leq t \leq n$.
z_{t-1}	Valor de la serie temporal en el instante t-1.
c	Constante, será nula si no existen sistematismos.
ϕ	Constante, tal que $\lvert\phi\rvert < 1$, para que el proceso sea estacionario.
a_t	Ruido aleatorio en el instante t.

Considerando, por tanto, un AR (1), las etapas para determinar los valores c y ϕ en la digitalización gráfica manual serían las siguientes:

— Digitalizar una línea y obtener una serie de distancias perpendiculares entre la línea digitalizada y la original (ver Figura 3.4). Cada distancia podrá ser positiva o negativa dependiendo del lado en que se encuentre la línea digitalizada respecto a la original.

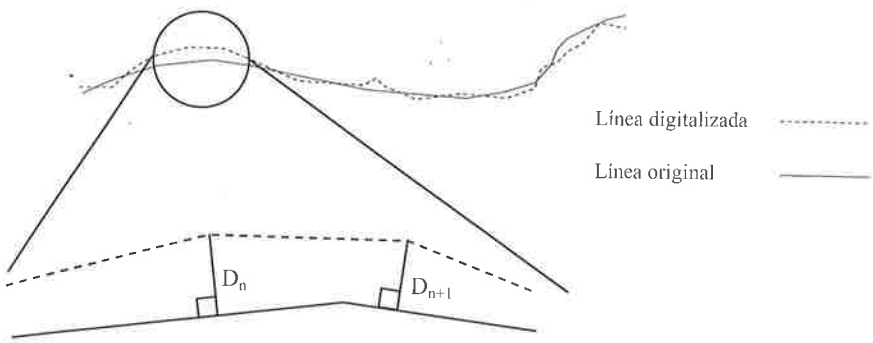


Figura 3.5 - Obtención de la distancia perpendicular entre la línea original y la digitalizada

— Ajustar un modelo AR(1) a la serie de n distancias perpendiculares, mediante mínimos cuadrados a través el sistema de ecuaciones ($A\,x - t = v$) de la Ec. 3.6.

$$\begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ 1 & \vdots \\ 1 & z_{n-1} \end{bmatrix} \cdot \begin{bmatrix} c \\ \phi \end{bmatrix} - \begin{bmatrix} z_2 \\ z_3 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} v_2 \\ v_3 \\ \vdots \\ v_{n-1} \end{bmatrix} \tag{3.6}$$

Tras el ajuste se ha de realizar el análisis de los resultados obtenidos. En la digitalización gráfica manual interesará que c tenga un valor muy próximo a cero, que indicará que no han existido sistematismos. El valor de ϕ deberá ser lo más pequeño posible, ya que así el error cometido en un punto se amortigua al transmitirlo al siguiente, observándose globalmente un menor periodo en las oscilaciones de la serie de distancias perpendiculares respecto a la línea original. Esto será un indicador de que la superficie entre ambas líneas será menor (ver Figuras 3.6 y 3.7 con ejemplos de modelos con distinto valor de ϕ).

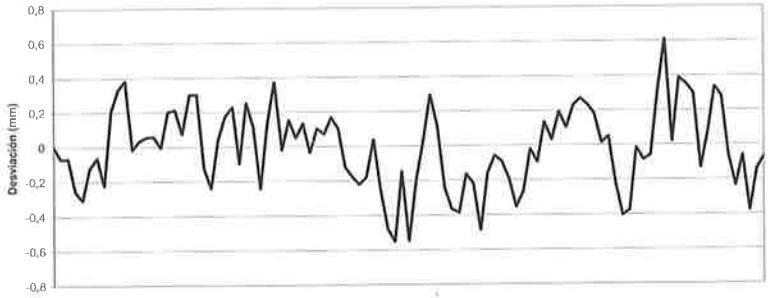


Figura 3.6 - Modelo AR(1) con $c = 0$ y $\phi = 0{,}6$, para un error aleatorio de 0,5 mm al 95 % de confianza

Figura 3.7 - Modelo AR(1) con $c = 0$ y $\phi = 0{,}9$, para un error aleatorio de 0,5 mm al 95 % de confianza

Además, para estudiar la magnitud de estos errores aleatorios, se puede analizar el valor medio, desviación típica y máximo de los valores absolutos de la serie que, lógicamente, será conveniente sean reducidos.

TRABAJO A REALIZAR

En un pliego de condiciones referente a la digitalización mediante tablero de una serie de planos catastrales se especifica cuál debe ser la discrepancia entre las líneas originales y las digitalizadas. La empresa receptora del trabajo ha realizado un control sobre una muestra del lote de planos y ha detectado fallos por encima de los límites de aceptación del pliego.

Antes esta situación, la empresa encargada de la digitalización quiere conocer y cuantificar la destreza de sus operarios encargados de estas tareas. Por ello decide someterlos a una prueba consistente en la digitalización de una muestra de líneas de diferentes orientaciones, grosores, sinuosidades, etc., y el posterior cálculo de algunos parámetros estadísticos que permitan extraer conclusiones objetivas.

En las Tablas 3.3 y 3.4 se presentan dos series, ejecutadas por dos operarios, y procedentes de la digitalización de una misma línea. Se pide ajustar un modelo autorregresivo AR(1) a cada serie y calcular otros parámetros estadísticos (media, desviación típica...) para, a continuación, analizar los resultados.

Tabla 3.3 - Serie procedente de la digitalización del operador A

1	0,0000	16	-0,0166	31	-0,1125	46	0,0451	61	-0,0452	76	0,0092	91	-0,0279
2	0,0616	17	-0,0412	32	0,0117	47	-0,0836	62	-0,0610	77	-0,0882	92	0,0480
3	-0,1624	18	0,0234	33	-0,0630	48	0,1577	63	0,0706	78	0,0063	93	0,1603
4	-0,2579	19	-0,0902	34	-0,0324	49	0,1318	64	0,2271	79	-0,1447	94	0,2046
5	-0,1077	20	-0,0310	35	0,1091	50	-0,0122	65	0,0929	80	0,0693	95	0,1312
6	-0,0712	21	0,0172	36	0,2140	51	-0,0971	66	0,1831	81	-0,0953	96	0,3210
7	-0,0385	22	0,1817	37	0,2777	52	-0,1427	67	0,2289	82	-0,1063	97	0,2148
8	-0,0261	23	0,0070	38	0,3054	53	0,0150	68	0,0742	83	-0,0880	98	0,1968
9	-0,2009	24	0,1302	39	0,3346	54	0,0710	69	0,0928	84	0,1033	99	0,1526
10	-0,0553	25	0,1863	40	0,3996	55	0,2379	70	0,3179	85	-0,1390	100	0,1022
11	-0,0229	26	0,1993	41	0,3940	56	0,0759	71	0,1109	86	-0,0780	101	0,0787
12	-0,0919	27	0,1641	42	0,0609	57	0,0678	72	0,1650	87	0,1012		
13	-0,0409	28	-0,0128	43	0,0930	58	-0,0528	73	0,1606	88	0,1863		
14	-0,0739	29	-0,1835	44	0,0700	59	-0,0539	74	0,1736	89	-0,0718		
15	0,0081	30	-0,1672	45	0,0740	60	0,0169	75	0,0540	90	-0,1989		

Tabla 3.4 - Serie procedente de la digitalización del operador B

1	0,0000	16	0,2454	31	-0,0326	46	0,4331	61	0,2001	76	0,2500	91	0,3408
2	0,0136	17	0,3056	32	0,1762	47	0,3567	62	0,2273	77	0,2170	92	0,1473
3	-0,0400	18	0,4536	33	-0,0760	48	0,2796	63	0,3415	78	0,1480	93	0,0262
4	0,0609	19	0,3411	34	-0,1334	49	0,1702	64	0,2101	79	0,1286	94	0,0653
5	-0,0363	20	0,0910	35	-0,2002	50	0,1360	65	0,1387	80	-0,0623	95	0,0423
6	-0,0932	21	0,1339	36	-0,2842	51	0,1027	66	0,0649	81	-0,0127	96	-0,0606
7	-0,1450	22	-0,0576	37	-0,0095	52	0,0392	67	0,0807	82	-0,0489	97	0,0091
8	-0,0463	23	0,0790	38	0,0419	53	-0,0165	68	0,1261	83	0,1209	98	0,1630
9	-0,0447	24	0,0040	39	0,1979	54	-0,1214	69	0,2254	84	0,0770	99	-0,0233
10	-0,2151	25	-0,0450	40	0,3756	55	0,0151	70	0,1599	85	0,0934	100	-0,0309
11	-0,2903	26	0,0295	41	0,4454	56	0,1016	71	0,2832	86	0,2648	101	0,0165
12	0,0308	27	0,1403	42	0,2182	57	-0,0179	72	0,3443	87	0,2904		
13	0,1820	28	0,0344	43	0,3239	58	0,1065	73	0,2590	88	0,3679		
14	0,0936	29	-0,0407	44	0,4315	59	0,2707	74	0,2365	89	0,5121		
15	0,2767	30	0,1642	45	0,4612	60	0,2019	75	0,3049	90	0,4453		

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

3.4. CASO 13

Empleo de cuantiles en la determinación del tamaño de celda

EL TAMAÑO DE LA CELDA EN LA RASTERIZACIÓN

La importancia del proceso de rasterización es grande dado que la conversión de información vectorial a raster es un proceso bastante común y, algunas veces, indispensable en los flujos de información propios de los SIG (Ariza, 2002). El tamaño de la celda aparece como parámetro fundamental en todos los estudios del error. Como idea general, su repercusión está ligada al tamaño relativo respecto al elemento a rasterizar. Ocurre pues, que cuanto más pequeño es el tamaño de la celda respecto al de la figura geométrica a la que va a representar, tanto menor será el error que se comete.

EJEMPLO DE UTILIZACIÓN DE CUANTILES EN LA ESTIMACIÓN DEL TAMAÑO DE CELDA

En este ejemplo se va a proceder a la determinación del tamaño de celda para la rasterización de un fichero de superficies en formato vectorial de una base de datos geográfica (BDG) y su incorporación a un SIG raster. Se considera un fichero con 844 superficies determinadas por polígonos vectoriales procedente de la cobertura de usos del suelo de Andalucía (CMA, 1995) que se muestra en la Figura 3.8

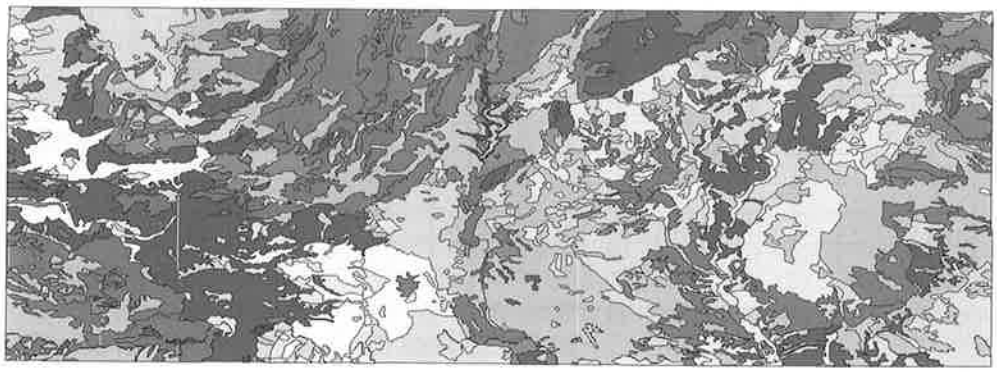


Figura 3.8 - Fichero vectorial de la cobertura de usos del suelo de Andalucía

Las coordenadas de la ventana geográfica encuadrante del fichero de la Figura 3.7 son las siguientes:

X mínima	490915	Y mínima	4169047
X máxima	527665	Y máxima	4187583

En primer lugar se planteará una aproximación basada en el menor objeto existente en la base de datos. Para ello en la Tabla 3.5 se han listado los 10 elementos de menor tamaño superficial. Como se puede observar en dicha tabla, el primer valor que aparece es realmente pequeño, de tan sólo unos 2 m^2 de superficie, siendo el siguiente valor de algo más de 87 m^2, un tamaño unas 42 veces superior al primero.

Se ha de comprobar si el polígono de 2,089 m^2, y los restantes de valor bajo, tienen realmente sentido en la base de datos o si, por el contrario, son polígonos espúreos (polígonos que no tienen presencia física y se deben a los cómputos numéricos e imperfecciones en la base de datos; en algunos casos también se les denomina polígonos astilla). Tras localizar estos polígonos con unos de niveles de zoom considerables se decide que las manchas correspondientes a los tamaños 2,086 m^2 y 87,258 m^2 son polígonos espúreos, por lo que no se tendrán en cuenta.

Tabla 3.5 - Listado de los 10 elementos de menor tamaño superficial

ZONA_ID	CODIGO	PERIMETRO [m]	AREA [m^2]
36	358	6,920	2,086
845	358	47,337	87,258
33	358	119,943	199,336
586	358	89,187	266,449
293	358	87,952	335,930
29	358	131,043	343,781
51	358	102,227	520,219
574	358	179,916	624,629
416	358	119,895	682,297
4	358	382,679	1419,031

Según lo anterior, la mínima unidad a cartografiar es la de 199,336 m^2, es decir, de 200 m^2. Por ello, asemejándola a un cuadrado, éste debería tener unos lados de 14 metros, lo que lleva a considerar un tamaño máximo de celdilla de 7 m de lado según la aproximación por el teorema de Nyquist[1]. En la Tabla 3.6 aparecen, para cada tamaño superficial de la Tabla 3.5, todos los tamaños de celda estimados según el cuadrado equivalente y mediante Nyquist, así como el número de filas, columnas y celdas del fichero resultante.

Indudablemente, mejor que tomar la decisión sobre un único elemento de la base de datos, cosa que puede llevar a tamaños de celda muy reducidos y grandes volúmenes de ficheros, sería más conveniente tomar la decisión desde un planteamiento estadístico que permita decidir el tamaño con el conocimiento del error que se comete.

[1] El teorema de Nyquist establece que para muestrear adecuadamente un fenómeno de frecuencia f se requiere, al menos, una frecuencia doble de la propia del fenómeno de interés. Llevado esto a términos de resolución espacial en el modelo ráster, la interpretación es que para poder tener en consideración objetos de, por ejemplo, 1x1 m^2, es necesario considerar celdas de 0,5 × 0,5 m^2.

La idea básica del método que se va a aplicar consiste en rasterizar la BDG con distintos tamaños de celda y estudiar lo que ocurre con la población de los errores relativos, tomados en valor absoluto, cometidos para cada una de las instancias superficiales (polígonos) existentes en ella. Así, para cada tamaño de celda se procede a ordenar los errores relativos, localizando los valores correspondientes a los cuantiles de interés (por ejemplo el 99%, 97,5%, 95%, 90%, etc.). Estos valores de error areal *EA(n%)* indican que existe un *n%* de elementos cuyo *Error* <=*EA*. A pesar de necesitar el desarrollo de un análisis previo, este método permite determinar para una BDG concreta el tamaño adecuado para el nivel de error que se desee.

Tabla 3.6 - Tamaños de celdilla estimados

AREA [m²]	**Lado cuadrado equivalente [m]**	**Lado cuadrado Nyquist [m]**	**Tamaño Fichero (Nyquist) [filas × columnas = nº celdas]**
m^2	m	m	filas × columnas = celdas
2,086	1,444	0,722	50902 × 25674 = 1306857948
87,258	9,341	4,607	7870 × 1985 = 15621950
199,336	**14,118**	**7,059**	**5208 × 2626 = 13676208**
266,449	16,323	8,161	4504 × 2272 = 10233088
335,930	18,328	9,164	4012 × 2024 = 8120288
343,781	18,541	9,270	3966 × 2000 = 7932000
520,219	22,808	11,404	3224 × 1626 = 5242224
624,629	24,992	12,496	2942 × 1484 = 4365928
682,297	26,121	13,060	2814 × 1420 = 3995880
1419,031	37,670	18,835	1952 × 986 = 1924672

En este caso se ha procedido a realizar rasterizaciones con celdas cuadradas con los siguientes tamaños de lado: 1, 5, 10, 20 y 30 metros; y se ha creado una tabla en la que se incorporan los valores de la superficie vectorial original y la asignada por el programa ArcInfo tras la rasterización según los tamaños indicados anteriormente. Con estos valores se ha determinado el error relativo en área en valor absoluto. Esta distribución de 844 errores se ha ordenado de menor a mayor y se han detectado los niveles de error correspondientes a los percentiles del 99%, 97,5%, 95%, 90%, etc., tal como se muestran en la Tabla 3.7.

En ella, junto con los valores de error relativo para cada uno de los percentiles marcados, se presenta un resumen estadístico (valor medio, mediana, desviación típica, etc.) correspondiente a la distribución poblacional (N = 844 casos o polígonos) del error cuando se rasteriza a cada tamaño de celda. En la Tabla 3.8 aparece el volumen de almacenamiento, medido en número de píxeles, necesario para la rasterización del fichero vectorial original con cada tamaño de celda.

Tabla 3.7 - Resumen estadístico de errores relativos en las pruebas de rasterización

	Error relativo en superficie según el tamaño de rasterización				
	1 metro	5 metros	10 metros	20 metros	30 metros
Media	$1{,}124 \cdot 10^{-3}$	$4{,}866 \cdot 10^{-3}$	$7{,}649 \cdot 10^{-3}$	$1{,}635 \cdot 10^{-2}$	$2{,}539 \cdot 10^{-2}$
Mediana	$5{,}660 \cdot 10^{-5}$	$6{,}317 \cdot 10^{-4}$	$1{,}733 \cdot 10^{-3}$	$4{,}921 \cdot 10^{-3}$	$7{,}436 \cdot 10^{-3}$
Desv. típica	$1{,}834 \cdot 10^{-2}$	$4{,}074 \cdot 10^{-2}$	$4{,}601 \cdot 10^{-2}$	$7{,}421 \cdot 10^{-2}$	$9{,}579 \cdot 10^{-2}$
Varianza	$3{,}363 \cdot 10^{-4}$	$1{,}660 \cdot 10^{-3}$	$2{,}117 \cdot 10^{-3}$	$5{,}507 \cdot 10^{-3}$	$9{,}175 \cdot 10^{-3}$
Percentiles 10	$6{,}978 \cdot 10^{-6}$	$9{,}562 \cdot 10^{-5}$	$2{,}080 \cdot 10^{-4}$	$6{,}094 \cdot 10^{-4}$	$9{,}541 \cdot 10^{-4}$
20	$1{,}483 \cdot 10^{-5}$	$1{,}823 \cdot 10^{-4}$	$4{,}452 \cdot 10^{-4}$	$1{,}455 \cdot 10^{-3}$	$2{,}272 \cdot 10^{-3}$
30	$2{,}502 \cdot 10^{-5}$	$3{,}129 \cdot 10^{-4}$	$7{,}575 \cdot 10^{-4}$	$2{,}283 \cdot 10^{-3}$	$3{,}498 \cdot 10^{-3}$
40	$3{,}906 \cdot 10^{-5}$	$4{,}560 \cdot 10^{-4}$	$1{,}232 \cdot 10^{-3}$	$3{,}195 \cdot 10^{-3}$	$5{,}152 \cdot 10^{-3}$
50	$5{,}660 \cdot 10^{-5}$	$6{,}317 \cdot 10^{-4}$	$1{,}733 \cdot 10^{-3}$	$4{,}921 \cdot 10^{-3}$	$7{,}436 \cdot 10^{-3}$
60	$7{,}897 \cdot 10^{-5}$	$9{,}528 \cdot 10^{-4}$	$2{,}498 \cdot 10^{-3}$	$6{,}818 \cdot 10^{-3}$	$1{,}048 \cdot 10^{-2}$
70	$1{,}202 \cdot 10^{-4}$	$1{,}453 \cdot 10^{-3}$	$3{,}351 \cdot 10^{-3}$	$9{,}843 \cdot 10^{-3}$	$1{,}576 \cdot 10^{-2}$
80	$1{,}943 \cdot 10^{-4}$	$2{,}166 \cdot 10^{-3}$	$5{,}373 \cdot 10^{-3}$	$1{,}420 \cdot 10^{-2}$	$2{,}463 \cdot 10^{-2}$
90	$4{,}429 \cdot 10^{-4}$	$3{,}882 \cdot 10^{-3}$	$8{,}980 \cdot 10^{-3}$	$2{,}322 \cdot 10^{-2}$	$4{,}100 \cdot 10^{-2}$
95	$8{,}376 \cdot 10^{-4}$	$7{,}303 \cdot 10^{-3}$	$1{,}676 \cdot 10^{-2}$	$4{,}059 \cdot 10^{-2}$	$6{,}658 \cdot 10^{-2}$
97,5	$2{,}314 \cdot 10^{-3}$	$1{,}426 \cdot 10^{-2}$	$3{,}183 \cdot 10^{-2}$	$7{,}079 \cdot 10^{-2}$	$9{,}985 \cdot 10^{-2}$
99	$8{,}165 \cdot 10^{-3}$	0,1151	0,1171	0,2428	0,3980

Tabla 3.8 - Tamaño del fichero con cada tamaño de celda

Tamaño del lado	1 metro	5 metros	10 metros	20 metros	30 metros
Nº de píxeles	681.198.000	27.247.920	6.811.980	1.702.995	756.887

Los valores anteriores representados gráficamente se presentan en la Figura 3.9. En el eje horizontal aparece la evolución de los tamaños de celdas utilizados en la rasterización y en el eje vertical los errores relativos (en valor absoluto) en una escala logarítmica. Las curvas correspondientes a cada uno de los percentiles considerados se presentan en un color distinto. Se observa, pues, que la evolución del error relativo forma una especie de familia de curvas respecto a los percentiles, y que los valores se elevan conforme se incrementa el tamaño de celda, como era de esperar.

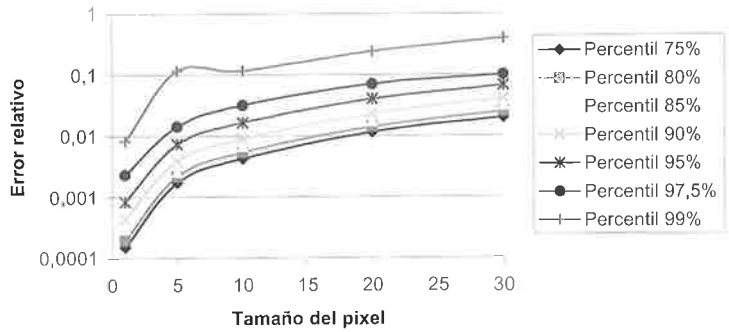


Figura 3.9 - Error relativo cometido con cada percentil en función del tamaño de celda

Según se deduce de la Figura 3.9, si nuestro interés es acotar el error relativo en valores inferiores al 1% de la superficie de los polígonos, y deseamos que esto ocurra para un 95% de los casos de la base de datos, se toma la línea del percentil 95% y se observa que el tamaño más adecuado de rasterización es el de 5 metros, ya que el error relativo alcanzado es muy próximo e inferior al 1%. Si se observa la Tabla 3.9, para este tamaño de celda, el error relativo medio es del 0,5%, con una desviación típica del 4%. Sea ahora otro ejemplo: si lo que se desea es que en el 75% de los polígonos el error sea inferior al 1%, se podría optar por un tamaño de píxel de aproximadamente 20 metros, y se alcanzaría un error relativo medio del 1,6% con desviación típica del 7,4%.

TRABAJO A REALIZAR

Dada una BDG_1 de carácter vectorial, se extraerá una ventana geográfica BDG_2 ($BDG_2 \in BDG_1$) que pueda considerarse representativa de la misma en cuanto a la morfología y tamaño de los polígonos que la conformen. Esta BDG_2 se rasterizará con diversos tamaños de celda según lo expuesto anteriormente, para obtener una serie de curvas como las presentadas en la Figura 3.8. Sobre la base de estos resultados se discutirá y decidirá el tamaño adecuado de celda para el proceso de rasterización.

Además, se compararán los resultados con el tamaño que se hubiera tomado en caso de utilizar el criterio de Nyquist.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar el material necesario para este caso.

3.5. CASO 14

Evaluación de la calidad geométrica de un escáner

LA DIGITALIZACIÓN MEDIANTE ESCÁNER

La digitalización con escáner consiste en la generación de un fichero raster de imagen mediante un sistema de barrido óptico. Esta técnica de digitalización presenta gran interés en el ámbito de la Cartografía debido a que (Deseilligny, 1994):

- Es de rápida ejecución.
- Permite obtener geometrías tras un proceso adecuado de vectorización.
- Permite obtener atributos tras un proceso de clasificación de las manchas.
- Permite la recuperación de textos mediante un sistema de reconocimiento óptico de caracteres adecuado a las condiciones de los textos presentes en la cartografía.

Además, no se puede olvidar que un mapa digitalizado mediante escáner es una copia del original y, como tal, puede distribuirse o venderse, si se cuida su calidad geométrica.

La calidad de un escaneado y la facilidad de su postproceso dependen tanto del escáner como de las condiciones físicas del material que se digitaliza. Los aspectos clave de esta operación son la resolución y las distintas calibraciones. El proceso de calibración completo debe abarcar tanto el comportamiento radiométrico, a menudo subestimado, como el geométrico.

Sin embargo, en el caso de un escáner no fotogramétrico, de coste muy inferior, no se suele incorporar software para la calibración, y si se hace, los patrones y procedimientos son de poca exactitud. Además, los proveedores de este tipo de dispositivos raramente informan al usuario acerca de las especificaciones técnicas y de error, como por ejemplo, el error medio cuadrático y máximo que puede ocurrir en diferentes casos (Baltsavias y Waegli, 1996), algo lamentable, ya que los errores de posición son muy variables, incluso dentro de una misma digitalización. Así, los principales problemas de este tipo de aparato son la exactitud posicional, la uniformidad y repetitividad, las inestabilidades mecánicas (vibraciones) y las distorsiones de la lente o las imperfecciones de otros elementos ópticos (espejos, filtros) (Baltsavias, 1994).

EVALUACIÓN DE LA CALIDAD GEOMÉTRICA

La calibración geométrica es, lógicamente, más importante que la radiométrica para su utilización en la digitalización de cartografía a escalas grandes y medianas, sobre todo en los denominados mapas topográficos. En Baltsavias y Käser (1998) y Haering et al. (1998) se realizan las siguientes pruebas para la evaluación del rendimiento geométrico:

— Evaluación de la exactitud geométrica: Se calcula por mínimos cuadrados una transformación afín de seis parámetros entre las coordenadas (x_c, y_c) de una malla calibrada trazada sobre una placa de vidrio y las coordenadas píxel (x_p, y_p) obtenidas tras su digitalización (Ec. 3.7).

$$\begin{aligned} x_c &= a\,x_p + b\,y_p + c \\ y_c &= d\,x_p + e\,y_p + f \end{aligned} \tag{3.8}$$

Será conveniente considerar tres transformaciones, una para cada banda de color (rojo, verde, azul), sobre todo si el escáner realiza una pasada para cada una de ellas. Si la digitalización se produce en una única pasada, es de esperar que no haya diferencias sustanciales.

El tamaño del píxel real se puede obtener como la componente cuadrática de los parámetros a y b (tamaño del píxel en X) y de los parámetros d y e (tamaño del píxel en Y). Cualquier diferencia respecto al valor teórico supondrá un desajuste en el tamaño del píxel. Además, si ambas dimensiones del tamaño del píxel real no fueran iguales, esto supondría la existencia de un factor de escala entre ambos ejes (el cociente entre ambos tamaños), ya que cualquier programa considerará que las dos dimensiones del tamaño del píxel son iguales.

En el caso de que el escáner realice una pasada por cada color, es útil comparar los valores obtenidos para los parámetros c y f en cada transformación. En caso de existir diferencias significativas, habrá un desajuste entre las bandas y éste puede provocar una pérdida de nitidez en las imágenes en color.

No obstante, este tipo de transformación no permite la corrección de errores de tipo no lineal, y que son frecuentes en la digitalización como, por ejemplo, problemas de arrastre del sensor, distorsiones radiales del sistema óptico del sensor, etc., que deben ser corregidos mediante expresiones más complejas, que no suelen estar implementadas en los programas habituales de tratamiento digital de imágenes (Delgado y Cardenal, 1998)

Tras ajustar la transformación afín se podrán obtener las coordenadas calculadas (x_c, y_c) de la malla en el sistema de referencia de la placa a partir de las coordenadas píxel. Seguidamente se calcula, para cada vértice de la malla, las diferencias en X e Y entre las coordenadas placa calibradas y las transformadas (residuos). Después se pueden calcular diferentes medidas estadísticas de los errores (media, desviación típica, mínimo, máximo...). Es también conveniente trazar el vector diferencia de cada vértice (a la escala adecuada para poder visualizarlo) para detectar posibles comportamientos diferentes en la superficie de digitalización, como se observa en el caso de la Figura 3.10.

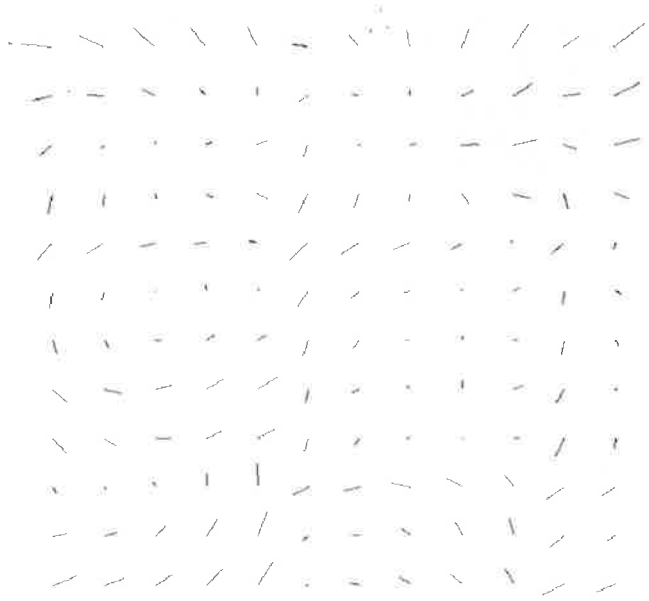


Figura 3.10 - Ejemplo de trazado de los vectores diferencia tras la transformación afín

— Diferencias entre los canales de color: Es interesante comparar las coordenadas píxel de los vértices de la malla entre cada pareja de canales (rojo - verde, rojo - azul y verde - azul), obteniendo su media, mínimo, máximo, etc. Esto interesa en el caso de que la digitalización tenga lugar en tres pasadas, una por cada color, mediante un sensor CCD. Si el escáner sólo realiza una pasada, las diferencias halladas suelen ser más elevadas en la dirección de la digitalización, debido a una errónea calibración de la distancia existente entre las tres líneas de color del sensor.

— Repetitividad: Se comparan los resultados obtenidos en múltiples digitalizaciones. Por ejemplo, se pueden analizar los valores máximos y mínimos de las medidas estadísticas de la exactitud geométrica que se han obtenido.

— Resolución geométrica: Se efectúa una inspección visual de la digitalización de un patrón, como el USAF (Figura 3.11) sobre vidrio de diferentes resoluciones. Es de esperar que el contraste de las líneas verticales (dirección de la digitalización) sea mayor que las horizontales (dirección del sensor), debido al rápido movimiento del sensor, por encima de sus posibilidades, que impone el fabricante para proporcionar un mayor atractivo comercial.

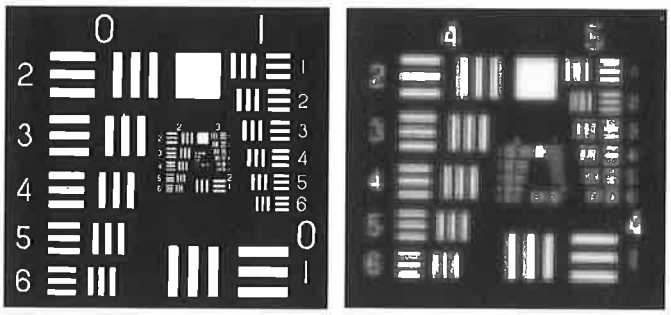


Figura 3.11 - Patrón de resoluciones USAF y sección de su digitalización con escáner

TRABAJO A REALIZAR

Supóngase que se trabaja en el Departamento de Cartografía de un organismo público se ha recibido un escáner de sobremesa no fotogramétrico (ver sus datos en la Tabla 3.9), del que se desconoce cualquier parámetro referente a su calidad geométrica.

Por ello, antes de emplearlo para cualquier trabajo, se ha decidido verificar su calidad geométrica, evaluando la exactitud geométrica y las diferencias entre los canales de color, ya que efectúa una pasada para la digitalización de cada uno de los tres colores.

Por tanto, habrá que resolver la transformación afín de seis parámetros entre cada pareja de sistemas de coordenadas (calibradas - rojo, calibradas - verde, calibradas - azul) y obtener sus residuos (realizando un gráfico con la distribución de los vectores de los mismos), analizando pormenorizadamente todos los resultados obtenidos. También habrá que analizar si existen diferencias entre las distintas bandas de color.

Tabla 3.9 - Características del escáner

Tipo de escáner	Escáner plano con sensor CCD de 15.000 píxeles. Digitalización en 3 pasadas.
Resolución óptica máxima	1200 ppp
Superficie máxima de digitalización	A3

Para realizar el estudio, se ha digitalizado, a la máxima resolución óptica (1200 ppp), una placa calibrada de vidrio de tamaño 23 cm × 23 cm, que incluye una retícula de 5 × 5 cruces con una separación de 50 mm × 50 mm con una precisión de 1 µm. Las coordenadas calibradas y las coordenadas imagen medidas en cada banda de color se presentan en la Tabla 3.10.

Tabla 3.10 - Coordenadas calibradas e imagen de los puntos de control de una placa utilizada para calibrar un escáner

	Coordenadas calibradas		Coordenadas imagen (Rojo)		Coordenadas imagen (Verde)		Coordenadas imagen (Azul)	
Punto	x_c	y_c	x_r	y_r	x_v	y_v	x_a	y_a
1	-100	100	-4742,270	-4803,900	-4742,780	-4803,850	-4742,380	-4806,430
2	-50	100	-2377,750	-4790,210	-2377,890	-4790,230	-2379,040	-4792,130
3	0	100	-0,902	-4777,040	-1,758	-4777,020	-1,579	-4778,770
4	50	100	2378,316	-4764,970	2378,213	-4764,600	2377,344	-4765,380
5	100	100	4748,442	-4753,630	4748,111	-4753,540	4746,925	-4753,160
6	-100	50	-4748,570	-2440,790	-4749,350	-2441,660	-4749,860	-2443,630
7	-50	50	-2384,170	-2426,990	-2384,030	-2427,960	-2385,310	-2429,380
8	0	50	-7,699	-2414,330	-7,264	-2414,920	-8,729	-2415,630
9	50	50	2371,823	-2402,820	2371,195	-2402,970	2370,911	-2403,340
10	100	50	4741,734	-2390,580	4741,248	-2391,110	4740,763	-2390,920
11	-100	0	-4755,200	-80,390	-4756,400	-80,428	-4757,330	-83,950
12	-50	0	-2390,590	-66,915	-2391,900	-66,915	-2393,000	-69,393
13	0	0	-13,224	-53,828	-15,054	-53,844	-16,257	-55,483
14	50	0	2365,569	-42,005	2364,565	-41,647	2363,067	-42,721
15	100	0	4735,542	-30,471	4733,845	-29,579	4732,462	-30,376
16	-100	-50	-4760,640	2281,812	-4762,460	2280,176	-4764,020	2277,446
17	-50	-50	-2396,060	2294,896	-2397,980	2293,898	-2399,890	2291,715
18	0	-50	-19,383	2307,811	-21,096	2306,667	-23,288	2305,565
19	50	-50	2359,942	2319,893	2358,044	2319,739	2356,026	2318,437
20	100	-50	4729,511	2331,564	4727,458	2331,690	4725,754	2330,894
21	-100	-100	-4766,480	4643,950	-4768,340	4641,559	-4771,640	4638,361
22	-50	-100	-2402,240	4657,881	-2403,970	4655,718	-2407,420	4653,223
23	0	-100	-25,382	4670,630	-26,784	4669,417	-30,431	4667,315
24	50	-100	2353,833	4682,418	2350,767	4681,909	2348,482	4680,252
25	100	-100	4723,675	4694,336	4721,280	4693,464	4718,625	4692,886

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

3.6. CASO 15

Control de una impresión sobre papel

EL CONTROL DE LA IMPRESIÓN SOBRE PAPEL

A pesar del auge de la informática y de la información digital, el papel sigue siendo un medio de gran importancia para los productos cartográficos. El papel es un soporte de gran nobleza y calidad que permite su uso en circunstancias en las que la información digital no resulta todavía del todo competitiva. Las grandes series nacionales, las guías turísticas, mapas de carreteras, los atlas, etc., son ejemplos de productos cartográficos impresos que requieren la aplicación de técnicas de control para determinar si cumplen o no con las especificaciones marcadas en el diseño.

En el supuesto de que la información a imprimir es correcta, el control de una impresión debe atender tanto a los propios soportes (papel y tintas) como al trabajo de impresión y acabado. Este caso práctico se centrará fundamentalmente en los dos últimos aspectos destacados.

PROGRAMA DE ASEGURAMIENTO DE LA CALIDAD

A modo de ejemplo, y para entender cómo puede ser un programa completo de control de los trabajos de impresión y acabado, se presenta aquí un escueto resumen del sistema que utiliza la *U.S. Goverment Printing Office* (GPO) de los Estados Unidos. Esta Oficina estableció hace varias décadas el Programa de Aseguramiento de la Calidad Mediante Atributos (QATAP, *Quality Assurance Through Attributes Program*). Este programa se incluye dentro de las especificaciones de los contratos que firma dicha agencia con sus proveedores, estableciendo un marco para la determinación y cuantificación de los defectos y de las penalizaciones. Los principales objetivos del QATAP son:

— Establecer, previamente a la producción de un documento, los requerimientos de calidad para su impresión.

— Facilitar la determinación objetiva de la conformidad con estos requerimientos una vez que el producto ha sido impreso o recibido.

Para ello se distinguen cinco niveles de calidad del producto (NCP) (Tabla 3.11), basados en la intención de uso del mismo, que deberán ser fijados por el cliente. El nivel más exigente es el Nivel I y el de menos calidad el Nivel V. Respecto al primero,

cada nivel permite, sucesivamente, mayores desviaciones respecto a las especificaciones establecidas a priori. La elección del NCP se debe efectuar atendiendo a las necesidades de fidelidad de la reproducción, la calidad estética y la durabilidad.

Tabla 3.11 - Niveles de calidad del producto establecidos en el QATAP

NCP	Descripción general
Nivel I	"Mejor calidad", "mayor calidad". La información ha de mantener la máxima fidelidad en detalles, color y resolución. Se ha de mantener la máxima calidad en los materiales, métodos de producción y de fabricación. El acabado ha de ser de la mayor exactitud y durabilidad, y de la mejor apariencia. Ejemplos: libros de arte, catálogos de exposiciones importantes, revistas médicas...
Nivel II	"Mejor calidad, "calidad de prestigio", "calidad de biblioteca". Se requiere una elevada fidelidad a los originales. La calidad de los materiales, métodos de producción y de fabricación ha de ser elevada. El acabado ha de ser de elevada exactitud y durabilidad, y de buena apariencia. Ejemplos: anuarios, revistas profesionales ilustradas...
Nivel III	"Buena calidad", "calidad sobre la media". Las ilustraciones han de transmitir información precisa, aunque la fidelidad al más mínimo detalle no sea requerida. La calidad de los materiales, métodos de producción y de fabricación ha de estar por encima de la media. El acabado ha de ser de exactitud, durabilidad y apariencia por encima de la media. Ejemplos: informes anuales, presupuestos, libros de texto...
Nivel IV	"Calidad básica", "calidad útil". La fidelidad y resolución respecto a la copia original ha de ser de calidad media. La calidad de los materiales, métodos de producción y de fabricación ha de ser media. El acabado ha de ser de exactitud, durabilidad y apariencia media. Ejemplos: guías telefónicas, índices, informes o manuales técnicos.
Nivel V	"Calidad funcional", "calidad de utilidad reducida". Se acepta si al menos no hay pérdida de información. La calidad de los materiales, métodos de producción y de fabricación son adecuados si no se pierde información y no se limita la función del producto. Solo se incluyen trabajos de un color, de texto y líneas. Ejemplos: formatos internos, folletos...

Fuente: GPO (1979)

Para especificar (o evaluar) cuáles han de ser las características de un producto de un determinado nivel de calidad, el QATAP distingue 30 atributos, clasificados en: a) 11 atributos de impresión, b) 18 atributos de acabado, y c) 1 atributo para el papel (color, opacidad...) (ver Tabla 3.12).

Tabla 3.12 - Atributos de impresión y de acabado establecidos en el QATAP

Atributos de impresión	Atributos de acabado
1. Manchas y suciedades. 2. Marcas extrañas. 3. Moiré. 4. Registro. 5. Posición del texto e ilustraciones. 6. Anillos de Newton. 7. Calidad y uniformidad de los tipos de letra. 8. Exactitud de los semitonos. 9. Exactitud de los colores planos. 10. Exactitud de los colores de cuatricromía. 11. Resistencia al roce de la imagen impresa.	1. Tamaño de corte. 2. Emplazamiento incorrecto y desalineación de la imagen de cubierta. 3. Posición de la cubierta. 4. Posición de doblez y desvío. 5. Duración del encuadernado en estado perfecto. 6. Pérdida de cubierta, páginas y encuadernado. 7. Exceso de cola. 8. Páginas dañadas. 9. Aristas o bordes dañados. 10. Deformación de libros de tapa dura (cubierta y páginas). 11. Cubiertas dañadas. 12. Falta de páginas. 13. Cubierta cabeza abajo. 14. Página cabeza abajo. 15. Páginas en blanco (distintas a las especificadas). 16. Paginación equivocada. 17. Pérdida de información. 18. Serio desencaje en cuatricromía.

Fuente: GPO (1979)

Para cada atributo el QATAP indica su descripción, cómo se ha de medir o evaluar, y especifica los intervalos que van a determinar la importancia del valor del atributo en función del NCP fijado a priori para el producto, ya que obviamente, un mismo valor de un determinado atributo no significará lo mismo en un producto de Nivel I que en otro de Nivel IV.

No obstante, la forma de evaluar los atributos de impresión y de acabado es diferente. Los valores de los primeros son clasificados a partir de medidas objetivas, según una tabla de deméritos. En esta tabla se distinguen tres intervalos para cada NCP, cuantificando cada uno en 4, 20 o 100 deméritos. Los atributos de impresión se miden para cada página, considerándose que existe un "defecto mayor" si el nivel de demérito medio de ese atributo, evaluado para todas las páginas del producto, excede de 4,0. No existen defectos críticos en este tipo de atributos. En la Tabla 3.13 se presenta un ejemplo del atributo de impresión "error en la colocación de una figura".

En cambio, los atributos de acabado son simplemente clasificados como "defecto mayor" o "defecto crítico" en función del NCP. Por ejemplo, los atributos de "páginas invertidas", "páginas en blanco" o "error en la paginación" son considerados defectos críticos. Además pueden existir tablas de tolerancia, que distingan o no distintos NCP, que especifiquen el umbral mínimo para considerar que existe un defecto. Un ejemplo del atributo "tapas dañadas" se presenta en la Tabla 3.14.

Tabla 3.13 - Atributo de impresión "error en la colocación de una imagen"

Definiciones:

Error en la colocación de una imagen: Desplazamiento lineal de la imagen entre la posición especificada y la actual.

Instrumentos:

Examen visual.
Regla graduada hasta 0,1 mm.

Procedimiento:

Por medio de la regla, determinar la desviación horizontal y vertical de la posición de la imagen impresa respecto a la posición especificada. Asignar los deméritos a la mayor de ambas.

Tabla de deméritos

NCP	Medida	Deméritos		
		4	20	100
Nivel I	[0,8 - 1,6) mm	X		
	[1,6 - 3,2] mm		X	
	> 3,2 mm			X
Nivel II	[0,8 - 1,6) mm	X		
	[1,6 - 3,2] mm		X	
	> 3,2 mm			X
Nivel III	[1,6 - 3,2) mm	X		
	[3,2 - 4,8] mm		X	
	> 4,8 mm			X
Nivel IV	[3,2 - 4,8) mm	X		
	[4,8 - 6,4] mm		X	
	> 6,4 mm			X
Nivel V	No aplicable al Nivel V			

Tabla 3.14 - Atributo de acabado "tapas dañadas"

Definiciones:

Las tapas dañadas pueden tener uno de los siguientes problemas: arrugas, burbujas (elevaciones), cortes o rasgaduras.

Instrumentos:

Examen visual.
Plástico transparente con una malla de 6,35 mm de lado.

Procedimiento:

Colocar la malla sobre la tapa y contar cuántos cuadros cubren el área dañada. A continuación calcular cuál es el porcentaje de estos cuadros respecto al total que cubre toda la tapa. Clasificar el valor obtenido en función de la tabla de tolerancia.

Tabla de tolerancia

NCP	Medida	Clasificación	
		Mayor	Crítico
Nivel I	Cualquier arruga, burbuja o rasgadura	X	
	Cualquier corte		X
Nivel II	Cualquier arruga, burbuja o rasgadura	X	
	Cualquier corte		X
Nivel III	10% de la superficie con arrugas, burbujas o rasgaduras	X	
	Cualquier corte	X	
Nivel IV	20% de la superficie con arrugas, burbujas o rasgaduras	X	
	Cualquier corte	X	
Nivel V	30% de la superficie con arrugas, burbujas o rasgaduras	X	
	Cualquier corte	X	

Así, una vez evaluados todos los atributos se podrá establecer cuántos defectos mayores, críticos y totales (esto es, la suma de los críticos y los mayores) existen en un determinado producto.

Lógicamente, no es viable inspeccionar todos los productos de un lote de producción, ya que el coste sería prohibitivo. En su lugar, se inspecciona una muestra y se establece un umbral de aceptación. Este umbral especifica el máximo de defectos que se admitirán por cada 100 unidades consideradas satisfactorias en el lote. El QATAP emplea el procedimiento de muestreo ANSI/ASQC Z 1.4-1993, *(Sampling Procedures and Tables for Inspection by Attributes)* para determinar el tamaño de la muestra apropiado según el tamaño del lote y el umbral de aceptación. El QATAP ha fijado un umbral de 1,0 para los defectos críticos y de 6,5 para los defectos totales.

En caso de superar alguno, o ambos umbrales, el cliente tiene derecho a que se reemplace todo el lote por otro, o a que se corrijan los errores si esto es posible. En caso de que no se pueda efectuar una nueva impresión del lote, y sea necesario usarlo tal y como está, el cliente tiene derecho a un descuento económico, con un máximo del 25%, en función del número de defectos mayores y críticos detectados, según unas tablas predefinidas por la GPO.

TRABAJO A REALIZAR

Considérese que se trabaja en una empresa dedicada al negocio editorial cartográfico, y que se recibe la impresión de un lote de 10.000 ejemplares de un atlas de Andalucía formado por mapas topográficos E100K. Se ha de idear un programa de control de calidad de este producto, vinculante en el contrato, que permita decidir si se acepta o rechaza el trabajo. Las especificaciones deberán ser rigurosas, ya que se pretende poner en el mercado un producto de elevada calidad.

Este programa ha de incluir:

— Una clasificación y enumeración de las variables (o atributos) a controlar.

— Para cada variable a verificar, una breve descripción y cómo se ha de evaluar, incluyendo valores cuantitativos y/o cualitativos.

— Un método de muestreo del lote, especificando cuál o cuáles serán los umbrales de rechazo del mismo.

La especificación del procedimiento a seguir en caso de que el lote no pase el control de calidad, detallando una o varias alternativas, los plazos a cumplir y, en caso de establecer descuentos en el precio, señalar los criterios para fijar el descuento del lote.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

BLOQUE 4

CONTROL DE CALIDAD POSICIONAL

Este bloque está dedicado al control de calidad posicional. Esta componente de la calidad hace referencia a las exactitudes planimétricas y altimétricas de la misma. La exactitud posicional es indispensable en bases de datos geográficas de carácter básico (o topográfico) y aún más cuanto mayor es la escala, dado que esta información es un apoyo esencial para los trabajos de ingeniería. Una BDG de mala calidad puede causar serios perjuicios, por lo que el control de calidad posicional es ineludible en la producción cartográfica.

Todas las instituciones generadoras de cartografía deben disponer de métodos adecuados para garantizar la calidad de sus productos, que estarán diseñados y generados con un uso predefinido que determinará sus valores de exactitud. En el caso de que el producto cubra varios usos, deberá cumplir con los requerimientos del uso más restrictivo.

En el proceso cartográfico de generación de una BDG los errores pueden provenir tanto de los datos como de los procesos propiamente dichos, pero, en cualquier caso, el error se transmite y aumenta, en la mayoría de las veces, a lo largo de la cadena productiva. Existen dos formas básicas de estudiar el error final de una cartografía; utilizando una similitud con otras ciencias, llamaremos a estas formas de estudio:

— **Física.** Consistirá en estudiar todos y cada unos de los procesos de la cadena productiva, intentando comprender y acotar el error que se produce. Previamente a la realización de este tipo de estudios hace falta contar con trabajos de investigación básica en cartografía, tales que definan adecuadamente los rangos de error que se introducen en cada operación cartográfica. Un ejemplo se puede encontrar en Maling (1989) (ver Tabla 4.1). También hace falta una formulación matemática de lo que supone cada proceso que interviene en la producción.

— **Estadística.** Consiste en observar el mapa definitivo y contrastar su calidad mediante un test estadístico sobre una muestra de datos de comprobación que se consideran representativos, con casi total despreocupación de lo que ocurre por dentro de los procesos (caja negra). Ejemplos son el test NMAS (USBB, 1947) o el test EMAS (ASCI, 1983, ASP, 1985, Veregin, 1989). Los métodos estadísticos son los únicos que permiten establecer niveles de confianza. Todos los test empleados se basan en la comparación del mapa con una fuente de mayor exactitud. Ésta podrá ser bien un mapa de mayor exactitud (cartografía a escala significativamente superior y/o con una mayor calidad posicional controlada previamente), o bien datos tomados sobre el terreno con una exactitud superior a la de la cartografía a analizar. En este último aspecto, el empleo de los actuales sistemas de posicionamiento global (GPS) reducen considerablemente los costes en esta fase del proceso.

Tabla 4.1 - Componentes de error en distintas fase de la producción cartográfica

Operación	Menor valor (+/-) mm	Mayor Valor (+/-) mm
Levantamiento de control	0,005	
Trazado de control	0,01	0,32
Trazado de detalle	0,15	0,30
Compilación	0,30	0,32
Dibujo en limpio	0,06	0,18
Reproducción	0,10	0,20
Registro de color	0,17	0,30

Fuente: Maling (1989).

En suma, la posición geoespacial de todo dato geográfico tendrá una cierta incertidumbre *I* para un nivel de probabilidad o confianza $1-\alpha$. Tanto si se aplica un método físico, como si se escoge uno estadístico, es posible obtener la dispersión (mediante el error medio cuadrático, EMC) de las tres componentes posicionales: EMC_X, EMC_Y, EMC_Z. Seguidamente, considerando un determinado nivel de probabilidad o confianza, se estimará la incertidumbre; por ejemplo, para la componente X se obtendrá como $I_X = K_{(1-\alpha)} \times EMC_X$, en planimetría como $I_C = K'_{(1-\alpha)} \times EMC_C$, y en un caso tridimensional como $I_E = K''_{(1-\alpha)} \times EMC_E$. Para ello los coeficientes K, K' y K'' se determinarán respectivamente a partir de las distribuciones normal unidimensional (Figura 4.1), normal circular (Figura 4.2), y normal esférica.

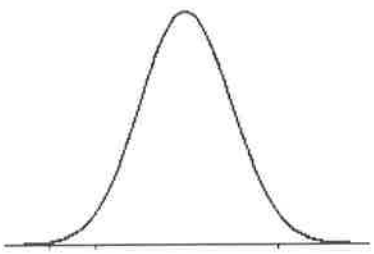


Figura 4.1 - Distribución normal unidimensional

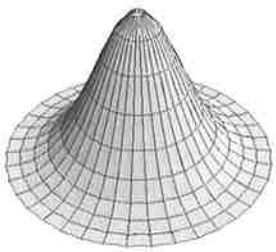


Figura 4.2 - Distribución normal circular

Así, para una componente aislada, la distribución normal N (0,1) unidimensional da las siguientes incertidumbres para las probabilidades indicadas:

— Probabilidad = 50% (error probable): $0,6745 \cdot EMC.$
— Probabilidad = 68,27% (error estándar): $1,0000 \cdot EMC.$
— Probabilidad = 90%: $1,6449 \cdot EMC.$
— Probabilidad = 95%: $1,9600 \cdot EMC.$
— Probabilidad = 99%: $2,5758 \cdot EMC.$

Cuando se combinan las componentes X e Y en una única planimétrica, se emplea la normal N (0,1) circular sobre el EMC_C (error circular estándar), definido éste como (Greenwalt y Shultz, 1962):

— $EMC_C = EMC_X = EMC_Y$ en el caso de que los errores estándar de las dos componentes sean iguales.

— $EMC_C = 1/2\ (EMC_X + EMC_Y)$, si el cociente del mayor entre el menor de las dos sea superior o igual a 0,2 y menor a 1.

Así, la distribución normal N (0,1) circular da las siguientes incertidumbres para las probabilidades indicadas:

— Probabilidad = 39,35% (error circular estándar): $1,0000 \cdot EMC_C.$
— Probabilidad = 50% (error circular probable): $1,1774 \cdot EMC_C.$
— Probabilidad = 90%: $2,1460 \cdot EMC_C.$
— Probabilidad = 95%: $2,4477 \cdot EMC_C.$
— Probabilidad = 99%: $3,0349 \cdot EMC_C.$

Cuando se combinan las componentes X, Y, Z en una única tridimensional, se emplea la normal N (0,1) esférica sobre el EMC_E (error esférico estándar), definido éste como (Greenwalt y Shultz, 1962):

— $EMC_C = EMC_X = EMC_Y = EMC_Z$ en el caso de que los errores estándar de las tres componentes sean iguales.

— $EMC_C = 1/3\ (EMC_X + EMC_Y + EMC_Z)$, si el cociente del mayor entre el menor de las tres sea superior o igual a 0,35 y menor a 1.

Así, la distribución normal N (0,1) esférica da las siguientes incertidumbres para las probabilidades indicadas:

— Probabilidad = 19,9% (error esférico estándar): $1,0000 \cdot EMC_C.$
— Probabilidad = 50% (error esférico probable): $1,5382 \cdot EMC_C.$
— Probabilidad = 90%: $2,5003 \cdot EMC_C.$
— Probabilidad = 95%: $2,7955 \cdot EMC_C.$
— Probabilidad = 99%:. $3,3683 \cdot EMC_C.$

El primer caso de este bloque está dedicado al control de calidad posicional de carácter físico (estudio teórico), mientras las tres restantes tienen por objetivo aplicar diferentes estándares de carácter estadístico:

— Caso 16: Estudio teórico de la calidad posicional de una cartografía. Se estudia la transmisión del error posicional en un proceso cartográfico muy común hasta hace pocos años, que parte de la captura del dato en campo, hasta su digitalización en tablero. Para ello se emplea la expresión diferencial para la propagación de la varianza.

— Caso 17:.Aplicación del estándar NMAS. Se aplica este estándar para estudiar la calidad posicional de una cartografía a partir de una serie de coordenadas del mapa y del terreno, procediendo estas últimas de una fuente de mayor exactitud.

— Caso 18: Aplicación del estándar EMAS. Con un discurso paralelo al caso anterior, aquí se aplica este estándar sobre los mismos datos. El EMAS favorece un análisis más profundo de las características estadísticas de la información geográfica.

— Caso 19: Aplicación del estándar NSSDA. Presenta uno de los últimos estándares que se están empleando, que no consiste en una prueba de aceptación o rechazo, sino en un indicador de la calidad posicional, que deja la decisión de aceptación al usuario final.

— Caso 20: Control de una rectificación. Se expone la problemática de la georreferenciación mediante rectificación polinómica y de su posterior control para determinar la bondad del ajuste, empleando métodos de validación cruzada.

Junto a las sencillas medidas de tendencia central y de dispersión, las herramientas estadísticas a utilizar en este bloque son aquellas relativas a la inferencia estadística, tanto paramétrica (contrastes de hipótesis sobre la media, sobre la varianza, bondad del ajuste...) como no paramétrica (pruebas de aleatoriedad...), eliminación de outliers, etc. También se emplearán algunas distribuciones de probabilidad (Normal, t de Student, χ^2 de Pearson) y la bondad del ajuste a la normal. Todas ellas se pueden consultar en el Anejo de Estadística. Las fuentes de este bloque son las distintas publicaciones en las que se han publicado cada uno de los test: USBB (1947), ASCI (1983), FGDC (1998), que quedan recopiladas en Ariza (2002).

4.1. CASO 16

Estudio teórico de la calidad posicional de una cartografía

LA PROPAGACIÓN DEL ERROR

El planteamiento general de propagación del error posicional va a venir dado por la estimación del valor de la varianza de la coordenada X ó Y.

Así, dada una función $A = f(t_1, t_2, t_3, ..., t_n)$, su varianza puede tomarse como:

$$S_A^2 = \frac{1}{n}\left[\left(\frac{\partial A}{\partial t_1}\right)_{t_1}\Delta t_1 + \left(\frac{\partial A}{\partial t_2}\right)_{t_2}\Delta t_2 + \cdots + \left(\frac{\partial A}{\partial t_n}\right)_{t_n}\Delta t_n\right]^2$$

Operando se tiene que:

$$S_A^2 = \frac{1}{n}\left(\frac{\partial A}{\partial t_1}\right)_{t_1}^2 \Delta t_1^2 + \frac{1}{n}\left(\frac{\partial A}{\partial t_2}\right)_{t_2}^2 \Delta t_2^2 + \cdots + \frac{1}{n}\left(\frac{\partial A}{\partial t_n}\right)_{t_n}^2 \Delta t_n^2 + \cdots$$

$$\cdots + \frac{2}{n}\left[\left(\frac{\partial A}{\partial t_1}\right)_{t_1}\left(\frac{\partial A}{\partial t_2}\right)_{t_2}\Delta t_1 \Delta t_2 + \left(\frac{\partial A}{\partial t_1}\right)_{t_1}\left(\frac{\partial A}{\partial t_3}\right)_{t_3}\Delta t_1 \Delta t_3 + \cdots\right]$$

Si se supone que existe independencia entre las distintas variables, las covarianzas serán nulas y se llegará a siguiente la expresión, que será la que utilizaremos para elaborar este caso:

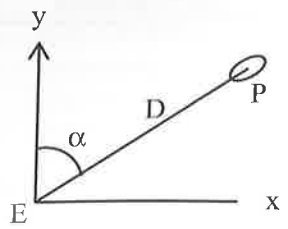
$$S_A^2 = \left(\frac{\partial A}{\partial t_1}\right)_{t_1}^2 S_{t_1}^2 + \left(\frac{\partial A}{\partial t_1}\right)_{t_2}^2 S_{t_2}^2 + \cdots + \left(\frac{\partial A}{\partial t_n}\right)_{t_n}^2 S_{t_n}^2$$

PROCESO CARTOGRÁFICO

Se va a considerar un proceso cartográfico muy común hasta hace poco, consistente en el conjunto de operaciones que van desde la captura en campo del dato, hasta su digitalización en tablero, pasando por un proceso previo de impresión en offset[1]. Este caso ejemplifica muy simplificadamente el proceso seguido por muchos datos e

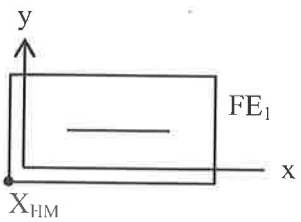
[1] Los mapas impresos que se suelen adquirir normalmente están impresos mediante la técnica denominada offset. Esta técnica permite la impresión económica, y en color, de grandes tiradas y por ello se utiliza en cartografía, en la impresión convencional de libros y otros productos editoriales.

información que se han incorporado a los SIG procedentes de soportes analógicos. Para facilitar los cálculos sólo se trabajará con la componente X. Los pasos son:



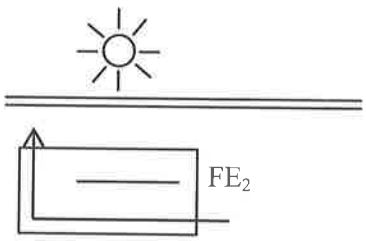
a) **Topografía.** Se toma la posición de un punto P correspondiente al centro geométrico del brocal de un pozo mediante la medida de un acimut α y una distancia D. La ecuación que permite determinar su coordenada X_P, a partir de la coordenada X_E del punto de estación es:

$$X_P = X_E + D\ Sen\ (\alpha)$$



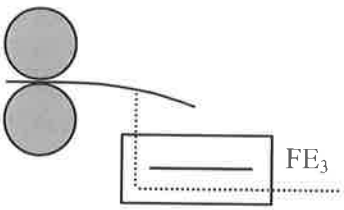
b) **Puesta en escala y referencia al origen del mapa.** Se calcula el incremento de coordenada ΔX_P respecto al origen de la hoja del mapa sobre la que se representa X_{HM}, lo cual se puede expresar como una transformación lineal dependiente de un factor de escala FE_1:

$$\Delta X_P = (X_P - X_{HM})\ FE_1$$



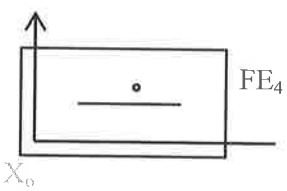
c) **Paso a la plancha de impresión.** Mediante un proceso de manipulación fotomecánico, y sin uso alguno de técnicas de generalización, se pasa de la minuta de compilación anterior a una plancha de impresión según una transformación lineal como la siguiente:

$$X_{Pl} = \Delta X_P\ FE_2$$



d) **Impresión en papel.** También puede considerarse una transformación lineal:

$$X_{Im} = X_{Pl}\ FE_3$$



e) **Digitalización manual.** Finalmente, para obtener las coordenadas de P en formato digital X_D, se calibra el proceso con el origen inferior izquierdo del mapa X_o, y se procede a la digitalización manual sobre papel, lo que también puede considerarse una transformación lineal:

$$X_D = X_{Im}\ FE_4 + X_o$$

TRABAJO A REALIZAR

Aplicar el desarrollo matemático necesario para calcular la desviación típica del error total del proceso cartográfico presentado en el epígrafe anterior. Supóngase que los errores que se producen en los distintos procesos son independientes. Los valores y desviaciones de todas las variables y factores de escala a utilizar figuran en la Tabla 4.2.

Tabla 4.2 - Valores y desviaciones de las distintas variables y factores de escala

Variable	Significado	Valor	Desviación
X_o	Origen inferior izquierdo de la hoja del mapa, digitalizado con tableta.	123500,00 m	0,02 m
X_E	Coordenada X del punto de estación.	123456,00 m	0,03 m
D	Distancia medida en levantamiento.	172,50 m	0,07 m
α	Acimut medido en levantamiento.	0,5236 rad	0,00002 rad
X_{HM}	Origen inferior izquierdo de la hoja del mapa. Decisión arbitraria.	123500,00 m	0,0 m
FE_1	Factor de escala en el paso del terreno al mapa	0,0004	0,0000002
FE_2	Factor de escala en el paso del mapa a la plancha de impresión	1,000	0,0002
FE_3	Factor de escala de paso de plancha a papel	1,000	0,0005
FE_4	Factor de escala paso de mapa a terreno	2500,000	0,0005

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

4.2. CASO 17

Aplicación del estándar NMAS

EL ESTÁNDAR NMAS

El test NMAS (*National Map Accuracy Standard*) (USBB, 1947) ha sido el método de control empleado por las administraciones americanas productoras de cartografía desde 1947 hasta hace pocos años. Esto ha implicado la extensión de su uso a otras muchas instituciones y organismos oficiales dedicados a la cartografía, de muchos otros países. La mayor ventaja del test NMAS es su simplicidad de cálculo y su facilidad de entendimiento dado que los resultados se expresan de la forma cumple/no cumple, de cómoda interpretación por el usuario.

El test comprueba si el 10% de los puntos de la muestra tiene un error horizontal (componentes x,y conjuntas) mayor de 0,085 cm (1/30") en mapas de escala mayor de E20K ó de 0,05 cm (1/50") en mapas a escala menor o igual a E20K.

También se puede aplicar sobre la componente vertical. En este caso se comprueba si el 10% de los puntos de la muestra tienen un error vertical mayor de la mitad del intervalo de las curvas de nivel. Un aspecto controvertido es el poder corregir el error vertical de un punto actuando sobre su componente horizontal, modificándola en una cantidad igual al error horizontal aceptable. Esto hace que el estándar de exactitud sea más fácilmente alcanzable en superficies de fuerte pendiente.

No obstante, las desventajas del empleo de NMAS son numerosas. La principal es la falta de indicación de una medida de la inexactitud presente en el mapa. No se sabe nada sobre el comportamiento estadístico de los errores y, por tanto, es difícil estimar cuánto puede costar obtener realmente una BDG de mayor exactitud. Además, es un test ligeramente permisivo dado que los umbrales estimados en la aceptación o no de la cartografía son realmente amplios (Atkinson, García y Ariza, 2001b).

El proceso para aplicar el test NMAS consta de los siguientes pasos:

1. Seleccionar una muestra de n puntos, siendo $n \geq 20$.

2. Calcular el error de cada punto en cada componente:

$$e_{x_i} = x_{t_i} - x_{m_i} \qquad e_{y_i} = y_{t_i} - y_{m_i} \qquad e_{z_i} = z_{t_i} - z_{m_i}$$

donde:

$x_{t_i}, y_{t_i}, z_{t_i}$ Coordenadas sobre el terreno (u otra fuente de mayor precisión).

$x_{m_i}, y_{m_i}, z_{m_i}$ Coordenadas sobre el mapa.

3. Calcular la componente horizontal de los errores en x, y en cada punto:

$$e_{H_i} = \sqrt{e_{X_i}^2 + e_{Y_i}^2}$$

4. Establecer cuáles son los errores máximos tolerables en las componentes horizontales y verticales:
   - Horizontal: 0,085 cm (1/30") en mapas de escala mayor de E20K o de 0,05 cm (1/50") en mapas a escala menor o igual a E20K.
   - Vertical: La mitad de la equidistancia entre curvas de nivel.

5. Contar cuántos puntos tienen un error horizontal superior al error máximo tolerable para determinar si se ha alcanzado el estándar de exactitud posicional predefinido.

   El test es superado en la componente horizontal si como máximo el 10% de los puntos tienen un error superior al error máximo tolerable.

6. Contar cuántos puntos tienen un error vertical, e_z, superior al error máximo tolerable para determinar si se ha alcanzado el estándar de exactitud posicional predefinido.

   El test es superado en la componente vertical si como máximo el 10% de los puntos tienen un error superior al error máximo tolerable.

Además, dado que para el desarrollo de los métodos estadísticos se toma como premisa la normalidad y aleatoriedad de la distribución de los errores, siempre son recomendables los estudios estadísticos previos que corroboren o no dichas presunciones.

EJEMPLO DE APLICACIÓN DEL NMAS

Se tienen dos conjuntos de coordenadas (x,y) en la Tabla[2] 4.3, el primero procedente de una serie de puntos de una cartografía a escala E10K, y el segundo procedente de sus homólogos sobre el terreno. Por tanto el límite de error horizontal será de 0,085 cm a escala E10K, esto es, 8,5 m.

2 Nota importante: por comodidad se ha utilizado un total de 20 puntos (mínimo marcado por el estándar), aunque el número de puntos debería ser algo mayor, del orden de unos 50 puntos.

Tabla 4.3 - Conjunto de puntos a analizar (ejemplo)				
Punto	**X terreno [m]**	**Y terreno [m]**	**X mapa [m]**	**Y mapa [m]**
1	1689,369	878,295	1690,299	879,472
2	1680,428	861,671	1678,944	864,046
3	1792,548	880,128	1795,534	874,487
4	1835,400	733,400	1837,636	737,800
5	3263,700	750,700	3259,424	746,327
6	2424,000	1639,200	2419,734	1634,246
7	2424l,500	1648,900	2416,279	1645,183
8	1842,171	1679,171	1842,480	1672,258
9	1858,348	1537,628	1861,937	1535,444
10	2001,400	1281,200	2000,769	1285,554
11	1996,969	1271,504	1995,005	1267,338
12	1874,400	1184,500	1873,557	1183,637
13	962,500	1381,800	963,643	1378,497
14	1108,700	1559,500	1104,826	1558,200
15	533,804	1497,930	530,917	1495,828
16	197,128	1294,148	197,259	1293,236
17	406,700	1020,900	404,128	1017,812
18	3123,200	1891,500	3124,926	1906,140
19	3152,700	1883,400	3153,457	1889,217
20	3190,200	1889,300	3183,901	1893,882

Primeramente se determinan las diferencias entre las coordenadas del mapa y del terreno, y se calcula la componente horizontal de estas diferencias (Tabla 4.4).

Tabla 4.4 - Diferencias entre coordenadas-mapa y coordenadas-terreno para los datos de ejemplo mostrados en la Tabla 4.3			
Punto	**e_X [m]**	**e_Y [m]**	**e_H [m]**
1	-0,930	-1,177	1,500
2	1,484	-2,375	2,801
3	-2,986	5,641	6,383
4	-2,236	-4,400	4,936
5	4,276	4,373	6,116
6	4,266	4,954	6,538
7	7,221	3,717	8,122
8	-0,309	6,913	6,920
9	-3,589	2,184	4,201
10	0,631	-4,354	4,399
11	1,964	4,166	4,606
12	0,843	0,863	1,206
13	-1,143	3,303	3,495
14	3,874	1,300	4,086
15	2,887	2,102	3,571
16	-0,131	0,912	0,921
17	2,572	3,088	4,019
18	**-1,726**	**-14,640**	**14,741**
19	-0,757	-5,817	5,866
20	6,299	-4,582	7,789

A la vista de los errores horizontales, se verifica que sólo un único punto (punto 18) supera el error máximo tolerable. Por la tanto se supera el test de exactitud posicional NMAS.

TRABAJO A REALIZAR

Para una cartografía E50k de la que se han extraído los puntos de control cuyas coordenadas UTM se incluye en la Tabla 4.5, se pide determinar si superará el test NMAS, mediante el siguiente proceso completo de análisis:

- Eliminación de *outliers*.
- Estudio de la aleatoriedad de la muestra.
- Estudio de la normalidad de los datos.
- Si la muestra es aleatoria y se ajusta a una distribución normal, aplicación del NMAS.

Una vez concluido el proceso de cálculo se ha de emitir un juicio crítico, aunque, como se ha comentado será más limitado al no disponer de una medida de la exactitud de la muestra de puntos.

Tabla 4.5 - Coordenadas mapa y terreno del conjunto de puntos a analizar

Punto	X terreno [m]	Y terreno [m]	X mapa [m]	Y mapa [m]
1	429689,369	4201878,295	429690,299	4201875,472
2	429680,428	4201861,671	429678,944	4201860,046
3	429792,548	4201880,128	429795,534	4201870,487
4	429835,400	4201733,400	429837,636	4201733,800
15	431263,700	4201750,700	431259,424	4201742,327
17	430424,000	4202639,200	430419,734	4202630,246
18	430424,500	4202648,900	430416,279	4202641,183
19	429842,171	4202679,171	429842,480	4202668,258
20	429858,348	4202537,628	429861,937	4202531,444
21	430001,400	4202281,200	430000,769	4202281,554
22	429996,969	4202271,504	429995,005	4202263,338
23	429874,400	4202184,500	429873,557	4202179,637
24	428962,500	4202381,800	428963,643	4202374,497
25	429108,700	4202559,500	429104,826	4202554,200
26	428533,804	4202497,930	428530,917	4202491,828
27	428197,128	4202294,148	428197,259	4202289,236
28	428406,700	4202020,900	428404,128	4202013,812
121	431123,200	4202891,500	431124,926	4202902,140
122	431152,700	4202883,400	431153,457	4202885,217
123	431190,200	4202889,300	431183,901	4202889,882
124	431354,829	4203012,370	431356,120	4203010,263
125	431226,371	4203504,671	431224,635	4203500,006
126	431236,700	4203544,700	431232,621	4203536,733
127	431405,828	4203501,751	431369,801	4203512,623
128	431775,551	4203573,271	431777,038	4203571,233

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

4.3. CASO 18

Aplicación del estándar EMAS

EL ESTÁNDAR EMAS

El test EMAS (*Engineering Map Accuracy Standard*) (ASCI, 1983, ASP, 1985, Veregin, 1989) especifica la exactitud de los mapas topográficos a gran escala. Compara coordenadas del mapa con fuentes de mayor exactitud evaluando las componentes X, Y, Z separadamente. Para ello se sirve de una expresión estadística de exactitud sobre una muestra de, al menos, 20 puntos.

Efectúa un test sobre la media, bajo el supuesto de varianza poblacional desconocida, para determinar si el error sistemático se encuentra entre los límites aceptables, y un test sobre la varianza para determinar si el error aleatorio es aceptable.

El proceso para aplicar el test EMAS consta de los siguientes pasos:

1. Seleccionar una muestra de n puntos, siendo $n \geq 20$.

2. Calcular el error de cada punto en cada componente:

$$e_{x_i} = x_{t_i} - x_{m_i} \qquad e_{y_i} = y_{t_i} - y_{m_i} \qquad e_{z_i} = z_{t_i} - z_{m_i}$$

donde:

$x_{t_i}, y_{t_i}, z_{t_i}$ Coordenadas sobre el terreno (u otra fuente de mayor precisión).

$x_{m_i}, y_{m_i}, z_{m_i}$ Coordenadas sobre el mapa.

3. Calcular el error medio en cada componente:

$$\bar{e}_x = \frac{1}{n}\sum_{i=1}^{n} e_{x_i} \qquad \bar{e}_y = \frac{1}{n}\sum_{i=1}^{n} e_{y_i} \qquad \bar{e}_z = \frac{1}{n}\sum_{i=1}^{n} e_{z_i}$$

4. Calcular la desviación típica muestral en cada componente:

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(e_{x_i} - \bar{e}_x\right)^2} \quad S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(e_{y_i} - \bar{e}_y\right)^2} \quad S_z = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(e_{z_i} - \bar{e}_z\right)^2}$$

5. Efectuar, para cada componente, el test de cumplimiento del estándar para determinar si el error sistemático se encuentra entre los límites aceptables.

   Para ello se realizará un test sobre la media, bajo el supuesto de varianza poblacional desconocida, estableciendo las siguientes hipótesis entre las medias:

$$H_0: \mu_p = 0 \qquad H_1: \mu_p \neq 0$$

El mapa superará el test con un nivel de significación α si se cumple que:

$$|t_x| \leq t_{n-1,\,\alpha/2} \qquad |t_y| \leq t_{n-1,\,\alpha/2} \qquad |t_z| \leq t_{n-1,\,\alpha/2}$$

donde:

$t_{n-1,\,\alpha/2}$ Valor teórico de la distribución t de Student, con $n-1$ grados de libertad.

t_x, t_y, t_z Resultado de calcular los siguientes estadísticos:

$$t_x = \frac{\bar{e}_x \sqrt{n}}{S_x} \qquad t_y = \frac{\bar{e}_y \sqrt{n}}{S_y} \qquad t_z = \frac{\bar{e}_z \sqrt{n}}{S_z}$$

6. Efectuar, para cada componente, el test de cumplimiento del estándar para determinar si el error aleatorio se encuentra entre los límites aceptables.

   Para ello se realizará un test sobre la varianza, estableciendo las siguientes hipótesis, en relación a un máximo valor de varianza, a^2, preestablecido y concretado sobre cada componente: a_x^2, a_y^2, a_z^2.

$$H_0: \sigma^2_p \leq a^2 \qquad H_1: \sigma^2_p > a^2$$

   El mapa superará el test con un nivel de significación α si se cumple que:

$$\chi^2_x \leq \chi^2_{n-1,\,\alpha} \qquad \chi^2_y \leq \chi^2_{n-1,\,\alpha} \qquad \chi^2_z \leq \chi^2_{n-1,\,\alpha}$$

   siendo:

$\chi^2_{n-1,\,\alpha}$ Valor teórico de la distribución Chi cuadrado, con $n-1$ grados de libertad.

$\chi^2_x, \chi^2_y, \chi^2_z$ Resultado de calcular los siguientes estadísticos:

$$\chi^2_x = \frac{S_x^2 (n-1)}{a_x^2} \qquad \chi^2_y = \frac{S_y^2 (n-1)}{a_y^2} \qquad \chi^2_z = \frac{S_z^2 (n-1)}{a_z^2}$$

De esta forma, en un control exclusivamente planimétrico, bajo el supuesto de independencia de las componentes X e Y, y un nivel de significación $\alpha = 5\%$, dado que se efectúan cuatro contrastes de hipótesis, sólo pasarán conjuntamente un 81.5% de los casos [$0{,}95 \times 0{,}95 \times 0{,}95 \times 0{,}95 = 0{,}815$]. Este comportamiento hace que el test EMAS sea demasiado restrictivo según el planteamiento propuesto, lo que puede generar problemas tanto al productor como al consumidor. Para el primero se pueden rechazar un gran número de productos que son correctos y el segundo tiene problemas en aceptarlos por lo que deberá pagar más para obtener la calidad que desea.

Una racionalización del comportamiento descrito del EMAS pasa por la aplicación de niveles de significación en las diferentes hipótesis que tengan en cuenta una significación global $\alpha = 5\%$. Así pueden adoptarse los criterios de Bonferroni (García y Martín, 1997).

No hay que olvidar que, como en el caso práctico anterior, dado que para el desarrollo de los métodos estadísticos se toma como premisa la normalidad y aleatoriedad de la distribución de los errores, son recomendables los estudios estadísticos que corroboren o no dichas presunciones.

EJEMPLO DE APLICACIÓN DEL EMAS

Se aplicará el test EMAS sobre la misma serie de coordenadas x,y del caso práctico anterior (Tabla 4.3), que procedían de una cartografía a escala E10K. Suponiendo que el máximo error gráfico sobre el mapa impreso se establece en 0,2 mm, el error radial máximo de los datos es de 2 m. Tomando un nivel de confianza del 95%, en la distribución de la Normal (0,1) circular se obtiene un valor de 2,4477, con lo que el error medio cuadrático en cada componente, a_x, a_y, debería ser de 2 / 2,4477 = 0,817 m. Seguidamente se aplicará el test EMAS con un nivel de confianza del 95% en cualquier hipótesis estadística.

Primeramente se obtienen las diferencias entre ambos conjuntos de coordenadas y se calculan algunas medidas estadísticas básicas útiles, como la media, mediana, desviación típica muestral, máximo y mínimo (Tabla 4.6).

Tabla 4.6 - Diferencias entre coordenadas-mapa y coordenadas-terreno (ejemplo)

Punto	e_X [m]	e_Y [m]
1	-0,930	-1,177
2	1,484	-2,375
3	-2,986	5,641
4	-2,236	-4,400
5	4,276	4,373
6	4,266	4,954
7	7,221	3,717
8	-0,309	6,913
9	-3,589	2,184
10	0,631	-4,354
11	1,964	4,166
12	0,843	0,863
13	-1,143	3,303
14	3,874	1,300
15	2,887	2,102
16	-0,131	0,912
17	2,572	3,088
18	-1,726	-14,640
19	-0,757	-5,817
20	6,299	-4,582

n	20	20
Media	1,126	0,309
Mediana	0,737	1,701
S muestral	3,015	5,130
Máximo	7,221	6,913
Mínimo	-3,589	-14,640

Se puede observar cómo las medidas de tendencia central (media, mediana) se mantienen por debajo de 2 m, pero con una desviación típica muy elevada, lo cual nos anuncia la existencia de errores de tipo sistemático, moderados, y de errores aleatorios elevados. Los máximos y mínimos advierten de la existencia de puntos del mapa que van a diferir en gran medida con los del terreno.

Seguidamente se aplica el test sobre la media para los sistematismos:

$$t_x = \frac{\bar{e}_x \sqrt{n}}{S_x} = 1,6693862$$

$$t_y = \frac{\bar{e}_y \sqrt{n}}{S_y} = 0,26900146$$

La distribución de la t de student para un contraste bilateral, con 19 grados de libertad y con un nivel de confianza del 95% es:

$$t_{19;0,025} = 2,0930247$$

Luego se supera el test de sistematismos porque:

$$|t_x| \leq t_{19,\ 0,025}$$
$$|t_y| \leq t_{19,\ 0,025}$$

A continuación de aplica el test sobre la varianza, para los errores aleatorios.

$$\chi_x^2 = \frac{S_x^2 (n-1)}{a_x^2} = 258,692965$$

$$\chi_y^2 = \frac{S_y^2 (n-1)}{a_y^2} = 748,936015$$

De la distribución Chi cuadrado para un contraste de hipótesis de una cola, con 19 grados de libertad y con un nivel de confianza del 95% se obtiene un valor de:

$$\chi^2_{19;0,05} = 30,1435051$$

Luego no se supera el test de errores aleatorios porque:

$$\chi^2_x > \chi^2_{19;0,05}$$
$$\chi^2_y > \chi^2_{19;0,05}$$

Por tanto, no se supera el test EMAS debido a la presencia de fuertes errores aleatorios, muy por encima del umbral de aceptación del test. Puede pensarse que esta

muestra no pertenece realmente a una población con una desviación típica en cada componente de 0,817 m.

Estos resultados contrastan con los resultados del test NMAS del ejemplo del caso práctico anterior, ya que el estándar de exactitud NMAS se alcanzaba sin problemas.

Además, se detectan grandes discrepancias entre los estadísticos calculados para la componente X y la componente Y. La componente X tiene una presencia de errores sistemáticos mucho mayor que la componente Y, y a su vez la componente Y tiene una presencia de errores aleatorios mucho mayor que la componente X. Sin embargo, en un mapa la distribución del error posicional ha de ser homogénea. Por tanto estos resultados provocan desconfianza acerca de los métodos empleados para la elaboración de la cartografía bajo estudio.

TRABAJO A REALIZAR

Dada la serie de coordenadas UTM propuesta en el caso práctico anterior (Tabla 4.5), se pide determinar si superará el test EMAS. La cartografía está a escala E50k y se considera el error gráfico, y por tanto el error máximo admisible, igual a 0,2 mm (al 95% de confianza). En cualquier hipótesis estadística que se considere se establecerá el nivel de confianza en el 95%.

Una vez concluido el proceso de cálculo se dispone de gran cantidad de información para emitir un juicio crítico. Es conveniente estudiar la aleatoriedad, la normalidad, el error sistemático y el aleatorio (en su caso), para cada componente analizada, observando todos los parámetros estadísticos que se crean convenientes (media, desviación típica...), la homogeneidad entre las dos componentes, etc.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

4.4. CASO 19

Aplicación del estándar NSSDA

EL ESTÁNDAR NSSDA

Hasta hace poco, los estándares de exactitud posicional, como el NMAS o EMAS, estaban concebidos para su aplicación sobre cartografía en papel, no sobre información digital. El NSSDA (*National Standard for Spatial Data Accuracy*) es un estándar reciente, propuesto por el *Federal Geographic Data Committee* (FGDC, 1998) y puede emplearse para datos cartográficos analógicos y/o digitales (Minnesota Planning, 1999). Este estándar viene a sustituir al NMAS y al EMAS. Esta norma es de obligado cumplimiento para las agencias federales de los Estados Unidos que realicen labores de producción cartográfica.

Este método permite obtener un índice de calidad posicional en unidades reales, sobre el terreno, pero no dice si el mapa es aceptado o rechazado, tal y como hacían los test anteriores. El resultado determinado con el NSSDA deberá aparecer como metadato en el producto. En este caso, le corresponde al usuario de la cartografía, o agencia que solicita la cartografía, el establecer los límites de aceptación en función de sus necesidades, lo cual constituye su mayor ventaja. Como inconveniente, presupone que se han eliminado los errores sistemáticos, hecho que no podemos constatar en una cartografía si no verificamos la presencia de este tipo de errores (por ejemplo, mediante el test sobre la media que se realiza en el EMAS).

El NSSDA analiza tanto la componente horizontal (x,y de forma conjunta, como el NMAS), como la componente vertical. Se fundamenta en el cálculo del error medio cuadrático (EMC) de la muestra. Con la obtención del EMC horizontal, por un lado, y vertical, por otro, calcula el error de la muestra analizada en función de un determinado nivel de confianza impuesto por el usuario (generalmente el 95 %).

El proceso para aplicar el test NSSDA consta de los siguientes pasos:

1. Seleccionar una muestra de n puntos, siendo $n \geq 20$.

2. Calcular el error de cada punto en cada componente:

$$e_{x_i} = x_{t_i} - x_{m_i} \qquad e_{y_i} = y_{t_i} - y_{m_i} \qquad e_{z_i} = z_{t_i} - z_{m_i}$$

donde:

$x_{t_i}, y_{t_i}, z_{t_i}$ Coordenadas sobre el terreno (u otra fuente de mayor precisión).

$x_{m_i}, y_{m_i}, z_{m_i}$ Coordenadas sobre el mapa.

3. Calcular el error medio cuadrático de cada componente:

$$EMC_X = \sqrt{\frac{\Sigma e_{X_i}^2}{n}} \qquad EMC_Y = \sqrt{\frac{\Sigma e_{Y_i}^2}{n}} \qquad EMC_Z = \sqrt{\frac{\Sigma e_{Z_i}^2}{n}}$$

5. Obtener el valor $NSSDA_H$ horizontal según una de las siguientes expresiones:

si $EMC_X = EMC_Y$	$NSSDA_H = 2{,}4477^{1/2} \cdot EMC_r = 2{,}4477 \cdot EMC_X$ donde: $EMC_r = \sqrt{EMC_X^2 + EMC_Y^2}$
si $EMC_X \neq EMC_Y$ y $0{,}6 < (EMC_{min} / EMC_{max}) < 1{,}0$	$NSSDA_H = 2{,}4477 \cdot 0{,}5 \cdot (EMC_X + EMC_Y)$

6. Obtener el valor $NSSDA_Z$ vertical según la siguiente expresión:

$$NSSDA_Z = 1{,}9600 \cdot EMC_Z$$

7. Incluir en junto a la leyenda del mapa:

   "Se ha verificado una exactitud horizontal de _ metros al 95% de nivel de confianza".

   "Se ha verificado una exactitud vertical de _ metros al 95% de nivel de confianza".

De nuevo, como en los estándares NMAS y EMAS, conviene verificar que se cumplen las condiciones de aleatoriedad y normalidad de manera previa a la aplicación del NSSDA.

EJEMPLO DE APLICACIÓN DEL NSSDA

Se aplicará el test NSSDA sobre la serie de coordenadas de la Tabla 4.3 empleada en los ejemplos de los casos anteriores relativos a los estándares NMAS y EMAS.

El primer paso es determinar las diferencias entre las coordenadas del mapa y del terreno (Tabla 4.6). A continuación se calculan el EMC_X y el EMC_Y, para derivar el EMC_r:

$$EMC_X = \sqrt{\frac{\Sigma_{i=1}^{n} e_{X_i}^2}{n}} = 3{,}147\ m \qquad EMC_Y = \sqrt{\frac{\Sigma_{i=1}^{n} e_{Y_i}^2}{n}} = 5{,}009\ m$$

$$EMC_r = \sqrt{EMC_X^2 + EMC_Y^2} = 5{,}916\ m$$

Dado que $EMC_X \neq EMC_Y$ y $0,6 < (EMC_{min} / EMC_{max}) < 1,0$, el valor de $NSSDA_H$ se calcula de la siguiente forma:

$$NSSDA_H = 2,4477 \cdot 0,5 \cdot (EMC_X + EMC_Y) = 9,982\ m$$

Por tanto, a la vista del valor del $NSSDA_H$ obtenido, hay que cuestionar la calidad posicional del mapa analizado, que recordemos se correspondía con una cartografía impresa a escala E10k. Errores de la magnitud obtenida son perfectamente visibles a esa escala. No obstante, como se ha referido, la utilidad de estos datos vendrá dada por el uso al que vayan destinados.

TRABAJO A REALIZAR

Dada la serie de coordenadas UTM de la Tabla 4.5, se pide aplicar el estándar NSSDA a esta cartografía en su componente horizontal.

Una vez concluido el proceso de cálculo se ha de emitir un juicio crítico, comentando además para qué escalas serían validos estos datos (en función del error gráfico máximo considerado, que vendrá determinado por el uso al que esté destinada esta información).

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

4.5. CASO 20

Control de una rectificación

RECTIFICACIÓN POLINÓMICA

Se denomina rectificación al proceso de someter una BDG (mapa vectorial o imagen) a una transformación polinómica, de carácter global[3], que permite pasar de un espacio o sistema de coordenadas A, a otro espacio o sistema B. La rectificación puede ser el simple registro de una imagen (coordenadas de la imagen en el espacio imagen, es decir, en filas y columnas) a un mapa (coordenadas en una proyección cartográfica), de una imagen a otra imagen, etc.

Un caso particular de la rectificación es la georreferenciación o acción de georreferenciar, que se refiere a cualquier transformación de carácter geométrico, a la que se somete una BDG, para dotarla de un sistema de referencia cartográfico.

La importancia de este proceso es grande, dado que permite la integración de informaciones procedentes de diferentes sensores o fuentes, de épocas distintas, facilitando el análisis de cambios del territorio y la sinergia por la integración de fuentes de información. Una mala georreferenciación degrada el uso del contenido de la información dado que impide la correcta superposición y mosaico de las imágenes, o de éstas con otras BDG.

Respecto a los polinomios a utilizar para la rectificación, es de gran interés práctico el saber qué tipo y grado de polinomio utilizar en estas transformaciones. En primer lugar conviene tener claro que los métodos polinómicos sólo son adecuados para gestionar BDG con poca distorsión[4]. Por lo general, en las transformaciones globales siempre se recomiendan los polinomios de menor grado, tanto por la economía de puntos como por evitar la aparición de los extraños que generan los polinomios de mayor grado (tercero o superior). En relación a esto último, siempre se ha de cuidar que el área de interés esté en el interior del perímetro convexo formado por los puntos de control para evitar los problemas de las extrapolaciones. De cualquier forma, el ajuste polinómico es empírico y, por tanto, posee poco o ningún sentido físico. Los coeficientes se determinan, por ajuste mínimo cuadrático, a partir de las parejas de

[3] La transformación es global por que el mismo polinomio se aplica a todo el dominio espacial de la BDG. No existen transformaciones particulares para cada una de las zonas que pudieran considerarse en dicho espacio.

[4] Ello implica que no son adecuados, por ejemplo, para su uso con imágenes procedentes de sensores aerotransportados.

puntos conocidos, identificados inequívocamente y con posiciones conocidas en ambos espacios, el de partida y el de llegada.

El número mínimo, N_{min}, de puntos de control necesarios para determinar todos los coeficientes del polinomio, igual al número de incógnitas, es de $N_{min} = (g+1)(g+2)/2$, donde g es el grado del polinomio. Cuando el número N de puntos de control es igual a N_{min} no hay error residual en los puntos de control. Cuando el número de puntos de control es superior al de incógnitas, los coeficientes del polinomio son aproximaciones estadísticas, existen $F = N - N_{min}$ grados de libertad y hay errores residuales en cada uno de los puntos de control utilizados. Se suele recomendar utilizar en los ajustes un número de puntos que sea de dos a tres veces el número mínimo marcado por el grado del polinomio. Cuanto mayor sea el orden del polinomio, mayor capacidad del mismo para modelar comportamientos complejos, como ocurre con los polinomios bivariados.

CONTROL DE LA RECTIFICACIÓN

Respecto al uso de los puntos de control como forma de determinar la bondad del ajuste existe cierta controversia. La actuación usual es intentar ajustar el modelo de manera tal que el EMC global sea inferior a un umbral prestablecido (por ejemplo, para imágenes, el tamaño nominal de la celda). Además, en este caso, en el que sólo se considera como parámetro de bondad del ajuste el EMC global, se suelen eliminar de dicho ajuste aquellos puntos con mayor error residual. Según Fogel (1996) la eliminación de un punto mejorará el ajuste en algunas zonas, pero habrá mayores distorsiones locales en la zona donde se eliminó el punto, si éste estaba correctamente identificado. Para un ajuste dado, un valor elevado de error residual en un punto informa de una gran variación local que no es capaz de captar el modelo que se está ajustando[5]. Todo esto indica que los coeficientes calculados para la transformación no son independientes de los puntos utilizados. Además, los errores residuales se reducen al disminuir el número de puntos que se emplean, existiendo una convergencia entre los valores reales y predichos cuando el número de grados de libertad aumenta (McGwire, 1996).

Por tanto, la evaluación de los resultados debería hacerse no tanto mediante la consideración del EMC global obtenido en el ajuste, sino mediante el uso de una serie de puntos de comprobación independientes, a los que se les aplique la transformación ajustada. Con el uso de puntos utilizados en el ajuste de la función de transformación se obtienen siempre subestimaciones del EMC global, es decir, exactitudes mayores

[5] Se puede considerar que los errores residuales se dividen pues en dos categorías, aquellos que pertenecen a los puntos que se ajustan y aquellos que pertenecen a los puntos de las zonas deformadas y que no se ajustan. Los primeros presentarán valores reducidos y distribuidos como una normal, y los segundos presentarán un comportamiento significativamente distinto, tal que pueden ser considerados como *outliers* de la distribución de los primeros si el porcentaje de este tipo de puntos está acotado dentro de ciertos límites. Ésta es la idea que lleva a autores como Li y *col.* (2001) a considerar que las deformaciones locales pueden ser detectadas como *outliers* de las observaciones de los procesos de ajuste.

que en el caso de uso de puntos independientes (McGwire, 1996). Cuando no hay puntos independientes de comprobación realmente no se puede saber si el polinomio se ajusta simplemente a los desplazamientos puntuales o a las verdaderas tendencias globales y locales presentes.

La toma u obtención de puntos de control es siempre costosa, máxime en imágenes o cartografías correspondientes a zonas donde no existan infraestructuras desarrolladas por el hombre. Por ello, si se sigue la regla anterior de contar como puntos válidos para el ajuste final con 2 ó 3 veces el número mínimo, la toma de una muestra independiente, y estadísticamente adecuada de puntos de comprobación, puede resultar en un sobrecoste importante.

No obstante, los métodos de validación cruzada[6], en algunos casos también denominados *Prediction Error Sum Of Squares* (PRESS), evitan la toma de una muestra de puntos de comprobación independientes, aprovechando todos los puntos adquiridos para el ajuste. Se trata de un proceso iterativo de sustitución y cálculo de los mejores ajustes polinómicos. El proceso consiste en tomar el conjunto total M de puntos disponibles y reservar uno de ellos. Por mínimos cuadrados, se procede al mejor ajuste polinómico con el subconjunto formado por los $M-1$ puntos. Dado que en cada ajuste se elimina un punto, sobre el mismo se obtiene la discrepancia entre la interpolación que para esa posición da el modelo y su valor real (no es un error residual del ajuste, pues el punto no se considera en los cálculos). Se reitera el proceso M veces y con los M valores de discrepancia obtenidos, uno por punto, se calculan su valor medio y su EMC, que sirven como estimaciones del EMC global del ajuste en puntos de comprobación independientes.

El valor promedio del cuadrado de las M discrepancias es el valor del índice PRESS. La raíz cuadrada del índice PRESS es un valor en las mismas unidades de medida que los errores residuales.

No obstante, como polinomio de ajuste global se toma el que incluye a todos los puntos del conjunto. Aunque este polinomio no ha servido realmente en ningún caso, pues en los ajustes se habrán derivado M polinomios distintos, debe ser sensiblemente igual a todos ellos. En general, todos los M polinomios ajustados con los $M-1$ puntos tendrán unos coeficientes muy similares salvo que el punto eliminado en cada paso tenga una gran preponderancia sobre los valores de dichos coeficientes.

Otra de las técnicas que se pueden aplicar en estos casos es la de detección de puntos de gran influencia, que son aquellos que para un modelo de regresión presentan una influencia desproporcionada en los coeficientes obtenidos para dicho modelo. Tanto es así que, en casos extremos, los valores estimados pueden depender más de un subcon-

[6] El uso de este método lo propone ya de manera operativa el SAI (1999) en su *Guía para el control de calidad de ortoimágenes*, de aplicación en los trabajos que contrata para el seguimiento de la Política Agraria Comunitaria (PAC).

junto de datos con gran influencia que de la mayoría. Para determinar estos puntos existen algunos índices estadísticos (Montgomery, Peck y Vining, 2001) cuya metodología de cálculo es similar al proceso al descrito anteriormente para el PRESS. Se calculan los estadísticos correspondientes al modelo completo (con todos los datos) y dejando el punto de interés fuera, y así se analiza la influencia mediante diversos estadísticos (por ejemplo: Distancia de Cook, COVRATIO, DBETAS, DFFITS, etc.

TRABAJO A REALIZAR

El trabajo a realizar va a consistir en utilizar la técnica clásica y la de validación cruzada para conocer su uso, los resultados que aportan y compararlas.

Se propone realizar un ajuste global, con polinomios de diverso grado, entre las dos series de datos que se adjuntan en la Tabla 4.7, donde las coordenadas imagen constituyen las filas y columnas en el espacio origen (columnas 3ª y 4ª) y las coordenadas mapa se corresponden con las coordenadas UTM del espacio de llegada (columnas 1ª y 2ª).

Se pide aplicar la metodología de validación cruzada y analizar, punto a punto, las diferencias obtenidas entre los residuos generados con los polinomios de ajuste global sobre cada punto, y las discrepancias de la interpolación que se obtienen en cada uno de esos puntos por el método de la validación cruzada. Se comparará el EMC de los ajustes globales, con el PRESS de las discrepancias.

También se debe generar un gráfico de desplazamientos en los puntos para cada uno de los casos anteriores. Mediante una herramienta informática adecuada, se realizará un curvado con estos datos. Con estas informaciones se procederá a interpretar la distribución espacial y a analizar las discrepancias entre métodos, comentando si los valores globales de un ajuste permiten llegar a estas conclusiones.

Se propone el cálculo del índice DFFITS y de la distancia de Cook para determinar qué puntos son de gran influencia sobre el modelo. Se deberá analizar cuál es el comportamiento particular de estos puntos.

Tabla 4.7 Coordenadas de un conjunto de puntos en un espacio origen y en un espacio destino

Punto	Coordenadas Mapa		Coordenadas Imagen		Punto	Coordenadas Mapa		Coordenadas Imagen	
ID	X	Y	X	Y	ID	X	Y	X	Y
1	411044,00	4204878,00	643,00	2723,00	25	519316,00	4217717,00	4118,50	1650,00
2	415021,00	4202796,00	786,00	2768,00	26	514047,00	4209622,00	3994,00	1947,25
3	421345,00	4203243,00	990,00	2715,00	27	538984,00	4205812,00	4837,00	1923,00
4	434449,00	4199456,00	1442,00	2761,00	28	535856,00	4190713,00	4824,50	2437,50
5	436672,00	4201575,00	1503,00	2677,50	29	508841,00	4241889,00	3629,00	919,00
6	446529,00	4201400,00	1826,50	2623,50	30	506376,00	4255672,00	3465,50	481,50
7	452436,00	4200586,00	2025,50	2614,50	31	517991,00	4258675,00	3829,00	313,00
8	451360,00	4197093,00	2011,50	2736,00	32	525227,00	4243734,00	4156,50	760,00
9	456323,00	4199424,00	2160,50	2629,50	33	548313,00	4238145,00	4950,50	803,50
10	462020,00	4195532,00	2370,50	2723,00	34	481569,00	4244622,00	2717,00	993,50
11	469609,00	4196483,00	2614,00	2646,50	35	449391,00	4250899,00	1623,00	980,50
12	477318,00	4192797,00	2889,00	2721,00	36	430271,00	4252825,00	984,00	1031,50
13	481049,00	4192785,00	3012,00	2698,00	37	446794,00	4214884,00	1754,50	2179,00
14	507725,00	4180180,00	3963,00	2952,00	38	467912,00	4208515,00	2485,50	2261,00
15	528062,00	4186341,00	4593,50	2627,00	39	469436,00	4243781,00	2324,00	1093,50
16	409310,00	4223656,00	473,00	2115,50	40	499887,00	4233965,00	3382,50	1233,50
17	414882,00	4243102,00	537,50	1442,50	41	491711,00	4218756,00	3205,50	1781,50
18	418670,00	4247163,00	638,00	1286,50	42	491813,00	4209569,00	3263,50	2082,50
19	408384,00	4247058,00	301,50	1352,50	43	445389,00	4236012,00	1581,00	1493,50
20	401888,00	4211959,00	300,00	2545,00	44	402410,00	4204976,00	360,00	2770,00
21	399417,00	4210283,00	229,50	2615,50	45	400855,00	4208853,00	285,50	2652,00
22	421138,00	4237448,00	777,00	1591,00	46	401024,00	4200248,00	342,50	2934,50
23	422191,00	4222116,00	904,00	2088,50	47	416031,00	4198911,00	842,50	2887,50
24	505742,00	4196625,00	3799,00	2423,00	48	492647,00	4190867,00	3403,50	2691,50

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

BLOQUE 5

CONTROL DE CALIDAD TEMÁTICA

Este bloque aborda el control de calidad temática. La componente temática no es exclusiva de los tradicionalmente denominados mapas temáticos, pues cualquier elemento representado en un mapa topográfico pertenece a un tema.

La evaluación de la exactitud temática suele hacerse en forma de matriz de error, también denominada tabla de contingencia o matriz de confusión (ver ejemplo en Tabla 5.1). Su ordenamiento suele ser tal que la verdad del terreno aparece en columnas, mientras que las unidades cartografiadas (mapa) aparecen en las filas. La tabla así formada presenta una visión general de las asignaciones correctas (elementos de la diagonal) y de las migraciones o fugas (elementos fuera de la diagonal), recogiendo estas últimas los denominados errores de comisión y de omisión. Las comisiones son los elementos que no perteneciendo a una clase aparecen en ella, mientras que las omisiones son los que perteneciendo a esa clase no aparecen por estar mal clasificados en otra. Cuando los errores de comisión y omisión se presentan en términos porcentuales se denominan riesgos del usuario y del productor, respectivamente.

Tabla 5.1 - Matriz de confusión de la clasificación de una imagen

	V1	V2	V3	V4	V5	V6	V7	V8	V9	Total
C1	238051	1	939	0	0	5	0	29	115	239140
C2	7	4086	5082	0	48	151	105	36	2	9517
C3	132	188	51817	5	4	119	601	280	0	53146
C4	0	0	0	11148	834	135	110	0	4	12231
C5	0	4	34	1618	2853	726	174	0	124	5533
C6	24	16	500	78	340	6774	155	6	595	8488
C7	9	45	1867	0	32	75	8257	5	0	10290
C8	2	1	325	0	0	1	8	2993	0	3330
C9	189	0	17	0	197	553	0	0	4374	5330
Total	238414	4341	60581	12849	4308	8539	9410	3349	5214	347005

Nota: "V" se refiere a la verdad terreno y "C" a las unidades cartografiadas

La formación de la matriz a partir de una muestra requiere de ciertas condiciones:

— Las clases que se establezcan deben ser independientes, mutuamente excluyentes, exhaustivas y en número suficiente.

— Deben usarse métodos de muestreo que no introduzcan autocorrelación.

— Conviene el uso de métodos estratificados para asegurar la presencia de clases extrañas o minoritarias.

No importa el origen de la matriz de error (fotointerpretación, teledetección, etc.), ya que el tratamiento estadístico es el mismo.

Si bien existen diversos métodos para la evaluación de la exactitud temática, en este bloque sólo se atenderá a los basados en la matriz de confusión, por ser los más desarrollados y permitir tanto técnicas de análisis descriptivo como analítico.

También se incluye un caso práctico dedicado al modelado cartográfico, enfocada al seguimiento de la propagación del error. Se entiende por modelado cartográfico a aquellas transformaciones que permiten, mediante relaciones funcionales o empíricas (conocimiento), derivar nueva información mediante un proceso en el que se hacen explícitas ciertas relaciones o informaciones que estaban implícitas en una base de datos geográfica. La inexactitud de un mapa es algo que no sólo afecta a la posición o a un tema en sí, sino que se propaga en los procesos de modelado cartográfico, es decir cuando se combina la información de diversos mapas; cabe pues hablar de un modelado de la propagación de errores. El Modelado Cartográfico basado en herramientas estadísticas, en modelos de pertenencia, etc., es una herramienta fundamental en el desarrollo de cartografía temática.

El contenido de este bloque se resume en los siguientes casos prácticos:

— Caso 21: Parámetros de exactitud temática en la clasificación. Aquí se aborda el cálculo de los parámetros de control de la calidad temática, globales y por clase, sobre una matriz de error procedente de la clasificación de una imagen de satélite. El caso se orienta a entender que la labor de análisis de los resultados obtenidos es más importante que el propio cálculo de los mismos.

— Caso 22: Normalización de la matriz de confusión y cálculo de pseudoceros. Se introducen estos dos procesos que permiten, por un lado, facilitar la comparación celda a celda y, por otro, eliminar los valores nulos de la matriz, sustituyéndolos por pequeñas probabilidades que realmente pueden existir. Lógicamente, estos procesos de cálculo deben ser realizados mediante la elaboración de un programa informático.

— Caso 23: Comparación de dos matrices de confusión. Se presenta un test estadístico para la comparación global de dos matrices de confusión. Este caso se

complementa con el Caso 28 (perteneciente al bloque de Simulación), en la que se aplican técnicas de *bootstrapp* para derivar valores medios y varianzas que podrán ser utilizados en test para analizar la semejanza estadística, celda a celda, de dos matrices de confusión.

— Caso 24: Exactitud temática tras la generalización de la leyenda. Aquí se pretende evidenciar las implicaciones que conlleva la generalización, es decir, la reducción del número de clases, en la componente temática de la calidad, también a partir del análisis de la matriz de confusión.

— Caso 25: Estimación de incertidumbre en el modelado cartográfico. Se muestra un modelo cartográfico en el que se usan varias entradas temáticas. Se calcula y analiza cómo se transmite la incertidumbre del valor proporcionado por el modelo, evidenciando la exactitud temática de la cartografía temática que pudiera derivarse del mismo.

Son numerosos los trabajos en forma de artículos que se dedican al análisis de la calidad de la componente temática mediante el empleo de matrices de confusión. En los manuales de Giordano y Veregin (1994), Congalton y Green (1998) y Ariza (2002), pueden encontrarse los aspectos más fundamentales de estas técnicas. Para facilitar el trabajo relativo a este bloque, en el Anexo de Estadística que se incluye en este manual se pueden consultar los aspectos relacionados con las distribuciones de probabilidad y la inferencia estadística paramétrica que se utiliza en los casos prácticos de este bloque.

5.1. CASO 21

Parámetros de exactitud temática en la clasificación

PARÁMETROS DE EXACTITUD TEMÁTICA

Los parámetros de exactitud temática derivados de la matriz de confusión pretenden eliminar la incomodidad del uso de la misma resumiendo su contenido sin reducir demasiado la información presente en ella, es decir, sin perder calidad en la valoración global (Ariza y Pinilla, 1996). Sin embargo, pueden existir casos en los que una clase concreta merezca, por parte de un usuario, una especial atención y mayores niveles de conocimiento sobre su calidad, lo cual requiere de índices parciales o de clase. Es por ello, que para cuantificar la bondad de la clasificación, de forma general, o atendiendo a clases concretas, se utilizan diversos parámetros que veremos a continuación.

Estos índices parciales pueden clasificarse de diversas formas atendiendo a múltiples criterios como, por ejemplo, la facilidad de cálculo, la forma de cálculo, el conocimiento apriorístico de probabilidades, etc. La clasificación que aquí se recoge se basa, fundamentalmente, en las relaciones de cálculo de dichos índices con respecto a la matriz de error, lo que permite deducir de forma indirecta su sentido (global, clase) y complejidad de cálculo. La Tabla 5.2 recoge la nomenclatura que se seguirá en esta exposición, la Tabla 5.3 aplica esta nomenclatura a la estructura de la matriz de confusión y la Tabla 5.4 contiene los parámetros globales y por clase más comunes.

Tabla 5.2 - Notación seguida

N	número de casos (celdas-píxeles)
M	número de clases
V	número de clases "verdad terreno" ($V = M$)
C	número de clases "unidades cartográficas" ($C = M$)
i	filas, unidades del mapa o producto
j	columnas, verdad-terreno
V_j	Clase verdad-terreno
C_i	Clase unidad cartográfica
x_{ij}	número de casos en C_i, V_j
n_{i+}	total casos C_i
n_{+j}	total casos V_j
P_{ij}	proporción en C_i, V_i $(= x_{ij}/N)$
P_{i+}	proporción C_i $(= n_{i+}/N)$
P_{+j}	proporción V_i $(= n_{+j}/N)$

Tabla 5.3 - Ejemplo de matriz de confusión

Unidad Cartográfica	Clase (Verdad) Terreno 1	2	....	v-1	v	Total
1	x_{11}	x_{12}	....	$x_{1,v-1}$	x_{1v}	n_{1+}
2	x_{21}	x_{22}	....	$x_{2,v-1}$	x_{2v}	n_{2+}
.	.	.	....	.	.	.
.	.	.	....	.	.	.
.	.	.	....	.	.	.
c-1	$x_{c-1,1}$	$x_{c-1,2}$	....	$x_{c-1,v-1}$	$x_{c-1,v}$	$n_{c-1,+}$
c	$x_{c,1}$	$x_{c,2}$	....	$x_{c,v-1}$	x_{cv}	n_{c+}
Total	n_{+1}	n_{+2}	....	$n_{+,v-1}$	$n_{+,v}$	N

Tabla 5.4 - Parámetros de calidad temática

Índices Globales	
$Pa = \frac{1}{N}\sum_{i=1}^{M} x_{i,i} = \sum_{i=1}^{M} P_{i,i}$ $\sigma^2_{Pa} = \sigma^2(Pa) = \frac{Pa(1-Pa)}{N}$	***Pa*** *Porcentaje de Acuerdo.* Es considerado como la probabilidad global de estar bien clasificado, por lo que se puede considerar distribuido según una binomial. Sus valores se encuentran en el intervalo [*0, 1*], siendo el valor 1 el caso de total acuerdo.
$Ca_{ps} = \sum_{i=1}^{M} P_{i+} P_{+i} = \frac{1}{N^2}\sum_{i=1}^{M} n_{i+} n_{+i}$	***Ca_{ps}*** *Coeficiente de Acuerdo Aleatorio a Posteriori.* Se basa en la probabilidad a posteriori de cada clase. Representa el porcentaje de acuerdo que cabe esperar al azar por cruce de clases.
$Ca_{pr} = \frac{1}{M}$	***Ca_{pr}*** *Coeficiente de Acuerdo Aleatorio a Priori.* Es similar al anterior, pero en este caso las probabilidades consideradas son a priori a la clasificación. Cuando todas son iguales, se cumple que $Ca_{pr} = 1/M$.
$K = \frac{Pa - Ca_{ps}}{1 - Ca_{ps}}$ $\sigma^2_K = \frac{Pa(1-Pa)}{N(1-Ca_{ps})^2}$	***K*** Coeficiente *Kappa* de ajuste. Su valor da idea del porcentaje de acuerdo obtenido en una clasificación una vez se ha eliminado la parte que se debería al azar. Se basa en la probabilidad a posteriori de pertenencia a un grupo. Sus valores se encuentran en el intervalo [*0, 1*], siendo el valor 1 el caso de total acuerdo.
$T = \frac{Pa - Ca_{pr}}{1 - Ca_{pr}}$ $\sigma^2_T = \frac{Pa(1-Pa)}{N(1-Ca_{pr})^2}$	***T*** Coeficiente *Tau* de ajuste. Su valor da idea de cuánto ha mejorado el sistema de clasificación respecto a una clasificación aleatoria de los *N* elementos en *M* grupos. Se basa en la probabilidad a priori de pertenencia a un grupo.
Índices por clase	
$EU(i) = \frac{x_{i,i}}{n_{i+}}$	***EU*** *Exactitud del Usuario* También llamado pureza de la unidad cartográfica, es el porcentaje de elementos de una clase que están bien clasificados en la cartografía.
$RP(i) = 1 - EU(i)$	***RP*** *Riesgo del Productor.* Es el complementario a la unidad del índice anterior.
$EP(j) = \frac{x_{j,j}}{n_{+j}}$	***EP*** *Exactitud del Productor* para una clase. Es el porcentaje de elementos del muestreo de verificación que se han clasificado correctamente dentro de la unidad cartográfica equivalente.
$RU(j) = 1 - EP(j)$	***RU*** *Riesgo del Usuario.* Es el complementario a la unidad de este índice anterior

EJEMPLO DE ANÁLISIS DE UNA MATRIZ DE CONFUSIÓN

Se tiene la matriz de confusión presentada en la Tabla 5.6, con las clases especificadas en la Tabla 5.5, procedente de la clasificación realizada por un programa de análisis de imagen, sobre una escena del Delta del Ebro (6.000 × 6.000 píxeles) en la que se tomaron 347.005 píxeles distribuidos en 120 campos de verificación. No obstante, la unidad de muestreo de la matriz es el píxel, lo que permitirá conocer las superficies relativas de verificación entre las clases. Se pretende obtener sus índices globales, así como sus índices por clase. De los resultados deberán extraerse conclusiones acerca de la calidad temática de la clasificación realizada.

Tabla 5.5 - Clases de la imagen

Clase	Descripción
1	agua
2	cultivos
3	suelo improductivo
4	arroz
5	frutales
6	matorral
7	olivos y algarrobos
8	salinas
9	juncal

Tabla 5.6 - Matriz de confusión de la clasificación de una imagen

	V1	V2	V3	V4	V5	V6	V7	V8	V9	Total
C1	238051	1	939	0	0	5	0	29	115	239140
C2	7	4086	5082	0	48	151	105	36	2	9517
C3	132	188	51817	5	4	119	601	280	0	53146
C4	0	0	0	11148	834	135	110	0	4	12231
C5	0	4	34	1618	2853	726	174	0	124	5533
C6	24	16	500	78	340	6774	155	6	595	8488
C7	9	45	1867	0	32	75	8257	5	0	10290
C8	2	1	325	0	0	1	8	2993	0	3330
C9	189	0	17	0	197	553	0	0	4374	5330
Total	238414	4341	60581	12849	4308	8539	9410	3349	5214	347005

Nota: "V" se refiere a la verdad terreno y "C" a las unidades cartografiadas

Primeramente es especificado el número de clases y de casos presentes en la matriz:

— Número de Clases en la Matriz (M): 9

— Número de Casos (n): 347005 (0,96%)

A continuación se calculan los índices globales de la matriz:

— Porcentaje de acuerdo (Pa): 0,952
Varianza del Pa (σ^2_{Pa}): 0,000

- Coeficiente Acuerdo Aleatorio a Priori (Ca_{pr}): 0,1111
- Coeficiente Acuerdo Aleatorio a Posteriori (Ca_{ps}): 0,5038
- Coeficiente de Acuerdo Tau (T): 0,94601
  Varianza de T (σ^2_T): 0
- Coeficiente de Acuerdo Kappa (K): 0,90329
  Varianza de K (σ^2_K): 0,000001

Seguidamente se obtienen los índices por cada clase:

Clase	Exactitud Productor	Riesgo Usuario	Exactitud Usuario	Riesgo Productor
1	0,998	0,002	0,995	0,005
2	0,941	0,059	0,429	0,571
3	0,855	0,145	0,975	0,025
4	0,868	0,132	0,911	0,089
5	0,662	0,338	0,516	0,484
6	0,793	0,207	0,798	0,202
7	0,877	0,123	0,802	0,198
8	0,894	0,106	0,899	0,101
9	0,839	0,161	0,821	0,179

El primer comentario se puede realizar en cuanto al tamaño de la muestra de verificación. En principio, dado que la superficie verificada constituye el 0,96% de la imagen, puede decirse que el muestreo de verificación tiene un tamaño adecuado, ya que está muy próximo al 1%, valor sugerido por algunos autores como Congalton (1988).

No obstante, si atendemos al número de píxeles empleados, 347005, se puede decir que la muestra es excesivamente grande, muy por encima de los 50 píxeles por clase (Congalton, 1991), que daría lugar a una muestra de 450 píxeles, o del valor sugerido por una distribución multinomial o binomial. Por ejemplo, para éste último caso, para un 95% de confianza, con un error admitido del 5% y suponiendo que a priori la probabilidad de acertar es igual a la de errar, se obtendría un tamaño de 384 píxeles. Por tanto, se puede decir que, a priori, el esfuerzo realizado para efectuar la verificación es excesivo.

Así, la recomendación de verificar el 1% de la imagen queda en entredicho, ya que este porcentaje no debería ser fijo, sino dependiente del tamaño de la población, o número total de píxeles de la imagen.

Por otro lado, dado que no se tiene información acerca del tipo de muestreo ni del número de muestras de cada clase, no se puede saber si todas ellas están suficientemente representadas. Lo que sí se deduce es que, por término medio, los campos de verificación tienen un gran número de píxeles (el promedio es de casi 2900 cada uno).

Hay que indicar que es adecuado considerar como unidad de muestreo una entidad mayor al píxel, ya que puede ser muy difícil hallarlo en el terreno, incluso empleando un receptor GPS, o en una cartografía, ortofotografías, etc.

Las conclusiones que se obtienen de los resultados de los parámetros globales son, en principio, claras. El valor de *Pa* indica que, en primera instancia, la clasificación ha sido excelente, dado que la mayoría de celdillas de la imagen han sido correctamente asignadas. Los valores de *T* y *K* también son muy elevados, evidenciando la bondad de la clasificación.

La gran diferencia entre el Ca_{pr} y el Ca_{ps} indica que existen clases que acumulan un número mucho mayor de celdillas que otras, por lo que la probabilidad que cabría esperar a priori se incrementa a posteriori. Concretamente se pasa del 11% al 50% de probabilidad, siendo éste un valor bastante elevado.

Sin embargo, éste es un caso en el que los valores globales *Pa*, *T* y *K* sobrestiman en gran medida la bondad de la clasificación. Baste para ello observar los resultados ofrecidos por los índices por clase. El buen resultado del *Pa* viene dado por la óptima clasificación de la clase 1 (agua), que es la que contiene un mayor número de celdillas de la imagen. Pero si se examinan otras clases, se comprueba que hay algunas muy mal clasificadas.

Por ejemplo, observando los valores de riesgo del usuario destacan, por ser muy elevados, los de las clases 5 (frutales) y 6 (matorral). Esto indica que hay muchas celdas que, perteneciendo a estas clases, han sido asignadas a otras clases tras la clasificación (ver matriz de confusión).

Si se atiende a los valores de riesgo del productor existen resultados no satisfactorios, sobre todo los de las clases 2 (cultivos) y 5 (frutales). En la clase 2 un 57% de las celdas asignadas en realidad pertenecen a otras clases, comprobando en la matriz de confusión que no se ha podido discernir bien de la clase 3 (suelo improductivo), quizá por el pequeño tamaño de las plantas de los cultivos que apenas hace cambiar la respuesta espectral del suelo desnudo. En la clase 5, la mitad de las celdas asignadas pertenecen en realidad a las clases 4 (arroz) y 6 (matorral), posiblemente debido al hecho de que tanto los frutales como el arroz son de regadío y a que el tamaño de las copas de los frutales provoca un porcentaje de recubrimiento del suelo similar al de los matorrales.

Se puede decir que esta clasificación no es en absoluto adecuada, dado que el objetivo pretendido es distinguir los diferentes usos del suelo. Únicamente se han conseguido resultados óptimos en la clasificación de la clase 1, debido a la muy diferente signatura espectral del agua.

TRABAJO A REALIZAR

Se tomarán los datos procedentes de la clasificación realizada por un programa de análisis de imagen, sobre una escena correspondiente a una imagen Landsat-TM de la provincia de Jaén (Figuras 5.1 y 5.2), cuyos datos se incluyen en la Tabla 5.7. En la Tabla 5.8 aparece una breve descripción de la zona que a la que pertenece la escena y en la Tabla 5.9 una descripción de la imagen. Con esta información se pretende trasladar al lector una carga semántica importante que facilite la interpretación y análisis de los resultados.

Tabla 5.7 - Datos de la imagen	
Plataforma	Landsat
Sensor	TM
Resolución espectral (Nº Bandas)	7
Resolución espacial	30 metros (bandas 1-5 y 7) 120 metros (banda 6)
Subescena	T20033-95
Dimensiones (filas, columnas, bandas)	5706, 3201, 7
Tamaño (bytes)	127.854.342
Formato de grabación	BSQ
Fecha de Adquisición	7 de julio de 1995

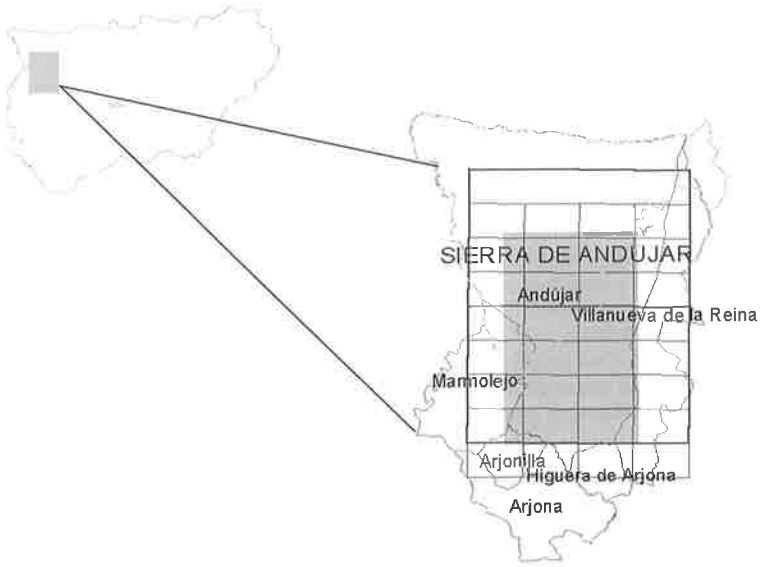


Figura 5.1 - Localización de la escena dentro de la provincia de Jaén

Tabla 5.8 - Descripción de la zona	
LOCALIZACIÓN Y DESCRIPCIÓN GENERAL	La zona bajo estudio afecta a los términos municipales (Jaén) de: Villanueva de la Reina, Marmolejo, Arjonilla, Higuera de Arjona y Arjona, pero sobre todo al de Andújar (ver Figura 5.1). La escena reúne cierto grado de complejidad debido a la variabilidad en cultivos, suelos, geología y topografía. Existen dos zonas de marcada diferenciación. La zona norte, dentro de los límites del parque natural Sierra de Andújar, y la zona sur en la Vega del Guadalquivir.
TOPOGRAFÍA	La topografía es la primera diferencia que se observa entre las dos zonas, en el norte el relieve es montañoso con pendientes que pasan de moderadas a fuertes. En el sur predominan las terrazas, lomas, vegas aluviales y llanuras de inundación.
GEOLOGÍA Y LITOLOGÍA	La heterogeneidad se manifiesta también en los suelos. En la zona norte se encuentran pizarras, granitos, y cuarcitas que sustentan suelos como Litosoles, Inceptisoles, Entisoles, etc. En la zona sur prevalecen las areniscas, margas, arenas, gravas y arcillas, con suelos del tipo Anfisoles (suelos rojos), Vertisoles y Mollisoles.
VEGETACIÓN	La zona norte presenta abundante vegetación natural con formaciones boscosas de especial relevancia, como alcornocales, melojares y encinares, aunque la mayor extensión de la sierra la ocupan madroñales, coscojales y jarales, a veces con coberturas de hasta el 90%. En zonas más antropizadas desaparece el matorral y quedan pastizales aunque se mantienen alcornoques y encinas dispersos, y amplias extensiones de repoblación de pinares con cantuesales y tomillares. En los tramos altos y medios de los ríos se hallan alisedas y fresnedas, y en los cursos bajos de los ríos y arroyos tamujares y adelfares. La zona sur queda diferenciada por su carácter agrícola con cultivos de regadío de porte herbáceo y de porte arbóreo (frutales), además de cultivos de secano y olivares.
PROPIEDAD DE LA TIERRA	Igualmente, la distribución de la propiedad es diferenciable entre las dos zonas, con grandes fincas de caza o ganadería (con cubiertas homogéneas) en el norte, y pequeñas superficies con distintos cultivos (pequeñas propiedades) en el sur.

Fuente: Belart (1999).

Tabla 5.9 - Descripción de la imagen
En el norte de la imagen se observa una franja de bosques perennifolios, en tonos verdes, intercalados con manchas, que van de marrón a rosa, que se corresponden con montes densos a claros y dehesas, los pastos aparecen en ocres a blanco.
Dada la fecha de la imagen, conviene conocer que, tras un periodo de varios años de sequía, en el momento de toma aún no habían comenzado las lluvias. La zona norte está formada por fincas de caza mayor y las especies de la zona, al carecer de suficiente comida, dejaron los campos mermados de cualquier vegetación inferior a dos metros de altura, por lo que en numerosas fincas, especialmente en las adehesadas, la aportación del suelo a la imagen es más intensa que la aportación de la vegetación.
Los pantanos del Encinarejo y Jándula se muestran en tonos que van del azul oscuro en los bordes, al negro en el centro, debido a la profundidad de las aguas. El escaso nivel de dichos pantanos es muy notable en los márgenes, con tonos de violeta a morado para los lodos.
El centro de la imagen aparece con grandes manchas verdes correspondientes a pinares de repoblación de gran cobertura. En el margen derecho del centro de la imagen, los pinares se intercalan con jóvenes plantaciones de olivar que aparecen en tonos blancos a ocres si no tienen sistema de riego y verdes si lo tienen.
El sur aporta el mayor cromatismo a la imagen, ya que en él se encuentra el río Guadalquivir, en tonos azul claro y a ambos márgenes, verdes muy intensos correspondientes a zonas de regadío, y a zonas con vegetación de ribera. Al norte del río, hay tierras que aunque estuvieran consideradas regadío, no se muestran en la imagen como tal por las restricciones de agua del año y el cierre del canal de riego del Rumblar. Estas zonas aparecen en tonos marrones y ocres igual que los olivares de este sector que no mantienen riego.
El sur de la imagen presenta tonos que van de gris a morado y corresponden a suelos margosos con cultivos de secano, y de gris a azulado para olivares sobre margas. En esta parte de la imagen se diferencian claramente anomalías en el suelo debida a materiales ferrosos y los tonos van del amarillo en los márgenes a distintas tonalidades de marrón en el centro.
La zona urbana, al norte del río Guadalquivir, corresponde a Andújar y es la parte de mayor confusión en tonos de toda la imagen, pues se mezclan píxel morados del asfalto, con blancos por terrenos sin construcción, verdes por las zonas ajardinadas, marrones, o azules.

Fuente: Belart (1999).

Figura 5.2 - Composición RGB de la escena

La fuente utilizada mayoritariamente como verdad-terreno fue un conjunto de fotografías aéreas en color, E60k del año 1996, escaneadas, ortorectificadas y georreferenciadas.

El muestreo necesario para construir la matriz de confusión se realizó tomando como unidad de muestreo el píxel. El número de muestras fue de 385 muestras (Figura 5.3), correspondientes al número sugerido por una distribución binomial al 95% de confianza, siguiendo un esquema aleatorio simple. En la Tabla 5.10 aparecen las clases consideradas y en la Tabla 5.11 la matriz de confusión.

El trabajo consistirá en:

— Calcular y analizar los distintos índices globales (Pa, Ca_{pr}, Ca_{ps}, T, K) y sus varianzas, así como los índices por clase.

— Efectuar un comentario general de los parámetros de exactitud temática y una exposición razonada de los principales problemas de clasificación presentes en este estudio, así como sus posibles causas.

— Estimar cual habría sido el número de celdas necesarias para el muestreo de la clasificación según las distribuciones binomial y multinomial (ver Caso 6). Se considerará un nivel de confianza del 95% y un error aceptable del 5%. En cuanto a las probabilidades de asignación correcta se considerará el peor caso ($p = q = 50\%$) y un caso basado en el Pa obtenido ($p = Pa$).

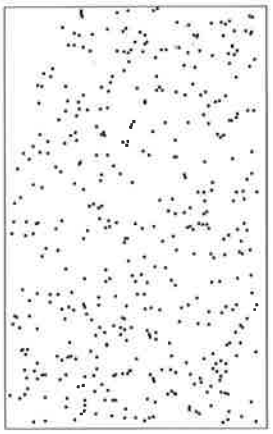


Figura 5.3 - Muestro aleatorio simple de la imagen Landsat-TM (385 muestras)

Tabla 5.10 - Clases de la imagen

Clase	Descripción	
1	Secano	Cubiertas correspondientes a terrenos agrícolas de labor con cultivos herbáceos en secano. También se incluyen en esta clase los barbechos.
2	Vegetación de ribera	Formada por distintos tipos de vegetación riparia, chopos, alisos, etc., casi siempre en los márgenes de los ríos, aunque estos no mantengan agua en el momento de la imagen.
3	Regadío	Terrenos que han presentado en la fecha de la imagen, indicios de tener algún sistema de riego en plantaciones herbáceas.
4	Olivar	Incluye todo tipo de olivar, en distintos estados de crecimiento (a excepción del primer estado), y con diversos marcos de plantación y riego (ausente o presente).
5	Urbano	Espacios estructurados por edificios, calles y zonas de superficie artificial, incluyendo la edificación discontinua, urbanizaciones, zonas industriales, redes viarias, autopistas, autovías, así como espacios con alpechines.
6	Bosque y arbolado	Formado por especies arbóreas con cobertura superior al 70% formada por pináceas (especies del genero pinnus), quejigales, encinares, alcornocales, etc.
7	Embalses	Espacios artificiales de acopio de agua.
8	Matorral denso	Formado por matorral subarbustivo (tomillares, romerales, lavandulares, etc.) y matorral arbustivo como: lentisco, madroño, jaras, aulagas, romeros, retamas, y encinas de porte menor (chaparros). Se han agrupado distintas densidades de cobertura, desde el 95% presente en algunas zonas hasta el 20%.
9	Pastizales y matorral claro	Incluyen los pastizales (con especies herbáceas) que pueden tener arbolado y/o matorral pero con cobertura subarbustiva inferior al 10%.
10	Río	Cursos de agua naturales o artificiales.

Tabla 5.11 - Matriz de confusión

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
C1	25	0	0	4	1	0	0	0	2	0
C2	0	4	5	1	0	2	0	0	0	0
C3	0	0	6	0	0	0	0	0	1	0
C4	1	4	0	52	0	8	0	3	7	1
C5	1	0	0	2	7	0	0	0	1	0
C6	0	0	0	0	0	42	0	3	0	0
C7	0	0	0	0	0	0	2	0	0	0
C8	0	0	0	1	0	39	0	80	5	0
C9	10	0	0	4	0	0	0	1	49	0
C10	0	2	0	0	0	0	1	0	0	8

Nota: "V" se refiere a la verdad terreno y "C" a las unidades cartografiadas

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

5.2. CASO 22

Normalización de la matriz de confusión y cálculo de pseudoceros

NORMALIZACIÓN DE LA MATRIZ DE CONFUSIÓN

Las matrices de error pueden presentarse normalizadas[1], lo cual facilita la labor de comparación de una clasificación de una población realizada por diversos métodos o a partir de fuentes diferentes. Así, tras la normalización, las celdas obtenidas son directamente comparables, ya que todas están expresadas en tanto por uno, y los valores obtenidos fuera de la diagonal dan una mejor idea de los errores de omisión y comisión presentes. Esto presenta varias ventajas:

— Para todas las clases, los valores presentes fuera de la diagonal representan directamente los errores de comisión y omisión, dado que tras la normalización las marginales valen la unidad.

— Facilita la labor de comparación de matrices obtenidas por diversos métodos o fuentes.

— Facilita la comparación, celda a celda, entre distintas matrices.

Para realizar la normalización se puede utilizar el algoritmo propuesto por Fienberg (1970) que consiste en un cálculo iterativo en el que se ajustan a la unidad las marginales de todas las filas (columnas), se realiza la misma labor en las columnas (filas), y así sucesivamente hasta alcanzar un número de iteraciones que se considere adecuado, o unas discrepancias, respecto a la unidad, inferiores a un umbral preestablecido. Es, por tanto, un proceso de aproximación numérica dado que no se puede establecer un sistema de ecuaciones que resuelva tal planteamiento. Un ejemplo de la primera iteración en la normalización de una matriz de confusión (Tabla 5.12) se presenta en las tablas 5.13 y 5.14.

[1] Recientemente, Stehman (2004) critica fuertemente este método en su aplicación al seguimiento de la calidad, llegando a desaconsejarlo por completo. Su trabajo demuestra que este proceso introduce sesgo e imprecisión en parámetros como las exactitudes de usuario y productor debido a que la normalización genera una distribución uniforme de probabilidad marginal, hecho que evidentemente no se da en la realidad. Sin embargo, dado que la práctica de normalización está asumida y recomendada por otros autores también de importancia (por ejemplo, Congalton y Green, 1999), se ha decidido incluirla aquí mientras que la comunidad científica y técnica adopte una postura más definitiva.

Tabla 5.12 - Ejemplo de matriz de confusión a normalizar			
7	1	0	8
0	8	2	10
1	1	5	7
8	10	7	25

Tabla 5.13 - Ejemplo de primer paso de la primera iteración (por columnas)			
0,875	0,1	0	0,975
0	0,8	0,2857	1,0857
0,125	0,1	0,7143	0,9393
1	1	1	3

Tabla 5.14 - Ejemplo de segundo paso de la primera iteración (por filas)			
0,8974	0,1026	0	1
0	0,7368	0,2632	1
0,1331	0,1065	0,7605	1
1,0305	0,9459	1,0236	3

Hay que recalcar que los dos pasos de las tablas 5.13 y 5.14, constituyen una única iteración. Aunque se puede establecer un número de iteraciones a priori (por ejemplo 100 iteraciones), dado que el cálculo es convergente, es más adecuado controlar la normalización mediante un umbral (por ejemplo, una diferencia de una milésima respecto a la unidad en el sumatorio por fila o por columna)

Como ejemplo se muestra en la Tabla 5.15 el resultado de la normalización de la matriz de confusión del ejemplo de la caso práctico anterior.

Tabla 5.15 - Normalización de la matriz de confusión (Tabla 5.6) (100 iteraciones)										
	V1	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**Total**
C1	0,990	0,000	0,003	0,000	0,000	0,000	0,000	0,000	0,003	0,997
C2	0,000	0,913	0,061	0,000	0,007	0,011	0,006	0,002	0,000	1,000
C3	0,003	0,058	0,858	0,000	0,001	0,012	0,045	0,022	0,000	1,000
C4	0,000	0,000	0,000	0,828	0,152	0,013	0,007	0,000	0,001	1,001
C5	0,000	0,002	0,001	0,164	0,703	0,092	0,016	0,000	0,023	1,001
C6	0,001	0,005	0,009	0,007	0,076	0,785	0,013	0,001	0,103	1,001
C7	0,000	0,021	0,046	0,000	0,010	0,012	0,911	0,001	0,000	1,000
C8	0,000	0,001	0,022	0,000	0,000	0,000	0,002	0,974	0,000	1,000
C9	0,006	0,000	0,000	0,000	0,051	0,074	0,000	0,000	0,870	1,000
Total	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	
Nota: "V" se refiere a la verdad terreno y "C" a las unidades cartografiadas										

CÁLCULO DE PSEUDOCEROS EN LA MATRIZ DE CONFUSIÓN

Uno de los aspectos que más pueden afectar a la normalización es la presencia de numerosas celdas con valor cero. Así, previamente a la normalización, se suele proceder a la sustitución de estos ceros por pseudoceros. La presencia de ceros de tipo aleatorio puede deberse a un muestreo deficiente o bien a una clasificación excepcionalmente bue-

na (Congalton y Green, 1998). La introducción de pseudoceros pretende introducir las pequeñas probabilidades que realmente existen. Para ello se pueden utilizar las siguientes expresiones de la Ec. 5.1 (Fieinberg, 1970).

$$P_S x_{i,j} = \frac{N}{N+K}\left(x_{i,j} + K\lambda_{i,j}\right) \qquad K = \frac{N^2 - \Sigma x_{i,j}^2}{\Sigma\left(N\lambda_{i,j} - x_{i,j}\right)^2} \qquad \lambda_{i,j} = \frac{n_{i+}n_{+j}}{N^2} \qquad (5.1)$$

donde:

- $P_s x_{i,j}$ Valor del pseudocero en la celdilla (i,j).
- $x_{i,j}$ Valor original de la celdilla (i,j)
- N Número total de casos.
- n_{i+} Valor de la marginal por fila.
- n_{+j} Valor de la marginal por columna.
- $\lambda_{i,j}$ Parámetro que pretende la incorporación de las probabilidades que cabe esperar en cada celda bajo la hipótesis de independencia.
- K Total alcanzado por los pseudovalores que se incorporan a la matriz.

Como ejemplo se muestra en la Tabla 5.16 el resultado de la aplicación del método de pseudoceros a la misma matriz (Tabla 5.6), previamente a la normalización.

Tabla 5.16 - Resultado de la aplicación de pseudoceros según el método de Fienberg a la matriz de la Tabla 5.6

	V1	V2	V3	V4	V5	V6	V7	V8	V9	Total
C1	238049,8	1,047762	939,6518	0,141419	$4{,}75\cdot10^{-2}$	5,093902	0,103667	29,0364	115,0555	239140
C2	7,104316	4085,937	5081,945	$5{,}62\cdot10^{-2}$	48,00113	151,0013	105,0024	36,00089	2,002252	9517
C3	132,5811	188,0076	51816,32	5,031354	4,010497	119,019	610,0133	280,0037	$1{,}27\cdot10^{-2}$	53146
C4	0,134307	$2{,}44\cdot10^{-3}$	$3{,}41\cdot10^{-2}$	11147,83	842,989	135,0027	110,0035	$1{,}88\cdot10^{-3}$	4,002873	12231
C5	$6{,}02\cdot10^{-2}$	4,001041	34,01488	1617,977	2852,956	725,9906	173,9996	$8{,}52\cdot10^{-4}$	123,9993	5533
C6	24,09275	16,00144	500,0157	78,00378	339,9962	6773,895	155,0012	6,001213	594,9926	8488
C7	9,112766	45,00134	1866,999	$6{,}08\cdot10^{-3}$	32,00153	75,00285	8256,873	5,001506	$2{,}46\cdot10^{-3}$	10290
C8	2,036507	1,000649	325,0041	$1{,}96\cdot10^{-3}$	$6{,}61\cdot10^{-4}$	1,001293	8,001316	2992,953	$7{,}99\cdot10^{-4}$	3330
C9	189,0555	$1{,}83\cdot10^{-3}$	17,01459	$3{,}15\cdot10^{-3}$	196,9979	552,9933	$2{,}31\cdot10^{-3}$	$8{,}21\cdot10^{-4}$	4373,932	5330
Total	238414	4341	60581	12849	4308	8539	9410	3349	5214	

TRABAJO A REALIZAR

El trabajo a realizar consistirá en:

— Proceder a la aplicación del método de pseudoceros.

— Acometer la normalización de la matriz resultante del paso anterior.

Para ello se utilizará la matriz de confusión del caso práctico anterior (tablas 5.10 y 5.11). Dado el proceso de cálculo a realizar, se recomienda recurrir a la elaboración de un programa informático o al empleo de una hoja de cálculo.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

5.3. CASO 23

Comparación de dos matrices de confusión

COMPARACIÓN DE MATRICES DE CONFUSIÓN

Como ya se refirió en el caso práctico anterior, la normalización facilita las labores de comparación celda a celda entre distintas matrices de confusión. Los valores normalizados son más fáciles de comparar que los valores originales, si bien no se conoce la función de distribución de los mismos, por lo que no existe método estadístico paramétrico para decidir si existen o no diferencias celda a celda.

Por otro lado, se puede aplicar un contraste de hipótesis de igualdad sobre dos valores globales representativos de cada una de las matrices de error, por ejemplo dos valores de *Pa* o *Kappa*. Así, si se quieren contrastar los valores de Kappa (K_1 y K_2), las hipótesis son:

$$H_0: K_1 - K_2 = 0; \ (K_1 = K_2)$$
$$H_1: K_1 - K_2 \neq 0;$$

El estadístico a utilizar es (Ec. 5.2):

$$z = \frac{|K_1 - K_2|}{\sqrt{\sigma_{K_1}^2 + \sigma_{K_2}^2}} \qquad (5.2)$$

donde:

σ_1^2 Varianza de K_1.
σ_2^2 Varianza de K_2.

La regla de decisión es que H_0 se rechaza si $z > |Z_{\alpha/2}|$, siendo éste último el valor correspondiente de la distribución normal N(0,1).

TRABAJO A REALIZAR

Se pretende comparar la matriz de confusión A, empleada en los casos prácticos anteriores (Tabla 5.17), con otra matriz de confusión B de la misma escena, pero obtenida a partir de un muestreo más intenso (Tabla 5.18). El proceso de cálculo que se sugiere es el siguiente:

— Aplicar el método de pseudoceros y el de normalización a ambas matrices.

- Calcular los índices globales (y sus varianzas, en su caso) y por clase.
- Someter los índices globales a un contraste de hipótesis de igualdad, siendo el nivel de confianza del 95%.

Una vez efectuados los cálculos se deben analizar las diferencias halladas entre ambas matrices, tanto celda a celda en las matrices normalizadas, como a partir de los índices y de los contrastes de hipótesis.

Tabla 5.17 - Matriz de confusión A (385 muestras)

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
C1	25	0	0	4	1	0	0	0	2	0
C2	0	4	5	1	0	2	0	0	0	0
C3	0	0	6	0	0	0	0	0	1	0
C4	1	4	0	52	0	8	0	3	7	1
C5	1	0	0	2	7	0	0	0	1	0
C6	0	0	0	0	0	42	0	3	0	0
C7	0	0	0	0	0	0	2	0	0	0
C8	0	0	0	1	0	39	0	80	5	0
C9	10	0	0	4	0	0	0	1	49	0
C10	0	2	0	0	0	0	1	0	0	8

Nota: "V" se refiere a la verdad terreno y "C" a las unidades cartografiadas

Tabla 5.18 - Matriz de confusión B (788 muestras)

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
C1	48	0	0	6	5	0	0	0	4	0
C2	0	11	7	3	0	3	0	0	0	0
C3	0	0	15	0	0	0	0	0	1	0
C4	1	13	5	108	0	9	0	7	11	1
C5	5	0	0	3	21	0	0	0	2	0
C6	0	0	0	0	0	85	0	4	0	0
C7	0	0	0	0	0	0	7	0	0	0
C8	0	0	0	1	0	75	0	176	11	0
C9	18	0	0	7	0	0	0	8	88	0
C10	0	2	0	0	0	0	3	0	0	14

Nota: "V" se refiere a la verdad terreno y "C" a las unidades cartografiadas

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

5.4. CASO 24

Exactitud temática tras la generalización de la leyenda

GENERALIZACIÓN DE LEYENDAS

La generalización de la leyenda de una determinada información geográfica puede venir motivada por un cambio de escala, o por un cambio en el propósito del mapa, y por ello es un proceso que debe ser muy bien estudiado. Esta generalización es una operación inmediata cuando la leyenda original está perfectamente estructurada y jerarquizada, pues en este caso simplemente basta con determinar el nivel al que se quiere llegar y cortar las ramas que sobran del árbol. Por este motivo, Congalton (1999) recomienda que el esquema de clasificación adoptado tenga una estructura jerárquica. Un ejemplo es la leyenda del *Corine Landcover*, como puede observarse en las Tablas 5.19 y 5.20, en las que el paso de 5 a 3 niveles es inmediato.

Un aspecto importante en la generalización de los elementos de una leyenda por agregación es que conforme disminuye el número de clases presentes en la misma, también lo hace la confusión general. El error de confusión que se produce entre las clases que se agrupan se elimina, por lo que cabe esperar una mayor exactitud que en el mapa original (Ariza, 2002). Si se utiliza el *Pa* (porcentaje de acuerdo) como índice de calidad, este aumentaría tras la reclasificación.

Valga como ejemplo el siguiente caso: sea un trabajo con 10 clases en cuyo control de calidad por matriz de confusión se ha obtenido un *Pa = 75%* y, no estando satisfechos con el resultado, se desea agregar un cierto número de clases para obtener un *Pa* más favorable. Si se establece el supuesto de que las fugas fuera de la diagonal de la matriz van a ser homogéneas, esto es, de igual valor, una generalización desde 10 a 5 clases haría que aumentara automáticamente el *Pa* hasta el 83%, y una reducción a 3 clases ya alcanzaría el 90%. La Figura 5.4 muestra un ejemplo numérico en el paso de 3 a 2 clases; se han remarcado los recuadros que engloban a las celdas que se agrupan.

	V1	V2	V3
C1	4	2	1
C2	0	12	2
C3	1	3	9

Pa = 25 / 34 = 74%

	V1	V2+V3
C1	4	3
C2+C3	1	26

Pa = 30 / 34 = 82%

Figura 5.4 - Aumento del porcentaje de acuerdo por agregación de clases

Tabla 5.19 - Leyenda del Corinne Landcover a cinco niveles

NIVEL 1	NIVEL 2	NIVEL 3	NIVEL 4	NIVEL 5
3. ZONAS FORESTALES Y SEMINATURALES	3.1. Bosques	3.1.1 Bosque de frondosas	3.1.1.1 Perennifolia y quejigal	3.1.1.1.1 Perennifolias y esclerófilas y quejigales
				3.1.1.1.2 Laurisilva macaronésica
			3.1.1.2 Caducifolia y rebollar	
			3.1.1.3 Otras frondosas y plantación	
		3.1.2 Bosque de coníferas	3.1.2.1 Pináceas	
			3.1.2.2 Sabinales y enebral	
		3.1.3 Bosque mixto		
	3.2 Espacios de vegetación arbustiva y/o herbácea	3.2.1 Pastizales naturales	3.2.1.1. Pastizales y suprafort	
			3.2.1.2 Otros pastizales	
		3.2.2 Landas y matorrales	3.2.2.1 Landas y matorrales templado oceánicos	
			3.2.2.2 Fayal-brezal macaronésico	
		3.2.3 Vegetación esclerófila	3.2.3.1 Grandes formaciones matorral denso o medianamente denso	
			3.2.3.2 Matorrales subarbustivo o arbustivos muy poco densos	
			3.2.3.3 Matorrales xerófilos macaronésicos	
		3.2.4 Matorral boscoso de transición		
	3.3 Espacios abiertos con poca o sin vegetación	3.3.1 Playas, dunas y arenales		
		3.3.2 Roquedo		
		3.3.3 Espacios con vegetación escasa	3.3.3.1 Xeroestepa subdesértica	
			3.3.3.2 Cárcavas y/o zonas en proceso de erosión	
			3.3.3.3 Espacios orófilos altitudinales con vegetación escasa	
			3.3.3.4 Espacios orófilos altitudinales y otras zonas de roquedo con vegetación escasa	
		3.3.4 Zonas quemadas		
		3.3.5 Glaciares y nieves permanentes		

Tabla 5.20 - Leyenda del Corinne Landcover a tres niveles		
NIVEL 1	**NIVEL 2**	**NIVEL 3**
3. ZONAS FORESTALES Y SEMINATURALES	3.1. Bosques	3.1.1 Bosque de frondosas
		3.1.2 Bosque de coníferas
		3.1.3 Bosque mixto
	3.2 Espacios de vegetación arbustiva y/o herbácea	3.2.1 Pastizales naturales
		3.2.2 Landas y matorrales
		3.2.3 Vegetación esclerófila
		3.2.4 Matorral boscoso de transición
	3.3 Espacios abiertos con poca o sin vegetación	3.3.1 Playas, dunas y arenales
		3.3.2 Roquedo
		3.3.3 Espacios con vegetación escasa
		3.3.4 Zonas quemadas
		3.3.5 Glaciares y nieves permanentes

TRABAJO A REALIZAR

Tomando como leyenda la que se considera en la Tabla 5.21 y como matriz de confusión la que presenta la Tabla 5.22, el trabajo consistirá en:

— Calcular y analizar los distintos índices globales (Pa, Ca_{pr}, Ca_{ps}, T, K), así como los índices por clase de la matriz, antes de generalizar la leyenda.

— Generalizar la leyenda, argumentando las agrupaciones de clases realizadas.

— Establecer las hipótesis necesarias para realizar una estimación, a priori, de la mejora del Pa tras la reducción del número de clases.

— Calcular y analizar los distintos índices globales (Pa, Ca_{pr}, Ca_{ps}, T, K), así como los índices por clase de la matriz, tras la generalización de la leyenda.

— Comentar y analizar los resultados obtenidos.

Tabla 5.21 - Leyenda	
Clase	**Descripción**
1	Urbano
2	Secano sobre pizarra
3	Secano sobre marga
4	Regadío
5	Ribera
6	Olivar sobre marga
7	Olivar sobre pizarra en secano
8	Olivar sobre pizarra en regadío
9	Perennifolias y quejigales
10	Pinar
11	Monte denso
12	Monte claro
13	Pastizal
14	Río
15	Pantano

Tabla 5.22 - Matriz de confusión de la imagen

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
C1	25			1										1	6
C2		102	85			6	12					5			
C3		42	98			5	11					7			
C4				79	1								22	2	2
C5				2	12			1		3	7			6	1
C6		8	10			325	102	50	11		8	19			
C7		8	6			158	535	85	15		10	13			
C8				2		40	52	235	9		1				
C9						10	15	9	47	15	12	18			
C10					6				14	153	15	12			
C11					5	8	5	2	18	21	93	8		8	
C12		2	4			21	30		21			58			
C13				9									35		
C14				1	4						8	2		9	1
C15	3													2	42

Nota : "V" se refiere a la verdad terreno y "C" a las unidades cartografiadas

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

5.5. CASO 25

Estimación de incertidumbre en el modelado cartográfico

MODELO DE CAPACIDAD DE PASTO DEL SUELO

El modelo de capacidad de pasto del suelo (CPS) que se considerará (Drummond y Ramlal, 1992) se desarrolla sobre el conocimiento cartográfico de ciertas informaciones edáficas que, mediante funciones de pertenencia, permiten clasificar los suelos. Se basa en tres fuentes de información sobre el suelo:

1. Tipo de drenaje del suelo (5 clases).
2. Capacidad de sustento (3 clases).
3. Capacidad de retención de agua (5 clases).

Hay que reseñar que algunas de estas clasificaciones pueden plasmarse como verdaderos submodelos específicos.

Por tanto, *CPS = f (tipo de drenaje, capacidad de retención de agua, capacidad de sustento)*, lo que significa que la asignación o modelo es una función de pertenencia conjunta, y que incluye tantas pertenencias como parámetros lo caracterizan. Según se muestra en la Tabla 5.23, el modelo general de CPS establece 3 tipos o clases de suelo. Bajo el supuesto de un conocimiento perfecto de los parámetros que lo caracterizan, este modelo presenta una probabilidad de asignación correcta del 95% de los casos.

Tabla 5.23 - Modelo de Capacidad de Pasto del Suelo (CPS)

Tipo de drenaje (TD)	D1		D2		D3		D4			**D5**	
Capacidad sustento (CS)	S1	S2	S1	S2	S1	S2	S1	S2	S3	**S2**	S3
Capacidad de Retención de agua (CR)											
R1	1	1	1	1	1	1	1	2	3	2	3
R2	1	1	1	1	1	1	1	2	3	2	3
R3	2	2	2	2	2	2	2	2	3	3	3
R4	3	3	3	3	3	3	3	3	3	**3**	3
R5	3	3	3	3	3	3	3	3	3	3	3

La Tabla 5.23 presenta las distintas combinaciones entre las características que definen el modelo, que para un caso concreto se interpreta como sigue: Si TD = D5, CS = S2 y CR = R4, la función de pertenencia conjunta indica que CPS = 3.

El *tipo de drenaje* está relacionado con la profundidad del nivel freático, según las clases que se muestran en la Tabla 5.24.

Tabla 5.24 - Modelo para el tipo de drenaje (TD)	
Tipo de drenaje (TD)	**Profundidad máxima (media) [m]**
D1	> 0,80
D2	[0,40 – 0,80]
D3	[0,25 – 0,40)
D4	[0,15 – 0,25)
D5	< 0,15

Por otra parte, la *capacidad de sustento del suelo* se relaciona con la profundidad del nivel freático y el tipo de suelo según el modelo establecido en la Tabla 5.25.

Tabla 5.25 - Modelo para la capacidad de sustento (CS)					
Profundidad máxima (media) [cm]	**Tipo de suelo (TS)**				
	T1	**T2**	**T3**	**T4**	**T5**
[0 – 12)	S3	S3	S3	S3	S3
[12 – 24)	S3	S3	S3	S3	S3
[24 – 33)	S3	S2	S2	S3	S2
[33 – 40)	S2	S1	S3	S2	S1
[40 – 60)	S2	S2	S2	S2	S1
[60 – 80)	S1	S1	S2	S2	S2
[80 – 140)	S1	S1	S1	S2	S1

Además, el tipo de suelo se relaciona con la textura del suelo mediante el contenido de materia orgánica y el porcentaje de arcillas presente en su composición según el modelo expresado en la Tabla 5.26.

Tabla 5.26 - Modelo de tipo de suelo (TS) en función de su composición		
Tipo de suelo (TS)	**Materia orgánica (%)**	**Arcillas (%)**
T1. Turba	15 - 100	0 - 8
T2. Arcilla sobre turba	25 - 70	8 - 100
T3. Arcilla	0 - 15	25 - 100
T4. Arcillo - arenoso	0 - 2,5	8 - 25
T5. Arenoso	0 - 2,5	0 - 8

Finalmente, respecto a la capacidad de retención de agua se distinguen 5 clases según recoge la Tabla 5.27.

Tabla 5.27 - Modelo para la capacidad de retención de agua (CR)	
Clase	**Capacidad de retención [mm]**
R1	> 200
R2	[150 – 200]
R3	[100 – 150)
R4	[50 – 100)
R5	< 50

ESTIMACIÓN DE LA FIABILIDAD DEL MODELO

El modelo que se ha presentado, como otros muchos existentes en los ámbitos medioambiental o de las ingenierías, supone la clasificación final de un elemento, en este caso la capacidad de pasto del suelo, según una función de pertenencia conjunta a varias subclases (profundidad del nivel freático, retención de agua...). Esto es lo que se denomina una función de pertenencia conjunta. La fiabilidad de la asignación final dependerá de las fiabilidades de las asignaciones particulares realizadas a cada una de las subclases que intervienen en la función de pertenencia conjunta. En el caso de que los factores que definen las subclases sean independientes, bajo la teoría estadística la probabilidad del suceso conjunto es el producto de las probabilidades individuales.

No obstante, y dado que el modelo de asignación a CPS es un modelo empírico, que no siempre es adecuado, el valor anterior habrá que matizarlo con la propia fiabilidad del modelo global (en este caso 0,95). Lo mismo puede ocurrir para cualquier subclase que se exprese en mayor detalle como un submodelo particular.

EJEMPLO DE APLICACIÓN

Supóngase que se quiere saber la capacidad de pasto (1, 2 ó 3) y la fiabilidad de este valor, según el modelo descrito en la Tabla 5.23, de un suelo con las siguientes características:

- Suelo de tipo T2. Se ha comprobado una fiabilidad del 95% en la estimación de la proporción de cada componente del suelo (materia orgánica, arcillas).
- Profundidad nivel freático de 0,5 m, con desviación de 0,1 m.
- Capacidad de retención de agua 120 mm, con desviación de 15 mm.

En primer lugar se establecerá cuál es la capacidad de pasto del suelo que se ha especificado. Observando la Tabla 5.24, se establece el tipo de drenaje como D2. Para obtener la capacidad de sustento se emplea la Tabla 5.25 y se concluye que es de tipo S2. A partir de la Tabla 5.27 se comprueba que la capacidad de retención de agua es de tipo R3. Finalmente, entrando en la tabla del modelo (Tabla 5.23) con los valores obtenidos de D2, S2 y R3, se concluye que *la capacidad de pasto del suelo es 2*.

No obstante, dado que existen lógicas imprecisiones en los valores de profundidad del nivel freático, capacidad de retención de agua y composición del suelo, el valor de capacidad de pasto obtenido tendrá un nivel de incertidumbre que habrá que estimar.

Primeramente se puede estudiar la probabilidad de que el *tipo de drenaje* sea D2. Para determinar su probabilidad de pertenencia disponemos del valor de desviación

típica, de 0,1 m. Con ello se puede calcular la probabilidad de que el valor dado de profundidad del nivel freático (0,5 m) pertenezca al intervalo de la clase D2 [0,40-0,80], según una distribución normal N(0,5;0,1):

$$P[0,4 < X < 0,8] = P\left[\frac{0,4-0,5}{0,1} < Z < \frac{0,8-0,5}{0,1}\right] = P[-1 < Z < 3] = 0,999 - 0,159 = 0,840$$

Por tanto, la probabilidad de pertenencia al tipo de drenaje D2 es $P_{TD} = 0,840$.

La probabilidad de que la *capacidad de retención de agua* sea R3, se obtiene a partir del valor de capacidad de retención, de su desviación (120 y 15 mm, respectivamente) y de los extremos del intervalo R3 [100-150]:

$$P[100 < X < 150] = P\left[\frac{100-120}{15} < Z < \frac{150-120}{15}\right] = P[-1,33 < Z < 2] = 0,977 - 0,091 = 0,886$$

Así, la probabilidad de pertenencia a la capacidad de retención R3 es $P_{CR} = 0,886$.

La probabilidad de que la categoría de *capacidad de sustento* sea S2 depende de tres variables: de la profundidad del nivel freático, su inexactitud, y de la fiabilidad en la estimación de la composición del suelo. Así, habrá que efectuar una composición de la exactitud de las tres variables. La probabilidad de que la profundidad del nivel freático (50 cm) se sitúe en el intervalo [40-60] es del 68,2%. Luego la probabilidad de S2 será $P_{CS} = 0,682 \cdot 0,95 \cdot 0,95 = 0,616$ (61,6%).

Finalmente, la probabilidad de que la *Capacidad de Pasto del Suelo* sea 2 se obtiene a partir de las tres probabilidades calculadas y teniendo en cuenta que la fiabilidad del modelo es del 95%: $P = 0,95 \cdot P_{TD} \cdot P_{CR} \cdot P_{CS} = 0,436$ (43,6%), que evidencia una incertidumbre muy elevada en el valor de capacidad de pasto aportado por el modelo.

TRABAJO A REALIZAR

El trabajo consiste en calcular, mediante la función de pertenencia conjunta, la fiabilidad de la asignación del tipo de *Capacidad de Pasto del Suelo* de un conjunto de polígonos de especial interés de nuestra BDG a los que se les ha determinado adecuadamente las características que se especifican en la Tabla 5.28.

Tabla 5.28 - Datos para el cálculo del modelo de capacidad de pasto de un suelo

Pól.	Tipo de suelo	Profundidad nivel freático [m]		Capacidad retención agua [mm]	
221	T2	0,60	$\sigma = 0,14$	166	$\sigma = 17$
357	T3	0,50		125	
504	T3	0,32		121	
688	T1	1,05		225	

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

BLOQUE 6

SIMULACIÓN

La simulación es una herramienta de gran valor para el ingeniero e investigador. Sus campos de aplicación son múltiples (estudio del genoma, el estudio del desarrollo de plantas, catástrofes medioambientales, estudios de ingeniería, inversión en bolsa, etc.) y cada vez más importantes y frecuentes en el desarrollo de aplicaciones cartográficas de control debido a sus ventajas para el estudio deductivo de ciertas formulaciones que, de otra manera, resultarían inabordables. Su uso puede orientarse tanto sobre aspectos temáticos como en el ámbito posicional.

La simulación consiste en construir un programa de ordenador que describa el comportamiento del sistema de interés, o el modelo que lo representa. Con este programa se experimenta con el comportamiento del sistema para llegar a conclusiones que apoyen la toma de decisiones (Ríos, 1997). El proceso básico de la simulación consta de 4 fases:

- Obtener observaciones básicas mediante el uso de un generador de números aleatorios.
- Transformar las observaciones en entradas al modelo según las especificaciones del mismo.
- Utilizando el modelo implementado, transformar las entradas en salidas.
- Analizar estadísticamente las salidas para estimar el comportamiento del modelo.

En este bloque se exponen algunos ejemplos de empleo de técnicas de simulación aplicadas al campo de la Cartografía. En ellos se comprueba la utilidad de la simulación para la delimitación de errores existentes, tanto en el caso de calidad posicional como temática, así como para su propagación en el modelado:

— Práctica 26: Aproximación a la banda de error. Se aplica la simulación al estudio de las bandas de incertidumbre sobre diversos elementos lineales y poligonales. Como resultado del proceso de simulación se puede conocer la evolución del nivel de probabilidad de la pertenencia de los puntos del espacio al elemento considerado.

— Práctica 27: Simulación en el modelado de propagación del error. Se compara el estudio de la propagación del error según dos formas de trabajo, la representación matemática y la resolución por simulación. Para el desarrollo de la práctica se toma como ejemplo un modelo de uso muy extendido como es el caso de la cartografía de riesgos de erosión basada en el cálculo de la erosión potencial en parcela según la ecuación universal de pérdida de suelo (EUPS).

— Práctica 28: Simulación para la comparación de matrices de confusión. En este caso se aplica la simulación a las matrices de confusión procedentes del control de calidad temática para estimar la varianza de cada celda, y poder así efectuar comparaciones celda a celda mediante contrastes de hipótesis de igualdad.

Para facilitar la comprensión de los aspectos más cruciales de la simulación, en el Anexo de Estadística se ha incluido un apartado dedicado a la generación de número aleatorios, ya que esta fase es de vital importancia para llevarla a cabo. También son de interés aquí los apartados dedicados a la inferencia paramétrica.

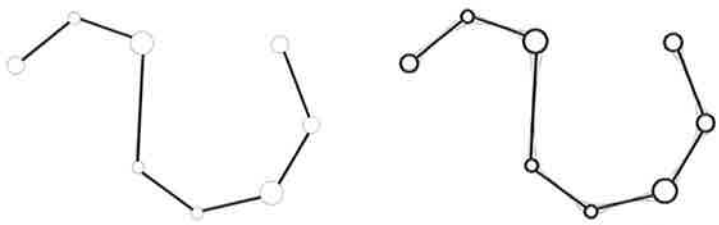
PRÁCTICA 26

Aproximación a la banda de error

LA EXACTITUD GEOMÉTRICA

Basada en la posicional, la exactitud geométrica constituye un mayor grado de complejidad, la entendemos como los aspectos de exactitud que atañen a la geometría que se conforma cuando varios elementos puntuales se ordenan según una disposición en el espacio, ya sea un segmento, una poligonal o un polígono (Ariza, 2002).

Dado que la mayor parte (80%) de la información existente en una base de datos geográfica es lineal (Cuening, 1971), resulta evidente el interés de conocer cómo afecta el error a estos elementos, máxime cuando los elementos areales se derivan de los mismos. Parámetros clásicos de error, como error medio cuadrático, error circular probable o error circular estándar, son propios de elementos puntuales y no aplicables directamente a líneas o áreas. No obstante, como apunta Leung (1998), parece poco intuitivo separar el estudio de los errores puntuales del estudio de los errores en líneas y otras geometrías cuyo elemento básico es el punto (Figura 6.1).



a) b)

Figura 6.1 - Banda de error de una poligonal: a) valores de los errores en los vértices; b) comportamiento de la banda de error

Por otro lado, en los estudios de la exactitud geométrica es fundamental el concepto de banda épsilon, banda de indeterminación, incertidumbre o de imprecisión geométrica. Así, se puede considerar la existencia de curvas de isoprobabilidad que permiten establecer áreas donde se puede afirmar, con una probabilidad dada, la presencia del elemento en cuestión. Por ejemplo, la posición de un segmento estará con una probabilidad del 95% dentro de la curva de isoprobabilidad que lo envuelve y cuya anchura es de dos veces la desviación típica.

Para conocer el comportamiento estadístico espacial de lo que ocurre en la banda de indeterminación lo más cómodo es recurrir a la simulación. Un método que ha sido experimentado consiste en generar un número suficientemente elevado de parejas

de puntos extremos a los que se les ha introducido una perturbación aleatoria. Posteriormente, cada uno de estos segmentos se rasteriza de manera que cada celdilla toma el valor del número de veces que ha caído un elemento rasterizado sobre la misma. Ejemplos de esta metodología pueden encontrarse en Caspary (1992), Shi, Ehlers y Tempfli (1994), Ariza (1997), Podobnikar (1999), etc.

Según Veregin (1994), los resultados de estas investigaciones pueden utilizarse para estimar la localización de las bandas de incertidumbre de líneas y polígonos sobre la base de la incertidumbre posicional de sus vértices, con la ventaja añadida de que las unidades utilizadas (píxeles) permiten un uso adimensional y, por tanto, no dependen de la escala del mapa digitalizado, pudiendo utilizar dichos parámetros en cualquier caso real.

PROCESO PARA LA SIMULACIÓN DE LA BANDA DE ERROR

El proceso de simulación que se considera será el de perturbación de los extremos de un segmento, polilínea o polígono, mediante una componente aleatoria, para, posteriormente, proceder a la rasterización del número de veces que cae el elemento rasterizado sobre una determinada celdilla.

Estas perturbaciones han de ser generadas de forma aleatoria, de modo que pertenezcan a una distribución Normal de media cero y de desviación típica igual al error medio cuadrático considerado. No obstante, un ordenador genera números aleatorios según una distribución uniforme *U[0,1]*. Sin embargo, se puede emplear el método de Box-Muller, muy utilizado para generar valores que sigan una distribución normal *N(0,1)*, a partir de valores que sigan una distribución uniforme $r \approx U[0,1]$. Así, partiendo de dos valores r_1 y r_2 se hace la siguiente transformación:

$$x_1 = (-2\ Ln\ r_1)^{1/2}\ cos\ (2\pi r_2)$$
$$x_2 = (-2\ Ln\ r_1)^{1/2}\ sen\ (2\pi r_2)$$

Con los vértices perturbados se construyen los segmentos del elemento elegido y se procede a su rasterización acumulativa sobre una malla que se autoencaja sobre la ventana marcada por las dimensiones máximas del elemento aumentada en un tanto por ciento (rectángulo encuadrante). Este aumento de la ventana es necesario para que todas las realizaciones del elemento en cuestión entren dentro de la ventana.

Un aspecto fundamental del programa es la rasterización acumulativa. La rasterización de cada elemento correspondiente a cada iteración se realiza mediante el algoritmo de fuerza bruta. El término *acumulativa* se refiere a que cada vez que sobre una celdilla pasa un nuevo segmento se añade una unidad al valor de su atributo.

Después de la rasterización acumulativa se obtiene un archivo raster con una imagen como las de las Figuras 6.2 ó 6.3, en las que se aprecia la simulación de un segmento aplicando 2000 iteraciones y de un cuadrado aplicando 1000 iteraciones, respectivamente. Tras la rasterización, puede ser muy útil la elaboración de curvados que unan los puntos con igual número de casos (Figuras 6.4 y 6.5), o, si el número de casos de cada celdilla es dividido por el total de casos, curvas de isoprobabilidad. También puede realizarse una vista tridimensional del curvado (Figuras 6.6 y 6.7).

Figura 6.2 - Rasterización acumulativa de un segmento. 2000 iteraciones

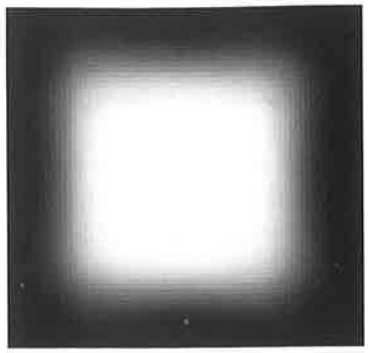


Figura 6.3 - Rasterización acumulativa de un polígono (cuadrado). 1000 iteraciones

Figura 6.4 - Curvado de la rasterización del segmento

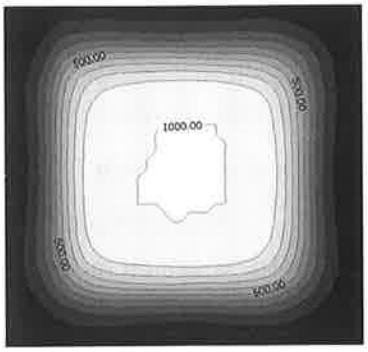


Figura 6.5 - Curvado de la rasterización del cuadrado

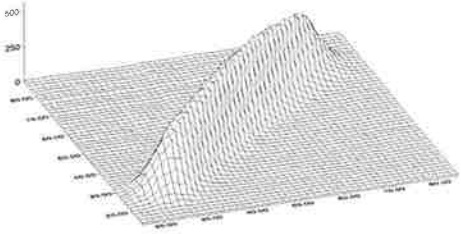


Figura 6.6 - Vista tridimensional del curvado de probabilidades para el caso del segmento

Figura 6.7 - Vista tridimensional del curvado de probabilidades para el caso del cuadrado

TRABAJO A REALIZAR

Consistirá en elaborar un programa que aplique el proceso de simulación descrito. Posteriormente se deberá utilizar el programa y:

— Aplicar el proceso de simulación sobre distintos elementos geométricos: segmentos, polilíneas y polígonos de formas diversas. También se deberá modificar el valor de desviación típica de las coordenadas de estas entidades geométricas para observar su influencia.

— A partir de los ficheros de salida, obtener, mediante cualquier programa informático de interpolación de curvas de nivel, una representación gráfica de la probabilidad de pertenencia a los puntos del plano a los segmentos de la figura bajo estudio.

— Hacer un análisis descriptivo de la morfología de las bandas encontradas.

— Elaborar una lista de posibles aplicaciones que se podrían dar a estos estudios, comentando y detallando cada una.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar material adicional para este caso.

PRÁCTICA 27

Simulación en el modelado de propagación del error

LA ECUACIÓN UNIVERSAL DE PÉRDIDA DEL SUELO (EUPS)

La EUPS (USLE, *Universal Soil Loss Equation*) (Wischmeier, 1959) estima la erosión potencial en parcela. Se plantea según la Ec 6.1.

$$A = K \cdot R \cdot LS \cdot C \cdot P \qquad (6.1)$$

donde (en el Sistema Internacional):

- A Pérdida de suelo en una parcela (Tm / Ha / año).
- K Erodibilidad del suelo (Tm · Ha · hora / Ha / MJ / mm).
- R Erosividad de la lluvia (MJ · mm / Ha / hora / año).
- LS Factor topográfico.
- C Factor de cultivo y ordenación de la rotación.
- P Prácticas de conservación.

El factor R (Ec 6.2) pretende ser una medida de la energía impactante del agua. Su valor es anual. Funcionalmente, está ligado a la intensidad máxima de lluvia con duración de, al menos, treinta minutos. No tiene en cuenta ni la escorrentía (ella también disipa energía de la lluvia que cae) ni la orientación de la lluvia.

$$R = \sum_{i=1}^{n} E_i \, I_{30} \qquad (6.2)$$

donde:

- E_i Energía de cada tormenta (chubasco) ocurrida durante el intervalo de tiempo considerado (un año en nuestro caso). Sus dimensiones son MJ / Ha.
- I_{30} Intensidad máxima de las lluvias (mm / hora), en treinta minutos, durante el intervalo de tiempo considerado.
- n Número de chubascos que se dan en el año.

El valor de la energía cinética de cada chubasco (Ec 6.3) se calcula mediante la integración de la energía cinética por unidad de lluvia (e_c[1] [MJ / Ha / mm]), por la cantidad de lluvia (r) y por el tiempo (Δt).

---

[1] El valor de e_c se calcula mediante las siguientes relaciones empíricas:

$0{,}119 + 0{,}0873 \log r$	$r \leq 76$ mm / h
$0{,}283$	$r > 76$ mm / h

$$E_i = \sum_{t=0}^{t_i} e_c r \Delta t \qquad (6.3)$$

El factor K (erodibilidad del suelo), al igual que el anterior, se obtiene por regresión. Mide la resistencia de la estructura del suelo al impacto de la lluvia, pero también frente a la escorrentía (capacidad de producirla o de mitigarla). Todo esto implica que dicho factor K sea dependiente de la estructura y textura del suelo, así como de la presencia de mejorantes (materia orgánica, etc.).

El factor topográfico LS es fundamental. Toma el valor de la unidad cuando la parcela tiene una longitud de 22,13 m y una pendiente del 9%. La relación funcional con la longitud L y pendiente S (en %) aparece en la Ec 6.4.

$$LS = \left(\frac{L}{22,13} \right)^m \frac{0,043\,S^2 + 0,3S + 0,43}{6,613} \qquad (6.4)$$

siendo m una función de la pendiente que, de forma aproximada, se expresa:

$m = 0,3$ $S <= 3\%$
$m = 0,4$ $3 < S <= 5\%$
$m = 0,5$ $S > 5\%$

El factor de cultivo C se encuentra actualmente en profunda revisión. Hoy en día se considera descompuesto en cuatro subfactores: Consolidación del suelo, Cubierta por dosel, Cubierta por rastrojo, Rugosidad. Otra posible modificación sobre C es su ponderación estacional, debido a su estrecha relación tanto con la fenología (estado de la planta) como con el valor de R.

El factor P hace referencia al uso de prácticas de conservación (aterrazamiento, abancalamiento, etc.) y supone una corrección sobre el valor de la erosión calculada.

MODELADO DE LA PROPAGACIÓN DEL ERROR

Un modelado de la propagación del error se realiza mediante representaciones matemáticas que dependerán del proceso involucrado, de las hipótesis de propagación del error, y de ciertos conocimientos previos sobre la calidad de los datos y el propio comportamiento de las funciones. Así, los valores obtenidos de la aplicación directa de modelos sobre parámetros ambientales complejos, como la erosión, estarán muy condicionados, por ello habrá que interpretar los resultados en su justa medida

A pesar de poder plantear fácilmente, para algunos casos, modelos matemáticos de propagación del error, este método no suele ser el más idóneo debido a que la propagación de errores está muy afectada por la intercorrelación entre las variables. Por otra parte, existen operaciones de modelado cartográfico para las cuales la plasmación matemática de la propagación del error es bastante difícil. En estos casos, se recurre a la simulación.

Como se observa en la expresión final del modelo EUPS, se adopta un modelo multiplicativo formado por cinco parámetros. Si bien cada uno de estos cinco parámetros (*R*, *K*, *LS*, *C* y *P*) supone un verdadero submodelo, para mayor simplicidad en la exposición obviaremos esto, atendiendo a la valoración final de cada uno de ellos.

Modelado mediante representación matemática

Imaginando un mapa de tipo raster, y una celda, o píxel, *i* del mismo con unos valores *R*, *K*, *LS*, *C* y *P* medios, el estudio de la variación de la erosión se puede plantear de la siguiente forma:

$$\Delta A_i = (A_i - A_m) = \left(\frac{\partial A}{\partial R}\right)_R \Delta R + \left(\frac{\partial A}{\partial K}\right)_K \Delta K + \left(\frac{\partial A}{\partial LS}\right)_{LS} \Delta LS + \left(\frac{\partial A}{\partial C}\right)_C \Delta C + \left(\frac{\partial A}{\partial P}\right)_P \Delta P$$

Por lo tanto, la varianza quedará expresada de la siguiente forma:

$$S_{A}^{2}{}_{i} = \frac{1}{n} \Sigma \Delta A_i^2 = \frac{1}{n} \Sigma (A_i - A_m)^2$$

es decir:

$$S_{Ai}^2 = \frac{1}{n}\left[\left(\frac{\partial A}{\partial R}\right)_R \Delta R + \left(\frac{\partial A}{\partial K}\right)_K \Delta K + \left(\frac{\partial A}{\partial LS}\right)_{LS} \Delta LS + \left(\frac{\partial A}{\partial C}\right)_C \Delta C + \left(\frac{\partial A}{\partial P}\right)_P \Delta P\right]^2$$

que desarrollando:

$$S_{Ai}^2 = \frac{1}{n}\left(\frac{\delta A}{\delta R}\right)_R^2 \Delta R^2 + \frac{1}{n}\left(\frac{\delta A}{\delta K}\right)_K^2 \Delta K^2 + \frac{1}{n}\left(\frac{\delta A}{\delta LS}\right)_{LS}^2 \Delta LS^2 + \cdots$$

$$\cdots + \frac{2}{n}\left[\left(\frac{\delta A}{\delta R}\right)_R \left(\frac{\delta A}{\delta K}\right)_K \Delta R \Delta K + \left(\frac{\delta A}{\delta R}\right)_R \left(\frac{\delta A}{\delta LS}\right)_{LS} \Delta R \Delta LS + \cdots\right]$$

puesto que los temas se suponen independientes, las covarianzas serán nulas, quedando la expresión como sigue:

$$S_{Ai}^2 = \left(\frac{\delta A}{\delta R}\right)_R^2 \frac{1}{n} \Delta R^2 + \left(\frac{\delta A}{\delta K}\right)_K^2 \frac{1}{n} \Delta K^2 + \left(\frac{\delta A}{\delta LS}\right)_{LS}^2 \frac{1}{n} \Delta LS^2 + \left(\frac{\delta A}{\delta C}\right)_C^2 \frac{1}{n} \Delta C^2 + \left(\frac{\delta A}{\delta P}\right)_P^2 \frac{1}{n} \Delta P^2$$

$$S_{Ai}^2 = \left(\frac{\delta A}{\delta R}\right)_R^2 S_R^2 + \left(\frac{\delta A}{\delta K}\right)_K^2 S_K^2 + \left(\frac{\delta A}{\delta LS}\right)_{LS}^2 S_{LS}^2 + \left(\frac{\delta A}{\delta C}\right)_C^2 S_C^2 + \left(\frac{\delta A}{\delta P}\right)_P^2 S_P^2$$

por lo que finalmente resulta la siguiente expresión, donde cada una de las varianzas deberían sustituirse por un estudio similar dentro de cada uno de los submodelos que definen el parámetro de interés:

$$S_{Ai}^2 = (K(LS)CP)^2 S_R{}^2 + (R(LS)CP)^2 S_K{}^2 + (KRCP)^2 S_{LS}{}^2 + (RK(LS)P)^2 S_C{}^2 + (RK(LS)C)^2 S_P{}^2$$

Hay que subrayar que al considerar este modelo en cada celdilla *i* aislada, se están obviando algunos efectos que pueden ser importantes. El modelo podría mejorarse en gran medida considerando una vecindad de cuatro vecinos o de ocho vecinos.

Modelado mediante simulación

Como ya se ha referido, el estudio por simulación puede facilitar en algunos casos la obtención de la solución final. Tras el establecimiento de las hipótesis de error sobre cada uno de los temas, el proceso de simulación en este estudio requiere una serie de pasos:

- Según las hipótesis establecidas, introducir errores aleatorios en cada una de las capas origen, que se consideran entradas al modelo.
- Aplicar las transformaciones SIG sobre las entradas perturbadas. El producto del modelado cartográfico es en este caso una realización de la EUPS.
- Comparar el resultado de la EUPS derivado de las entradas perturbadas con otro el derivado de los datos originales.
- Repetir los procesos 1, 2, y 3 un número *M* adecuado de veces.
- Calcular las características estadísticas de los *M* modelos espaciales de la EUPS así formados.

Indudablemente, si el número de iteraciones (*M*) de la simulación es suficientemente alto, la solución de este proceso debe coincidir con el que se obtendría del estudio matemático. Este flujo de operaciones queda descrito en la Figura 6.8.

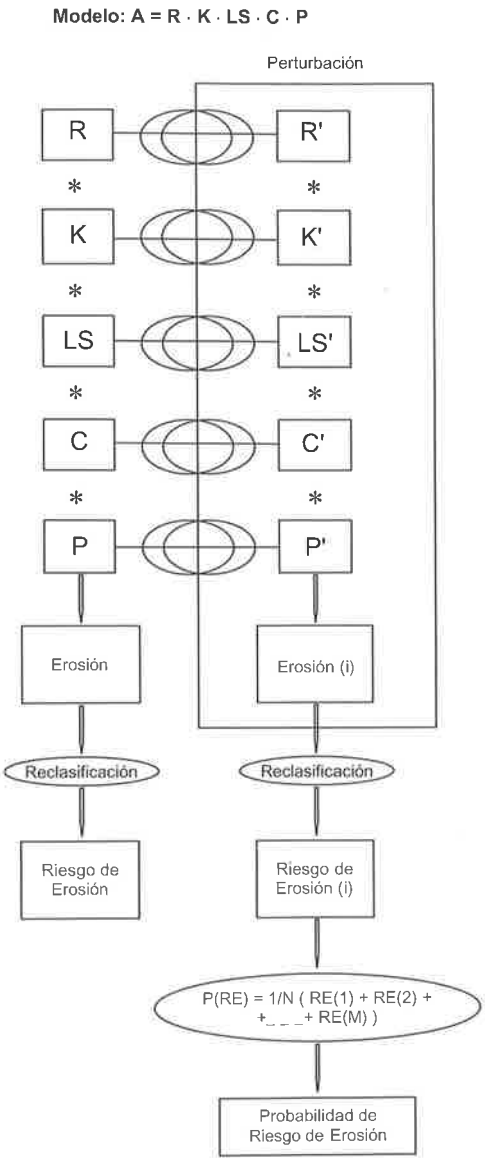


Figura 6.8 - Modelo de Erosión Potencial y de simulación del proceso

Ejemplo de simulación en el modelado del error

Se presenta aquí un ejemplo simplificado de simulación aplicada al modelado del error en un entorno de Sistema de Información Geográfica. En la Tabla 6.1 se muestra un código elemental desarrollado para el programa ArcInfo, pero que, con las modificaciones pertinentes, podría utilizarse en cualquier SIG que permita el desarrollo de *scripts*.

Para el desarrollo de este ejemplo se han tomado como capas las siguientes, cuyo tamaño de celda es de 80×80 m^2 (Figura 6.9):

- Capa de elevaciones del terreno, de la que se deriva el factor *LS* según las relaciones apuntadas previamente.
- Capa de cultivos y aprovechamientos de la que, por reclasificación, se deriva la cartografía de factor de cultivo *C* y de prácticas de conservación *P*. Se considera que *P* es igual a la unidad para todo el territorio.
- Capa edafológica, que mediante reclasificación, en función de unas tablas específicas, permite determinar el factor de erodibilidad del suelo *K*.
- El valor de *R* promedio para la zona es de 168.

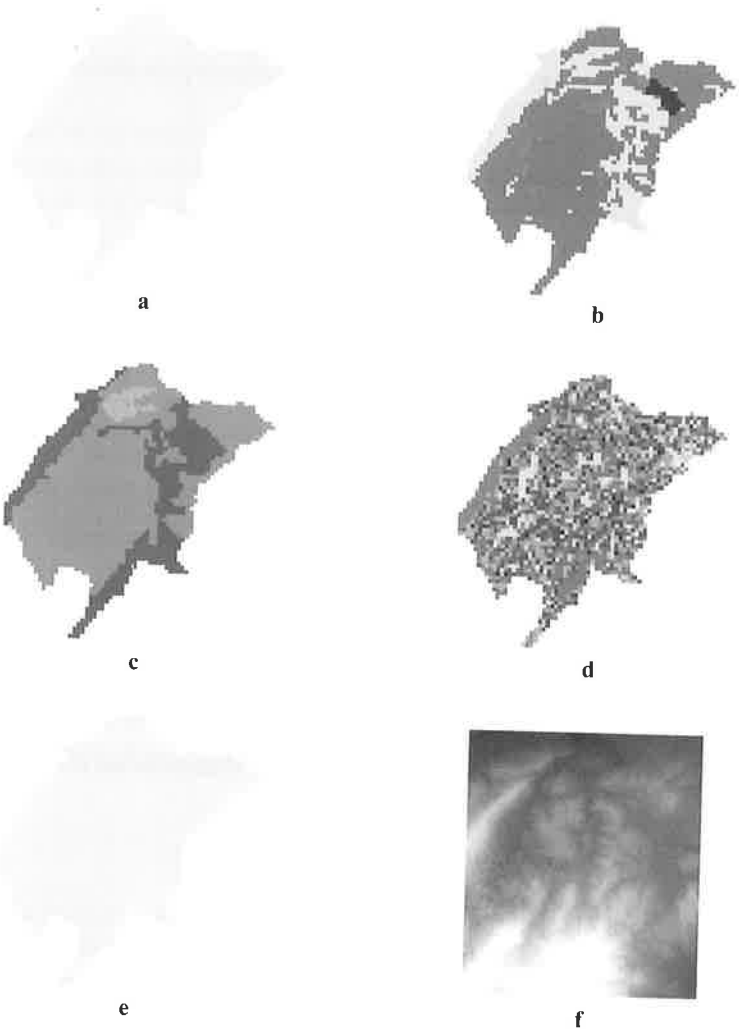


Figura 6.9 - a) Máscara de la zona de trabajo, b) Factor C, c) Factor K, d) Factor LS, e) Factor R, f) MDT de la zona

Utilizando toda esta información como verdad terreno, y asignándoles unos valores medios y desviaciones poblacionales (Tablas 6.2, 6.3 y 6.4), se ha procedido a la simulación tal como queda descrita en el siguiente programa, que se ejecuta dentro de *GridTools* de *Arc/Info*. Se alimenta con unas capas (*Grids*) correspondientes a cada uno de los factores (Figura 6.9) y realiza tantas ejecuciones de simulación como se indique. Como resultado el sistema genera dos ficheros, el primero con el resultado medio de erosión según la EUPS tras las iteraciones, y el segundo, una vez fijado un umbral de riesgo, la probabilidad de éste tras el proceso de simulación (Figura 6.10).

El programa se presenta en la Tabla 6.1 en letra `Courier`, y los comentarios y explicaciones aclaratorias del programa en letra *Times New Roman*. Finalmente, la Figura 6.10 presenta un ejemplo de los resultados obtenidos.

Tabla 6.1 - Código del programa de simulación para Arc/Info

```
/* PROGRAMA PARA LA SIMULACIÓN EN LA MODELIZACION DE LA PROPAGACIÓN
/* DEL ERROR EN LA ECUACION UNIVERSAL DE PERDIDA DE SUELO

&ECHO &OF
DISPLAY 9999
CLEAR
```

Se introducen los ficheros de cada factor, los valores de desviaciones de los mismos, el fichero de reclasificación 0/1 según exista riesgo o no, las iteraciones a realizar y una máscara de la zona.

```
/* IMÁGENES Y VALORES
&SV FR =              [RESPONSE ' Introduce el fichero de Factor R ' n ]
&SV DFR =             [RESPONSE ' Introduce la desviación de R ' n ]
&SV FK =              [RESPONSE ' Introduce el fichero de Factor K ' n ]
&SV DFK =             [RESPONSE ' Introduce las desviaciones de K 'n ]
&SV FLS =             [RESPONSE ' Introduce el fichero de Factor LS ' n ]
&SV DFLS =            [RESPONSE ' Introduce las desviaciones de LS ' n ]
&SV FFC =             [RESPONSE ' Introduce el fichero de Factor C ' n ]
&SV DFFC =            [RESPONSE ' Introduce las desviaciones de C ' n ]
&SV xxmascara =       [RESPONSE ' Introduce la máscara de la zona ' n ]
&SV xxITERACIONES =   [RESPONSE ' Introduce las iteraciones ' n ]
&SV RIESGO =          [RESPONSE ' Introduce fichero de reclasificación' n]
```

Se establece la ventana de análisis tomándola de una de las imágenes introducidas.

```
/* VENTANA DE ANALISIS
SETWINDOW %FLS%
SETCELL %FLS%
STATUS
```

Se crean ficheros intermedios.

```
/* CREACION DE GRIDS INTERMEDIOS Y PUESTA A CERO
xxSUMAMOD = %FK% * 0
xxSUMARIE = %FK% * 0
```

Para cada componente del modelo de la EUPS se crea un modelo perturbado según errores normales y desviaciones indicadas en el inicio del programa.

```
/* BUCLE DE CALCULO
&DO EJECUCIONES = 1 &TO %xxITERACIONES%
```

Tabla 6.1 - Código del programa de simulación para Arc/Info

```
/* GENERACION DE LOS MODELOS PERTURBADOS PARA CADA COMPONENTE
xxIMG-ALEA1 = normal ( ) * %DFR% + %FR%
xxIMG-ALEA2 = normal ( ) * %DFK% + %FK%
xxIMG-ALEA3 = normal ( ) * %DFLS% + %FLS%
xxIMG-ALEA4 = normal ( ) * %DFFC% + %FFC%
```

*Se aplica el modelo multiplicativo Erosión = K * R * LS * C * P.*

```
/* APLICACIÓN DEL MODELO
xxite-MODELO = xxIMG-ALEA1 * xxIMG-ALEA2 * xxIMG-ALEA3 * xxIMG-ALEA4
```

Se reclasifica el modelo xxite-MODELO utilizando el fichero ASCII indicado por la variable RIESGO, en el que se indican los intervalos de la reclasificación.

```
/* RECLASIFICACION SEGUN EL UMBRAL DE RIESGO
xxite-MAYOR = RECLASS (xxite-MODELO, %RIESGO%, DATA)
```

Se acumula el valor de la erosión para cada iteración.

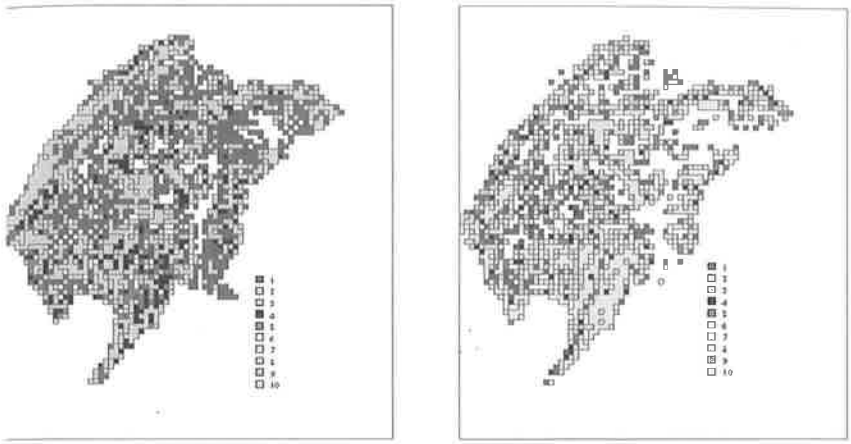
```
/* CALCULO DEL VALOR ACUMULADO EROSION
xxpar-mod = xxite-MODELO + xxSUMAMOD
KILL xxSUMAMOD ALL
xxSUMAMOD = xxpar-mod
```

Se acumula la presencia de riesgo para cada iteración.

```
/* CALCULO DEL VALOR ACUMULADO DE RIESGO
xxpar-may = xxite-MAYOR + xxSUMARIE
KILL xxSUMARIE ALL
xxSUMARIE = xxpar-MAY

/* BORRADO DE LOS RESULTADOS INTERMEDIOS
KILL xxIMG-ALEA1 ALL
KILL xxIMG-ALEA2 ALL
KILL xxIMG-ALEA3 ALL
KILL xxIMG-ALEA4 ALL
KILL xxpar-MOD ALL
KILL xxpar-MAY ALL
KILL xxite-MODELO ALL
KILL xxite-MAYOR ALL

/* FIN DEL BUCLE
&END
```

Calculo del valor medio. Se divide el valor acumulado entre el número de iteraciones. Se recorta con la máscara.

```
/* CALCULO DE LA EROSION MEDIA DE LA SIMULACION
ERO-MED = %mascara% * ( xxSUMAMOD / %xxITERACIONES%)
KILL xxSUMAMOD ALL
GRIDPAINT ERO-MED
&SV KK = [RESPONSE ' Esta es la Erosion media ']

/* CALCULO DE LA PROBABILIDAD DEL RIESGO EN LA SIMULACION
RIES-MED = %mascara% * (xxSUMARIE / %xxITERACIONES%)
KILL xxsumarie ALL
GRIDPAINT RIES-MED
&SV kk = [RESPONSE ' Este es el riesgo medio']

/* FIN DEL PROGRAMA
&RETURN
```



a) b)

Figura 6.10 - Resultados de la simulación: a) Erosión media según la EUPS para 10 iteraciones clasificada en el intervalo [0, 65], b) Probabilidades, en tantos por uno, del riesgo medio de erosión mayor de 10 Ton/ha/año para 10 iteraciones

Tabla 6.2 - Valores de C y P para los cultivos		
Cultivo	**Factor C**	**Factor P**
Monte alto y m. bajo	0,110	1
Olivar	0,420	1
Erial	0,038	1
Secano	0,520	1
Huerta	0,290	1

Tabla 6.3 - Principales parámetros del Suelo						
Tipo	**K (SI)**	**Arcilla %**	**Limo + Arena %**	**MO**	**Estructura**	**Permeab.**
13: cambisoles cálcicos	0,0355	25	72	3	4	4
15: fluvisoles calcáreos	0,0632	47	52	1	4	5
19: luvisoles crómicos	0,0382	39	55	3,5	4	4
21: regosoles calcáreos	0,0460	32	65	1,5	4	4

Tabla 6.4 - Desviaciones	
R	10,0
K	0,005
LS	2,0
C	0,05
P	0,0

Nota: Estos valores usados para las desviaciones son arbitrarios, no existiendo ningún estudio o referencia que los avale.

TRABAJO A REALIZAR

Se propone realizar un doble estudio, consistente en analizar cómo afecta la incertidumbre posicional existente en la coordenada altimétrica en:

— La determinación automática de cuencas hidrográficas.

— La determinación automática de la posición de los cursos de agua.

Para el desarrollo de esta práctica hace falta disponer de un programa SIG de tipo ráster, y de cierto conocimiento de programación sobre dicho programa. También hace falta disponer de un MDT que cubra una zona que presente suficiente variabilidad, y que se corresponda con la cuenca aportadora de un arroyo o río. Es conveniente que la ventana geográfica no sea muy extensa, al objeto de que el fichero no sea voluminoso.

Partiendo del MDT, cuya precisión altimétrica debe ser conocida, se escribirá un programa que, basado en los comandos internos del SIG, genere automáticamente perturbaciones sobre la cota de los puntos del MDT[2]. Los MDT así perturbados se incorporarán como entradas en los procesos de análisis SIG que permiten determinar cuencas aportadoras y líneas de corriente.

Para el análisis se propone seguir una metodología similar a la mostrada en el ejemplo incluido en esta práctica.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

puede encontrar el material necesario para este caso.

[2] El comportamiento real de las incertidumbres en un MDT está ligado, por autocorrelación espacial, con sus valores vecinos, y sus valores muy condicionados por la propia pendiente del terreno. No obstante, en la realización del ejercicio planteado, y en aras a simplificarlo, se propone obviar estas dos circunstancias.

PRÁCTICA 28

Simulación para la comparación de matrices de confusión

EL BOOTSTRAP

El *Bootstrap* es un método de simulación que permite estimar la distribución muestral de un estadístico cuando no existe un estimador paramétrico. Consiste en el remuestreo aleatorio de la muestra original un elevado número de veces, de forma que en cada iteración el estadístico de interés es calculado y almacenado. Así se consigue obtener una aproximación de la distribución, que vendrá dada por la distribución frecuencial del estadístico, y se podrá estimar, por ejemplo, la desviación típica.

El proceso de *Bootstrap* puede resumirse en los siguientes pasos (Hardin y Shumway, 1997)

- Extraer aleatoriamente una muestra de tamaño n a partir de la población utilizando una estrategia de muestreo apropiada.
- Extraer m muestras aleatorias a partir de la muestra original, siendo m elevado ($m > 200$). Cada muestra también debe ser de tamaño n.
- Determinar el estadístico de interés en cada muestra. Ordenar el conjunto de m valores del estadístico para crear una distribución, que será una aproximación a la distribución muestral del estadístico.

APLICACIÓN A LAS MATRICES DE CONFUSIÓN

Como se explica en el Bloque de Calidad Temática, disponer de los valores normalizados de la matriz facilita la comparación celda a celda. Sin embargo, dado que no se conoce la función de distribución de los mismos, no existe método estadístico paramétrico para decidir si las diferencias existentes son estadísticamente significativas (Ariza, 2002; Hardin y Shumway, 1997), por lo que el usuario queda limitado a la inspección visual.

La utilidad de la aplicación del método de *Bootstrap* a las matrices de confusión reside en la posibilidad de estimar la distribución de cada una de las celdas, con lo que ya se pueden contrastar posteriormente hipótesis estadísticas que permitan saber si es significativa la diferencia entre una misma celda en dos matrices diferentes.

El proceso a aplicar para una determinada celda de la matriz sería el siguiente:

— Obtener la matriz de confusión procedente de la elaboración de cartografía temática, utilizando una técnica de muestreo apropiada (ver ejemplo en Tabla 6.5).

Tabla 6.5 - Ejemplo de matriz de confusión.

	V1	V2	V3
C1	4	1	2
C2	1	5	0
C3	2	1	3

— Convertir la matriz de confusión en una lista donde el número de elementos es igual a *n*. Cada celda de la matriz introduce en la lista tantos elementos como el valor de la celda. Cada elemento debe tener asociado un indicador de la fila y columna a la que pertenece en la matriz de confusión (ver ejemplo en Tabla 6.6 donde cada elemento de la lista es de la forma fila-columna).

Tabla 6.6 - Matriz de confusión de la Tabla 6.5 en forma de lista.

1-1	1-1	1-1	1-1	1-2	1-3	1-3	2-1	2-2	2-2	2-2	2-2	2-2	3-1	3-1	3-2	3-3	3-3	3-3

— Extraer una muestra aleatoria con reemplazo de tamaño *n* de la lista y construir una nueva matriz de confusión sirviéndose de los indicadores de fila y columna.

— Aplicar a la matriz el método de pseudoceros de Fienberg (1970) (descrito en el bloque de calidad temática).

— Normalizar la matriz usando el método de Fienberg (1970) (descrito en el bloque de calidad temática).

— Guardar el valor de la celda.

— Repetir los pasos c) a f) *m* veces, siendo *m* un número suficientemente elevado.

— Ordenar los valores almacenados de la celda para construir el histograma de frecuencias, que será la estimación de la distribución.

— Contrastar la normalidad de la distribución estimada.

— Estimar la desviación típica de la celda.

Una vez se conocen las desviaciones típicas en dos celdas de dos matrices que se pretende comparar, se puede aplicar un contraste de hipótesis de igualdad de valores:

$$Z = \frac{x_{i,j_1} - x_{i,j_2}}{\sqrt{\hat{\sigma}^2_{i,j_1} + \hat{\sigma}^2_{i,j_2}}}$$

donde:

Z Estadístico para contrastar la hipótesis.

x_{i,j_1} Valor original de la celda (i,j) en la matriz 1.

x_{i,j_2} Valor original de la celda (i,j) en la matriz 2.

$\hat{\sigma}^2_{i,j_1}$ Desviación típica estimada para la celda (i,j) en la matriz 1.

$\hat{\sigma}^2_{i,j_2}$ Desviación típica estimada para la celda (i,j) en la matriz 2.

TRABAJO A REALIZAR

La práctica consistirá en aplicar el método de simulación de *Bootstrap* para poder comparar las celdas de dos matrices de confusión. Para ello se considerarán las matrices propuestas para su comparación en el bloque de calidad temática (tablas 5.17 y 5.18). El número de simulaciones deberá ser igual o superior a 200.

Una vez efectuado el *Bootstrap*, se efectuará un contraste de hipótesis de igualdad de medias de todas las celdas de ambas matrices. El nivel de confianza a considerar es el 95%.

Para todo ello se recomienda recurrir a elaboración de un programa informático[3] que efectúe los cálculos necesarios, o al uso de una hoja de cálculo.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

[3] Puede tratarse de un programa o de una macro desarrollada en una hoja de cálculo.

BLOQUE 7

CONTROL DE OTRAS COMPONENTES DE UNA BDG

En la actualidad, junto al uso creciente de las BDG, también existe un aumento del afán, por parte de los productores y usuarios, de incorporar informaciones de valor sobre la calidad de los datos. De no hacerse así, gracias a las facilidades de análisis mediante las herramientas informáticas, el dato geográfico puede convertirse en un verdadero peligro por las consecuencias que pudieran derivarse de su uso en la toma de decisiones incorrectas.

Así, en este bloque se proponen un conjunto de casos que se refieren a componentes de la calidad menos clásicas, sobre las que existe menor desarrollo y grado de aplicación, y que no pueden considerarse secundarias.

Los casos que se proponen son las siguientes:

— Caso 29: Control de la compleción y de la calidad temática. Se propone un ejercicio de elaboración de tablas de errores de comisión y omisión a partir de una representación gráfica de una BDG y de su verdad nominal. En cuanto a la calidad temática, de forma análoga a la estudiada en la clasificación de imágenes de satélite, se propondrá la elaboración de matrices de confusión entre objetos y atributos de la base de datos.

— Caso 30: Control de la coherencia lógica. Centrada en las leyes que han de cumplir el conjunto de datos de una BDG, la tarea propuesta es el control de la coherencia temática y la adherencia al formato del fichero a partir de un sencillo ejemplo con datos municipales de superficie y población.

— Caso 31: Elaboración de metadatos. Se considera la elaboración un informe de metainformación. El trabajo propuesto consiste en tomar la Norma MIGRA[1] (AENOR 1998) para elaborar un informe de una base de datos con abundante información disponible, como es la Carta Digital del Mundo (*Digital Chart of the World*).

Las componentes sobre las que se trabaja en estos casos son de muy reciente adopción. A nivel normativo son varias las normas que cubren estos aspectos, destacando entre ellas la ISO 19114, que presenta diversos ejemplos con elementos y subelementos propios de coherencia y compleción, y la ISO 19115, que se dedica a los metadatos y sobre la cual se están desarrollando las iniciativas de infraestructuras de datos espaciales (IDE's). Una visión más general se puede encontrar en el manual editado por Guptill y Morrison (1995) y en el trabajo de Ariza (2002).

[1] En la actualidad se están llevando acabo en España diversas iniciativas encaminadas a las infraestructuras de datos espaciales (IDEs), en las que se prevé aplicar la Norma ISO 19115. Sin embargo, aquí hemos preferido presentar MIGRA por ser su aplicación mucho más simple.

7.1. CASO 29

Control de la compleción y de la calidad temática

COMPLECIÓN

Según el Diccionario de la Lengua Española (RAE, 2001) la compleción es la cualidad de completo, por tanto, si se desea conocer si una cosa está/es completa se deberá comparar con un patrón. En el caso de las BDG la compleción debe entenderse frente a unas demandas o necesidades que ha de cubrir el producto cartográfico, pero para esto es necesario concretar el patrón de comparación, aspecto que se materializa en las especificaciones de producto. Por tanto, la compleción no es evaluada por comparación directa con el mundo real, sino frente a lo que se denomina terreno nominal, nivel que se establece sobre aquel tras el proceso de abstracción que se muestra en la Tabla 7.1, y que se materializa como el subconjunto de elementos del mundo real que cumplen con las especificaciones de producto.

Tabla 7.1. Proceso de abstracción: desde el mundo real al terreno nominal

Mundo real ⇓	Es la realidad en toda su complejidad. Es el conjunto de elementos, características, relaciones y hechos que pueden ser o no conocidos por el ser humano.
Mundo conceptual o universo abstracto ⇓	El discurso humano no tiene lugar al nivel del mundo real, puesto que mediante nuestro lenguaje natural, dando nombres a las cosas, abstraemos la realidad. De esta manera se conforma un mundo conceptual, o universo abstracto, como el conjunto de entidades percibidas de la realidad que sirve como marco de referencia para la obtención de un conjunto de datos.
Mundo Geoespacial ⇓	El universo abstracto no es suficientemente adecuado para su plasmación en BDG. El modelo geoespacial, con sus restricciones y peculiaridades, establece una nueva simplificación o abstracción que facilita el tratamiento de la información, conformándose así en mundo geoespacial
Universo de discurso ⇓	Visión del mundo real, o hipotético, que incluye todo lo que es de interés para una aplicación geoespacial concreta. Esta visión está condicionada por el proceso de abstracción que genera el mundo conceptual y por las restricciones que impone el modelo geoespacial y por los requisitos de la aplicación concreta (especificaciones de producto).
Terreno nominal	Concepción del mundo real derivada de las especificaciones de un conjunto de datos geográficos; por tanto es la plasmación del universo de discurso como selección de elementos del mundo real. El terreno nominal constituye el conjunto ideal de datos geográficos, con el que se compara el conjunto real, la BDG, para evaluar su calidad. En ocasiones puede venir definido por otro producto cartográfico, siendo recomendable que su calidad esté contrastada.

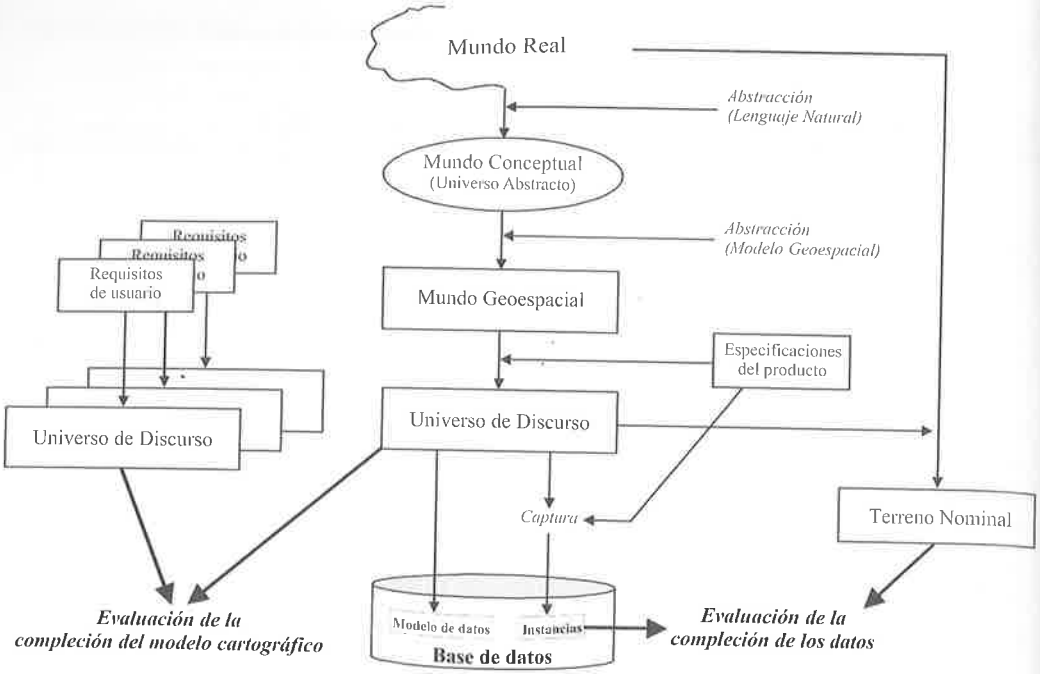


Figura 7.1 - Relaciones entre el proceso de abstracción y la compleción.

Como se observa en la Figura 7.1, se puede hablar de una compleción de los datos (calidad técnica), la cual vendría a indicar en qué medida, de omisión o comisión, los objetos presentes en las bases de datos cumplen con las prescripciones establecidas (Figura 7.2). Por otro lado se encuentra la compleción del modelo cartográfico (calidad teórica) que hace referencia a la inclusión de las clases, atributos y relaciones necesarias para plasmar adecuadamente la realidad en el modelo cartográfico.

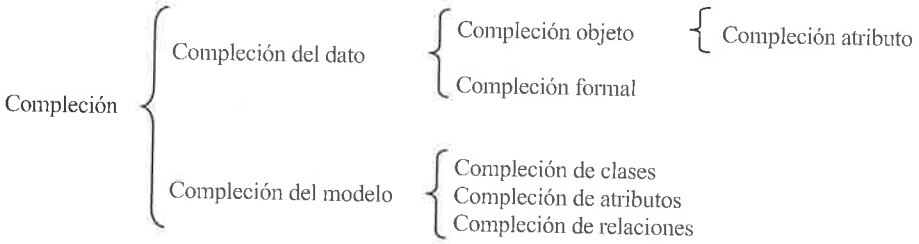


Figura 7.2 - Esquema con las acepciones de compleción

Tanto los errores de omisión de objetos (compleción del objeto) como la de partes de ellos (compleción del atributo) pueden afectar a la consistencia lógica de la base de datos. En estos casos se debe señalar tanto como un error de falta de compleción como de inconsistencia lógica a la hora de elaborar el informe de calidad. En la Tabla 7.2 se pre-

senta un resumen de los diferentes casos que se pueden presentar cuando existen problemas de compleción de objetos o atributos.

Tabla 7.2. Ejemplo de diferentes problemas de compleción de objetos o atributos

	ATR 1	ATR 2	ATR 3	ATR 4[a]	ATR 5	ATR 6	ATR[b]	...
OBJETO 1	✓	✓	✓	×		×[c]	×	...
OBJETO 2			×	×			×	...
OBJETO 3		✓/×[d]		×			×	...
OBJETO 4[e]	×	×	×	×	×	×	×	...
OBJETO ..				×			×	...
×[f]	×	×	×	×	×	×	×	...
...	...	...	...	...	...	...	...	...

[a] El valor del atributo se ha perdido para todos los objetos.
[b] Falta este tipo de atributo.
[c] Falta el valor del atributo para este objeto.
[d] El atributo compuesto está incompleto.
[e] Faltan los valores de todos los atributos para este objeto.
[f] Falta un objeto (este es el único caso de compleción de objetos en este ejemplo).

Fuente: Brassel y col. (1995).

Los diferentes tipos de falta de compleción de atributos de la Tabla 7.2 pueden resumirse en los siguientes:

— Ausencia de tipos de atributos, incluso sin especificar en los metadatos. La omisión de tipos de atributos sólo puede comprobarse con las referencias y definiciones del modelo cartográfico el cual establece la información de interés (por ejemplo, la "lista de atributos relevantes para objetos de este tipo").

— Los metadatos especifican que debe aparecer un tipo de atributo, pero no hay valor alguno codificado. Éste puede comprobarse con facilidad con el gestor de bases de datos.

— No aparecen valores de atributos en algunos objetos de un determinado tipo. Esto se comprueba de forma automática.

— Están incompletos los atributos complejos. Si la estructura del atributo es estática, y no afecta a la consistencia lógica, se puede comprobar de forma automática. En caso contrario, las omisiones sólo se hallarán con la revisión manual de los datos.

En cuanto a la falta de compleción de objetos, hay que comprobar que los datos están representando completamente todas las entidades del universo de discurso. Éste estará claramente definido en el caso de la existencia de una referencia explícita, en forma de tablas, inventarios o definiciones recogidas en un estándar. Si el universo de discurso viene definido de forma implícita, habrá que recurrir a la interpretación, el conocimiento de expertos y la opinión individual. Una vez que el universo de discurso queda definido, la compleción de objetos se establece simplemente mediante una comparación

de datos a través del el terreno nominal que aquel define. El informe de compleción debe incluir información exacta acerca de la derivación del universo de discurso, indicando las listas, estándares, mapas, inventarios, informaciones de experto, etc. La Tabla 7.3 muestra un ejemplo de informe de compleción de objetos y atributos.

Tabla 7.3 - Informe de compleción consistente en una serie temporal de observaciones sobre plantaciones de avellano

Especificaciones del conjunto de datos

Descripción general:

El conjunto de datos representa las fechas anuales de floración completa del avellano en 50 localizaciones concretas entre 1988 y 1993 y es actualizado cada año. El muestreo de los datos y su manipulación son realizados por el Dpto. de Geografía de la Universidad de Bern (GIUB). Los lugares de observación pueden variar ligeramente de un año a otro. Los emplazamientos y atributos requeridos están especificados en un informe interno en el GIUB desde el 23 de septiembre de 1987.

Nombre del conjunto de datos:	phaeno_hazel.88/93
Productor de los datos:	Universidad de Bern, Departamento de Geografía.
Catálogo de referencia:	Informe interno en el GIUB desde el 23 de septiembre de 1987.
Tipo de entidad:	Fecha de la floración completa del avellano.
Período de tiempo:	1988-actualidad, actualización anual.
Número de objetos:	50
Atributos requeridos:	coord-x; coord-y; elevación; pendiente; orientación; fecha; nombre de localización.

Informe de compleción

El productor de los datos es responsable de proporcionar una actualización del informe de compleción.
El informe de compleción está disponible en copia impresa o archivo *Post-Script*.
Todos los tipos de compleción están estrictamente evaluados según las especificaciones del conjunto de datos.
Los registros sin fecha son presentados como omisiones de objeto.
Compleción de objeto (C.O.):

Se refiere al número de omisiones de objeto por año.
Cada omisión ha de ser listada al final (entre paréntesis los años correspondientes).

Compleción de atributo (C.A.): distinción entre omisiones globales y locales en cada año.
Omisiones globales (valores de un atributo inexistentes en un año), aparecen con "x".
Las omisiones locales son dadas como un porcentaje para cada atributo y año.

Descripción general

El conjunto de datos representa las fechas anuales de floración completa del avellano en 50 localizaciones concretas entre 1988 y 1993 y es actualizado cada año. El muestreo de los datos y su manipulación son realizados por el Dpto. de Geografía de la Universidad de Bern (GIUB). Los lugares de observación pueden variar ligeramente de un año a otro. Los emplazamientos y atributos requeridos están especificados en un informe interno en el GIUB desde el 23 de septiembre de 1987.

Año:	1988	1989	1990	1991	1992	1993	año
C.O.	2	4	0	1	0	1	
C.A.							
Coord-x	2	4	10	0	0	2	
Coord-y	2	0	10	8	0	0	
Elevacion	40	28	20	22	8	12	
Orientacion	x	x	50	10	8	8	
Pendiente	x	x	60	22	22	36	
Nombre loc	20	16	24	8	14	6	

Lista de omisiones de objetos [nombre de la localización (años)]:
Boudry (1988, 1989, 1991); Moutier (1989); St. Imier (1988, 1989); Bassecourt (1989); Grandval (1993).

Fuente: Brassel y col. (1995).

CALIDAD TEMÁTICA

La compleción está ligada a los errores de omisión y comisión por lo que las matrices de error son una herramienta válida también para este caso. Por ejemplo, en la Tabla 7.4 se presenta un ejemplo de matriz de confusión[2] obtenida tras la comparación de la BDG con su correspondiente terreno nominal. Entre paréntesis aparece el número de casos de cada entidad o categoría del atributo. En la Tabla 7.5 se presenta un ejemplo de matriz de confusión entre clases de un mismo tipo de atributo.

Tabla 7.4 - Matriz de confusión entre entidades

	Sendero (5)	Carretera (7)	Árbol (26)	Industria (3)	Casa (10)
Sendero (7)	5/7 = 71%	2/7 = 28,5%	0%	0%	0%
Carretera (5)	0%	5 / 5 = 100%	0%	0%	0%
Árbol (25)	0%	0%	26/25 = 104%	0%	0%
Industria (4)	0%	0%	0%	3/4 = 75%	1/4 = 25%
Casa (10)	0%	0%	0%	0%	9 / 10 = 90%

Tabla 7.5 - Matriz de confusión entre categorías de atributo de una entidad (árbol)

	Clase A 1 a 3 metros (6)	Clase B 3 a 5 metros (8)	Clase C 5 a 10 metros (5)	Clase D > 10 metros (4)	Sin determinar (1)
Clase A (5)	4/5 = 80%	1/5 = 20%	0%	0%	0%
Clase B (8)	1/8 = 12,5%	5/8 = 62,5%	0%	0%	1/8 = 12,5%
Clase C (10)	1/10 = 10%	2/10 = 20%	5/10 = 50%	4/10 = 40%	0%
Clase D (2)	0%	0%	0%	2/2 = 100%	0%
Sin determinar (0)	0%	0%	0%	0%	0%

TRABAJO A REALIZAR

Se pretende cuantificar la exactitud de un conjunto de datos presentes en una BDG frente a sus especificaciones, con el objetivo de incluir los resultados en los metadatos, y en un informe adicional de la calidad de los datos. Para ello se deberán realizar varias tareas:

— Elaborar una lista de errores. Se comparará minuciosamente el universo de discurso (Figura 7.3) con el conjunto de datos (Figura 7.4).

— Establecer los errores de omisión y comisión que se han producido para cada tipo de objeto. Para ello se propone seguir un esquema como el que figura a continuación. Seguidamente elaborar la matriz de confusión entre los objetos.

[2] Siguiendo el mismo criterio que se apuntó anteriormente para la construcción de las matrices de confusión, las columnas se refieren a la verdad del terreno, en este caso, el universo de discurso, y las filas a las unidades plasmadas en realmente en la cartografía, es decir, en la BDG.

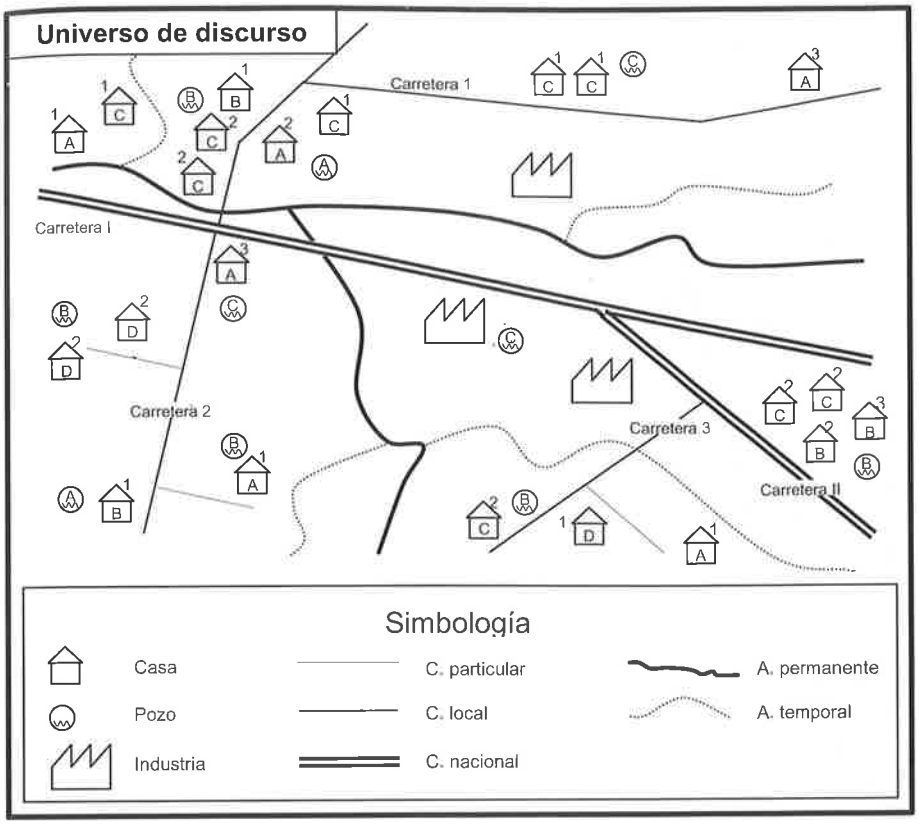
Objeto	nº casos en universo de discurso	Error de comisión	Error de comisión (%)	Error de omisión	Error de omisión (%)
Casa					
...					

— Elaborar las tres matrices correspondientes a las confusiones existentes entre las categorías de los atributos "número de ocupantes", "número de plantas" y "profundidad media".

Las especificaciones de producto definiendo el universo de discurso aparecen en la Tabla 7.6. Las especificaciones describen aquellas entidades, atributos y relaciones que ocurren en la realidad y que son consideradas importantes para el producto. El universo de discurso que se ajusta a las especificaciones aparece gráficamente en la Figura 7.3.

Tabla 7.6 - Especificaciones de producto

Objeto	**Atributos**	**Tipo de variable**	**Dominio**
Casa	número de plantas número de ocupantes	entero cadena de texto	1 - 3 A: 0; B: 1-3; C: 4-5; D: >5
Pozo	profundidad media (m)	cadena de texto	A: <2; B: 2-5; C: >5
Industria	-	-	-
Carretera nacional	nombre	cadena de texto	-
Carretera local	nombre	cadena de texto	-
Carretera particular	nombre	cadena de texto	-
Arroyo permanente	nombre	cadena de texto	-
Arroyo temporal	nombre	cadena de texto	-
Nota: El nombre de las carreteras y arroyos puede no existir o ser desconocido.			



La letra dentro del símbolo de la casa se corresponde con el número de ocupantes.
El dígito sobre el símbolo de la casa es el número de plantas.
El dígito dentro del símbolo de pozo es la profundidad media, en metros.
El nombre de la carretera aparece junto a ella.

Figura 7.3 - Universo de discurso

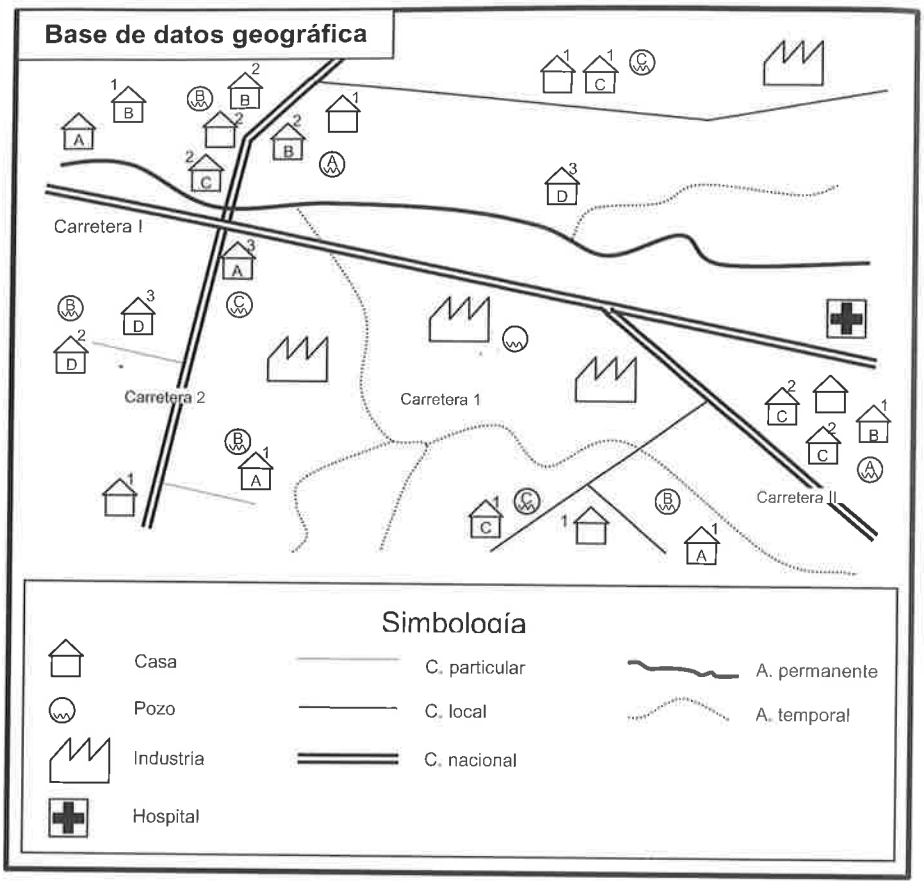


Figura 7.4 - Base de datos geográfica

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

7.2. CASO 30

Control de la coherencia lógica

LA COHERENCIA LÓGICA

Según la Norma UNE-ENV 12656 (AENOR, 2000), la coherencia, o consistencia lógica, hace referencia al grado de conformidad de un conjunto de datos geográficos con respecto a la estructura interna descrita por sus especificaciones (estructura, atributos, las relaciones de compatibilidad entre los datos espaciales, etc.). La coherencia se puede considerar desde varios puntos de vista, entre otros:

- Geométrico (coherencia geométrica). Tiene en cuenta las especificaciones que sólo afectan a los aspectos geométricos de la información geográfica.
- Temático[3] (coherencia temática). Considera las especificaciones que sólo afectan a los aspectos temáticos de la información geográfica.
- Topológico (coherencia topológica). Considera las especificaciones que sólo afectan a los aspectos geométricos y topológicos de la información geográfica.

Así, las leyes de la topología son las que se han de cumplir desde el punto de vista de la geometría y topología, mientras que, desde el punto de vista de los atributos, se ha de procurar que los valores estén dentro de un rango, o que coincidan con unos valores establecidos, que exista codificación para la ausencia de valor y para valores nulos, que exista consistencia referencial, etc. No existe consistencia lógica en una base de datos cuando se encuentran contradicciones lógicas entre los elementos contenidos en la misma. La Tabla 7.7 presenta algunos ejemplos de posibles medidas de nivel de coherencia en distintos ámbitos.

Tabla 7.7 - Indicadores y medidas de la coherencia de una BDG

Indicador	Medida	Comentario
Consistencia topológica	Porcentaje de uniones que se formarán entre líneas	Valor dentro de un rango
	Porcentaje de polígonos correctamente formados	Valor dentro de un rango
	Tolerancias usadas en limpiar líneas o unir áreas adyacentes	Vector de medidas indicando tolerancias para cada clase
Adherencia a la estructura del fichero	El fichero es conforme a las especificaciones	Valor booleano indicando si cumple o no
Validez de los valores	Porcentaje de conformidad	Valor dentro de un rango de aceptación
	Conformidad	Valor booleano indicando si cumple o no

[3] Las normas también se refieren a esta componente como semántica (coherencia semántica).

Coherencia temática

Desde el punto de vista temático existe incoherencia cuando a un determinado grupo de atributos se les asigna valores en contradicción, por ejemplo: si el atributo *A* representa la población de un barrio y los atributos *B* y *C* el número de familias y el número medio de componentes por familia, se verifica un error lógico cuando el producto de *B* por *C* sea distinto al valor *A* (Veregin, 1994). Un caso de especial cuidado es aquel en el que se utilizan atributos derivados de las propiedades geométricas de los elementos geográficos de la base de datos, en estos casos, otro ejemplo claro de inconsistencia es cuando la superficie, perímetro, etc. almacenado en la base temática no coincide con las propiedades del objeto gráfico al que se le asocia.

Coherencia topológica

Para el caso de una topología planar la consistencia supone que se cumplen reglas como:

- Toda celda de dimensión 1 está cerrada por 2 celdas de dimensión 0.
- Toda celda de dimensión 1 tiene dos celdas de dimensión 2, una a cada lado.
- Toda celda de dimensión 2 está rodeada por un ciclo de celdas de 1 y 0 dimensiones.
- Toda celda de dimensión 0 está cerrada por un ciclo de celdas de dimensiones 1 y 2.
- No hay intersección entre celdas de dimensión 1 si no hay una de dimensión 0.

Para comprobar estas reglas existen diferentes alternativas: test de intersecciones, aplicación de la ecuación de Euler, test de cadenas de nodos (cadenas de celdas alternativas de 1 y 2 dimensiones alrededor de cada nodo), test de cadenas de polígonos (ciclos de nodos y celdas de una dimensión que forman el cerramiento de una celda de dos dimensiones) (Kainz, 1995).

En un grafo plano conexo las uniones o aristas del grafo particionan el espacio en regiones disjuntas, una de las cuales no está cerrada. Estas regiones se denominan caras. Según el teorema de Euler, si *G(V, U)* es un grafo plano y conexo, se cumple:

$$N_V - N_U + N_R = 2$$

donde:

N_V Número de vértices (V).
N_U Número de aristas (U)
N_R Número de caras o regiones (R), incluyendo la región externa circundante.

Por ejemplo, en la Figura 7.5 hay un problema de inconsistencia topológica, ya que falta un nodo en la intersección entre dos celdas de dimensión uno. Esto se detecta aplicando el teorema de Euler, donde $N_V = 6$, $N_U = 11$, $N_R = 6$, y por tanto $N_V - N_U + N_R = 1$.

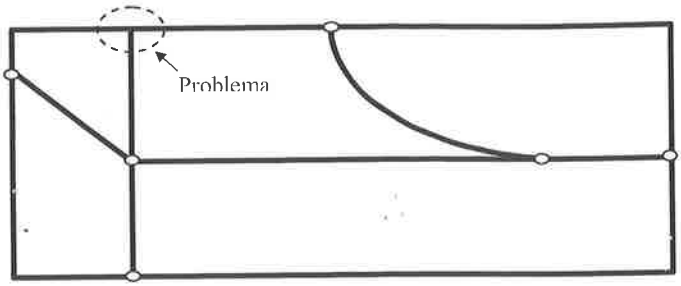


Figura 7.5 - Ejemplo de inconsistencia topológica

Coherencia temporal

Dado que el atributo temporal se puede gestionar por medio de un modelo topológico, es posible valorar su coherencia lógica de forma análoga a la topológica. Una entidad tiene su topología correcta, desde el punto de vista del atributo temporal, cuando la entidad posea una única versión precedente y una única versión futura (Langran, 1992). En otras palabras, debe subsistir contigüidad y no superposición, condiciones que se gestionan fácilmente haciendo uso de la topología. Otro aspecto de interés es la coherencia temporal entre distintos elementos de una base de datos cuando poseen distintas inercias temporales y la fase de campo o captura se ha dilatado, lo que puede generar discontinuidades y artefactos en los bordes de áreas temporalmente disjuntas.

Adherencia al formato o estructura del fichero

Las pruebas de adherencia al formato o estructura del fichero pueden ser implementadas de forma automática dado que las especificaciones estarán claramente especificadas. El proceso de comprobación puede consistir en diversas acciones, a modo de ejemplo, la Tabla 7.8 presenta algunas de las que se proponen en la Norma GDF (prENV 14825 de CEN, 1996) para el chequeo de la calidad del formato.

Otro ejemplo en esta línea es el que presentan Morant y Lerma (1999) sobre el control de la cartografía en la empresa GrafCan que se presenta en la Tabla 7.9.

Tabla 7.8 - Posibles errores en la adherencia a un formato

Grupos de errores	Posibles errores	Grupos de errores	Posibles errores
Errores de sintaxis	Uso de caracteres no especificados Longitud de registro incorrecta Longitud de campo incorrecta Violación de tipo de campo Errores de justificación	Errores de topología	Elementos lineales inconexos Ausencia de nodos
Errores en los valores	Descriptor de registro incorrecto Códigos de registro incorrectos Códigos de elementos incorrectos Códigos de atributos incorrectos Códigos de relación incorrectos Valores fuera de rango (min, max)	Errores de integridad de los valores	Relación elemento-atributo incorrecta Relación elemento-relación incorrecta Relación atributo-relación incorrecta
Errores de integridad de la base de datos	Punteros incorrectos. Valor de contador de campos incorrecto.		

Fuente: prENV 14825 (CEN, 1996)

Tabla 7.9 - Control de la cartografía en la empresa GrafCan

Coherencia de codificación con el catálogo de elementos.
Se verifica que todas las entidades gráficas se corresponden con tipos de elementos especificado en el catálogo. Se verifica el tipo de elemento gráfico (shape, texto, line-string), su codificación (nivel, color, estilo, texto), para los símbolos el nombre y para los textos su fuente, tamaño, justificación y offset.

Continuidad de elementos entre hojas.
Se comprueba la continuidad geométrica verificando que los elementos que pertenecen a varias hojas posean un punto sobre el marco de las mismas y que las coordenadas de este punto coincidan en las hojas colindantes. También se atiende a la comprobación de la continuidad estructural y de la codificación.

Curvas de nivel.
Se comprueba la continuidad analítica de los tramos tanto en planimetría como en altimetría, que tienen cota constante y que la altura se corresponde con valores adecuados al intervalo marcado, que no existe cruce con otras curvas de nivel y que los finales ocurren sobre unos elementos concretos (otro tramo de curva de nivel, borde de hoja, etc.).

Coordenada Z de elementos tridimensionales.
Se comprueba que los elementos que son tridimensionales según el catálogo posean su coordenada Z dentro del rango de la cota de las curvas de nivel entre las que se encuentra.

Cota de superficies de agua libre.
Se comprueba que este tipo de elementos (balsas, embalses, etc.) poseen su línea de máximo embalse con cota constante.

Pendiente de cauces.
Se verifica que los cauces de agua sólo tengan una dirección en sus pendientes (suben o bajan).

Asignación de texto.
Se verifica que los textos que deben tener asignado un código, según el catálogo, realmente lo tienen.

Textos y elementos geográficos.
Se comprueba la existencia de posibles incoherencias entre la presencia/ausencia de textos y el elemento gráfico que rotula. Se puede detectar si para cada texto existe un elemento gráfico con asignación del código del texto, si todos los elementos que deberían estar asignados realmente lo están, etc.

Fuente: Morant y Lerma (1999).

TRABAJO A REALIZAR

Considérese que se ha realizado una subcontratación de trabajos, y que se necesita comprobar la coherencia[4] de la base de datos que se presenta en la Tabla 7.11. Las especificaciones establecidas inicialmente son las que se recogen la Tabla 7.10.

El trabajo consistirá en detectar los errores existentes y proponer índices de error, de carácter parcial y global para esta BDG. Los primeros harán referencia toda la base de datos, los segundos a atributos concretos, y para cada uno se ha de indicar su nombre, definición métrica, interpretación, valor asumible y rango.

Tabla 7.10 - Especificaciones de producto

Objeto	Atributos	Tipo variable	Unidad de medida	Rango
Municipio	Código	Entero positivo 5 dígitos		
	Nombre	Cadena de 75 caracteres		
	Extensión superficial	Entero positivo	Km^2	(5 - 1000)
	Nº de habitantes (mujeres)	Entero positivo	Habitante	(100 - 60000)
	Nº de habitantes (varones)	Entero positivo	Habitante	(100 - 60000)
	Nº de habitantes totales	Entero positivo	Habitante	(200 - 120000)
	Densidad de población	Real positivo (1 decimal)	Hab/Km^2	(5 - 375)

Observaciones:

Código: Sólo números. Sus dos primeras cifras serán "23".

Nombre: Sólo letras, en mayúsculas, sin tildes, y espacios. No comenzará con "LA", "EL", "LAS", "LOS", que irán al final de la cadena, entre paréntesis.

Nº de hab.: La diferencia entre mujeres y varones no podrá ser superior al 15% en relación al menor de los dos.

Densidad: Habrá de coincidir con el cociente (Nº de habitantes totales / Extensión superficial)

Tabla 7.11 - Base de datos

Código	Nombre	Extensión	Mujeres	Varones	Total	Densidad
23001	ALBANCHEZ DE ÚBEDA	39	738	764	1502	38,5
23002	ALCALA LA REAL	261	10901	10620	21521	82,5
23003	ALCAUDETE	238	5766	5638	11404	47,9
23004	ALDEAQUEMADA	121	300	327	627	5,2
23005	ANDUJAR	963	19316	18938	38254	39,7
23006	ARJONA	158	2821	2874	5695	36,0
26007	ARJONILLA	43	1978	1963	3941	91,7
23008	ARQUILLOS	65	975	961	1936	29,8
23009	BAEZA	193	8077	7764	15841	82,1
23010	BAILEN	118	8568	8913	17441	148,1
23011	BAÑOS DE LA_ENCINA	392	1372	1411	2783	7,1
23012	BEAS DE SEGURA	216	3979	4156	8135	37,7

[4] Este ejemplo muestra un caso muy simple de incoherencias y de falta de sujeción a las prescripciones. Dado que las bases de datos permiten la consideración de reglas elementales de validación, por medio de máscaras de entrada, muchos de los fallos que aparecen en el ejemplo se habrían evitado.

Tabla 7.11 - Base de datos

Código	Nombre	Extensión	Mujeres	Varones	Total	Densidad
23014	BEGIJAR	43	1585	1586	3171	73,7
23015	BELMEZ DE LA MORALEDA	49	1005	1050	2055	41,9
23016	BENATAE	45	271	310	581	12,9
23017	CABRA DE SANTO CRISTO	186	1105,5	1170	2275	12,2
23018	Cambil	140	1519	1664	3183	22,7
23019	CAMPILO DE ARENAS	114	1084	1137	2221	19,5
23020	CANENA	14	1102	1037	2139	152,8
23021	CARBONEROS	58	344	337	681	11,7
23024	LA CAROLINA	201	7661	7460	15121	75,2
23025	CASTELLAR DE SANTISTEBAN	159	1822	1874	3696	23,2
23026	CASTILLO DE LOCUBIN	104	2692	2585	5277	50,7
23027	CAZALILLA	47	416	424	840	17,9
23028	CAZORLA	0,304	4452	4191	8643	28,4
23029	CHICLANA DE SEGURA	236	697	693	1390	5,9
23030	CHILLUEVAR	38	900	881	1781	146,9
23031	ESCAÑUELA	14	480	471	951	67,9
23032	ESPELUY	26	402	389	791	30,4
23033	FRAILES	41	925	969	1894	46,2
23034	FUENSANTA DE MARTOS	54	1689	1683	3372	62,4
23035	FUERTE DEL REY	35	587	569	1156	33,0
23037	GENAVE	64	339	369	708	11,1
23038	GUARDIA DE JAÉN (LA)	38	1009	1097	2106	55,4
23039	GUARROMAN	96	1429	1402	2831	29,5
23040	HIGUERA DE ARJONA	44	983	954	1937	44,0
23041	HIGUERA DE CALATRAVA	39	341	365	706	18,1
2342	HINOJARES	41	261	247	508	12,4
23043	HORNOS	118	383	361	744	6,3
23044	HUELMA	250	2965	3082	6047	24,2
23045	HUESA	137	1367	1368	2735	20,0
23046	IBROS	55	1523	1533	3056	55,6
23047	IRUELA, (LA)	124	983	1005	1988	16,0
23048	IZNATORAF	86	614	606	1220	14,2
23049	JABALQUINTO	72	1280	1272	2552	35,4
23050	JAEN	426	56520	52727	109247	256,4
23051	JAMILENA	nueve	1635	1692	3327	369,7
23052	JIMENA	48	819	766	1585	33,0
23053	JODAR	149	9097	5954	12051	80,9
23054	LARVA	42	252	266	518	12,3
23055	LINARES	198	30036	28686	58722	296,6
23056	LOPERA	67	2023	1959	3982	59,4
23057	LUPION	24	570	514	1084	45,2
23058	MANCHA REAL	1097	4424	4643	9067	8,3
23059	MARMOLEJO	178	3823	3864	7687	43,2
23060	MARTOS	260	11708	10929	22637	87,1
23061	MENGIBAR	62	4244	4291	8535	137,7
23062	MONTIZON	211	985	1032	2017	9,6
23063	NAVAS-DE-SAN-JUAN-(LAS)	175	2648	2603	5251	30,0
23064	NOALEJO	50	1145	1153	2298	46,0

Tabla 7.11 - Base de datos

Código	Nombre	Extensión	Mujeres	Varones	Total	Densidad
23065	ORCERA	126	1125	1134	2259	17,9
23066	PEAL DE BECERRO	147	2709	-2670	39	36,6
23067	PEGALAJAR	80	1515	1611	3126	39,1
23069	PORCUNA	177	3489	3480	6969	39,4
23070	POZO ALCON	139	2912	2947	5859	42,2
23071	PUENTE DE GENAVE	38	1012	1036	2048	53,9
23072	PUERTA DE SEGURA (LA)	98	1365	1665	2730	27,9
23O73	QUESADA	328	3142	3016	6158	18,8
23074	RUS	47	1890	1903	3793	80,7
23075	SABIOTE	112	2095	2104	4199	37,5
23076	SANTA ELENA	144	476	544	1020	7,1
23077	SANTIAGO DE CALATRAVA	48	462	465	927	19,3
23079	SANTISTEBAN DEL PUERTO	373	2484	2558	5042	13,5
23080	SANTO TOME	73	1180	1182	2362	32,4
23081	SEGURA DE LA SIERRA	224	1105	1106	2211	1,0
23082	SILES	175	1360	1368	2728	15,6
23084	SORIHUELA DEL GUADALIMAR	-54	604	634	1238	22,9
23085	TORREBLASCOPEDRO	60	1511	1490	3001	50,0
23086	Torre del Campo	182	6409	6457	12866	70,7
23087	TORREDONJIMENO	158	7033	6806	13839	87,6
23088	TORREPEROGIL	91	3799	3692	7491	82,3
23090	TORRES	81	894	965	1859	23,0
23091	TORRES DE ALBANCHEZ	65	513	513	1026	15,8
23092	UBEDA	402	16767	15972	32739	81,4
23093	VALDEPEÑAS DE JAEN	183	2239	2281	4520	24,7
23094	VILCHES	274	2558	2531	5089	18,6
23095	VILLACARRILLO0	239	5609	5545	11154	46,7
23096	VILLANUEVA DE LA REINA	209	1664	1658	3322	15,89
23097	VILLANUEVA DEL ARZOBISPO	178	4320	4324	8644	48,6
23098	VILLARDOMPARDO	17	639	622	1261	74,2
23099	VILLARES (LOS)	88	2430	2448	4878	55,4
023101	VILLARRODRIGO	78	290	283	573	7,3
23901	CARCHELES (LOS)	43	793	777	1570	36,5
23902	BEDMAR Y GARCIEZ	119	1610	1629	3239	27,2
2390b	VILLATORRES	73	2244	2295	4539	62,2
23904	SANTIAGO PONTONES	684	2217	2303	4520	6,6

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

7.3 CASO 31

Elaboración de metadatos

LOS METADATOS

Los metadatos son datos que, referidos a otros datos a los que acompañan, permiten conocer la bondad de éstos y entenderlos con todas sus limitaciones. Es decir, se trata de información que documenta otra información y que permite su uso adecuado. Por ejemplo, los metadatos en una cartografía son las informaciones referidas a las fuentes utilizadas, a los métodos de levantamiento y tratamiento, a las fechas de realización, a las empresas o personas que participaron, etc. Los metadatos no debe ser una meta en sí mismos, sino una ayuda para la gestión y mejor uso de la información geográfica (Ariza, 2002).

Una de las partes más importantes de los metadatos es el informe relativo a la calidad de los datos.

Algunos estudios (Miller, 1994) indican que este tipo de información trata sobre diez grandes tópicos o temas: nombre del producto, resumen del contenido, palabras clave, institución de contacto, formato del producto, forma de los datos, extensión espacial, validez de los datos, organización depositaria y forma de adquisición. En el fondo se trata de contestar a preguntas básicas como: ¿de qué información se trata?, ¿cómo se ha producido?, ¿dónde está disponible?, ¿cómo se puede adquirir?, etc.

LA NORMA MIGRA

Para aprovechar al máximo los metadatos es necesario que exista una normalización mínima en la definición de los contenidos tal que establezca un marco común y se facilite el uso.

En nuestro país la norma experimental MIGRA[5] (Mecanismo de Intercambio de Información Geográfica Relacional formado por Agregación, UNE 148001-Ex) (AENOR 1998; Rodríguez, 1998), cuyo objetivo es servir al intercambio de datos, establece el informe de metadatos que se presenta en la Tabla 7.12. Los grandes apartados del informe se refieren a la especificación del productor y distribuidor, a los datos geográ-

[5] En la actualidad es la Norma ISO 19115 la que define el estándar internacional para los metadatos, de hecho es la que se está utilizando en nuestro país en todos los proyectos de definición de infraestructuras de datos espaciales (IDEs). Sin embargo, se prefiere utilizar la Norma MIGRA como base de la práctica por su mayor simplicidad dado que el objetivo de esta práctica no es más que introductorio en este ámbito.

ficos, de forma general, y al contenido especificado en cada uno de los ficheros que forman la transmisión, llegando a incluir para cada atributo sus unidades, exactitud y fecha.

Tabla 7.12 - Componentes del informe de metainformación en la Norma UNE 148001

Versión de MIGRA Versión de MIGRA **Productor** (de los datos a intercambiar) Organismo: Acrónimo: Nombre: Dirección: Código Postal: Localidad: Provincia: País: Dirección URL: Contacto: Puesto de trabajo: Nombre: Dirección: Código Postal: Localidad: Provincia: País: Teléfono: Telefax: Correo electrónico: **Distribuidor** (si es distinto del productor) Organismo: Acrónimo: Nombre: Dirección: Código Postal: Localidad: Provincia: País: Dirección URL: Contacto: Puesto de trabajo: Nombre: Dirección: Código Postal: Localidad: Provincia: País: Teléfono: Telefax: Correo electrónico:	**Datos** Nombre del conjunto de datos (Capa, zona, escala): Sistema de Referencia: Elipsoide: Datum: Sistema de coordenadas: (UTM huso 30, Lambert,etc.) Nº de dimensiones: (2 ó 3) Unidades x,y: Unidades z: Unidades altura y anchura de textos: (metros terreno, cm papel, etc.) Escala: Zona: Esquina1: Esquina2: Esquina3 Esquina4: Modelo: (vectorial, raster, matriz) Estructura topológica: (espagueti, cadena-nodo, topología completa, parcial) Criterio de creación de tramos: Tramos sueltos: (si/no) Tipos de nodo: (extremos, conectados, intermedios, extremos intermedios) Juego de Caracteres: Tabla de códigos: **Contenido** Fecha de creación: Fecha de última actualización: Descripción textual del contenido: Directorio de ficheros. Incluyendo para cada fichero : - Número secuencial del fichero - Nombre del fichero MIGRA - Nombre físico del fichero - Número de registros - Tamaño en Bytes	**Atributos** 1) A nivel del conjunto de ficheros de atributos: - Fecha de creación - Fecha de la última actualización - Número de ficheros de atributos 2) A nivel de cada fichero de atributos: - Número secuencial del fichero - Nombre físico del fichero - Descripción - Elemento MIGRA al que corresponden los atributos del fichero (objeto compuesto, lineal, tramo, etcétera) - Código o códigos de los elementos a los que se refieren los atributos del fichero, en el caso de que no se refieran a todos los posibles códigos de un elemento - Número de atributos del fichero - Longitud del registro - Número de registros - Tamaño en Bytes 3) A nivel de cada atributo incluido en un fichero: - Nombre del atributo - Formato - Unidades - Exactitud - Fecha **Notas: (aclaraciones)**

Fuente: MIGRA (AENOR, 1998).

La información relativa a la calidad se incluye de forma complementaria a la anterior, dado que posee un carácter opcional en MIGRA. En este informe se incluyen referencias a las certificaciones del productor y a las propias cualidades de la calidad de los datos. La Tabla 7.13 presenta un ejemplo obtenido de la propia norma donde las componentes de la calidad incluidas son la posicional y temática, la coherencia geométrica, topológica y semántica y la genealogía propiamente dicha. La norma permite distinguir entre valores teóricos y determinados, así como un tratamiento particular de cada uno de los paquetes o grupos de datos que se transfieren. Al igual que en el caso anterior, se pueden incluir notas aclaratorias y definir las métricas que convengan a cada caso mediante el uso de las explicaciones pertinentes.

Tabla 7.13 - Ejemplo de informe de calidad en MIGRA (UNE 148001)

[CERTIFICACIONES_DE_CALIDAD]
NUMERO_DE_CERTIFICACIONES=2
CERTIFICACION1=ISO 9001
CERTIFICACION2=PECAL 1

[EXACTITUD_POSICIONAL_TEORICA]
ABSOLUTA_X_Y=1 metro
ABSOLUTA_Z=NA
RELATIVA_X_Y=ND
RELATIVA_Z=NA

[EXACTITUD_POSICIONAL_DETERMINADA]
ABSOLUTA_X_Y=1.2 metros
ABSOLUTA_Z=NA
A_X_Y_MUESTRA=10%
A_X_Y_NIVEL_DE_CONFIANZA=95%
A_X_Y_LINEAS_DE_DESCRIPCION_DEL_METODO=2
A_DESCRIPCION_1=Comparación con el mapa digital 1:25.000
A_DESCRIPCION_2=del IGN
A_Z_MUESTRA=NA
A_Z_NIVEL_DE_CONFIANZA=NA
A_Z_LINEAS_DE_DESCRIPCION_DEL_METODO=0
RELATIVA_X_Y=ND
RELATIVA_Z=ND
R_X_Y_MUESTRA=ND
R_X_Y_NIVEL_DE_CONFIANZA=ND
R_X_Y_LINEAS_DE_DESCRIPCION_DEL_METODO=0
R_Z_MUESTRA=ND
R_Z_NIVEL_DE_CONFIANZA=ND
R_Z_LINEAS_DE_DESCRIPCION_DEL_METODO=0

[EXACTITUD_SEMANTICA]
EN_NOMBRES=4%
NOMBRES_MUESTRA=10%
NOMBRES_NIVEL_DE_CONFIANZA=95%
NOMBRES_DETERMINACION=Comprobación en campo sobre 1.000 nombres
EN_CODIGOS=1%
CODIGOS_MUESTRA=6%
CODIGOS_NIVEL_DE_CONFIANZA=95%
CODIGOS_DETERMINACION=Comparación con datos

[COHERENCIA_SEMANTICA]
ESPECIFICACIONES_SEMANTICAS=2
ESP1=Los objetos de código 05* están comunicados por algún objeto 06*
CUMPLIMIENTO1=100%
ESP2=Los puertos están en la línea de costa
CUMPLIMIENTO2=100%

[GENEALOGIA]
NUMERO_DE_GRUPOS_DE_DATOS=3

[GRUPO_DE_DATOS_1]
NOMBRE=Datos Geograficos
DESCRIPCION=Contenido de los Mapas provinciales 1:200000
PRODUCTOR=IGN
FECHA=1986-1992
FIN= Cartografía Básica
FUENTE= Mapas provinciales 1:200000 del IGN de los años 85-89
NUMERO_DE_PROCESOS=6
PROCESO_1=Digitalización manual mapas provinciales 1:200.000
PROCESO_2=Filtrado 24 metros (algoritmo Douglas-Peucker)
PROCESO_3=Tratamiento geométrico automático+interactivo
PROCESO_4=Tratamiento semántico interactivo
PROCESO_5=Tratamiento topológico automático
PROCESO_6=Actualización automática+interactiva

[GRUPO_DE_DATOS_2]
NOMBRE= Líneas Límite
LINEAS_DE_DESCRIPCION=3
DESCRIPCION_1=Lineas limite que describen la división administrativa española a cuatro DESCRIPCION_2=niveles: Municipio, Provincia, Autonomía y Nación. Comprende todo el DESCRIPCION_3=tema 1
PRODUCTOR=IGN
FECHA=1986
FIN= Cartografía Básica Digital
LINEAS_DE_FUENTE=2
FUENTE_1=Minutas originales 1:25.000 del IGN, del año 1986, actualizadas posteriormente FUENTE_2=con

Tabla 7.13 - Ejemplo de informe de calidad en MIGRA (UNE 148001)

de la DG del Catastro 1:1000

[COHERENCIA_GEOMETRICA]
ESPECIFICACIONES_GEOMETRICAS=2
ESP1=No hay vértices repetidos
CUMPLIMIENTO1=100%

ESP2=No hay trios de vértices consecutivos que subtiendan una flecha < 50 metros
CUMPLIMIENTO2=90%

[COHERENCIA_TOPOLOGICA]
ESPECIFICACIONES_TOPOLOGICAS=3
ESP1=Los tramos pertenecientes a un mismo objeto lineal están ordenados y conectados.
#Esta especificación es especialmente útil, ya que permite comprobar la continuidad de los #objetos lineales.
CUMPLIMIENTO1=100%
ESP2=No hay perimetros abiertos
CUMPLIMIENTO2=100%
ESP3=Un tramo empieza en un nodo y acaba en otro
CUMPLIMIENTO3=100%

tamientos de campo.
NUMERO_DE_PROCESOS=6
PROCESO_1=Digitalización manual
PROCESO_2=Asignación de nombres y Codigos
PROCESO_3=Tratamiento topológico automático
PROCESO_4=Integración con fuente de datos 1
PROCESO_5=Actualización automática+interactiva año 1994

PROCESO_6= Actualización automática+ interactiva 1995

[GRUPO_DE_DATOS_3]
NOMBRE= Datos temáticos de población
LINEAS_DE_DESCRIPCION=2
DESCRIPCION_1=Censo de población del Instituto Nacional de Estadística desglosado por DESCRIPCION_2=entidades de población.
PRODUCTOR=INE
FECHA=1981
FIN= 2Ejemplo de SIG de población
LINEAS_DE_FUENTE=2
FUENTE_1= Censo del INE del año 1986
NUMERO_DE_PROCESOS=2
PROCESO_1=Enganche de los datos sobre entidades de población del 1:200.000
PROCESO_2=Revisión y corrección de errores.

TRABAJO A REALIZAR

Se pide hacer un informe, siguiendo la especificación MIGRA, de la Carta Digital del Mundo (*Digital Chart of the World*). Para ello, primeramente, habrá que recopilar suficiente información acerca de esta base de datos. Dada la abundante información existente en Internet sobre la misma, se propone la utilización de este medio y de buscadores.

Material en línea

En el espacio web

http://www.ujaen.es/dep/ingcar/publicaciones/pqcarto

también puede encontrar el material necesario para este caso.

ANEXO

ESTADÍSTICA

1. INTRODUCCIÓN

La Estadística[1] siempre ha sido base para el desarrollo de la cartografía, tanto de mapas topográficos como temáticos. Sin embargo, su uso ha estado muy limitado desde el punto de vista del control de calidad de los productos y procesos cartográficos. La Estadística ofrece una gran cantidad de herramientas (por ejemplo: análisis de la varianza, análisis cluster, análisis de componentes principales, series temporales, etc.), cuyo uso está extendido en otros muchos ámbitos del control de la producción pero que no han sido aplicadas profusamente en el cartográfico.

Para dar una base adecuada a las prácticas que se plantean en este manual, se han incluido varios bloques:

— Estadística Descriptiva: Su objetivo es resumir un conjunto de datos en uno o varios valores que los sinteticen.

— Probabilidad: Permite modelizar las variables utilizadas.

— Inferencia Estadística: Pretende generalizar o extender los resultados obtenidos en una muestra a toda la población.

— Estadística circular: Su objetivo es modelizar datos cuya distribución es del tipo circular.

[1] Se puede decir que la Estadística es la ciencia que estudia las leyes de comportamiento de los fenómenos aleatorios para, en una segunda fase, hacer predicciones sobre resultados. Dentro de ella, la Estadística Descriptiva, estudia y describe los fenómenos; mientras que la Inferencia Estadística se encarga de las predicciones.

— Análisis Cluster: Su objetivo es segregar una población en varios grupos de elementos tal que posean cierto comportamiento estadístico común.

— Simulación: La base de la simulación es la generación de valores aleatorios y la replicación de procesos mediante programas de ordenador, tal que, de una forma cómoda y controlada, se pueda estudiar el comportamiento de modelos y sistemas reales.

— Muestreo: La teoría de muestras es una parte de la Estadística de gran importancia en el control del proceso y base de la correcta inferencia. Los aspectos básicos a determinar en un muestreo serán el propio método, el tamaño de la muestra y los parámetros a estimar.

— Análisis univariante de Series Temporales: Permitirá construir un modelo para explicar la estructura y evolución de una variable observada a lo largo del tiempo.

Cada uno de los bloques anteriores ocupa un apartado en el que se han incluido los conceptos más importantes para su aplicación práctica. Se quiere destacar aquí que, por tanto, no se pretende la profundidad de un manual de Estadística, sino ofrecer al estudiante y lector de esta obra un complemento que le permita refrescar muchos de los conocimientos previos que ha de poseer, y acercarse a los nuevos de una forma aplicada.

Existen numerosos tratados de Estadística en los que se profundiza, con gran detalle, en cada uno de los bloques que aquí se presentan. A este respecto, en cada apartado se referenciarán las fuentes que se recomiendan.

Finalmente, recordar que esta obra posee un glosario donde se incluyen diversas definiciones de carácter estadístico.

2. ESTADÍSTICA DESCRIPTIVA

Como se ha comentado supra, el objetivo de la Estadística Descriptiva es resumir o generalizar grandes cantidades de datos en uno o varios valores, de forma que, aunque de manera sintética, se ofrezca la máxima información posible de todo el conjunto. Si bien se trata de la parte más elemental de la Estadística es de gran utilidad práctica.

En este apartado se van a tratar las distribuciones de frecuencias unidimensionales y bidimensionales y tras ellas, las medidas de posición y dispersión más generalizadas, así como la manera de tipificar una variable. Debido al carácter básico de estos conocimientos, existe abundante bibliografía al respecto, tanto general como específica, entre ella cabe destacar Hermoso y Hernández (1994) por su claridad

desde el punto de vista teórico y Álvarez (1994) por estar enfocado desde un punto de vista práctico.

2.1. Distribución de frecuencias

La Estadística analiza los datos obtenidos y saca conclusiones a partir de éstos, teniendo en cuenta las variaciones de los datos. Una distribución de frecuencias consiste en un ordenamiento de los datos según su magnitud, de manera que cada modalidad que toma la variable estudiada viene representada por su frecuencia absoluta: el número de veces que se ha obtenido en la muestra estudiada el valor o conjunto de valores representado por la modalidad.

2.1.1. DISTRIBUCIÓN DE FRECUENCIAS UNIDIMENSIONAL

Si se dispone de una muestra procedente de una variable unidimensional A, la distribución de sus frecuencias consiste en un ordenamiento de los datos según su magnitud, de manera que cada modalidad que toma la variable estudiada viene representada por su frecuencia absoluta (n_i), es decir, el número de veces que se ha obtenido en la muestra estudiada el valor o conjunto de valores representado por la modalidad. También se puede considerar su frecuencia absoluta acumulada (N_i), es decir, el número de valores muestrales con modalidad menor o igual que la de la modalidad estudiada.

Estas dos frecuencias se pueden considerar también como frecuencias relativas (f_i y F_i, respectivamente), simplemente dividiendo la frecuencia entre el tamaño muestral, por lo que se tiene la frecuencia expresada en tanto por uno.

2.1.2. DISTRIBUCIÓN DE FRECUENCIAS BIDIMENSIONAL

En este caso se quieren estudiar dos caracteres A y B, de manera simultánea, dentro de una población, para lo que se considera una muestra con n individuos. A la variable que representa dicho estudio se le llama variable bidimensional (A,B), es decir, cada individuo viene determinado por el par $\{(a_i,b_j);\ i = 1,...,k\ ;\ j = 1,...,p\}$,

donde:

- a_i Modalidad de la variable A que tiene el individuo.
- b_j Modalidad de la variable B que tiene el individuo.
- k Número de modalidades de la variable A.
- p Número de modalidades de la variable B.

Se definen:

— La frecuencia absoluta n_{ij} del par (a_i,b_j), que indica el número de individuos que presentan el valor a_i de la variable estadística unidimensional A y b_j de la variable B.

— La frecuencia absoluta n_{i+} de los individuos con valor a_i de la variable A:

$$n_{i+} = \sum_{j=1}^{p} n_{ij}$$

— La frecuencia absoluta n_{+j} de los individuos con valor b_j de la variable B:

$$n_{+j} = \sum_{i=1}^{k} n_{ij}$$

Se verifica que:

$$n = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} = \sum_{i=1}^{k} n_{i+} = \sum_{j=1}^{p} n_{+j}$$

— De manera análoga al caso unidimensional se pueden definir las frecuencias relativas y las frecuencias acumuladas.

Se denomina distribución bidimensional de frecuencias al conjunto de valores $\{(a_i,b_j),n_{ij}\}$ con $i = 1,...,k\ ;\ j = 1,...,\ p$.

La representación numérica se realiza de la siguiente forma: Dada una población estudiada según 2 caracteres A y B, con distribución bidimensional de frecuencias $\{(a_i,b_j),n_{ij}\}$; entonces una forma de representar los resultados es la tabla de doble entrada o tabla de contingencia (Tabla A.1).

Tabla A.1.- Tabla de contingencia o distribución bidimensional

A/B	b_1	b_2	...	b_j	...	b_p	n_{i+}
a_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	n_{1+}
a_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	n_{2+}
...	...	...	...	...	...	...	...
a_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	n_{i+}
...	...	...	...	...	...	...	...
a_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{kp}	n_{k+}
N_{+j}	n_{+1}	n_{+2}	...	n_{+j}	...	n_{+p}	n

A partir de una distribución bidimensional pueden obtenerse distribuciones unidimensionales:

a) Si de la variable bidimensional (A,B) se estudia sólo una de las dos variables sobre el grupo de individuos que forman la muestra, los valores de dicha variable junto con sus frecuencias definen una distribución unidimensional, denominada distribución marginal de A o de B (según la variable que se estudie). Las distribuciones marginales de frecuencias se presentan en la Tabla A.2.

Tabla A.2 - Distribuciones marginales

A	n_{i+}
a_1	n_{1+}
...	...
a_i	n_{i+}
...	...
a_k	n_{k+}
	n

B	n_{+j}
b_1	n_{+1}
...	...
b_j	n_{+j}
...	...
b_p	n_{+p}
	n

b) Las distribuciones condicionadas expresan cómo se distribuyen, según una de las dos variables unidimensionales, el conjunto de individuos que cumplen una condición, que viene expresada por un valor o conjunto de valores que presenta la otra variable. Las distribuciones condicionadas se presentan en la Tabla A.3.

Tabla A.3 - Distribuciones condicionadas

$A \mid B = b_j$

A	n_{ij}
a_1	n_{1j}
...	...
a_i	n_{ij}
...	...
a_k	n_{kj}
	n_{+j}

$B \mid A = a_i$

B	n_{ij}
b_1	n_{i1}
...	...
b_j	n_{ij}
...	...
b_p	n_{ip}
	n_{i+}

2.2. Resumen numérico de los datos muestrales

Para resumir un conjunto de datos muestrales las medidas más utilizadas de la Estadística Descriptiva son las medidas de posición y las de dispersión. Las primeras pretenden informar sobre la situación de los valores de la muestra y las segundas sobre la variación, rango o dispersión de los elementos de la misma con relación a una medida de tendencia central. A continuación se presentan las medidas mas usuales para ambos casos.

2.2.1. MEDIDAS DE POSICIÓN

Hay dos tipos básicos:

— Medidas de tendencia central o promedios, que dan un valor central representativo.

— Cuantiles, que suministran valores representativos de particiones de la distribución.

Si $a_1,a_2,...,a_n$ son las observaciones de una muestra procedente de una variable A, las medidas se definen de la siguiente forma:

— Media muestral: Valor promedio de la variable estudiada según:

$$\overline{a} = \frac{1}{n}\sum_{i=1}^{n} a_i$$

La media muestral representa el centro de masas o centro de gravedad de los datos muestrales y es la medida de tendencia central más importante.

Una propiedad bastante útil es que la media muestral de una combinación lineal de variables es la combinación lineal de las medias, es decir:

$$\overline{(aA_1 + bA_2)} = a\overline{A}_1 + b\overline{A}_2$$

— Mediana: Valor muestral que deja a ambos lados el 50% de los datos muestrales, es decir, es el valor de la variable cuya frecuencia relativa acumulada es la primera que iguala o sobrepasa el valor 0,5. Por tanto, se determina ordenando los valores y estudiando la distribución de frecuencias.

La mediana se utiliza para estudiar cuál es el error situado en el centro, es decir, el error tal que por debajo de él están la mitad de los errores medidos y tal que por encima de él está la otra mitad de los errores medidos.

— Moda: Es el valor más frecuente de la variable estudiada, es decir, el valor más repetido en la muestra.

Para el caso en el que se tienen los datos de manera aislada y no existen valores repetidos, se toma un número de $\sqrt{n}$ intervalos de la misma amplitud (b) que incluya a todos los datos muestrales obtenidos. Entonces, el intervalo en el que está la moda (e_i, e_{i+1}) será el que tenga una mayor frecuencia absoluta (n_{Mo}) y la moda será el valor:

$$Mo = e_i + \frac{(n_{Mo} - n_{Mo-1})b}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})}$$

La moda se utiliza para estudiar cuál es el error más repetido en la muestra.

— Cuantil de orden α, con $\alpha \in (0,1)$: Es el valor de la variable (a_α) que deja por debajo de él a una proporción igual a α de los datos muestrales, es decir la frecuencia relativa acumulada en este valor es la primera que iguala o sobrepasa el valor α. Según lo dicho, la mediana es el cuantil de orden $\alpha = 0,5$.

Se utiliza, por ejemplo, para calcular el intervalo de valores en el que se encuentran la mitad de los errores centrales, que viene dado por los cuantiles de orden 0,25 y 0,75. Cuando la distribución se divide en cuatro partes, cada una se corresponde con un 25% de las frecuencias, en este caso los cuantiles se denominan cuartiles.

2.2.2. MEDIDAS DE DISPERSIÓN

Dado que la medida de tendencia central más importante es la media muestral, aquí se presentan los índices de dispersión que se basan sobre la misma y que, por otra parte, son los más usados; entre ellos:

— La varianza muestral: Determina la dispersión o variabilidad de los datos muestrales respecto de la media muestral. Es el valor medio de los cuadrados de las diferencias de todos los elementos de la muestra respecto a su propia media:

$$S_A^2 = \frac{\sum_{i=1}^{n}(a_i - \bar{a})^2}{n-1} = \frac{\sum_{i=1}^{n} a_i^2 - \frac{\left(\sum_{i=1}^{n} a_i\right)^2}{n}}{n-1}$$

Si no hay variabilidad en la muestra, entonces todas las observaciones muestrales son iguales y la varianza vale cero. Generalmente, cuanto más grande sea la varianza muestral, tanto mayor será la dispersión de los datos respecto de la media.

La unidad de medida de la varianza muestral es el cuadrado de la unidad general de los datos. Muchas veces esto es un inconveniente y resulta difícil de interpretar, por lo que se suele preferir utilizar como medida de dispersión la

raíz cuadrada positiva de la varianza muestral, llamada desviación típica muestral, es decir:

$$S_A = +\sqrt{S_A{}^2}$$

La ventaja principal de S_A es que se expresa en la unidad original de las mediciones. Un punto importante es que la desviación típica no refleja la magnitud de los datos muestrales, sino solamente su dispersión alrededor de la media.

Con la desviación típica se puede estudiar la dispersión de los errores respecto de su media, puesto que se sabe que, aproximadamente, el 95% de los datos muestrales se encuentran situados dentro del intervalo ($\overline{a} \pm 2S_A$).

Una propiedad importante de la varianza muestral si las variables A y B son independientes, es decir, el valor que toma una variable no influye en el valor que toma la otra variable, es la siguiente:

$$Var(aA \pm bB \pm c) = a^2 Var(A) + b^2 Var(B)$$

Es decir, para una combinación lineal de variables independientes, la varianza resultante es la suma de los cuadrados de los coeficientes de la combinación lineal por la varianza correspondiente a cada una de las variables que intervienen.

— Covarianza: Se define cuando se trabaja con dos variables. Por tanto, para una muestra $(a_1,b_1),...,(a_m,b_n)$ procedente de una variable bidimensional (A,B), la covarianza de A y B es una medida de la variabilidad conjunta de dichas variables. Su formulación es pues muy similar a la de la varianza, tal como se puede observar:

$$S_{AB} = \frac{1}{n}\sum_{i=1}^{n}(a_i - \overline{a})(b_i - \overline{b}) = \frac{1}{n}\sum_{i=1}^{n} a_i b_i - \overline{a}\cdot\overline{b}$$

Si las dos variables son independientes, entonces la covarianza entre ellas vale cero, por lo que si las dos variables de la propiedad de la varianza comentada anteriormente no son independientes, se tiene que:

$$Var(aA \pm bB \pm c) = Var(aA \pm bB) = a^2 Var(A) + b^2 Var(B) \pm 2abS_{AB}$$

— Coeficiente de correlación: Mide la relación lineal que existe entre las dos características A y B de una variable bidimensional (A,B). Se cuantifica de la siguiente manera:

$$r_{AB} = \frac{S_{AB}}{S_A S_B}, \text{ para } r_{AB} \in [-1, 1]$$

Cuando $r_{AB} < 0$ indica que las variables varían en sentido opuesto (si una crece, la otra disminuye). Cuanto más próximo a la unidad esté el valor absoluto del coeficiente de correlación, mayor será la relación lineal existente entre las variables. En la misma línea, un valor próximo a cero indica poca relación lineal entre ambas.

Cuando en lugar de una distribución bidimensional se dispone de una distribución k-dimensional, es decir, con k variables $(A_1,...,A_k)$, los conceptos y medidas anteriormente descritas pueden extenderse a este espacio. De esta forma, mediante la aplicación de las fórmulas expuestas, se puede definir:

— La matriz de varianzas/covarianzas de las variables: Matriz simétrica en la que se representan las varianzas (en la diagonal) y covarianzas (fuera de la diagonal).

$$\text{Matriz de Variazas-Covarianzas} = \begin{pmatrix} S_1^2 & S_{12} & \cdots & S_{1k} \\ S_{12} & S_2^2 & \cdots & S_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{1k} & S_{2k} & \cdots & S_k^2 \end{pmatrix}$$

— La matriz de correlación lineal: La diagonal presenta valores unidad y las demás posiciones valores los coeficientes de correlación lineal de las parejas bajo consideración.

$$\text{Matriz de Correlación} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{12} & 1 & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ r_{1k} & r_{2k} & \cdots & 1 \end{pmatrix}$$

2.3. Tipificación de una variable

La tipificación de una variable cualquiera A consiste en considerar una nueva variable Z, a partir de la anterior, con media igual a cero y desviación típica igual a la unidad. La tipificación consiste pues en una traslación (hacer la media cero) y un escalado (hacer la desviación igual a la unidad), de los valores de una distribución dada, tal como se muestra en la Figura A.1, en la que se traslada una distribución Normal N(5;2,5) hasta una N (0;2,5) y a continuación se escala para lograr una N(0;1).

El interés de esta operación es muy grande dado que permite el uso de tablas y la comparación entre distribuciones de distintos parámetros. Esta propiedad es válida para cualquier variable utilizada.

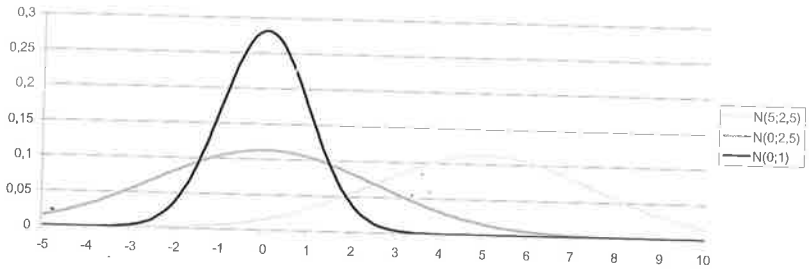


Figura A.1 – Tipicación de una distribución N(5;2,5)

La forma de materializar la tipificación de cualquier valor A es la siguiente:

$$Z = \frac{A - \overline{a}}{S_A}$$

2.4. Regresión y Correlación simple

Uno de los objetivos de toda ciencia es investigar relaciones entre los distintos hechos que estudia e intentar traducirla en estructuras manejables, mediante el uso de las Matemáticas. Esto es lo que se va a realizar a lo largo de este apartado.

Hay variables entre las que se puede definir una función matemática que las relaciona perfectamente, como por ejemplo el espacio y el tiempo a una velocidad constante ($e = v \cdot t$); por otro lado, existen variables que están relacionadas, pero no de forma exacta, ya que no se puede definir una función matemática que las relacione exactamente, como son el peso y la altura, oferta y demanda, etc. Se va a estudiar el grado de dependencia y la relación entre este segundo tipo de variables:

a) La relación entre dos variables viene dada por la regresión.

b) El grado de dependencia entre las variables viene dado por la correlación.

c) Finalmente se van a dar predicciones de una variable.

La variable que se quiere predecir se denomina variable dependiente o explicada, mientras que la variable o variables que se utilizan para estudiar la anterior se llaman variables independientes o explicativas.

2.4.1. REGRESIÓN

La regresión consiste en ajustar mediante una función una serie de datos observados para poder predecir el valor de la variable explicada conocido un valor de la explicativa. Se llama regresión de B sobre A (B|A) a la función que explica la variable B a partir de A, siendo A la variable explicativa y B la variable explicada; mientras que la regresión de A sobre B (A|B) informa del comportamiento de la variable A a partir de B, siendo, por tanto, B la variable explicativa y A la explicada.

Sea (A,B) una variable estadística bidimensional y se considera su distribución de frecuencias asociada. Para elegir la función matemática que se ajuste lo mejor posible a la distribución de frecuencias se utiliza el criterio de mínimos cuadrados.

Para hallar la regresión de B sobre A, se considera $b' = h(a)$ la mejor predicción de B en función de A si minimiza las diferencias al cuadrado entre los valores reales y los estimados por dicha función, es decir:

$$\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\left(b_j - h(a_i)\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}{e_{ij}}^2 \ ,$$

donde:

$(b_j - h(a_i)) = e_{ij}$ Es el residuo o error para el par (a_i,b_j).

A continuación se estudian distintas funciones de regresión:

Regresión lineal (recta de regresión):

Se pretende ajustar una recta a la distribución de frecuencias, por lo que la regresión se va a denominar lineal. Se debe escoger de entre todas las rectas posibles, la mejor, para lo cual se utiliza el criterio de mínimos cuadrados como ya se ha indicado anteriormente. Se va a trabajar con la regresión de B sobre A; la regresión de A sobre B será totalmente análoga.

La recta de mínimos cuadrados será la recta con expresión analítica:

$$b' = h(a) = la + m$$

que minimice la suma de los cuadrados de los errores:

$$S(l,m) = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\left(b_j - (la_i + m)\right)^2$$

Se van a buscar los valores de l y m que minimicen dicha expresión. Para ello se deriva la expresión respecto de los dos parámetros, se igualan dichas derivadas a cero y se obtienen los valores de los parámetros:

$$\begin{cases} l = \dfrac{S_{AB}}{S_A^{\,2}} \\ m = \overline{b} - \dfrac{S_{AB}}{S_A^{\,2}}\overline{a} \end{cases}$$

, que son los valores que minimizan la expresión $S(l,m)$.

Por consiguiente, la recta de regresión de B sobre A es:

$$b' = h(a) = \frac{S_{AB}}{S_A^{\,2}}a + \overline{b} - \frac{S_{AB}}{S_A^{\,2}}\overline{a}$$

que se puede escribir de forma más cómoda como:

$$b' - \overline{b} = \frac{S_{AB}}{S_A^{\,2}}(a - \overline{a})$$

La recta de regresión de A sobre B no se obtiene despejando a de la expresión anterior, sino que hay que seguir el mismo razonamiento, ya que el único punto común de ambas rectas es el par $(\overline{a},\overline{b})$, obteniéndose la siguiente recta de regresión:

$$a' - \overline{a} = \frac{S_{AB}}{S_B^{\,2}}(b - \overline{b})$$

Los coeficientes de regresión:

$$\begin{cases} B \mid A: & \dfrac{S_{AB}}{S_A^{\,2}} \\ X \mid Y: & \dfrac{S_{AB}}{S_B^{\,2}} \end{cases}$$

miden el grado de aumento de una variable al aumentar la otra. Ambos coeficientes tienen el mismo signo, que es el de la covarianza:

— Si la covarianza es positiva, ambos son positivos y las dos variables varían en el mismo sentido.

- Si la covarianza es negativa, ambos son negativos y las variables varían en sentido opuesto.
- Si la covarianza es nula, las rectas de regresión son paralelas a los ejes y perpendiculares entre sí.

Regresión no lineal

Aunque la regresión lineal tiene gran cantidad de aplicaciones, hay veces en las que la función puede no ser una recta. Se va a estudiar la función de regresión de B sobre A, ya que el caso de A sobre B es totalmente análogo. Se consideran las siguientes funciones de regresión no lineales:

Regresión Hiperbólica:

La función de regresión viene dada por:

$$b' = h(a) = l/a + m$$

Para su resolución se considera el cambio de variable $C = 1/A$, por lo que se ha linealizado dicha función mediante la recta:

$$b' = lc + m$$

Por tanto, la regresión hiperbólica se reduce al caso lineal, pero con las variables B y C, por lo que:

$$l = \frac{S_{BC}}{S_C^{\,2}} \; ; \; m = \overline{b} - l\overline{c}$$

y la función de regresión viene dada por:

$$b' - \overline{b} = \frac{S_{BC}}{S_C^{\,2}}\left(\frac{1}{a} - \overline{c}\right)$$

Regresión Potencial:

La función de regresión viene dada por:

$$b' = h(a) = m \cdot a^{l}$$

Para linealizarla, se toman logaritmos en ambos miembros, con lo que se obtiene:

$$ln(b)' = ln(m \cdot a^l) = ln(m) + ln(a^l) = ln(m) + l \cdot ln(a)$$

y con los cambios de variable $\tilde{B} = ln(B)$, $\tilde{A} = ln(A)$, $\tilde{m} = ln(m)$, se tiene la recta $\tilde{b}' = l \cdot \tilde{a} + \tilde{m}$, por lo que se aplica la regresión lineal a las variables $\tilde{A}$ y $\tilde{B}$, obteniendo la recta:

$$\tilde{b}' - \overline{\tilde{b}} = \frac{S_{\tilde{A}\tilde{B}}}{S_{\tilde{A}}^2}(\tilde{a} - \overline{\tilde{a}})$$

que despejando $\tilde{b}'$ da los valores l y $\tilde{m}$, por lo que m va a venir dado por $m = e^{\tilde{m}}$.

Regresión Exponencial:

La función de regresión viene dada por:

$$b' = h(a) = m \cdot l^a$$

Para linealizarla, se toman logaritmos en ambos miembros, obteniéndose:

$$ln(b)' = ln(m \cdot l^a) = ln(m) + ln(l^a) = ln(m) + a \cdot ln(l)$$

y con los cambios de variables $\tilde{B} = ln(B)$, $\tilde{m} = ln(m)$, $\tilde{l} = ln(l)$, se tiene la recta $\tilde{b}' = \tilde{l}\, a + \tilde{m}$, por lo que se aplica la regresión lineal a las variables A y $\tilde{B}$, obteniendo:

$$\tilde{b}' - \overline{\tilde{b}} = \frac{S_{A\tilde{B}}}{S_A^2}(a - \overline{a})$$

donde, despejando $\tilde{b}'$ se obtienen los valores de $\tilde{l}$ y $\tilde{m}$, por lo que l y m van a venir dados por $l = e^{\tilde{l}}$ y $m = e^{\tilde{m}}$.

Regresión Parabólica y Polinomial:

La función de regresión es:

$$b' = h(a) = la^2 + ma + q$$

que no se puede reducir al caso lineal como en los casos anteriores, por lo que hay que tomar mínimos cuadrados directamente.

Al igual que en el caso de la recta, se minimiza la expresión:

$$S(l,m,q)=\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}e_{ij}^{\ 2}=\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}(b_j-la_i^{\ 2}-ma_i-q)^2$$

derivando respecto de los coeficientes l, m y q:

$$\begin{cases}\dfrac{dS}{dl}=0\\ \dfrac{dS}{dm}=0\\ \dfrac{dS}{dq}=0\end{cases}\Rightarrow\begin{cases}\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}a_i^{\ 2}e_{ij}=0\\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}a_ie_{ij}=0\\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}e_{ij}=0\end{cases}\Rightarrow$$

$$\Rightarrow\begin{cases}\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}a_i^{\ 2}b_j=q\sum_{i=1}^{k}n_{i\cdot}a_i^{\ 2}+m\sum_{i=1}^{k}n_{i\cdot}a_i^{\ 3}+l\sum_{i=1}^{k}n_{i\cdot}a_i^{\ 4}\\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}a_ib_j=q\sum_{i=1}^{k}n_{i\cdot}a_i+m\sum_{i=1}^{k}n_{i\cdot}a_i^{\ 2}+l\sum_{i=1}^{k}n_{i\cdot}a_i^{\ 3}\\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p}n_{ij}b_j=qn+m\sum_{i=1}^{k}n_{i\cdot}a_i+l\sum_{i=1}^{k}n_{i\cdot}a_i^{\ 2}\end{cases}$$

Los valores desconocidos se obtienen resolviendo el sistema de tres ecuaciones con tres incógnitas.

En general, si se quiere ajustar un polinomio de grado n superior a 2:

$$y'=l_0+l_1a+l_2a^2+\ldots+l_na^n$$

el método de mínimos cuadrados lleva a un sistema análogo al anterior de $n+1$ ecuaciones con $n+1$ incógnitas.

2.4.2. CORRELACIÓN

La correlación se encarga de estudiar el grado de asociación de las variables, por lo que indica en qué medida la función obtenida mediante la regresión explica la variable dependiente en función de la independiente. Por esta razón, se habla de correlación o bondad de ajuste de una curva, por lo que no se puede entender la correlación sin la regresión. De nuevo, se va a estudiar la correlación de Y sobre X, ya que igual que en el apartado anterior, la de X sobre Y es totalmente análoga.

El método de mínimos cuadrados que se ha utilizado para la regresión mide el error cometido por la función de regresión que se ha ajustado mediante la suma de los cuadrados de los residuos:

$$\sum_{i=1}^{k}\sum_{j=1}^{p}(b_j - h(a_i))^2 n_{ij}$$

cantidad que, considerada con las frecuencias relativas de la distribución, se usa para medir la bondad del ajuste y se conoce con el nombre de Varianza Residual[2]:

$$S_{rB}{}^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}(b_j - h(a_i))^2 f_{ij} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p}(b_j - h(a_i))^2 n_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{p} e_{ij}{}^2 f_{ij}$$

ésta indica lo que queda sin explicar de la variable *B* después de realizar la regresión; por tanto, cuanto menor sea la varianza residual, mejor será el ajuste.

Para ver a partir de qué valores es bueno o malo el ajuste se define el coeficiente de determinación, que se basa en la descomposición de la varianza marginal de la variable dependiente:

$$S_B{}^2 = S_{eB}{}^2 + S_{rB}{}^2$$

donde:

$S_{eB}{}^2 = \sum_i \sum_j (h(a_i) - \bar{b})^2 f_{ij}$ Varianza explicada por la regresión.

Pero dicha descomposición solo es válida para funciones lineales en los parámetros y no lo es para las demás curvas, por lo que el coeficiente de determinación solo va a tener sentido para las funciones de la primera clase.

Se define el Coeficiente de Determinación como la proporción de la varianza total de *Y* que aparece explicada por la regresión:

$$R^2 = \frac{S_{eB}{}^2}{S_B{}^2} = 1 - \frac{S_{rB}{}^2}{S_B{}^2}$$

[2] En funciones lineales en los parámetros, como la recta, hipérbola equilátera y los polinomios, la media de los residuos es cero, por lo que se obtiene la varianza de los residuos; no obstante, en funciones no lineales en los parámetros, como la potencial o la exponencial, la media de los residuos no es cero, aunque se mantiene el mismo nombre.

que es un valor en el intervalo [*0, 1*] que mide la fiabilidad del ajuste ($1-R^2$ mide la proporción de varianza no explicada).

Por tanto, despejando de la expresión anterior, la varianza residual vale:

$$S_{rB}{}^2 = S_B{}^2(1-R^2)$$

La interpretación de los valores del coeficiente de determinación es:

— R^2 vale cero si y solo si $S_{rB}{}^2 = S_B{}^2$, si y solo si la varianza explicada vale cero, es decir, la regresión no explica nada de B a partir de A, por lo que el ajuste es el peor posible.

— R^2 vale uno si y solo si $S_{eB}{}^2 = S_B{}^2$, si y solo si la varianza residual vale cero, es decir, todos los residuos son nulos, por lo que el ajuste es perfecto.

— Si $0<R^2<1$, el ajuste será mejor, cuanto más próximo sea el coeficiente a uno.

El ajuste realizado es aceptable para valores del coeficiente de determinación mayores de 0,49, es decir, para fiabilidad superior al 49%.

La correlación en cada uno de los casos estudiados viene dada por:

Regresión lineal

En el caso de la recta de regresión, tras desarrollar el valor de la varianza residual, el coeficiente de determinación es:

$$R^2 = \frac{S_{AB}{}^2}{S_B{}^2 S_A{}^2}$$

Si se calcula el coeficiente de determinación de la recta de regresión de *A* sobre *B*, como caso excepcional, es exactamente el mismo que el de *B* sobre *A*.

Se define el Coeficiente de correlación lineal como:

$$r = \frac{S_{AB}}{S_A S_B}$$

es decir, $\sqrt{R^2}$ con el signo de la covarianza, y mide la correlación lineal entre las variables. Toma valores en [-1, 1]. El significado de los valores que puede tomar es:

— Si $r = -1$, existe correlación lineal perfecta negativa, es decir, la dependencia está totalmente explicada por la recta de regresión, pero las variables varían en sentido opuesto.

— Si $-1<r<0$, la correlación será más intensa cuanto más próximo esté el índice a -1 y la correlación será negativa.

— Si $r = 0$, las variables están incorreladas, es decir, no existe ninguna relación lineal entre dichas variables.

— Si $0<r<1$, la correlación es positiva y más intensa cuanto más próximo esté a uno.

— Si $r = 1$, existe correlación lineal perfecta positiva, es decir, la dependencia está totalmente explicada por las rectas de regresión y las variables varían en el mismo sentido.

Por tanto, r indica además el signo de la recta.

Se considera que un ajuste lineal es aceptable si $|r|$ es mayor o igual que 0,7, dependiendo además del tamaño n de las observaciones tomadas.

Además, R^2 se puede calcular como el producto de los coeficientes de regresión:

$$R^2 = \frac{S_{AB}^{\ 2}}{S_A^{\ 2} S_B^{\ 2}} = \frac{S_{AB}}{S_A^{\ 2}} \frac{S_{AB}}{S_B^{\ 2}}$$

Regresión Hiperbólica

Como la hipérbola es lineal en los parámetros, se puede definir el coeficiente de determinación de la hipérbola, que, tras su desarrollo, es igual a:

$$R_{B/A}^{\ 2} = \frac{S_{BC}^{\ 2}}{S_C^{\ 2} S_B^{\ 2}}$$

y toma valores en [0, 1].

Regresión Parabólica

Como es lineal en los parámetros tiene sentido calcular su coeficiente de determinación y se hace a partir del cálculo de la varianza residual.

Regresión Potencial y Exponencial

En estos casos se han obtenido los parámetros de la regresión tras realizar un ajuste mediante mínimos cuadrados de una recta de regresión después de realizar ciertas transformaciones.

Si se calcula el coeficiente de correlación de dicha recta, se obtiene la parte de regresión explicada por el ajuste entre las variables transformadas ($\tilde{A}$ y $\tilde{B}$), pero no las originales (A y B), por lo que con esto no se puede decir si el ajuste de las variables es bueno o no.

Además, estas funciones no son lineales en los parámetros, por lo que no tiene sentido calcular el coeficiente de determinación. Por ello, lo único que se puede hacer es un estudio comparativo sobre la bondad del ajuste mediante la varianza residual en cada uno de los casos.

Por consiguiente, dados varios ajustes por regresión de mínimos cuadrados de dos variables dadas (B/A), el mejor ajuste será aquel cuya varianza residual sea menor; pero si no se compara ni la regresión exponencial ni la potencial, se puede comparar con el coeficiente de determinación.

2.4.3. PREDICCIÓN

La predicción consiste en tratar de determinar los valores de la variable dependiente o explicada en función de valores conocidos de la variable explicativa o independiente.

Teniendo la ecuación de regresión, se sustituye el valor de la variable independiente en la ecuación, obteniéndose el valor correspondiente de la variable dependiente. La predicción será tanto más fiable cuanto mejor sea el ajuste.

2.4.4. MODELO LINEAL GENERAL DE REGRESIÓN

No siempre los valores de la variable dependiente están relacionados con los de una única variable. Se va a ver cómo se obtienen funciones multilineales de ajuste, es decir:

$$B' = l + m_1 A_1 + ... + m_k A_k + \varepsilon,$$

donde ε es el error que se comete en la estimación.

En forma matricial:

$$\hat{B} = AM + \varepsilon$$

donde:

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix},\ X = \begin{pmatrix} 1 & a_{11} & a_{12} & \dots & a_{1k} \\ 1 & a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix},\ M = \begin{pmatrix} l \\ m_1 \\ \vdots \\ m_k \end{pmatrix},\ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

A partir del método de mínimos cuadrados, los coeficientes l y m_i se obtienen del siguiente modo:

— Los coeficientes m_i se obtienen resolviendo el siguiente sistema de ecuaciones:

$$\begin{cases} m_1 \cdot n \cdot S_{a1}^2 & + m_2 \cdot n \cdot S_{A1,A2} & + \dots & + m_k \cdot n \cdot S_{A1,Ak} & = n \cdot S_{A1,B} \\ m_1 \cdot n \cdot S_{A1,A2} & + m_2 \cdot n \cdot S_{A2}^2 & + \dots & + m_k \cdot n \cdot S_{A2,Ak} & = n \cdot S_{A2,B} \\ \dots & + \dots & + \dots & + \dots & = \dots \\ m_1 \cdot n \cdot S_{A1,Ak} & + m_2 \cdot n \cdot S_{A2,Ak} & + \dots & + m_k \cdot n \cdot S_{Ak}^2 & = n \cdot S_{Ak,B} \end{cases},$$

donde S_i^2 y $S_{i,j}$ representan las varianzas y covarianzas muestrales de las diferentes variables.

— El coeficiente l se calcula a partir de la expresión:

$$l = \bar{b} - m_1 \bar{a}_1 - \dots - m_k \bar{a}_k$$

Los coeficientes de regresión m_i se expresan en las mismas unidades que sus variables asociadas (A_i). Para comparar dichos coeficientes, es necesario que estén referidos a la misma escala de medida:

$$\mu_i = m_i \sqrt{\frac{S_{Ai}^2}{S_B^2}}$$

En forma matricial, el vector de coeficientes se obtiene como $M = (A'A)^{-1}A'B$.

Igual que en el ajuste lineal, se tiene que analizar la bondad del ajuste de las observaciones experimentales al modelo construido, es decir, el grado de relación existente entre la variable dependiente B y las variables independientes A_i. Para ello, existen los siguientes coeficientes:

— Coeficiente general de determinación múltiple: La Varianza total de la variable B es la suma de la varianza de explicada por la regresión ($S_{B'}^2$) y la varianza residual o de los errores (S_e^2). Entonces, el coeficiente de determinación mide la proporción de varianza explicada por el modelo de ajuste:

$$R^2_{B,A1,...,Ak} = \frac{S^2_{B'}}{S^2_B} = 1 - \frac{S^2_e}{S^2_B} = \frac{m_1 \cdot S_{B,A1} + ... + m_k \cdot S_{B,Ak}}{S^2_B}$$

— Coeficiente general de correlación múltiple: Es la raíz cuadrada del coeficiente de determinación múltiple:

$$r_{B,A1,...,Ak} = \sqrt{R_{B,A1,...,Ak}}$$

— Coeficiente general de determinación múltiple corregido: El estimador insesgado del coeficiente de correlación múltiple poblacional se obtiene del muestral como:

$$\hat{R}^2_{B,A1,...,Ak} = 1 - \left(\frac{S^2_e}{S^2_B} \cdot \frac{n-1}{n-k-1} \right)$$

— Matriz de intercorrelaciones: Si se toman por parejas las k variables independientes (A_i), se puede obtener el coeficiente de correlación que mide la relación lineal de cada pareja de variables:

$$r_{Ai,Aj} = \frac{S_{Ai,Aj}}{S_{Ai} \cdot S_{Aj}}$$

que es el coeficiente de correlación lineal entre A_i y A_j.

Como claramente $r_{Ai,Aj} = r_{Aj,Ai}$ y $r_{Ai,Ai} = 1$, se pueden escribir los coeficientes de correlación lineal de cada par de variables en una matriz:

$$\begin{pmatrix} r_{A1,A1} & r_{A1,A2} & \cdots & r_{A1,Ak} \\ r_{A2,A1} & r_{A2,A2} & \cdots & r_{A2,Ak} \\ \cdots & \cdots & \cdots & \cdots \\ r_{Ak,A1} & r_{Ak,A2} & \cdots & r_{Ak,Ak} \end{pmatrix} = \begin{pmatrix} 1 & r_{A1,A2} & \cdots & r_{A1,Ak} \\ & 1 & \cdots & r_{A2,Ak} \\ & & \cdots & \cdots \\ & & & 1 \end{pmatrix}$$

— Matriz de cofactores: Un cofactor o menor complementario de un elemento de una matriz es el determinante que se obtiene al eliminar de la matriz la fila y columna en la que está el elemento.

— Coeficientes de correlación parcial: Cuando hay más de dos variables independientes, estos coeficientes miden la relación existente entre dos de ellas, eliminando la influencia de las demás y se obtiene a partir de la matriz de cofactores.

El coeficiente de correlación parcial entre las variables A_i y A_j, eliminando la influencia de las demás es:

$$r_{Ai,Aj;resto} = \frac{(-1)^{i+j} \cdot R_{Ai,Aj}}{\sqrt{R_{Ai,Ai} \cdot R_{Aj,Aj}}}$$

— Coeficientes de correlación múltiple: Mide la relación que existe entre una variable independiente y las restantes, cuando tenemos más de dos variables independientes, y vale:

$$r_{Ai;resto} = \sqrt{1 - \frac{|R|}{R_{Ai,Ai}}}$$

No siempre la relación funcional que existe entre dos variables es de tipo lineal. También puede ser de tipo polinómico:

$$B' = 1 + m_{1A} + m_{2A}{}^2 + ... + m_k A^k$$

En este caso se hace el siguiente cambio de variable: $A_1 = A$, $A_2 = A^2$, ... , $A_k = A^k$, con lo que realmente se tiene un ajuste multilineal entre la variable B y las variables A_i, que ya ha sido estudiado.

En el caso de que exista una variable producto de otras:

$$B' = 1 + m_1 A_1 + m_2 A_2 + m_3 A_1 A_2$$

Se hace también el cambio de variable $A_3 = A_1 A_2$ y se tiene de nuevo un ajuste multilineal.

2.4.4. DIAGNÓSTICO DE INFLUENCIA

Cuando se realiza una regresión puede ocurrir que un dato concreto tenga una influencia desproporcionada en los coeficientes obtenidos para el modelo. En un caso extremo los valores estimados pueden depender más de un subconjunto de datos con gran influencia que de la mayoría. Por tanto, es conveniente localizar estos puntos con gran

influencia sobre los valores estimados, errores y predicciones, y eliminar su impacto del modelo. Recordemos que el modelo lineal general de regresión se expresa como: $\hat{B} = AM + \varepsilon$. A partir de ahora el subíndice (i) va a indicar que para el cálculo de la medida indicada no se ha tenido en cuenta la muestra i-ésima. Algunos índices que pueden marcar la presencia de estos valores son:

— La matriz $H = A(A'A)^{-1}A'$ juega un papel importante en la detección de estos puntos, en particular sus elementos diagonales $h_i = a_i'(A'A)a_i$, donde a_i' es la fila i de la matriz A. La diagonal de esta matriz es una medida estandarizada de la distancia de la observación i-ésima del centro del espacio de A. Así, valores grandes de la diagonal indican observaciones que son potencialmente influyentes. La media de los elementos diagonales es p/n, donde p es el número de parámetros del modelo (incluido el término independiente) y n el número de observaciones usadas para ajustar el modelo, y si un valor diagonal excede dos veces dicha media debe ser investigada su posible influencia.

— Otra medida que también va a indicar los valores muestrales a investigar es el Residuo Studentizado (R-Student) definido como:

$$R\text{-}Student = \frac{b_i - \hat{b}_i}{S_{rB\,(i)}\sqrt{1-h_i}}$$

donde $\hat{b}_i$ es el valor estimado para a_i y $b - \hat{b}_i$ es el residuo de la observación i-ésima. Las observaciones con R-Student mayor que dos deben ser estudiadas.

— El estadístico COVRATIO representa el grado al que una observación tiene impacto sobre los errores estándar de los coeficientes de regresión, mide el cambio en el determinante de la matriz de covarianzas de los estimadores eliminando la observación i-ésima y está definido por:

$$COVRATIO = \frac{\det\left(S^2_{rB\,(i)}(A_{(i)}'A_{(i)})^{-1}\right)}{\det\left(S^2_{rB}(A'A)^{-1}\right)}$$

tal que las observaciones con $|COVRATIO - 1| \geq \frac{3p}{n}$ deben ser investigadas.

— La medida más representativa de la influencia global es la distancia de Cook, que mide la influencia de la observación i-ésima si se elimina de la muestra. Está definida por:

$$D_i = \frac{\sum_j (\hat{b}_j - \hat{b}_{j\,(i)})^2}{p \cdot S^2_{rA}}$$

Habitualmente de utiliza la distribución F de Snedecor con p y n-p grados de libertad (que se estudia en el apartado 3.3.7.) para ver la influencia que tiene el dato i-ésimo en el modelo. Si $D_i = F_{p,n-p,0.5}$ entonces el modelo de mínimos cuadrados obtenido es bastante sensible al dato i-ésimo. Como $F_{p,n-p,0.5}$ es próximo a uno, se suelen considerar como puntos con influencia aquellos que tienen $D_i \geq 1$, aunque McDonald (2002) considera el siguiente criterio para considerar un punto como influyente si $n > 15$:

$$D_i > 0{,}7 \text{ para } p = 2$$
$$D_i > 0{,}8 \text{ para } p = 3$$
$$D_i > 0{,}85 \text{ para } p > 3$$

— El impacto de la eliminación de una única observación sobre cada coeficiente de la regresión se observa mediante el DFBETAS, definido como:

$$DFBETAS_{j\,(i)} = \frac{m_j - m_{j\,(i)}}{\sqrt{S^2_{rA\,(i)}(A'A)_{jj}}}$$

donde $A'A_{jj}$ es el elemento (j,j) de la matriz $(A'A)^{-1}$.

En general, valores grandes de DFBETAS indican que las observaciones tienen influencia en la estimación del parámetro considerado, de manera que se utiliza el umbral $\pm 2/\sqrt{n}$.

— Para medir el cambio en el valor predicho en la observación i-ésima tras eliminar dicha observación del modelo se utiliza el DFFITS, definido como:

$$DFFITS_i = \frac{\hat{b}_i - \hat{b}_{i\,(i)}}{\sqrt{S^2_{rA\,(i)} h_i}}$$

En general, valores grandes de DFFITS indican que la observación tiene influencia, y se considera como umbral el valor de:

$$2\sqrt{p/n}$$

3. PROBABILIDAD

El Cálculo de Probabilidades se ocupa del estudio de los fenómenos aleatorios, es decir, de aquellos cuya ocurrencia está sujeta al azar. El objetivo es la modelización matemática de estos fenómenos, es decir, crear modelos matemáticos del comportamiento de los fenómenos aleatorios más usuales y de interés. El conocimiento de este comportamiento permitirá, entre otras, la estimación, la predicción e inclusión de los comportamientos aleatorios en procesos más complejos como el de simulación, etc. De nuevo, debido al carácter básico de estos conocimientos, existe abundante bibliografía al respecto, tanto general como específica, y entre ella cabe destacar Gutiérrez y otros (1993a) y Hermoso y Hernández (1994) por su claridad desde el punto de vista teórico y Álvarez (1994) por estar enfocado desde un punto de vista práctico.

Si se considera la realización de un experimento aleatorio en el que:

— Cada uno de los posibles resultados de dicho experimento recibe el nombre de suceso elemental.

— El conjunto formado por todos los sucesos elementales se denomina espacio muestral del experimento aleatorio (Ω).

— Un suceso de un espacio muestral es cualquier subconjunto del espacio muestral.

Se puede proceder a establecer una definición formal de Probabilidad[3] de la siguiente forma: Dado el espacio muestral Ω, y un suceso cualquiera E, se define la probabilidad sobre el espacio muestral Ω como una aplicación $P: \Omega \rightarrow \Re$ en la que se verifica:

1º: $P(E) \geq 0$

[3] A lo largo de la historia han existido distintos conceptos de Probabilidad hasta llegar a la definición formal. Bajo la "concepción clásica", se considera un experimento en el que existe un número finito de posibles resultados n que son igualmente factibles y excluyentes (no se puede obtener más de un resultado de forma simultánea), y sea A un resultado posible del experimento, entonces, la probabilidad del resultado A, que se denota como P(A), es, según la regla de Laplace, el cociente entre el número de casos favorables y el número de casos posibles. Bajo la "concepción frecuentista" se considera que, en muchas ocasiones, los posibles resultados de un experimento no son igualmente factibles y que, si se repite el experimento muchas veces bajo las mismas condiciones, la probabilidad de un resultado se aproxima por su frecuencia y, a medida que aumenta el número de veces que realizamos el experimento (N), esta frecuencia relativa se va a aproximar más a su verdadero valor. Por tanto se define la probabilidad de un resultado A como el límite cuando $N \rightarrow \infty$ de la frecuencia relativa de dicho resultado. Existe también una "concepción subjetiva" bajo la cual la probabilidad se puede interpretar como el grado de creencia o convicción respecto a la ocurrencia de un resultado.

2°: $P(\Omega) = 1$

3°: Si E_1, E_2, ... son infinitos sucesos con intersección vacía entre ellos ($E_i \cap E_j = \varnothing \ \forall i \neq j$), entonces se verifica que la probabilidad de la unión de los sucesos es la suma de las probabilidades de los distintos sucesos:

$$P(E_1 \cup E_2 \cup E_3 \cup ...) = P(E_1) + P(E_2) + P(E_3) + ...$$

Por tanto, se puede decir que la probabilidad de un suceso es un número real que mide la posibilidad colectiva de ocurrencia de los resultados de ese suceso. Las probabilidades de los distintos sucesos se van a asignar utilizando cualquiera de los métodos comentados en la nota a pie de página.

Como consecuencia de la definición, se tienen las siguientes propiedades:

— La probabilidad de cualquier suceso E está entre 0 y 1, ($P(\varnothing) = 0 \leq P(E) \leq 1 = P(\Omega)$).

— P es no decreciente, es decir, si $E_1 \subset E_2$ entonces $P(E_1) \leq P(E_2)$.

— Dado E un suceso cualquiera, entonces la probabilidad de su suceso contrario $\overline{E}$ vale $P(\overline{E}) = 1 - P(E)$.

3.1. Variables aleatorias

Se pretende estudiar el comportamiento de fenómenos aleatorios. Para modelizar estos fenómenos y poderlos tratar adecuadamente se introduce el concepto de variable aleatoria, que es una función que hace corresponder un número a cada suceso elemental del fenómeno. Por tanto, formalmente, se puede definir el concepto de variable aleatoria de la siguiente forma:

— Sea Ω un espacio muestral sobre el que existe una función de probabilidad P, y A una función de valor real sobre Ω que transforma los resultados del espacio muestral (sucesos) en valores reales. Entonces se dice que la función $A: \Omega \rightarrow \Re$ es una variable aleatoria. Además, cada variable aleatoria lleva asociada una función de probabilidad.

Una forma de caracterización de la variable aleatoria es mediante su función de distribución, definida para cualquier valor real como la probabilidad de que la variable aleatoria tome valores menores o iguales que el considerado, calculando probabilidades de que la variable aleatoria tome valores en determinados subconjuntos:

$$F(a) = P(A \leq a)$$

Las propiedades que caracterizan a toda función de distribución son:

— $F(-\infty) = 0$ y $F(\infty) = 1$

— F es una función no decreciente.

— F es una función continua a la derecha.

Además, toda función de distribución verifica las siguientes propiedades, muy útiles para calcular probabilidades:

— $P[a_1 < A \leq a_2] = F(a_2) - F(a_1)$

— $P[A=a] = F(a) - F(a-0)$

donde:

$F(a-0)$ Límite por la izquierda de $F(a)$ en el punto a.

— $P[a_1 < A < a_2] = F(a_2-0) - F(a_1)$

— $P[a_1 \leq A < a_2] = F(a_2-0) - F(a_1-0)$

— $P[a_1 \leq A \leq a_2] = F(a_2) - F(a_1-0)$

3.1.1. VARIABLES DISCRETAS

Se dice que una variable aleatoria A es discreta cuando sólo puede tomar un conjunto numerable de valores aislados.

Como ya se ha indicado anteriormente, toda variable aleatoria lleva asociada una función de probabilidad que asigna una probabilidad a cada valor de la variable. Para el caso de la variable aleatoria discreta A, se dice que $P(a) = P(A=a)$ es función de probabilidad de A si verifica:

$$P(a) \geq 0 \text{ para todo valor } a$$

$$\sum_{i=1}^{T} P(a_i) = 1$$

También se puede caracterizar la variable aleatoria discreta mediante su función de distribución:

$$F(a) = P(A \leq a) = \sum_{a_i \leq a} P(a_i)$$

que es una distribución escalonada y creciente.

3.1.2. VARIABLES CONTINUAS

Una variable aleatoria A es continua si puede tomar cualquier valor dentro de un intervalo. En este caso se define la función de densidad de una variable aleatoria como la función $f(a)$ tal que:

$$P[a \leq A \leq b] = \int_a^b f(x)dx$$

y además se verifica que:

$f(a) \geq 0$; para cualquier valor real de a

$$\int_{-\infty}^{\infty} f(a)da = 1$$

La función de distribución asociada a la variable aleatoria continua viene dada por:

$$F(a) = P[A \leq a] = \int_{-\infty}^{a} f(a)da$$

que es el área que deja la función de densidad $f(a)$ entre $-\infty$ y a. En este caso, la función de distribución es continua y creciente. Además, se verifica que $F'(a) = f(a)$.

La probabilidad de que la variable aleatoria continua tome un valor puntual vale cero, ya que:

$$P(A = a) = \int_a^a f(a)da = 0$$

3.2. Características de una distribución de probabilidad

Igual que se hacía en Estadística Descriptiva, se pueden definir una serie de características, medidas de posición y dispersión, que dan información resumida de la población, que en este caso es una distribución. Estas características son equivalentes a las presentadas en el apartado 2.2, por lo que tienen la misma interpretación y significado.

3.2.1. ESPERANZA MATEMÁTICA

También denominado momento de primer orden, la Esperanza matemática de una variable aleatoria A es el valor promedio de la variable aleatoria. En el caso de variables discretas, si existe, vale:

$$E[A]=\sum_{i=1}^{T} a_i p_i$$

y en el caso de variables continuas, si existe:

$$E[A]=\int_{-\infty}^{\infty} af(a)da$$

Como se puede observar, la Esperanza matemática no es una función, sino que es un escalar.

El concepto de Esperanza matemática también puede ser definido sobre una función $h(A)$ de una variable aleatoria. En este caso se define la Esperanza matemática de $h(A)$ como el valor promedio de dicha función. Para el caso discreto:

$$E[h(A)]=\sum_{i=1}^{T} h(a_i)p_i$$

y en el caso continuo:

$$E[h(A)]=\int_{-\infty}^{\infty} h(a)f(a)da$$

3.2.2. VARIANZA

Basada en la anterior medida, y por ello denominado momento de segundo orden, se define la varianza de la variable aleatoria A como:

$$Var(A) = \sigma_A^2 = E[A\text{-}E[A]]^2 = E[A]^2 - (E[A])^2$$

que como se puede observar de su definición, mide la dispersión de los valores que toma la variable aleatoria respecto de su valor esperado (momento de primer orden). La desviación típica de una variable aleatoria σ_A será la raíz cuadrada positiva de la varianza.

3.2.3. MEDIANA

Es el valor de la variable aleatoria que deja a ambos lados la misma probabilidad, es decir, el valor que verifica $F(a) = 1/2$. Igual que se define la mediana, se puede definir el cuantil de orden α como aquel valor de la variable aleatoria con función de distribución igual a α.

3.2.4. MODA

Es el valor más frecuente de la variable aleatoria, es decir, en el caso discreto, es el valor o valores con mayor probabilidad y en el caso continuo es el máximo de la función de densidad.

3.3. Distribuciones de probabilidad más importantes

Una gran cantidad de fenómenos naturales y de características de los objetos que se producen en un entorno industrial tienen comportamientos, marcado por sus distribuciones de probabilidad de ocurrencia, que se ajustan a unas funciones bien conocidas y caracterizadas, que se denominan "tipo" o "estándar". De entre estas funciones destaca la denominada "Normal", llamada así justamente por mostrar una forma de comportamiento muy extendida. En este apartado, se van a estudiar las principales distribuciones discretas (Uniforme, Binomial, Poisson e Hipergeométrica) y continuas unidimensionales (Uniforme y Normal) y las dos distribuciones multidimensionales más utilizadas, la Multinomial y la Normal multivariante.

3.3.1. DISTRIBUCIÓN UNIFORME DISCRETA SOBRE *N* PUNTOS

Sea A una variable aleatoria discreta que sólo toma valores en N puntos $(a_1,...,a_N)$ y, además, que todos ellos tengan la misma probabilidad. En estas condiciones, se dice que la variable aleatoria discreta A sigue una distribución Uniforme discreta sobre N puntos. Su valor medio es:

$$\frac{1}{N}\sum_{i=1}^{N} a_i$$

y su varianza:

$$\frac{1}{N}\sum_{i=1}^{N} a_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} a_i\right)^2$$

La función de distribución se obtiene, al igual que en todas las distribuciones de variables discretas, como:

$$F(a) = P(A \le a) = \sum_{a_i \le a} P(a_i)$$

que es muy sencilla de utilizar, por lo que no está tabulada.

Se utiliza en casos en los que el experimento tiene varios posibles resultados, todos ellos con la misma probabilidad de ocurrencia, como por ejemplo el resultado del lanzamiento de una moneda o un dado equilibrado.

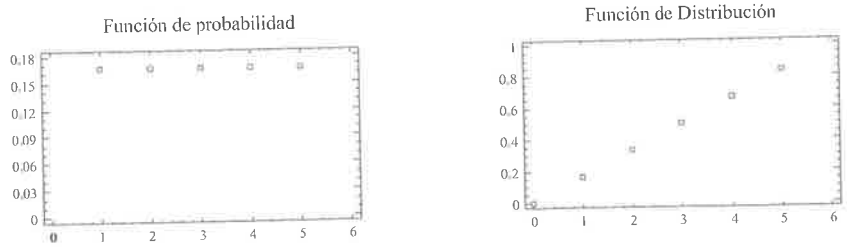


Figura A.2 – Funciones de probabilidad y distribución de las caras de un dado.

3.3.2. DISTRIBUCIÓN BINOMIAL

En este caso se considera un experimento en el que su resultado es la ocurrencia, o no, de un suceso B; tomándose como éxito la ocurrencia de este suceso.

Si p es la probabilidad de éxito ($p = P(B)$), y q la probabilidad de fracaso ($q = P(\overline{B})$), se verifica que la suma de ambas ha de ser la unidad: $p + q = 1$.

Bajo estos supuestos, si se realiza n veces el experimento, todas ellas de forma independiente, p será la probabilidad de éxito en cada una de las repeticiones de forma aislada, se puede definir una variable aleatoria A que mida el número de éxitos en las n pruebas.

Entonces la variable aleatoria A sigue una distribución Binomial de parámetros n y p, es decir, los parámetros son el número de veces que se realiza el experimento (n) y la probabilidad de éxito (p), por tanto se puede escribir: $A \sim B(n, p)$, que se lee como que A se distribuye como una Binomial de parámetros n (= *número de realizaciones*), y p (= *probabilidad de éxito*). La distribución $B(n = 1, p)$ recibe el nombre de distribución de Bernouilli (p). La Figura A.3 presenta gráficamente la función de probabilidad y de distribución del caso de una $B(n = 8, p = 0,4)$.

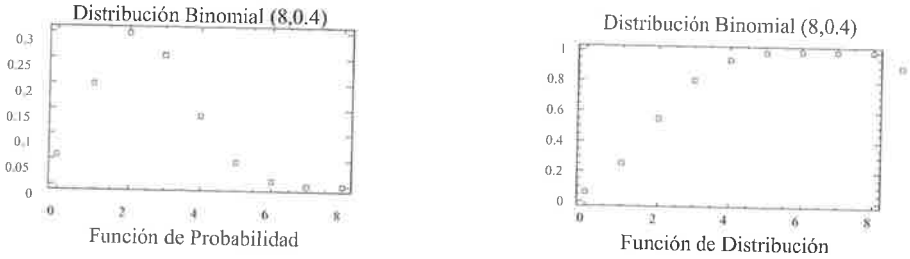


Figura A.3 – Funciones de probabilidad y de distribución de una Binomial (8,0.4)

Las principales características de esta distribución son:

$$E[A] = np \qquad Var(A) = n\,p\,(1 - p) = n\,p\,q$$

Dado que se trata de un comportamiento común, esta función es de gran aplicación y se encuentran tabuladas tanto su función de probabilidad ($p_i = P\,[A=i]$, $i = 0, 1, \ldots, n$) como su función de distribución ($F(a)$) (en la Tabla 1 del Apéndice A se presenta la de su función de distribución).

Una posibilidad cuando se dispone de una variable aleatoria discreta $A \sim B(n, p)$ que estudia el número de éxitos obtenidos en n realizaciones de un experimento es considerar la variable aleatoria que mide el número de fracasos en las n realizaciones del experimento. Dicha variable será igual al número de veces que se realiza el experimento menos el número de éxitos ($n - A$) y seguirá también una distribución Binomial, pero con parámetros (n, q).

Un resultado importante para la distribución Binomial es el Teorema de Bernouilli, según el cual la probabilidad de que la variable aleatoria $A \sim B(n, p)$

dividida por n (que es la frecuencia relativa de la ocurrencia del suceso éxito) difiera en valor absoluto de la probabilidad del suceso éxito (p) en más de una cantidad ε es menor o igual que:

$$\frac{pq}{n\varepsilon^2} \quad (*)$$

es decir:

$$P\left[\left|\frac{A}{n} - p\right| \geq \varepsilon\right] \leq \frac{pq}{n\varepsilon^2}$$

Además, si no se conoce p, será necesario calcular un extremo superior de (*) y, por lo tanto, se puede expresar el teorema como:

$$P\left[\left|\frac{A}{n} - p\right| \geq \varepsilon\right] \leq \frac{1}{4n\varepsilon^2}$$

ya que $p \cdot q$ está acotado por el valor máximo, que lo toma cuando $p = q = 1/2$, es decir, $p \cdot q = 1/4$.

3.3.3. DISTRIBUCIÓN DE POISSON

Se considera una variable aleatoria A que representa el número de sucesos aleatorios independientes que ocurren a velocidad constante sobre el tiempo o el espacio. El parámetro definitorio de esta distribución es λ, que se corresponde con el número medio de ocurrencias por unidad de tiempo o espacio.

En estas condiciones, la variable aleatoria A, que puede tomar valores enteros positivos ($a = 0, 1, 2, \ldots$), se dice que sigue una distribución de Poisson ($X \sim P(\lambda)$), y su función de probabilidad es:

$$P[A = a] = \frac{e^{-\lambda}\lambda^a}{a!}$$

Esta distribución (ver ejemplo en la Figura A.4) es de gran utilidad cuando se estudian sucesos poco frecuentes. Ejemplos pueden ser el número de accidentes en un fin de semana, el número de llamadas telefónicas en un cierto intervalo de tiempo, de ahí que también se le llama "ley de los sucesos raros". Sus funciones de probabilidad y de distribución también están tabuladas (en la Tabla 2 del Apéndice A se presenta su función de distribución) y sus principales características son:

$$E[A] = \lambda \qquad Var\,(A) = \lambda$$

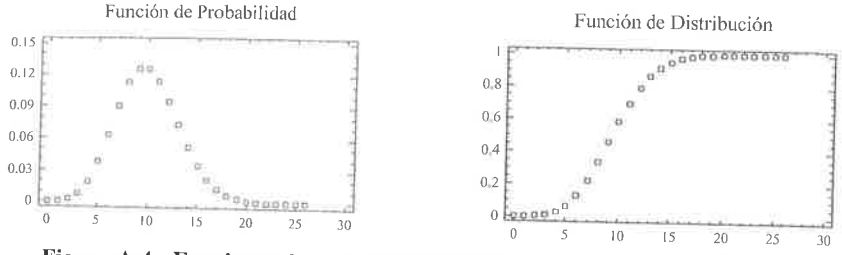


Figura A.4 - Funciones de probabilidad y distribución de Poisson ($\lambda = 10$)

3.3.4. DISTRIBUCIÓN HIPERGEOMÉTRICA

El experimento o situación que permite entender el comportamiento de la distribución Hipergeométrica es el siguiente: dada una población con N elementos, en la que hay k elementos cuya extracción significa la ocurrencia del suceso éxito, se toma una muestra de n elementos tal que se extraen de uno tras otro, es decir, el individuo extraído se deja fuera de la población, con lo cual las extracciones no son independientes y, por lo tanto, no es el caso de la distribución Binomial[4], se define la variable aleatoria A que estudia el número de éxitos en las n extracciones (ver ejemplo en la Figura A.5).

Entonces A ($a = 0, 1, \ldots Min\ (n, k)$) sigue una distribución Hipergeométrica de parámetros N, n y k, que se nota como $A \sim H(N, n, k)$. Su función de probabilidad viene dada por:

$$P[A = a] = \frac{\binom{k}{a}\binom{N-k}{n-a}}{\binom{N}{n}} \qquad a = 0,1,\ldots Min(n,k)$$

Las principales características de esta distribución son:

$$E[A] = \frac{nk}{N} \qquad Var\,[A] = \frac{nk(N-n)(N-k)}{N^2(N-1)}$$

La distribución Hipergeométrica tiene tabuladas tanto la función de probabilidad ($p_i = P[A=i]$, $i= 0, 1, \ldots, Min\ (n, k)$) como la función de distribución ($F(a)$) (ambas se presentan en la Tabla 3 del Apéndice A).

[4] Se tiene muestreo sin reemplazamiento, mientras que en la Binomial se tiene muestreo con reemplazamiento.

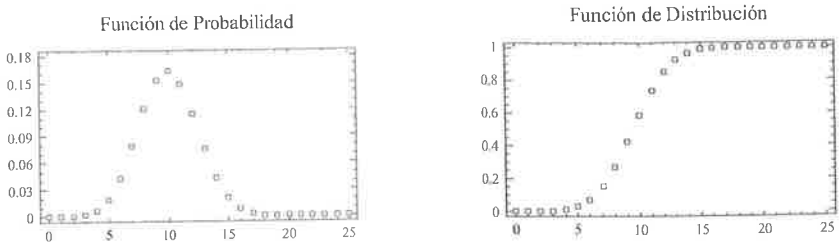


Figura A.5 - Función de probabilidad y distribución de Hipergeométrica ($N = 1000$, $n = 25$, $k = 400$)

Esta distribución se utiliza bastante en Control de Calidad y en la aceptación de muestreos.

3.3.5. LA DISTRIBUCIÓN UNIFORME

Sea A una variable aleatoria continua que toma valores en un intervalo finito, de forma que éstos se encuentran igualmente distribuidos sobre el intervalo, es decir, la probabilidad de que la variable aleatoria tome un valor en cada subintervalo de igual longitud es la misma (Figura A.6). Entonces se dice que A está uniformemente distribuida en su intervalo de valores (b, c) y se dice que $A \sim U(b, c)$. Su función de densidad está dada por la expresión:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & resto \end{cases}$$

Su función de distribución viene dada por:

$$F(a) = \begin{cases} 0 & a < b \\ \dfrac{a-b}{c-b} & b < a < c \\ 1 & a > c \end{cases}$$

Como es una distribución muy sencilla, no es necesaria la tabulación de su función de distribución. Sus principales características son:

$$E[A] = \frac{b+c}{2} \qquad Var[A] = \frac{(c-b)^2}{12}$$

Esta distribución se utiliza principalmente para dar una representación adecuada en el redondeo de las diferencias que surgen al medir cantidades físicas entre los valores observados y los reales.

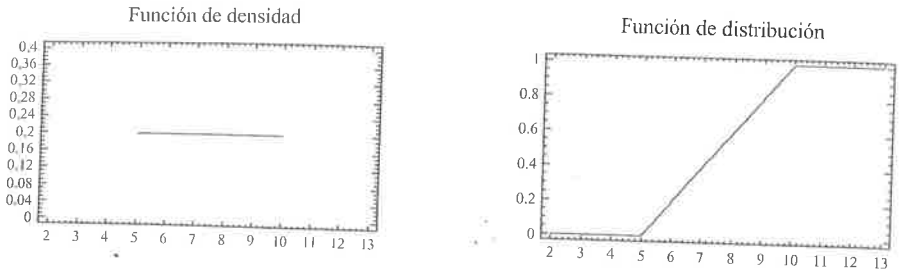


Figura A.6 - Función de densidad y de distribución de una distribución *U(5,10)*

3.3.6. LA DISTRIBUCIÓN NORMAL

Como se deduce de su nombre y de lo avanzado con anterioridad, la distribución Normal es la distribución más importante y más utilizada de las distribuciones de probabilidad, ya que una gran cantidad de fenómenos se modelizan adecuadamente con esta distribución: alturas, pesos, distribución de errores, calificación de exámenes, y un largo etcétera.

Se estudia en primer lugar la distribución Normal tipificada cuyos parámetros son media cero y desviación típica unidad ($N(0, 1)$), para tratar, a continuación, el caso general, ya que éste último siempre se va a poder reducir al primero mediante el proceso de tipificación ya comentado en el apartado 2.3.

Una variable aleatoria continua Z se dice que sigue una distribución $N(0,1)$ si, para todo valor $z \in \Re$, su función de densidad viene dada por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

cuyo comportamiento puede observarse en la Figura A.7.a, que adopta la forma de campana que posee un máximo en la posición media y que se extiende asintóticamente desde $-\infty$ a $+\infty$. Esta curva es simétrica respecto al eje de ordenadas, $f(z) = f(-z)$, y para $z > 0$ la curva es decreciente, y para $z < 0$ es creciente. Junto a lo anterior, analíticamente se demuestra que en $z = \pm 1$ existen puntos de inflexión.

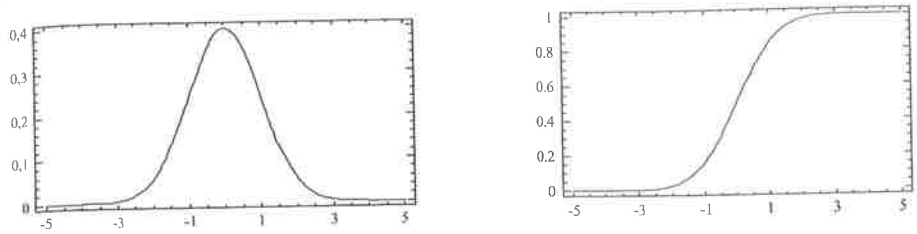


Figura A.7 - a) Función de densidad de la distribución N(0,1); b) Función de distribución de la *N(0,1)*

Las principales características son, como ya se ha comentado:

$$E[Z] = 0 \qquad Var[A] = 1$$

Para la tabulación de su función de distribución (Tabla 4 del Apéndice A), se nota mediante z_p al valor de la $N(0,1)$ que deja a su izquierda una probabilidad p, es decir $P[Z \leq z_p] = p$.

Sin embargo, lo más común es que los fenómenos bajo estudio no se distribuyan como una $N(0, 1)$, sino como una $N(\mu, \sigma)$, en la que $\mu \neq 0$ y/o $\sigma \neq 1$.

Una variable aleatoria continua A se dice que tiene una distribución Normal con media μ y desviación típica σ, $N(\mu,\sigma)$, si, para todo valor $a \in \Re$, su función de densidad viene dada por:

$$f(a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(a-\mu)^2/2\sigma^2}$$

cuyo comportamiento puede observarse en la Figura A.8, que adopta la forma de campana que posee un máximo en la posición media y que se extiende asintóticamente desde - ∞ a + ∞. Esta curva es simétrica respecto de la media y para $z > \mu$ la curva es decreciente, y para $z < \mu$ es creciente. Junto a lo anterior, analíticamente se demuestra que en $z = \mu \pm \sigma$ existen puntos de inflexión.

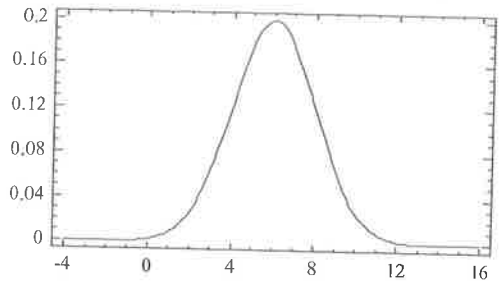


Figura A.8 - Función de densidad de la distribución *N(6,2)*

Según lo comentado anteriormente, sus principales características son:

$$E[A] = \mu \qquad Var[A] = \sigma^2$$

La distribución Normal tiene la siguiente propiedad, muy útil para trabajar con ella: Sea $A \sim N(\mu, \sigma)$; entonces la variable aleatoria Z definida como (ver apartado 2.3):

$$Z = \frac{A - \mu}{\sigma}$$

es decir, la variable aleatoria A tipificada, sigue una distribución $N(0, 1)$. Este proceso de tipificación permite pues utilizar las tablas existentes para la $N(0, 1)$ y aplicarlas a cualquier situación real en la que $\mu \neq 0$ y/o $\sigma \neq 1$.

3.3.7. DISTRIBUCIONES ASOCIADAS A LA DISTRIBUCIÓN NORMAL

Se van a presentar aquí algunas distribuciones continuas que están asociadas a la distribución Normal y que se utilizarán fundamentalmente en la Inferencia Estadística.

— **Distribución χ^2:** Se dice que una variable aleatoria continua A sigue una distribución $\chi^2(n)$ con n grados de libertad si, dadas n variables aleatorias e independientes ($\{A_i\}_{i=1,\ldots,n} \sim N(0,1)$), la variable aleatoria A es la suma de los cuadrados de las n variables aleatorias anteriores:

$$A = \sum_{i=1}^{n} A_i^2 \sim \chi^2(n)$$

por lo que toma cualquier valor real positivo. Sus principales características son:

$$E[A] = n \qquad Var(A) = 2\,n$$

En la tabulación de su función de distribución (Tabla 5 del Apéndice A), se utiliza la siguiente notación para indicar la probabilidad que deja por debajo un valor determinado: $P[A \leq \chi^2_{n,p}] = p$, es decir $\chi^2_{n,p}$ es el valor que deja por debajo una probabilidad p.

La función de densidad de la distribución $\chi^2(n)$ es asimétrica y positiva, como se puede observar en la Figura A.9.

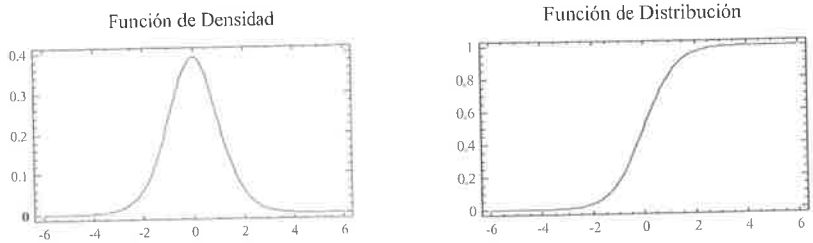


Figura A.9 - Función de densidad y distribución de la distribución $\chi^2(10)$

Se utiliza para estimar y contrastar la varianza de poblaciones Normales, además de resolver algunos contrastes no paramétricos.

— **La distribución *t* de Student:** dada una variable aleatoria $A \sim N(0,1)$ y una variable aleatoria $B \sim \chi^2(n)$ independientes, se puede definir una nueva variable aleatoria T definida como:

$$T = \frac{A}{\sqrt{B/n}} \sim t(n)$$

que puede tomar cualquier valor real, sigue una distribución t de Student con n grados de libertad. Sus principales características son:

$$E[T] = 0 \qquad Var(T) = \frac{n}{n-2}$$

La tabulación de su función de distribución (Tabla 6 del Apéndice A) es análoga a la de la distribución anterior.

Como se puede observar en la Figura A.10, la función de densidad es simétrica respecto al eje de ordenadas ($t_{n,1-p} = -t_{n,p}$). Su dispersión es mayor a la de la distribución normal estándar, aunque tiende a ésta rápidamente con n, siendo prácticamente idénticas para valores de $n >= 100$.

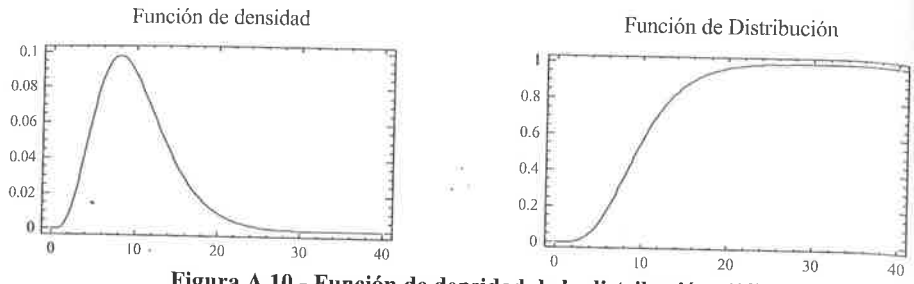


Figura A.10 - Función de densidad de la distribución t(10)

Se utiliza para estimar y contrastar medias de poblaciones Normales.

— **Función *F*:** dadas las variables aleatorias $A \sim \chi^2(m)$ y $B \sim \chi^2(n)$ independientes, la variable aleatoria F definida como:

$$F = \frac{A/m}{B/n} \sim F(m, n)$$

que toma cualquier valor real positivo, sigue una distribución F de parámetros m y n. Sus principales características son:

$$E[F] = \frac{n}{n-2} \; ; para \; n > 2 \qquad Var(F) = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)} \; ; para \; n > 4$$

La tabulación de su función de distribución (Tabla 7 del Apéndice A) es similar a las dos anteriores.

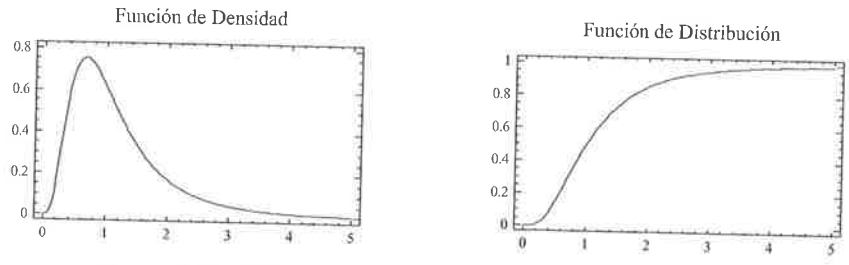


Figura A.11 - Función de densidad y de distribución de la *F(10,10)*

La distribución F es adecuada para comparar varianzas de poblaciones normales.

3.3.8. DISTRIBUCIONES MULTIDIMENSIONALES

Permiten trabajar con más de una variable aleatoria de manera simultánea. Se presentan la distribución Multinomial y la Normal multivariante, que generalizan respectivamente, al caso multidimensional, las distribuciones unidimensionales Binomial y Normal ya analizadas.

— **Distribución Multinomial:** a diferencia de lo que ocurría en la distribución binomial, en la que sólo existían dos alternativas (éxito o fracaso, con probabilidades p y q, respectivamente y tal que $p + q = 1$), en este caso existen k alternativas o sucesos (B_1, B_2, ..., B_k), tal que $p_i = P(B_i)$, para $i=1,...,k$ es la probabilidad de que ocurra el suceso B_i cada vez que se realiza el experimento. Se define ahora la variable aleatoria A_i que mide el número de veces que ocurre el suceso B_i en n realizaciones del experimento. A es pues una variable aleatoria k-dimensional $(A_1,...,A_k)$ que sigue una distribución Multinomial de parámetros n y $(p_1,...,p_k)$, donde:

$$\sum_{i=1}^{k} p_i = 1$$

es decir:

$$(A_1,...,A_k) \sim M\,(n; (p_1,...,p_k))$$

Su función de probabilidad está dada por la expresión:

$$P[A_1 = a_1,..., A_k = a_k] = \frac{n!}{a_1!\cdot...\cdot a_k!}\, {p_1}^{a_1}\cdot...\cdot {p_k}^{a_k}$$

donde:

$$\sum_{i=1}^{k} a_i = n$$

Sus principales características son:

$$E[A] = [EA_1,...,EA_k] = [n\,p_1,...,n\,p_k]$$

$$Var[A_i] = n\,p_i(1-p_i) \qquad Cov(A_i,A_j) = -n\,p_i p_j$$

Esta distribución no se encuentra tabulada debido a la gran cantidad de posibles combinaciones que se pueden presentar en los parámetros.

Se utiliza en los mismos casos que la distribución Binomial, pero en experimentos con más de dos posibles resultados.

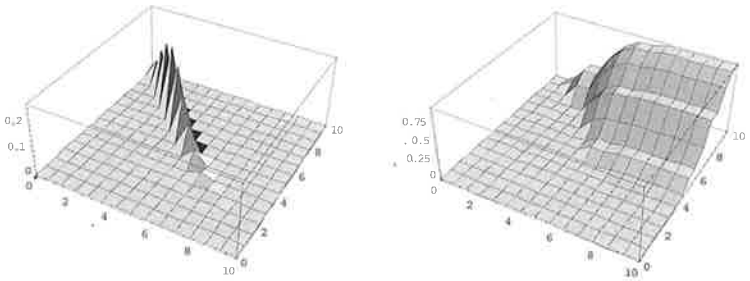


Figura A.12 - Función de densidad y de distribución de la $M(10;0,3;0,7)$

— **Distribución Normal multivariante:** dada una variable aleatoria continua de dimensión k, $A=(A_1,...,A_k)$ se dice que sigue una distribución Normal k-dimensional de parámetros:

$$\mu = (\mu_1, ..., \mu_k) \quad ; \quad S = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & ... & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & ... & \sigma_{2k} \\ ... & ... & ... & ... \\ \sigma_{k1} & \sigma_{k2} & ... & \sigma_k^2 \end{pmatrix}$$

donde μ es el vector de medias, es decir, la esperanza de cada variable unidimensional), y $[S]$ es la matriz de varianzas/covarianzas poblacional de la variable k-dimensional) si su función de densidad viene dada por:

$$f(a=(a_1,...,a_k)) = \frac{1}{(2\pi)^{k/2}(\det S^{-1})^{1/2}} e^{\left(\frac{-1}{2}(a-\mu)M(a-\mu)^T\right)}$$

Esta distribución no se encuentra tabulada debido a la gran cantidad de posibles combinaciones que se pueden presentar en los parámetros.

Se utiliza en los mismos casos que la Normal, puesto que es su extensión al caso multidimensional.

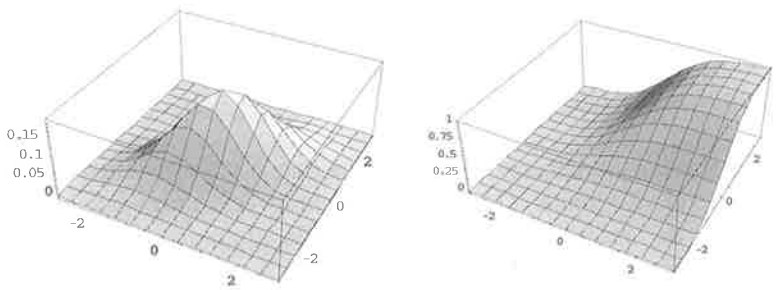


Figura A.13 - Función de densidad y de distribución de la Normal bivariante con $\mu = (0, 0)$, $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_{12} = \sigma_{21} = 1/3^{1/2}$

3.4. Teorema Central del Límite

El Teorema Central del Límite estudia el comportamiento asintótico de la suma de variables aleatorias. Se utiliza para aproximar distintas distribuciones por lo que tienen un marcado carácter práctico y es de continua aplicación pues permite el uso de la distribución Normal en múltiples casos.

Este teorema nos dice que dado un conjunto $\{A_i\}_{i=1,\ldots,n}$ de variables aleatorias independientes con media μ_i y varianzas σ_i^2 finitas, la variable aleatoria suma de todas ellas, definida como:

$$S = \sum_{i=1}^{n} A_i$$

se aproxima, una vez tipificada, a una distribución $N(0, 1)$:

$$\frac{S - E[S]}{\sqrt{Var(S)}} = \frac{\sum_{i=1}^{n} A_i - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \xrightarrow[n \to \infty]{} Z \sim N(0, 1)$$

es decir:

$$\sum_{i=1}^{n} A_i \approx N\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$$

y la aproximación mejora a medida que aumenta el valor de n.

Por tanto, cualquier distribución que se pueda escribir como suma de n variables aleatorias independientes con media y varianza finitas, donde n es un valor grande, se aproxima, tipificada, a una distribución $N(0, 1)$.

El valor a partir del cual se considera que n es suficientemente grande para poder utilizar el Teorema Central del Límite es muy subjetivo y depende principalmente de información disponible. Se empiezan a considerar valores adecuados de n a partir de *20*.

Un caso particular del teorema es cuando se consideran variables aleatorias independientes e idénticamente distribuidas, es decir, variables aleatorias con resultados no relacionados y todas ellas con la misma distribución (media μ y varianza σ^2 finitas).

$$\frac{\sum_{i=1}^{n} A_i - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{} Z \sim N(0,1), \text{ es decir, } \sum_{i=1}^{n} A_i \approx N(n\mu, \sigma\sqrt{n})$$

o

$$\frac{\overline{A}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{} Z \sim N(0,1)$$

donde:

$$\overline{A}_n = \frac{1}{n}\sum_{i=1}^{n} A_i \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$E\left[\sum_{i=1}^{n} A_i\right] = n\mu$$

$$Var\left[\sum_{i=1}^{n} A_i\right] = \sigma^2 n$$

En este caso, cualquier distribución con media y varianza finitas que se pueda escribir como suma de n variables aleatorias independientes e idénticamente distribuidas con n un valor grande se aproxima, tipificada, a una distribución $N(0, 1)$.

También se utiliza el Teorema Central del Límite para aproximar valores de distintas distribuciones de probabilidad, cuando los valores de sus parámetros son elevados y no se encuentran en las tablas. Algunos ejemplos son:

— Si A sigue una distribución Binomial $B(n, p)$, entonces:

$$\frac{A-np}{\sqrt{npq}} \xrightarrow[n\to\infty]{} Z \sim N(0,\ 1)$$

o lo que es lo mismo, $A \cong A' \sim N(\mu,\ \sigma)$ donde $\mu = n\ p$ y $\sigma = \sqrt{npq}$.

— Si A sigue una distribución de Poisson $P(\lambda)$, entonces:

$$\frac{A-\lambda}{\sqrt{\lambda}} \xrightarrow[\lambda\to\infty]{} Z \sim N(0,1)$$

es decir: $A \cong A' \sim N(\mu,\ \sigma)$ donde $\mu = \lambda$ y $\sigma = \sqrt{\lambda}$.

— Si A sigue una distribución t de Student $t(\lambda)$, con $\lambda \to \infty$, entonces

$$A \cong A' \sim N(0,1)$$

Hay que tener un especial cuidado al aproximar una distribución discreta mediante una continua, ya que se pasa de probabilidades puntuales a probabilidades en intervalos: si A es una variable aleatoria discreta y A' su variable aleatoria Normal aproximada y se desea conocer la probabilidad de que A tome un valor cualquiera a natural, entonces:

$$P(A = a) = P(A' = a) = 0$$

ya que A' es continua, lo cual no es posible, ya que todas las probabilidades de cualquier valor de X serían nulas. Por tanto, como las variables discretas sólo toman valores naturales, y las continuas toman cualquier valor dentro de un intervalo, se asigna a cada valor de A natural el intervalo correspondiente a $[a - 0,5\ ,\ a + 0,5)$ en la A' aproximada, de manera que todo valor real entre dos valores naturales consecutivos esté asignado a un valor. Por tanto:

$$P(X = a) \cong P(X' \in (a - 0,5\ ,\ a + 0,5))$$

4. INFERENCIA ESTADÍSTICA

Cuando se trataron las distribuciones de probabilidad (apartado 3.3), se presentó su uso como adecuado para modelizar o describir una variable aleatoria. En todos los casos se suponían conocidos los parámetros de la distribución de probabilidad; sin embargo, esto no suele ocurrir en la realidad, ya que los parámetros reales de las distribuciones suelen ser desconocidos y por ello se aproximan mediante muestreos. Por ejemplo, si se utiliza la distribución Normal para modelizar el error de una medida, se consideraba conocido el parámetro μ de la distribución Normal. La interpretación física de μ es que se trata del verdadero valor medio de la variable aleatoria estudiada. Sin embargo, si se conociera el verdadero valor de μ, entonces, se podría argumentar que los procedimientos formales de control de la calidad son innecesarios.

Por tanto, en general, se desconocen los parámetros de la distribución de probabilidad de la variable aleatoria de interés y, por ello, se han de desarrollar procedimientos para evaluar los parámetros de las distribuciones de probabilidad; de esto último se encarga la **inferencia paramétrica**.

Por otro lado, puede ocurrir que no se conozca la distribución de probabilidad, sino que sólo se conozcan propiedades generales de la variable aleatoria; en este caso se utilizará la **inferencia no paramétrica**.

El objetivo de la Inferencia Estadística es, pues, deducir o tomar decisiones con respecto a una población, a partir de los resultados obtenidos en una muestra seleccionada de la misma.

De nuevo existe una extensa bibliografía relativa a la Inferencia Estadística, tanto general, en la que también se estudian la Estadística Descriptiva y la Probabilidad, como son los trabajos de Canavos (1986) y Álvarez (1994), y específicos de la Inferencia Estadística, como Gutiérrez y otros (1993b).

Se supone que se usan muestras aleatorias en el análisis. El calificativo aleatorio se aplica a cualquier método de selección de una muestra que no tiene una orientación sistemática. Si se selecciona una muestra $A_1,...,A_n$, de manera que $\{A_i\}_{i=1,...,n}$ se distribuyen de forma independiente e idéntica, se tiene una muestra aleatoria de tamaño n. Esta definición es adecuada para muestras aleatorias obtenidas de poblaciones finitas o infinitas con reposición. En un muestreo sin reposición de una población finita de N elementos, se dice que una muestra de n elementos es una muestra aleatoria si todas las muestras posibles tienen igual probabilidad de ser escogidas.

Aunque la mayoría de los métodos que se estudian suponen el uso de un muestreo aleatorio, existen otras estrategias de muestreo que son útiles. Hay que tener cuidado y utilizar un método de análisis que sea congruente con el diseño del muestreo. Las técnicas inferenciales diseñadas para muestras aleatorias pueden llevar a errores grandes cuando se aplican a datos obtenidos mediante otras técnicas de muestreo.

En la Inferencia Estadística no se trabaja directamente a partir de la muestra, sino que se utilizan cantidades calculadas a partir de las observaciones de la muestra, que son los estadísticos muestrales. Un estadístico muestral se define como una función de los datos muestrales que no contiene a los parámetros desconocidos. Los estadísticos muestrales más usuales son la media y la varianza muestral.

Los estadísticos muestrales son variables aleatorias y, como tales, tienen su distribución de probabilidad. Si A es una variable aleatoria con esperanza igual a μ y varianza igual a σ_A^2, y se considera $A_1, A_2,..., A_n$ una muestra aleatoria simple de la que sean $\bar{A}$ y S_A^2 la media y varianza muestrales. Se tiene:

$$E[\bar{A}] = \frac{1}{n}\sum_{i=1}^{n} E[A_i] = \frac{n\mu}{n} = \mu$$

$$Var(\bar{A}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(A_i) = \frac{n\sigma_A^2}{n^2} = \sigma_A^2/n$$

$$E[S_A^2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(A_i - \bar{A})^2\right] = \sigma_A^2$$

Si se conoce la distribución de probabilidad de la población de la que se obtuvo la muestra, entonces muchas veces es posible determinar la distribución de probabilidad de los estadísticos muestrales. La distribución de probabilidad de un estadístico muestral se denomina distribución de muestreo.

La independencia de los estadísticos media y varianza muestrales viene dada por el lema de Fisher, que establece que si $A_1,A_2,...,A_n$ constituyen una muestra procedente de una población Normal, entonces $\bar{A}$ y S_A^2 son independientes, lo cual es muy útil para poder aplicar el resto de resultados.

La notación que se va a utilizar de aquí en adelante es la siguiente:

Población		Muestra	
μ	Media	n	Tamaño
σ^2	Varianza	$\overline{x}$	Media
π	Proporción	S^2	Varianza
		p	Proporción

4.1. Inferencia paramétrica

En este caso es conocida la distribución de probabilidad de la variable estudiada, pero se desconoce alguno de los parámetros de dicha distribución. Por lo tanto, interesa realizar inferencias respecto a los parámetros de la distribución. Para ello, se necesitan procedimientos de estimación a partir de los datos muestrales.

Se puede definir el estimador de un parámetro desconocido como un estadístico muestral que corresponde al parámetro. Un valor numérico particular de un estimador, calculado a partir de datos muestrales se llama estimación. Existen dos tipos de estimadores:

— Estimador puntual: Es un estadístico muestral que produce un único valor numérico como estimación del parámetro desconocido.

— Estimador por intervalo: Es un intervalo aleatorio que incluye el valor real del parámetro, con cierto nivel de probabilidad. Estos intervalos aleatorios se denominan normalmente intervalos de confianza.

En primer lugar, el estudio se centrará en la estimación puntual. Los estimadores puntuales adecuados requieren cierto número de propiedades importantes; entre éstas, dos de las más importantes son:

— Insesgadez: El estimador puntual debe ser insesgado; es decir, el valor esperado del estimador puntual debe ser igual al parámetro estimado.

— Varianza Mínima: El estimador debe tener varianza mínima y, como cualquier estimador puntual es una variable aleatoria, un estimador puntual de varianza mínima debe construirse de tal forma que tenga una menor varianza que cualquier otro estimador del parámetro.

Veamos a continuación los estimadores puntuales de los parámetros de las distribuciones más usuales.

Dada la variable aleatoria A con distribución de probabilidad $f(a)$, de la que se supone que se desconoce la media μ y la varianza σ^2, la estimación se realiza para la media mediante:

$$\hat{\mu} = \overline{A}$$

y para la varianza con:

$$\hat{\sigma}^2 = S^2$$

Se demuestra que $\overline{A}$ y S^2 son, respectivamente, estimadores puntuales insesgados de la media μ y la varianza poblacionales σ^2; sin embargo, S^2 no tiene varianza mínima.

No obstante, hay otras distribuciones cuyos parámetros no son la media y la varianza; por ejemplo, los parámetros de la distribución Binomial son n y π. En este caso, un buen estimador puntual del parámetro π de una distribución Binomial (n,π) es:

$$\hat{\pi} = p = \frac{A}{n}$$

4.1.1. CONTRASTES DE HIPÓTESIS PARAMÉTRICOS Y ESTIMACIÓN POR INTERVALOS

Un contraste de hipótesis es un procedimiento, prueba o test, por el cuál se decide si una propuesta sobre la población, denominada hipótesis nula, puede aceptarse o no. Veamos esto de una manera más formal: sea $A_1,A_2,...,A_n$ una muestra de tamaño n procedente de una población A con familia de distribuciones $\{F(A;\theta);\ \theta \in \Theta\}$, donde θ es un parámetro desconocido y Θ es el espacio paramétrico (conjunto de valores que puede tomar el parámetro).

Se llama hipótesis nula (H_0) a una afirmación relativa al parámetro θ desconocido, $H_0 : \theta \in \Theta_0$, es decir, la hipótesis nula consiste en afirmar que el parámetro toma determinados valores del espacio paramétrico. La hipótesis nula siempre contendrá el signo igual (igual, menor o igual, mayor o igual) y al valor del parámetro aceptado previamente como correcto.

Se llama hipótesis alternativa (H_1) a otra hipótesis relativa al parámetro desconocido, que generalmente es la hipótesis opuesta a la nula: $H_1 : \theta \in \Theta_1$ (en general $\Theta_1 = \Theta - \Theta_0$, aunque no siempre). Con ella, se pretende ver si el verdadero

valor del parámetro ha cambiado respecto del aceptado previamente en el sentido considerado en ella.

Dependiendo de que la hipótesis alternativa esté a uno o los dos lados de la hipótesis nula, se habla de hipótesis alternativa unilateral o bilateral. Una parte importante de cualquier problema de contraste de hipótesis es determinar los valores paramétricos especificados en las hipótesis nula y alternativa. En general, esto se realiza de una de las tres maneras siguientes:

— Los valores pueden resultar de pruebas o conocimientos anteriores.

— Los valorès pueden ser el resultado de alguna teoría o modelo del proceso.

— Los valores escogidos para el parámetro pueden ser el resultado de especificaciones de diseño. Pueden utilizarse procedimientos estadísticos de contraste de hipótesis para verificar la conformidad de los parámetros con sus valores especificados, o para ayudar a modificar un proceso hasta obtener valores requeridos.

Para realizar un contraste de hipótesis, se toma una muestra aleatoria de la población observada, se calcula un estadístico de contraste apropiado, y después, a partir de los resultados obtenidos en la muestra, se decide rechazar o no la hipótesis nula H_0. El conjunto de valores del estadístico de contraste que lleva al rechazo de H_0 se llama región crítica o región de rechazo del contraste.

Dado que la decisión sobre la hipótesis se toma en un ámbito estadístico, pueden ocurrir errores, que pueden ser de dos tipos:

— Error de Tipo I: Consiste en rechazar la H_0, tomándola como falsa, cuando realmente es verdadera.

— Error de Tipo II: Consiste en no rechazar H_0, tomándola como verdadera, cuando realmente es falsa.

Las probabilidades de que ocurran estos dos tipos de errores se denotan como:

$$\alpha = P[\textit{error tipo I}] = P[\textit{Rechazar } H_0 / H_0 \textit{ es verdadera}]$$

$$\beta = P[\textit{error tipo II}] = P[\textit{No rechazar } H_0 / H_0 \textit{ es falsa}]$$

Algunas veces conviene más trabajar con la potencia del contraste, definida como:

$$\textit{Potencia} = 1 - \beta = P[\textit{Rechazar } H_0 / H_0 \textit{ es falsa}]$$

Así, la potencia del contraste es la probabilidad de rechazar correctamente H_0.

Lo ideal sería encontrar una región crítica en la que ambos errores fueran nulos, pero lo que ocurre es que, para un tamaño muestral fijo, si se trata de disminuir un error, automáticamente aumenta el otro. Por ello, el procedimiento general para una prueba de hipótesis es especificar un valor de α, también llamado nivel de significación, y después diseñar un procedimiento para obtener el valor más pequeño posible de β. El valor de β es, por lo general, una función del tamaño muestral, y se controla indirectamente. Cuanto más grande sea el tamaño muestral para la prueba, menor será β, para un valor de α fijo.

Función de decisión

Para obtener el intervalo de confianza y resolver el contraste de hipótesis es necesaria la utilización de un estadístico o función de decisión. Con el valor obtenido para el estadístico en función de los datos muestrales y con un valor derivado del intervalo de confianza y que se extrae de las tablas adecuadas, se establece lo que se denomina una regla de decisión. Esta regla de decisión no es más que una comparación tras la que se dictamina si se cumple o no la H_0.

Las funciones de decisión dependen del parámetro que se esté contrastando y del conocimiento o no del resto de parámetros. A continuación se presentan las reglas para los casos más usuales.

Media

Sea A una variable aleatoria que sigue una distribución $N(\mu, \sigma)$ de la cual se toma una muestra de tamaño n. La función de decisión para el parámetro desconocido depende de algunas condiciones tal como se observa en la Tabla A.4.

Tabla A.4 - Funciones de decisión sobre la media

Condición	Función de decisión
σ es conocido	$z = \dfrac{\overline{A} - \mu}{\sigma / \sqrt{n}} \approx N(0,1)$
σ es desconocido	$t = \dfrac{\overline{A} - \mu}{S / \sqrt{n}} \approx t_{n-1}$

Varianza

Sea A una variable aleatoria con distribución $N(\mu, \sigma)$, donde σ es un parámetro desconocido y μ puede ser conocido o desconocido, de la que se toma una muestra de tamaño n. La función de decisión en este caso es:

$$x^2 = \frac{(n-1)S^2}{\sigma^2}$$

que sigue una distribución χ^2_{n-1}.

Diferencias de medias

Sean dos variables aleatorias independientes A_1 y A_2 con distribuciones $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ respectivamente, de las que se toman muestras de tamaños n_1 y n_2, respectivamente. Las funciones de decisión, para la diferencia de las medias de ambas variables, dependen de las condiciones expuestas en la Tabla A.5.

Sean B_1 y B_2 dos variables aleatorias dependientes tomadas por pares (apareadas), es decir, en condiciones muy semejantes, ambas con tamaño n. Sea $D = B_1 - B_2$ la variable resultante de la diferencia entre los pares, que sigue una normal y también es de tamaño n. Las funciones para decidir sobre la diferencia de ambas variables dependientes aparecen en la Tabla A.5.

Tabla A.5 - Funciones de decisión para la diferencia de medias

Muestras independientes	
Condiciones	**Funciones de decisión**
σ_i conocidos	$z = \dfrac{(\bar{A}_1 - \bar{A}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \approx N(0,1)$
σ_i desconocidos e iguales	$t = \dfrac{(\bar{A}_1 - \bar{A}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \approx t_{n_1+n_2-2}$
σ_i desconocidos y distintos, con n_i iguales	$t = \dfrac{(\bar{A}_1 - \bar{A}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2 + S_2^2}{n}}} \approx t_{2n-2}$
σ_i desconocidos y distintos, con n_i distintos	$z = \dfrac{(\bar{A}_1 - \bar{A}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \approx N(0,1)$
Muestras dependientes apareadas	
Condiciones	**Funciones de decisión**
σ_D conocido	$z = \dfrac{\bar{D} - \mu_D}{\sigma_D/\sqrt{n}} \approx N(0,1)$
σ_D desconocido	$t = \dfrac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \approx t_{n-1}$

Cociente de varianzas

Sean A_1 y A_2 dos variables aleatorias independientes con distribuciones $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ respectivamente, de las que se toman muestras de tamaño n_1 y n_2 respectivamente. La función de decisión para el cociente de varianzas σ_2^2 / σ_1^2 es:

$$\frac{S'^2_1}{S'^2_2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1,n_2-1}$$

Si se quiere estudiar un contraste para la igualdad de varianzas, la función de decisión es en este caso:

$$\frac{S'^2_1}{S'^2_2} \sim F_{n_1-1,n_2-1}$$

Proporción Binomial

Sea A una variable aleatoria, de la que se toma una muestra de tamaño n. Se estudia la función de decisión para la proporción poblacional. Sea p la proporción muestral, y sea π la proporción poblacional, entonces la función de decisión será:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$

Diferencia de proporciones Binomiales

Sean A_1 y A_2 dos variables aleatorias independientes, de las que se toman muestras de tamaños n_1 y n_2, respectivamente. Se estudia la función de decisión para la diferencia de proporciones de ambas poblaciones. La función de decisión es:

$$z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} \sim N(0,1)$$

Intervalos de confianza

Se habla de estimación por intervalos de confianza cuando se asigna al parámetro un intervalo de valores. Si se estima un parámetro θ mediante el estimador $\hat{\theta}$ y, por ejemplo, dicho estimador se distribuye normalmente con media $E(\hat{\theta})=\theta$ y desviación típica $\sigma(\hat{\theta})$, se tiene:

$$P\left[-1{,}96 < \frac{\hat{\theta}-\theta}{\sigma(\hat{\theta})} < 1{,}96\right] = 0{,}95 \Rightarrow P\left[\hat{\theta} - 1{,}96 \cdot \sigma(\hat{\theta}) < \theta < \hat{\theta} + 1{,}96 \cdot \sigma(\hat{\theta})\right] = 0{,}95$$

lo que indica que la probabilidad de que el parámetro de la población se encuentre en el intervalo $\hat{\theta} \pm 1{,}96\sigma$, que se denomina nivel de confianza, y que se representa por $1\text{-}\alpha$, es del 95% en el ejemplo. El valor α es el margen de error que se permite en cada estimación (en el ejemplo 0,05 ó el 5%).

Los intervalos de confianza para un parámetro desconocido se obtienen a partir de una función de decisión para dicho parámetro que esté de acuerdo a la información disponible. Para ello basta, simplemente, despejar el parámetro de dicha función y considerar la distribución de probabilidad que sigue la función de decisión con los percentiles $\alpha/2$ y $1\text{-}\alpha/2$, obteniendo así los extremos del intervalo.

Contraste de hipótesis.

El planteamiento de un contraste de hipótesis paramétrico consiste en desarrollar los pasos siguientes:

— Plantear de forma explícita las hipótesis nula y alternativa del problema.

— Localizar las condiciones del problema para encontrar la función de decisión correspondiente al parámetro a contrastar bajo las condiciones planteadas en la hipótesis nula (H_0: $\theta = \theta_0$).

— Si h es la función de decisión, tal que sigue una distribución H ($h \sim H$), se ha de calcular el estadístico muestral h_{exp} sustituyendo su valor en la función de decisión junto con los valores de la hipótesis nula planteada.

— Se comparará el valor del estadístico muestral obtenido con el valor correspondiente de la distribución H. En este caso existen dos alternativas de contrastes, el bilateral y el unilateral, dependiendo de la formulación realizada en la H_0:

a) Caso bilateral:

$$\begin{Bmatrix} H_0: & \theta = \theta_0 \\ H_1: & \theta' \neq \theta_0 \end{Bmatrix}$$

En este caso, para no rechazar la hipótesis nula, el valor calculado para el estadístico muestral h_{exp} debe estar en la región de aceptación, situada entre los valores $H_{\alpha/2}$ y $H_{1-\alpha/2}$.

b) Caso Unilateral:

$$\begin{Bmatrix} H_0: & \theta = \theta_0 \\ H_1: & \theta < \theta_0 \end{Bmatrix} \text{ ó } \begin{Bmatrix} H_0: & \theta \geq \theta_0 \\ H_1: & \theta < \theta_0 \end{Bmatrix} \quad \text{A} \qquad \begin{Bmatrix} H_0: & \theta = \theta_0 \\ H_1: & \theta > \theta_0 \end{Bmatrix} \text{ ó } \begin{Bmatrix} H_0: & \theta \leq \theta_0 \\ H_1: & \theta > \theta_0 \end{Bmatrix} \quad \text{B}$$

El contraste informa si el parámetro es mayor-igual (menor-igual) o menor (mayor) que una determinada cantidad. Para no rechazar la hipótesis nula el valor del estadístico muestral h_{exp} debe de ser superior (inferior) al valor H_α ($H_{1-\alpha}$), en el caso A (B).

4.1.2. BONDAD DE AJUSTE

Este tipo de contrastes es de gran valor dado que permiten analizar si una muestra de datos se ajusta o no a una función de distribución dada. A estos contrastes también se los denomina test de adherencia. Es común considerar que los errores de medida, y otros muchos datos experimentales, se distribuyen según una Normal; en esta línea, este tipo de contrastes permite decir si eso es o no cierto desde un punto de vista estadístico. De esta forma, las hipótesis a plantear son:

H_0: A sigue una distribución $F(a)$
H_1: A no sigue una distribución $F(a)$

En la línea del comentario anterior, las herramientas que se muestran se van a centrar en los contrastes para la comprobación de la adherencia a la Normal. Existen numerosas técnicas, entre ellas:

— Test χ^2 de bondad de ajuste: Válido para cualquier tipo de distribución de probabilidad. Se trata de un test débil, lo que quiere decir que, si genera un rechazo de la H_0, esta no se cumple, pero si se acepta, no se tiene gran seguridad.

— Test de Kolmogorov-Smirnov: Más robusto que el anterior, es un método válido sólo para distribuciones de probabilidad continuas.

— Corrección por continuidad de Lilliefors: Es el método más adecuado para el estudio de adherencia a la Normalidad. Cabe destacar que su formulación sólo es válida para la distribución Normal.

Test χ^2 de bondad de ajuste

Se consideran las siguientes condiciones para su aplicación:

— Sea $A_1,...,A_n$ una muestra procedente de una variable aleatoria A, tal que su distribución es $F(a)$ completamente determinada, es decir, sin parámetros desconocidos (más adelante se verá qué ocurre si los parámetros son desconocidos).

— Sea $S_1,...,S_r$ una partición del espacio muestral, es decir, se divide el espacio muestral en r partes (si la variable aleatoria es continua, se suelen tomar cuatro intervalos que están delimitados por los cuartiles).

— Sea p_i la probabilidad de que un valor de la variable esté en la clase S_i, suponiendo cierta la hipótesis nula, para $i = 1, ..., r$, es decir:

$$p_i = P[A_i \in S_i] > 0$$

— Si np_i es la frecuencia esperada en la clase S_i.

— Si n_i es el número de valores muestrales en la clase S_i.

Entonces se cumple que:

$$\chi^2_{\exp} = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(r\text{-}1)$$

Si hay que estimar k parámetros de $F(a)$ mediante los estimadores puntuales correspondientes, el estadístico χ^2_{exp} se distribuirá según una $\chi^2(r\text{-}k\text{-}1)$.

A partir de todo lo anterior se puede construir el siguiente contraste: dada una muestra de la variable aleatoria A, para contrastar, a nivel de significación α, la hipótesis nula de que A se distribuye según $F(a)$, frente a la hipótesis alternativa de que no lo hace, se rechaza la hipótesis nula si el valor del estadístico muestral χ^2_{exp} es mayor o igual que el valor teórico $\chi^2_{1-\alpha}(r\text{-}k\text{-}1)$.

Test de Kolmogorov-Smirnov

Sea $A_1,...,A_n$ una muestra de una variable aleatoria continua A. Se quiere contrastar que A sigue una determinada distribución continua frente a la alternativa de que no la sigue. El procedimiento que sigue en este caso es:

— Se ordenan de menor a mayor los datos muestrales obtenidos.

— Si los parámetros de la distribución que se pretende contrastar son desconocidos, se estiman puntualmente.

— Se calculan las frecuencias relativas acumuladas observadas para cada valor (o intervalo) muestral según[1]:

$$G(a_i) = (\text{número de valores muestrales menores o iguales que } a_i) / n$$

— Se calcula, a partir de la función de distribución, el valor correspondiente de cada valor observado de la muestra, suponiendo cierta la hipótesis nula, es decir, $F(a_i) = P(A \leq a_i)$.

— Para cada valor muestral, se calcula la diferencia, en valor absoluto, entre la frecuencia relativa acumulada y la función de distribución de cada valor muestral, es decir: $D_i = | F(a_i) - G(a_i) |$.

— Se toma el máximo de las diferencias absolutas anteriores: $D_{max} = \text{Máx}(D_i)$.

— Se determina el nivel crítico de decisión en la tabla de Kolmogorov-Smirnov ($D_{n,\alpha}$) en función del nivel de significación α.

Entonces, se rechaza la hipótesis nula de la bondad del ajuste, si el estadístico muestral D_{max} es mayor que el valor teórico $D_{n,\alpha}$ de la tabla del test de Kolmogorov-Smirnov (Tabla 8 del Apéndice A 2).

Corrección de Lilliefors por continuidad

Sea $A_1,...,A_n$ una muestra de una variable aleatoria continua A. Se quiere contrastar que A sigue una distribución Normal frente a la alternativa de que no la sigue. El procedimiento es idéntico al del test de Kolmogorov-Smirnov, salvo que el valor teórico se obtiene de la tabla del test de Lilliefors (Tabla 9 del Apéndice A). Este test es específico para contrastar la normalidad, por lo que es mejor que los dos anteriores para contrastar la Normalidad.

[1] En el caso de intervalo muestral, se toma como a_i el extremo final del intervalo.

4.1.3. PRUEBAS DE HOMOGENEIDAD DE VARIANZAS

Este tipo de pruebas permiten determinar si distintas muestras proceden de diferentes poblaciones en lo relativo a su comportamiento variacional. Las hipótesis nula y alternativa serán pues:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots\ldots\ldots\ldots = \sigma_k^2$$
$$H_1: \sigma_i^2 \neq \sigma_j^2 \text{ para algún } i \text{ distinto de } j$$

Existen numerosas metodologías para contrastar esta hipótesis, entre ellas:

— Test de Cochran: Para su aplicación es necesario que todas las muestras sean del mismo tamaño y se basa en comparar la varianza más grande de todos los grupos a comparar con la sumatoria de las varianzas de todos los grupos, etc.

— Test de Barlett: No es necesario que todas las muestras sean del mismo tamaño, por lo que se puede utilizar en más casos que el método anterior, aunque su metodología es más compleja.

Test de Cochran

Partiendo de las observaciones muestrales procedentes de k poblaciones (categorías o tratamientos), la aplicación de esta prueba exige las siguientes condiciones:

— Que todas las muestras sean de igual tamaño.

— Que las observaciones muestrales sean independientes (para ello se aplicará el test de rachas que se estudia más adelante).

— Que las observaciones muestrales procedan, cada una de ellas, de poblaciones Normales.

Para su realización, se calculan las varianzas muestrales S_1^2, S_2^2, ..., S_k^2 y se nota por S^2 a la mayor de ellas y se calcula el siguiente estadístico experimental para el contraste:

$$C_{exp} = \frac{S^2}{\sum_{i=1}^{k} S_i^2}$$

La decisión a adoptar es el rechazo de la hipótesis de homogeneidad de las varianzas si el estadístico de contraste C_{exp} es mayor o igual que el nivel crítico obtenido de la tabla del test de Cochran (Tabla 10 del Apéndice A). Este estadístico se obtiene en función del número de tratamientos o poblaciones (k) y el tamaño muestral de las muestras (n): $C_{k,n,\alpha}$.

Test de Bartlett

Partiendo de las observaciones muestrales procedentes de k poblaciones, la aplicación de esta prueba exige las mismas condiciones que el test de Cochran, salvo que las muestras pueden tener de tamaños distintos. Calculadas las varianzas muestrales S_1^2, S_2^2, ..., S_k^2, se obtiene el estadístico de contraste:

$$\chi^2_{exp} = \frac{1}{A} \cdot \left(N \cdot \ln S^2 - \sum_{i=1}^{k} (n_i - 1) \cdot \ln S_i^2 \right) \sim \chi^2_{k-1}$$

donde:

$$N = \sum_{i=1}^{k} (n_i - 1)$$

$$S^2 = \frac{\sum_{i=1}^{k} (n_i - 1) \cdot S_i^2}{N}$$

$$A = 1 + \frac{1}{3(k-1)} \left(\left(\sum_{i=1}^{k} \frac{1}{n_i - 1} \right) - \frac{1}{N} \right)$$

La decisión a adoptar es el rechazo de la H_0 si el estadístico de contraste χ^2_{exp} es mayor o igual que el nivel crítico obtenido de la tabla correspondiente de la χ^2 con (k-1) grados de libertad ($\chi^2_{k-1,1-\alpha}$).

4.2. Inferencia no paramétrica

La Inferencia no paramétrica se utiliza en aquellos casos en los que no se conoce la distribución de las variables aleatorias bajo estudio. Esto supone un mayor grado de libertad sobre las hipótesis de partida, pero también una cierta pérdida de capacidad decisoria y de fiabilidad. Por otra parte, la Inferencia no paramétrica también es de especial aplicación cuando el tamaño de las muestras es reducido y no se puede asumir la normalidad.

Se presentan dos contrastes no paramétricos, uno orientado al estudio de la aleatoriedad de las muestras, de valor para todos los trabajos que se basan en esta hipótesis, y otro test de ajuste o adherencia, pero en este caso a cualquier distribución.

4.2.1. TEST DE ALEATORIEDAD (TEST DE RACHAS DE WALD-WOLFOWITZ)

Es el método más usual para contrastar hipótesis de que las observaciones muestrales son independientes (es decir, la muestra se toma al azar), siendo el número de muestras 1, 2 ó más, ya que en el caso de que haya más de una muestra, se colocan los datos de una muestra tras los de otra.

Aunque existen varios métodos para resolver este contraste, se va a incidir en el llamado de la Normal, pues sirve para cualquier tamaño de la muestra (o el tamaño conjunto de todas las muestras), ya que si algún tamaño muestral es mayor de 20, las tablas del test no las contemplan.

En primer lugar conviene explicar que se entiende en Estadística por racha. Para ello nos puede servir la definición que da el Diccionario de la Lengua Española: periodo breve de fortuna o desgracia en cualquier actividad. Pues bien, en un listado de valores, una racha será un conjunto de ellos, situados unos detrás de otros, que comparten una característica, por ejemplo, el mismo signo. Existen diversas maneras de detectar las rachas, entre ellas:

— Se calcula la mediana *Me* de las observaciones y asigna a cada observación un signo positivo si la diferencia $x_i - Me$ es positiva y un signo negativo en caso contrario. Las rachas las conformarán los paquetes de signos consecutivos "+" ó "–" que se formen.

— Otra opción consiste en comparar los valores que se siguen, asignando un signo positivo a la posición de un valor de la variable si la diferencia de cada valor con el anterior es positiva y un signo negativo si es negativa.

En el caso de diferencias nulas, no se considera ese valor muestral y se reajusta el tamaño de la muestra.

Utilizando cualquiera de los dos métodos anteriores, se calculan:

n^+ = número de signos positivos
n^- = número de signos negativos
r = número de rachas

Para resolver el contraste se pueden utilizar dos métodos:

— Si los valores de n^+ y n^- son ambos menores o iguales que 20, se puede utilizar la tabla correspondiente del Test de rachas (Tabla 11 del Apéndice A); el valor

experimental r se compara con los valores extremos de la tabla y si está entre ellos se acepta la independencia de las observaciones muestrales.

— Método basado en la aproximación a la distribución N(0,1): Para valores de n^+ o n^- superiores a 20 se construye el estadístico de contraste:

$$z = \frac{(r \pm 0.5) - E(r)}{\sqrt{V(r)}} \sim N(0,1)$$

donde:

$$r + 0{,}5, \quad si \quad r < E(r)$$
$$r - 0{,}5 \quad si \quad r > E(r)$$
$$E(r) = \frac{2n^+n^-}{n^+ + n^-} + 1$$
$$V(r) = \frac{2n^+n^-(2n^+n^- - n^+ - n^-)}{(n^+ + n^-)^2(n^+ + n^- - 1)}$$

Se rechaza la H_0 (observaciones muestrales independientes) si $z \geq z_{1-\alpha/2}$ ó $z \leq z_{\alpha/2}$, es decir, se acepta la hipótesis nula si el valor experimental z está dentro del intervalo de valores $(z_{\alpha/2}, z_{1-\alpha/2})$.

4.2.2. PRUEBA DE LA χ^2 (TABLAS DE CONTINGENCIA)

Se contrasta la hipótesis de que dos muestras (A_1 y A_2) proceden de una misma población A o de dos poblaciones con igual distribución, es decir, las categorías no dependen de la muestra escogida, por lo que también se puede hablar de independencia.

En este caso, las categorías o valores que pueden tomar las variables aleatorias pueden ser tanto cuantitativas como cualitativas. Por tanto:

$$H_0: F(A_1) = F(A_2) \qquad H_1: F(A_1) \neq F(A_2)$$

La distribución de frecuencias muestrales, recibe el nombre de tabla de contingencia, y su expresión es tal como la mostrada por la Tabla A.6.

Tabla A.6 - Tabla de contingencia

	Categorías				
	1	2	...	r	Totales
Muestra 1ª	n_{11}	n_{12}	...	n_{1r}	N_1
Muestra 2ª	n_{21}	n_{22}	...	n_{2r}	N_2
Totales	n_1	n_2	...	n_r	N

Para hallar el estadístico de contraste, se debe previamente calcular las frecuencias teóricas de cada par *i,j*:

$$t_{ij} = \frac{n_j}{N} \cdot N_i$$

y, con ellas calculadas, se define el citado estadístico como:

$$\chi_{exp}{}^2 = \sum_{i=1}^{2} \sum_{j=1}^{r} \frac{(n_{ij} - t_{ij})^2}{t_{ij}} ,$$

que se distribuye aproximadamente según una ley χ^2 con $(r\text{-}1)$ grados de libertad.

La decisión a adoptar sobre el contraste es la siguiente, se rechaza la hipótesis nula si se cumple que $\chi_{exp}{}^2 \geq \chi^2_{r\text{-}1,1\text{-}\alpha}$.

Es posible la generalización a más de dos muestras, siendo en este caso la misma regla de decisión. Ahora sigue, aproximadamente, una distribución χ^2 con $(n\text{-}1)\cdot(r\text{-}1)$ grados de libertad:

$$\chi^2_{exp} = \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{(n_{ij} - t_{ij})^2}{t_{ij}} \sim \chi^2((r\text{-}1)(n\text{-}1))$$

4.2.3. PRUEBA DE KOLMOGOROV-SMIRNOV

Se comprueba que, dadas dos muestras distintas, éstas proceden de dos poblaciones con la misma distribución, por lo que la H_0 va a ser la igualdad de las distribuciones de las dos poblaciones (A_1 y A_2) de las que se han tomado las muestras, frente a la hipótesis alternativa opuesta:

$$H_0\text{: } F(A_1) = F(A_2) \qquad H_1\text{: } F(A_1) \neq F(A_2)$$

en cuyo caso se tiene un contraste *Bilateral*.

La resolución del contraste es la siguiente:

a) Si las muestras son del mismo tamaño e inferiores a 40, se acumulan las frecuencias absolutas de ambas series (N_1 y N_2) para cada valor que toma cada una de las muestras, y se obtienen las diferencias en valor absoluto en cada valor:

$$D_i = \left| N_{1,i} - N_{2,i} \right| .$$

El estadístico de contraste es el valor máximo de estas diferencias:

$$D_{máx} = \text{Máx}(D_i).$$

Se rechaza la hipótesis nula si el valor experimental $D_{máx}$ es mayor o igual que el valor obtenido en la correspondiente tabla de Kolmogorov-Smirnov ($D_{n,1-\alpha}$) (Tabla 12 del Apéndice A).

b) Muestras grandes o de distinto tamaño: Se acumulan ahora las frecuencias relativas de ambas series (F_1 y F_2) para cada valor que toma cada una de las muestras, y se obtienen las diferencias en valor absoluto para cada valor:

$$D_i = \left| F_{1,i} - F_{2,i} \right|$$

El estadístico de contraste es el valor máximo de estas diferencias:

$$D_{máx} = \text{Máx}(D_i).$$

Se rechaza la hipótesis nula si el valor experimental $D_{máx}$ es mayor o igual que el valor obtenido en la tabla de Kolmogorov-Smirnov (Tabla 13 del Apéndice A) correspondiente a este supuesto ($D_{n,1-\alpha}$ ó $D_{n1,n2,1-\alpha}$ dependiendo de que las muestras tengan o no el mismo tamaño).

4.3. Análisis de la Varianza

Los contrastes de hipótesis analizados en un apartado anterior pueden realizarse para considerar la igualdad de la media de una muestra μ frente a un valor supuesto para una población μ_0, en este caso las hipótesis son:

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$$

El Análisis de la Varianza consiste en una extensión de dicho contraste realizando, de manera conjunta, el estudio de la igualdad de comportamiento de tres o mas medias muestrales μ_i, correspondientes a tres o más poblaciones distintas, o tres o más tratamientos sobre una misma población, es decir:

$$H_0: \mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_k$$
$$H_1: \mu_i \neq \mu_j \text{ para algún } i \text{ distinto de } j$$

Este tipo de análisis sirve para verificar estadísticamente si la presencia de un determinado factor va a influir o no en el resultado de un experimento. Por ejemplo, se

puede comprobar si la temperatura ambiente influye o no en la existencia de errores de medida.

Las condiciones para la aplicación rigurosa del análisis de la varianza exigen que:

— Las observaciones muestrales sean independientes. Las muestras han de ser aleatorias, para ello se han de tomar al azar, y para su verificación se aplica el Test de Rachas a todo el conjunto de datos, ordenados muestra a muestra.

— Normalidad: Se supone que las muestras proceden de poblaciones distribuidas según una Normal, para lo cual se aplica el Test de Lilliefors.

— Homogeneidad varianzas u Homoscedasticidad: Esta hipótesis supone que las poblaciones están igualmente dispersas. Para su contrastación se aplica el Test de Bartlett o de Cochran.

4.3.1. ANÁLISIS DE LA VARIANZA PARA UN FACTOR

Es el caso más simple dado que sólo se atiende a una característica (o factor) que se analiza en k poblaciones o tratamientos distintos. La hipótesis nula es la de igualdad de medias, la hipótesis alternativa H_1 supone que algún parámetro de la media poblacional sea distinto, es decir:

$$H_0: \mu_1 = \mu_2 = \ldots\ldots\ldots = \mu_k$$
$$H_1: \mu_i \neq \mu_j \text{ para algún } i \text{ distinto de } j$$

Para ello se seleccionan k muestras, que pueden tener o no tamaños[2] n distintos, y las observaciones se disponen como en la Tabla A.7.

Tabla A.7 - Disposición de las observaciones

Tratamiento1	Tratamiento 2	...	Tratamiento i	...	Tratamiento k
a_{11}	a_{21}	...	a_{i1}	...	a_{k1}
a_{12}	a_{22}	...	a_{i2}	...	a_{k2}
...	...	...	...	...	...
$a_{1\,n1}$	$a_{2\,n2}$	...	$a_{i\,ni}$	...	$a_{k\,nk}$

Sobre los datos así ordenados se realizan los cálculos que se plantean en la Tabla A.8, lo cual permitirá decidir sobre la aceptación o no de la hipótesis de igualdad de poblaciones (medias), según el nivel de significación establecido.

[2] Cuando todos los tamaños son iguales el modelo se denomina equilibrado.

Tabla A.8 - Análisis de varianza para un solo factor				
Fuente de variación (FV)	**Suma de cuadrados (SC)**	**Grados de libertad (GL)**	**Cuadrados medios (CM)**	**F experimental (F)**
Entre tratamientos	Q_1	k-1	$S_1^2 = \frac{Q_1}{k-1}$	$F = \frac{S_1^2}{S_2^2}$
Residual	$Q_2 = Q - Q_1$	N-k N = nº total de observaciones	$S_2^2 = \frac{Q_2}{N-k}$	
Total	Q	N-1		

donde la variabilidad entre los distintos tratamientos se mide por:

$$Q_1 = \sum_{i=1}^{k} n_i(\bar{a}_i - \bar{a}) = \left(\sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{ni} a_{ij} \right)^2}{n_i} \right) - \frac{\left(\sum_{i=1}^{k} \sum_{j=1}^{ni} a_{ij} \right)^2}{N}$$

y la variabilidad residual por:

$$Q_2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (a_{ij} - \bar{a}_i) = Q - Q_1$$

y la variabilidad total, por la siguiente formulación:

$$Q = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (a_{ij} - \bar{a}) = \left(\sum_{i=1}^{k} \sum_{j=1}^{ni} a_{ij}^{\ 2} \right) - \frac{\left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} \right)^2}{N}$$

Una vez realizados los cálculos anteriores, se acepta la hipótesis de igualdad de medias si el valor F_{exp} experimental calculado como estadístico de contraste es menor que el teórico derivado de las tablas, es decir:

$$F_{exp} < F_{k-1,N-k,1-\alpha}$$

En caso de rechazo de H_0 para ese nivel de significación α, se puede estudiar entre qué tratamientos hay diferencias significativas.

Se dice que no hay diferencias significativas entre los tratamientos genéricos p y q, si:

$$F_{pq} = \frac{(\overline{a}_p - \overline{a}_q)^2}{S_2^2 \dfrac{n_p + n_q}{n_p\ n_q}} < F_{k-1,N-k,1-\alpha}\ (k-1)$$

4.3.2. ANÁLISIS DE LA VARIANZA PARA DOS FACTORES

En este caso se comparan unas poblaciones en las que se puede analizar una cierta característica bajo dos aspectos o factores. Para mayor facilidad, este caso general se va a estudiar para modelos equilibrados; dentro de los mismos se considerarán los casos sin interacción y con interacción de factores.

Modelo equilibrado sin interacción

Se analiza una característica bajo dos factores, de manera que el primero presenta F tratamientos o niveles y el segundo presenta C tratamientos o niveles, contrastando la hipótesis nula de igualdad de medias en ambos factores frente a las hipótesis alternativas de que algún parámetro de las medias poblacionales de cualquier factor sea distinto, es decir:

$$H_{01}: \mu_{11} = \mu_{12} = \ldots\ldots\ldots\ldots = \mu_{1F}$$
$$H_{02}: \mu_{21} = \mu_{22} = \ldots\ldots\ldots\ldots = \mu_{2C}$$
$$H_{11}: \mu_{1i} \neq \mu_{1j} \quad \text{para algún } i \text{ distinto de } j$$
$$H_{12}: \mu_{2i} \neq \mu_{2j} \quad \text{para algún } i \text{ distinto de } j$$

Para ello se seleccionan muestras de tamaño n en cada parcela experimental (cruce de fila y columna), disponiendo las observaciones en una tabla tal y como aparece en la Tabla A.9.

Tabla A.9 - Disposición de las muestras en el modelo equilibrado sin interacción

	Columna 1	.....	**Columna j**	.....	**Columna C**
Fila 1	$A_{111}, \ldots, A_{11n}$	.....	$A_{1j1}, \ldots, A_{1jn}$	.....	$A_{1C1}, \ldots, A_{1Cn}$
.....	.....	.....	.....	.....	.....
Fila i	$A_{i11}, \ldots, A_{i1n}$	.....	$A_{ij1}, \ldots, A_{ijn}$	.....	$A_{iC1}, \ldots, A_{iCn}$
.....	.....	.....	.....	.....	.....
Fila F	$A_{F11}, \ldots, A_{F1n}$	.....	$A_{Fj1}, \ldots, A_{Fjn}$	.....	$A_{FC1}, \ldots, A_{FCn}$

Para resolver el problema, se calculan las medias de cada tratamiento del primer factor ($\overline{F}_i$), de cada tratamiento del segundo factor ($\overline{C}_i$) y global ($\overline{A}$). A partir de ellas se calculan las distintas sumas de cuadrados que descomponen la variabilidad total Q del sistema. Éstas son:

$$Q_1 = n\,C \sum_{i=1}^{F} (\overline{F}_i - \overline{A})^2$$

que mide la variabilidad entre las filas (la suma de cuadrados para las filas),

$$Q_2 = n\,F \sum_{j=1}^{C} (\overline{C}_j - \overline{A})^2$$

que mide la variabilidad entre las columnas (la suma de cuadrados para las columnas)

$$Q_3 = \sum_{i=1}^{F} \sum_{j=1}^{C} \sum_{k=1}^{n} (A_{ijk} - \overline{F}_i - \overline{C}_j + \overline{A})^2$$

que, al igual que las anteriores, mide la variabilidad residual de cada parcela experimental, y

$$Q = Q_1 + Q_2 + Q_3 = \sum_{i=1}^{F} \sum_{j=1}^{C} \sum_{k=1}^{n} (A_{ijk} - \overline{A})^2$$

que es la suma de cuadrados de la fuente de variación total.

Las decisiones que se adoptan son las siguientes:

— Se acepta la hipótesis de igualdad de comportamiento entre las filas (todos los promedios del factor fila son iguales) si el valor F_F experimental calculado como estadístico de contraste es menor que el teórico, es decir, $F_F < F_{(F-1),(N-F-C+1),1-\alpha}$.

— Se acepta la hipótesis de igualdad de comportamiento entre las columnas (todos los promedios del factor columna son iguales) si el valor F_C experimental calculado como estadístico de contraste es menor que el teórico, es decir, $F_C < F_{(C-1),(N-F-C+1),1-\alpha}$.

Al igual que en el caso de un único factor, el proceso se suele resumir como aparece en la Tabla A.10.

Tabla A.10 - Análisis de varianza para dos factores sin interacción

Fuente de variación (FV)	Suma de cuadrados (SC)	Grados de libertad (GL)	Cuadrados medios (CM)	F experimental (F)
Entre filas	Q_1	$F-1$	$S_1^2 = \frac{Q_1}{F-1}$	$F_F = \frac{S_1^2}{S_3^2}$
Entre columnas	Q_2	$C-1$	$S_2^2 = \frac{Q_2}{C-1}$	$F_C = \frac{S_2^2}{S_3^2}$
Residual	Q_3	$N-F-C+1$	$S_3^2 = \frac{Q_3}{N-F-C+1}$	
Total	Q	$N-1$		

Modelo equilibrado con interacción

El planteamiento base es similar al comentado en el apartado anterior: se analiza una característica bajo dos factores, de manera que el primero presenta F tratamientos o niveles y el segundo presenta C tratamientos o niveles (columnas), contrastando la hipótesis nula de igualdad de medias en ambos factores frente a las hipótesis alternativas de que algún parámetro de las medias poblacionales de cualquier factor sea distinto. Es decir:

$$H_{01}: \mu_{11} = \mu_{12} = \ldots\ldots\ldots\ldots = \mu_{1F}$$
$$H_{02}: \mu_{21} = \mu_{22} = \ldots\ldots\ldots\ldots = \mu_{2C}$$
$$H_{11}: \mu_{1i} \neq \mu_{1j} \text{ para algún } i \text{ distinto de } j$$
$$H_{12}: \mu_{2i} \neq \mu_{2j} \text{ para algún } i \text{ distinto de } j.$$

Sin embargo, en este caso se considera además la posible existencia de interacciones entre los factores; es decir, se estudia si el efecto de cada nivel de un factor depende de los niveles del otro. Para este caso las hipótesis son:

H_{03}: La interacción no es significativa
H_{13}: La interacción sí es significativa

Para llevar a cabo este análisis de la varianza se seleccionan muestras de tamaño n en cada parcela experimental (cruce de fila y columna), disponiendo las observaciones en una tabla de datos del análisis de varianza. A partir de ella se calculan las medias de cada parcela ($\overline{A}_{ij}$), de cada tratamiento del primer factor ($\overline{F}_i$), de cada tratamiento del segundo factor ($\overline{C}_i$) y global ($\overline{A}$). Las distintas sumas de cuadrados que descomponen la variabilidad total Q del sistema son las siguientes:

$$Q_1 = n\,C \sum_{i=1}^{F} (\overline{F}_i - \overline{A})^2$$

que mide la variabilidad entre las filas (la suma de cuadrados para las filas),

$$Q_2 = n\,F \sum_{j=1}^{C} (\overline{C}_j - \overline{A})^2$$

que mide la variabilidad entre las columnas (la suma de cuadrados para las columnas),

$$Q_3 = n \cdot \sum_{j=1}^{C} \sum_{i=1}^{F} (\overline{A}_{ij} - \overline{F}_i - \overline{C}_j + \overline{A})^2$$

que mide la interacción entre el factor fila y el factor columna,

$$Q_4 = \sum_{i=1}^{F} \sum_{j=1}^{C} \sum_{k=1}^{n} (A_{ijk} - \overline{A}_{ij})^2$$

que mide la variabilidad residual (de cada parcela experimental), y

$$Q = Q_1 + Q_2 + Q_3 + Q_4 = \sum_{i=1}^{F} \sum_{j=1}^{C} \sum_{k=1}^{n} (A_{ijk} - \overline{A})^2$$

que es la suma de cuadrados de la fuente de variación total.

Al igual que en los casos anteriores, todo el proceso se resumen como se indica en la Tabla A.11.

Tabla A.11 - Análisis de varianza para dos factores con interacción

Fuente de variación (FV)	Suma de cuadrados (SC)	Grados de libertad (GL)	Cuadrados medios (CM)	F experimental (F)
Entre filas	Q_1	$F\text{-}1$	$S_1^2 = \frac{Q_1}{F-1}$	$F_F = \frac{S_1^2}{S_4^2}$
Entre Columnas	Q_2	$C\text{-}1$	$S_2^2 = \frac{Q_2}{C-1}$	$F_C = \frac{S_2^2}{S_4^2}$
Interacción	Q_3	$(F\text{-}1)(C\text{-}1)$	$S_3^2 = \frac{Q_3}{(F-1)(C-1)}$	$F_I = \frac{S_3^2}{S_4^2}$
Residual	Q_4	$F{\cdot}C(n\text{-}1)$	$S_4^2 = \frac{Q_4}{F \cdot C(n-1)}$	
Total	Q	$N\text{-}1$		

Las decisiones que se adoptan son las siguientes:

— Se acepta la hipótesis de que las interacciones no son significativas si el valor F_I experimental calculado como estadístico de contraste es menor que el teórico, es decir, $F_I < F_{(F-1)(C-1),(F\cdot C(n-1)),1-\alpha}$. En el caso de que la interacción no sea significativa, se debe de considerar el modelo sin interacción, sumando la variación Q_3 a la residual Q_4.

— Se acepta la hipótesis de igualdad de comportamiento entre filas (los promedios del factor fila son iguales) si el valor F_F experimental calculado como estadístico de contraste es menor que el teórico: $F_F < F_{(F-1),(F\cdot C(n-1)),1-\alpha}$.

— Se acepta la hipótesis de igualdad de comportamiento entre columnas (los promedios del factor columna son iguales) si el valor F_C experimental calculado como estadístico de contraste es menor que el teórico: $F_C < F_{(C-1),(F\cdot C(n-1)),1-\alpha}$.

4.3.3. TEST DE KRUSKAL-WALLIS

Cuando no se pueda utilizar el Análisis de la Varianza debido al incumplimiento de alguna de sus condiciones previas, se puede utilizar un test no paramétrico para contrastar la hipótesis de que las muestras procedentes de más de dos poblaciones tienen la misma medida de centralización. Este contraste es el de Kruskal-Wallis.

Su planteamiento es el siguiente: sean k variables aleatorias independientes $A_1,...,A_k$, de las que se toman muestras aleatorias independientes de tamaño n_i $(i=1,...,k)$, respectivamente y se desea contrastar la hipótesis nula de que todas las poblaciones tienen la misma medida de centralización:

$$H_0 : M_1 = M_2 = M_3 = \ldots\ldots\ldots\ldots = M_k$$
$$H_1 : M_i \neq M_j \text{ para algún } i \neq j$$

El contraste se realiza de la siguiente forma:

— Se sustituyen los valores de las variables por sus correspondientes rangos, considerando conjuntamente todas las muestras, y se calcula la suma R_i de los rangos en cada tratamiento y su cuadrado R_i^2.

— Siendo $N = \sum_{i=1}^{k} n_i$ (número total de observaciones), el estadístico de contraste es:

$$H = \frac{12}{N(N+1)}\left(\sum_{i=1}^{k}\left(\frac{R_i^2}{n_i}\right)\right) - 3(N+1)$$

— Si al obtener los rangos se producen muchos empates, el estadístico de contraste es:

$$H' = \frac{H}{E}\ ;\ \text{con}\ \ E = 1 - \frac{\sum_{i=1}^{g}\left(e_i^3 - e_i\right)}{N^3 - N}$$

siendo:

g Número de empates.
e_i Número de valores en cada empate.

Se rechaza la hipótesis nula si el estadístico de contraste H (o H') es mayor o igual que el nivel crítico obtenido de la tabla χ^2 con $k\text{-}1$ grados de libertad, es decir, el valor $\chi^2_{k-1,1-\alpha}$.

4.4. Eliminación de outliers

En la Estadística tradicional se han considerado datos extraños a aquellos que presentan una probabilidad de ocurrencia muy pequeña. Estos datos se denominan outliers debido a que su posición respecto a una función de densidad, por ejemplo la Normal, se encuentra muy alejada de la central, en las colas de la distribución. De manera más formal, se considera que un dato muestral A_i procedente de una población A es un outlier si no está dentro del intervalo dado por $(\mu_A \pm k\sigma_A)$. La perspectiva tradicional considera que estos puntos se pueden eliminar de la muestra porque son datos extraños, seguramente obtenidos a partir de un error de medida o de transcripción[3]. El valor de k viene dado por el nivel de confianza α utilizado, pues $\alpha = 1 - P[-k \le Z \le k]$.

[3] Indudablemente, este criterio no parece muy lógico. Si los outliers son consecuencia de errores de trascripción son realmente errores groseros, y si son datos de una población, por muy pequeña probabilidad que tengan, pertenecen a la misma y surge la duda de por qué eliminarlos. Modernamente, la estadística robusta propone formas de incorporar este tipo de datos en los procesos de análisis estadístico.

5. ESTADÍSTICA CIRCULAR

Muchas variables geoespaciales, como la orientación, los vientos, o los errores de una lente, poseen una distribución que no es unidimensional, que ocurre en un plano y que posee un aspecto circular o, al menos, no unidireccional. Para tratar adecuadamente este tipo de datos es necesario disponer de una base estadística propia y de ello se encarga la Estadística circular. Realmente la Estadística circular no es más que un caso más complejo y general que el de la Estadística más convencional o unidimensional. La Estadística circular permite el desarrollo de análisis similares a los que se pueden plantear en la estadística convencional (p.e. contrastes, análisis cluster, etc.). En este apartado se atenderá exclusivamente a presentar las distribuciones más básicas de la misma. La inferencia estadística puede ser utilizarse también en este caso para plantear hipótesis sobre los parámetros característicos de las distribuciones y para contrastar las hipótesis que se consideren oportunas para un estudio concreto.

En cartografía la consideración de la distribución de los errores como funciones circulares ha sido tratada por numerosas fuentes (ACIC, 1962; Maling, 1989), y también existen ejemplos, aplicados a cartografías temáticas, de análisis de variables circulares y esféricas (por ejemplo, Davis, 1986). Dentro de los manuales estadísticos dedicados a esta materia cabe destacar el trabajo de Fisher (1995).

5.1. Principales distribuciones circulares

De manera análoga a la Estadística tradicional, se definen la función de densidad, la función de distribución y los momentos de la variable aleatoria circular:

— La función de densidad $f(\theta)$ de una variable aleatoria circular Θ se define como una función continua, no negativa, periódica con periodo 2π ($f(\theta+2\pi) = f(\theta)$) y en la que se verifica:

$$\int_0^{2\pi} f(\theta)d\theta = 1$$

— Su correspondiente función de distribución viene definida como:

$$F(\theta) = \int_0^{\theta} f(\theta)d\theta \quad \text{para} \quad 0 \leq \theta \leq 2\pi$$

— El momento trigonométrico de orden p de la variable aleatoria Θ se define como:

$$\mu_p' = \rho_p e^{i\mu_p'} = \rho_p \cos\mu_p' + i\,\rho_p \operatorname{sen}\mu_p' =$$
$$= \int_0^{2\pi} \cos(p\theta) f(\theta) d\theta + i \int_0^{2\pi} \operatorname{sen}(p\theta) f(\theta) d\theta = \alpha_p' + i\beta_p'$$

donde:

α_p' y β_p'	Momentos p-ésimos coseno y seno, respectivamente.
$\mu = \mu_1$	Dirección media.
$\rho = \rho_1$	Longitud media resultante.

— Se define el momento trigonométrico central de orden p de Θ como el momento trigonométrico de orden p de $\Theta - \mu$:

$$\mu_p = \rho_p e^{i\mu_p} = \rho_p \cos\mu_p + i\,\rho_p \operatorname{sen}\mu_p =$$
$$= \int_0^{2\pi} \cos p(\theta - v) f(\theta) d\theta + i \int_0^{2\pi} \operatorname{sen} p(\theta - \mu) f(\theta) d\theta = \alpha_p + i\beta_p$$

— Se define la varianza circular de Θ como $v = 1 - \rho$, que toma valores entre 0 y 1.

— La desviación típica circular no se define como la raíz cuadrada de la varianza circular, sino como $\sigma = (-2\log(1-v))^{1/2} = (-2\log\rho)^{1/2}$.

— Una medida asociada de extensión es la dispersión circular, definida como $\delta = (1-\rho_2)/(2\rho^2)$.

Las principales distribuciones de probabilidad circulares son:

— Uniforme circular: Es el paso a la Estadística circular de la distribución Uniforme de la Estadística tradicional.

— Distribuciones envueltas: Cualquier distribución de la Estadística tradicional se puede envolver y pasarse a la Estadística circular.

— Normal circular: Es el paso a la Estadística circular de la distribución Normal de la Estadística tradicional.

5.1.1. UNIFORME CIRCULAR

Con este modelo, todas las direcciones entre 0 y 360° tienen la misma probabilidad. Sus principales características son:

La función de densidad es:

$$f(\theta) = \frac{1}{2\pi} \quad \text{para} \quad 0 \leq \theta \leq 2\pi$$

Y la función de distribución:

$$F(\theta) = \frac{\theta}{2\pi} \quad \text{para} \quad 0 \leq \theta \leq 2\pi$$

Los principales momentos son:

- La dirección media μ no está definida.
- La longitud media resultante: $\rho = 0$.
- La dispersión circular: $\delta = \infty$.
- $\alpha_p = \beta_p = 0$.

5.1.2. DISTRIBUCIONES ENVUELTAS

A partir de cualquier variable aleatoria A de la recta real, con función de densidad $g(a)$, se puede obtener una variable aleatoria circular Θ definiéndola como $\Theta = A[\text{mod } 2\pi]$. Entonces, la función de densidad de la nueva variable circular obtenida es:

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2k\pi)$$

A partir de este método se pueden obtener las distribuciones circulares más usuales.

5.1.3. DISTRIBUCIÓN NORMAL CIRCULAR WN(μ, ρ)

Es una distribución simétrica unimodal con dos parámetros, que se obtiene envolviendo la distribución Normal de la recta real en el círculo. La función de densidad es:

$$f(\theta)=\frac{1}{2\pi}\left(1+2\sum_{p=1}^{\infty}\rho^{p^2}\cos p(\theta-\mu)\right) \text{ para } 0\le\theta\le 2\pi \text{ y } 0\le\rho\le 1$$

Los principales momentos son:

— La dirección media es el parámetro μ.

— La longitud media resultante es el parámetro ρ.

— La dispersión circular: $\delta=(1-\rho^4)/(2\rho^2)$.

— $\alpha_p=\rho^{p^2}$

— $\beta_p=0$ para $p\ge 1$.

Si la distribución se obtiene envolviendo una distribución Normal de varianza σ^2, entonces, $\rho=e^{-\frac{1}{2}\sigma^2}$. Si $\rho\to 0$, la distribución converge a la Uniforme circular; y si $\rho\to 1$, la distribución tiende a la distribución puntual concentrada en la dirección μ.

6. ANÁLISIS CLUSTER

El análisis Cluster es una técnica de análisis multivariante que sirve para clasificar un único conjunto de datos en una serie de grupos (clusters) que no están definidos a priori, pero que están sugeridos por la propia esencia o estructura de los datos, de manera que los individuos que pueden considerarse similares serán asignados a un mismo grupo o cluster, mientras que los individuos diferentes se localizarán en grupos distintos. Este tipo de análisis es el que da base a los procesos de clasificación en teledetección. En cierta medida el análisis cluster es el proceso complementario al análisis de varianza, pues en él se dispone de grupos y se comprueba si son homogéneos o no, y aquí se dispone de un conjunto y se trabaja para dividirlo en grupos más homogéneos.

Este análisis puede ser unidimensional o realizarse sobre un mayo número de variables (como en el caso de la Teledetección, donde se utilizan diversas bandas espectrales). Conviene advertir que si el análisis se realiza *n*-dimensionalmente, es común tipificar las variables.

La bibliografía para el estudio del análisis cluster es muy amplia, y está incluido, principalmente, en libros relativos al Análisis Multivariante, entre los que se destacan el manual ya clásico de Anderberg (1973), y otros más modernos como el de Hair y

col. (2000), que realiza un amplio estudio teórico, y Pérez (2001), que trata su utilización mediante el programa informático SPSS.

Existen dos tipos básicos de análisis Cluster:

— No jerárquico: Cada grupo no depende de otro en una jerarquía y se persigue minimizar la varianza residual (suma de varianzas dentro de cada grupo), asignando los casos a grupos diferenciados que el mismo análisis configura.

— Jerárquicos: Se configuran grupos con estructura arborescente, de forma que aquellos de niveles más bajos vayan siendo englobados en otros de niveles superiores. Son de interés cuando se pretende detectar la estructura de los datos.

La solución que suelen dar los expertos es utilizar un método jerárquico para decidir el número de grupos y después un algoritmo no jerárquico, como es el de las k-Medias, para realizar el agrupamiento. A continuación se presentan dos algoritmos básicos que permiten desarrollar cada una de estas alternativas.

6.1. Análisis Cluster no jerárquico

Este proceso clasifica todos los casos de una población en grupos separados que configura el propio mecanismo de análisis. El análisis necesita que se fije, en principio, el número de agrupaciones en las que pretende dividir los datos. Como no existe un número definido de grupos, se debe repetir con diferente número hasta ver la clasificación que mejor se ajusta al problema estudiado. No obstante, el estudio del dendograma resultante de un cluster jerárquico, la experiencia previa o el conocimiento que se tiene acerca del problema en cuestión, suele ser útil para fijar el número de grupos.

Una clasificación correcta debe ser aquella en la que la dispersión dentro de cada grupo sea la menor posible (criterio de la varianza) y lleva a seleccionar una configuración cuando la suma de varianzas dentro de cada grupo (varianza residual) sea mínima. Para ello, se utiliza el algoritmo de las K-medias (MacQueen, 1967).

El algoritmo de las K-medias parte de una primera configuración arbitraria de grupos, eligiendo un primer individuo como semilla de cada grupo y, mediante pruebas sucesivas, contrasta el efecto que sobre la varianza residual tiene la asignación de cada uno de los casos a cada uno de los grupos. El valor mínimo de la varianza determina la configuración de los grupos con sus respectivas medias. Se asignan otra vez todos los casos a estos nuevos centroides, en un proceso que se repite hasta que ningún cambio pueda ya disminuir la varianza residual o hasta que la diferencia entre los centroides de dos pasos consecutivos sea menor que una cantidad fijada. Entonces,

da una solución para el número de grupos elegidos, a la que se llegará con menor número de iteraciones cuanto más cerca estén las medias iniciales de las finales.

Los agrupamientos no jerárquicos se utilizan para grandes tablas de datos y son bastante útiles para detectar casos atípicos, ya que si se elige un número grande de grupos, los que contengan muy pocos datos sirven para detectar casos extremos que pueden distorsionar la configuración, por lo que se debería realizar el análisis definitivo sin ellos, para después asignar los atípicos al grupo adecuado, que ha sido formado sin su influencia.

6.2. Análisis Cluster jerárquico

El análisis comienza considerando tantos grupos como individuos se estudian, y se busca conseguir agrupaciones sucesivas entre ellos, de manera que se vayan integrando en grupos que se unirán entre sí en un nivel superior formando agrupaciones mayores que se unirán hasta llegar a un grupo final que contiene a todos los casos analizados. Se obtiene así un árbol invertido que se llama dendograma. Si en un paso se agrupan dos conjuntos, éstos quedan agrupados para el resto del estudio.

Si se dispone de n datos de una variable p-dimensional, el agrupamiento se hace en función de la similitud entre los grupos, por lo que hay que definir una medida para dicha similitud. La similitud se mide mediante una distancia que se define previamente al uso del propio algoritmo, por ejemplo: distancia euclídea, distancia euclídea al cuadrado, en bloque, Chebychev, etc. Con todos los puntos y las distancias entre los mismos se construye una matriz de orden $n \times n$. El proceso consiste en pasar de esta matriz de $n \times n$ y n grupos a una matriz de un único grupo y dimensiones 1×1. Existen muchas formas de medir las distancias multivariantes, de manera que la variedad de formas de medir proporciona diversas posibilidades de análisis; por ello el uso de varios métodos es imprescindible para contrastar los resultados y elegir el número óptimo de grupos.

Las distancias más utilizadas son:

— La distancia euclídea:

$$d(A,B) = \sqrt{\sum_{i=1}^{p} (a_i - b_i)^2} = |A - B|$$

— La distancia de bloques o Manhattan:

$$d(A,B) = \sum_{i=1}^{p} | a_i - b_i |$$

— La distancia de Chebychev:

$$d(A,B) = \underset{i}{Max} | a_i - b_i |$$

La segunda decisión a tomar es el algoritmo a utilizar para la formación de grupos. Existen varias posibilidades, entre ellas:

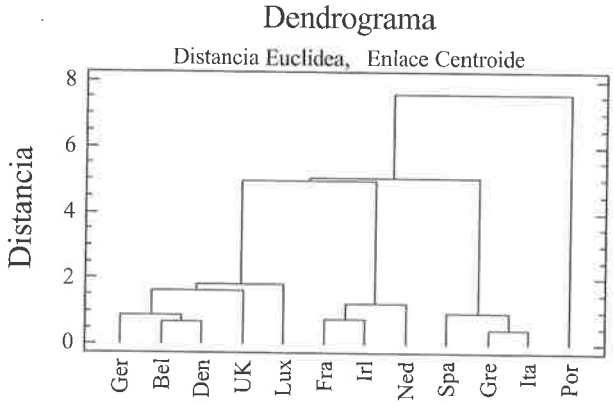


Figura A.14 - Ejemplo de Dendograma

— Enlace promedio: Considera como distancia entre dos grupos la distancia media entre todos los posibles pares de casos.

— Enlace centroide: Considera como distancia entre dos grupos la existente entre sus centros de gravedad, definidos por las medias aritméticas de las variables de los individuos que componen los grupos.

— Enlace por mínima varianza: Considera como distancia entre dos grupos el menor incremento de varianza residual global. En un nivel dado existirá un número de grupos de los que se deben elegir dos para una nueva fusión; se prueban todas las parejas posibles y se calcula la varianza residual global o intra-grupos con cada pareja unida y todos los demás grupos, de manera que la pareja de grupos que produzca el mínimo incremento de esta varianza residual será la elegida para su unión en un nuevo nivel. En el último nivel, con todos los individuos agrupados en un único grupo, la varianza residual es máxima y coincide con la varianza total.

7. SIMULACIÓN

La simulación es simplemente la construcción de un programa de ordenador que, describiendo el comportamiento de un sistema real mediante un modelo, permite obtener conclusiones de valor para apoyar la toma de decisiones (Ríos y Ríos, 1997). Se trata, pues, de una metodología de análisis basada en un soporte informático y en la teoría de sistemas. Independientemente del tipo de simulación que se realice, y del tipo de sistema que se esté modelizando (dinámico, estático, distribuido, continuo, etc.), la simulación estocástica se puede definir en una serie de pasos:

— Generación de unos valores aleatorios.

— Transformación de los valores aleatorios en entradas al modelo.

— Ejecución el modelo.

— Estudio estadístico de la salida para entender el comportamiento del modelo.

La simulación es pues el estudio del comportamiento del sistema real a través de la observación del modelo. La simulación es una herramienta de sumo interés dado que permite determinar si las suposiciones sobre el comportamiento del modelo son válidas. El análisis se ve favorecido dado que, al realizar las ejecuciones sobre el ordenador, se puede jugar con grupos de variables, fijar parámetros, modificar la importancia de las interacciones e incluso trabajar con distintos modelos del sistema, lo que permite profundizar en el conocimiento del mismo.

En el control de calidad de bases de datos geográficas la simulación es una herramienta cuya aplicación es muy válida, permite obtener resultados de valor tanto en aplicaciones sobre la componente posicional como de temática. En el ámbito de los SIG es también una herramienta de grandes posibilidades para determinar si los resultados de un análisis son más o menos robustos.

En este apartado se van a presentar las bases para la aplicación de esta potente herramienta. La bibliografía recomendada para un estudio más profundo sobre simulación es: Ríos (1997), Pardo (1987) y Ross (1987).

7.1. Generación de números aleatorios

Una parte muy importante de la Simulación es la generación de números aleatorios. Se dice que una secuencia infinita de números es aleatoria si cualquier secuencia finita seleccionada previamente a su diseño es igualmente factible de estar incluida en aquella. En un principio, mediante procedimientos físicos se crearon tablas

de números aleatorios intentando evitar el tenerlos que obtener mediante procedimientos de azar en su momento de uso. Estas tablas son eficientes para cálculos manuales, pero su uso con ordenadores es bastante problemático. Así, se han creado una serie de métodos que dan los números pseudo-aleatorios, que son aquellos que tienen un carácter aleatorio, pero que han sido obtenidos mediante fórmulas recursivas, aritméticas,...

Se van a obtener números pseudo-aleatorios procedentes de una distribución Uniforme en el intervalo [0, 1] y, a partir de ellos, se van a pasar a cualquier distribución de probabilidad, en especial a la $N(0, 1)$, de la que, deshaciendo la tipificación, se puede pasar a cualquier otra distribución Normal.

7.1.1. GENERACIÓN DE NÚMEROS PSEUDO-ALEATORIOS

Son muchos los métodos de generación de valores pseudo-aleatorios, siendo el método congruencial mixto uno de los más potentes, robustos y avanzados. Para su utilización es necesaria la utilización de módulos (ver glosario). Con este método, se obtienen una serie de números enteros positivos entre 0 y $m - 1$ de manera uniforme.

El procedimiento del método es el siguiente:

— Se fija el módulo m de trabajo.

— Se fijan a, u_0, y b enteros positivos menores que m.

— Cada valor es:

$$u_n = a \cdot u_{n-1} + b \ (mod\ m) = a^n \cdot u_0 + \left(\frac{a^n - 1}{a - 1} \right) b \ (\mathrm{mod}\, m).$$

En este caso, u_0 es la semilla, es decir, el valor del que se parte para generar los números pseudo-aleatorios.

Para obtener un periodo máximo, es decir, que la serie generada sea lo más larga posible antes de que los valores se vuelvan a repetir, se deben cumplir las siguientes condiciones:

— a debe de ser impar y distinto de 1.

— b debe de ser impar.

— La semilla u_0 puede ser cualquier valor entero positivo entre 0 y $m-1$.

— b y m deben de ser primos relativos.

— a debe ser congruente con 1 módulo p ($a=1$ ($mod\ p$)), para todo factor primo p del módulo m.

— a debe ser congruente con 1 módulo 4 ($a=1$ ($mod\ 4$)), si 4 es un divisor de m.

— m debe de ser mayor que el número de valores pseudo-aleatorios que se pretenden generar.

En caso de que no se cumplan todas estas condiciones, no se tiene garantizado el periodo máximo, y va a depender de la semilla elegida.

Si se pretende generar números pseudo-aleatorios de una distribución $U[0, 1]$, se dividen los valores obtenidos entre $m - 1$. Se puede comprobar la aleatoriedad de los números obtenidos utilizando el test de aleatoriedad ya estudiado y también se puede comprobar que sigue la distribución indicada utilizando el test de Kolmogorov-Smirnov.

7.1.2. GENERACIÓN DE VALORES DE VARIABLES ALEATORIAS DISCRETAS

Para generar números aleatorios procedentes de una distribución discreta que toma los valores a_i, se sigue el siguiente procedimiento:

— Se generan el número de valores buscados en la distribución $U[0, 1]$.

— Se calcula la función de distribución de la probabilidad buscada.

— Para cada valor uniforme u_i generado, $F(a_i) \leq u_i \leq F(a_{i+1})$, entonces, el valor generado es a_{i+1}.

7.1.3. GENERACIÓN DE VALORES DE VARIABLES ALEATORIAS CONTINUAS

Se pretenden generar números aleatorios procedentes de una variable aleatoria continua. Para ello, se tienen dos métodos de trabajo:

Método de la transformada inversa

Dada una variable aleatoria continua A con función de distribución F_A, se generan números aleatorios de una distribución $U[0, 1]$ y para cada valor generado, se

tiene el valor buscado mediante la función de distribución inversa de dicho valor: $a_i=F_A^{-1}(u_i)$.

Este método es bastante útil para las distribuciones continuas en las que es fácil obtener la inversa de la función de distribución, lo que no ocurre con la Normal.

Método de aceptación y rechazo

Sea A una variable aleatoria continua con función de densidad $f_A(a)$, que está dada por: $f_A(a) = c \cdot h(a)\ g(a)$, con $c \geq 1$, $h(a)$ función de densidad y $0 \leq g(a) \leq 1$. El algoritmo es el siguiente:

— Se generan valores procedentes de una variable aleatoria U con distribución $U[0,1]$

— Si el valor generado $u_i \leq g(v_i)$, se acepta el valor de V como valor de la variable A. En caso contrario, se rechaza el par (u_i,v_i) y se vuelve a generar.

Para que el método sea eficaz, hay que tener en cuenta que la variable aleatoria V sea fácil de generar.

Un caso particular de este método es la generación de la distribución Normal.

Generación de valores aleatorios de la distribución Normal

El primer paso es generar valores aleatorios de la $N(0, 1)$, para, a continuación pasarlos a la distribución Normal utilizada. Sea z una variable aleatoria con distribución $N(0, 1)$. Entonces, su función de densidad es:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Se aplica el método de aceptación y rechazo de la siguiente forma:

— La función de densidad de z se descompone como $f_Z(z) = c \cdot h(z)\ g(z)$, con $c \geq 1$, se toma:

$$h(z) = \frac{e^{-z}}{\left(1+e^{-z}\right)^2}$$

para cualquier valor real de z y $0 \leq g(z) \leq 1$.

— Se determina el valor de c a partir de $f_Z(z) \leq c \cdot h(z)$, es decir:

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2} \leq c \frac{e^{-z}}{\left(1+e^{-z}\right)^2}$$

que, desarrollada, equivale a:

$$\frac{-z^2}{2} + z + 2\ln\left(1+e^{-z}\right) \leq \ln\left(c\sqrt{2\pi}\right)$$

— Se calcula el máximo del primer miembro de la desigualdad, que se alcanza para $z = 0$, y vale $1{,}386294 = ln(4)$, con lo que se toma $c = 1{,}5958$.

— Mediante el método de la transformada inversa, se generan valores de la variable aleatoria con función de densidad $h(z)$, cuya función de distribución es:

$$H(z) = \frac{1}{1+e^{-z}}$$

— Se genera u_1^* de una distribución $U[0, 1]$.

— Entonces, $u_1^* = \frac{1}{1+e^{-z_1}}$ implica que $z_1 = -\ln\left(\frac{1-u_1^*}{u_1^*}\right)$.

— Se genera u_1 de una distribución $U[0, 1]$.

— Como $f_z(z) = c \cdot h(z)\, g(z)$, entonces $g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \frac{\left(1+e^{-z}\right)^2}{1{,}5958\, e^{-z}}$.

— Si el valor u_1 es menor que $g(z_1)$, se acepta z_1 como valor de Z. En caso contrario, no se acepta y se vuelve a empezar.

— Se repite el proceso hasta obtener el número de valores aleatorios deseados.

Sea A una variable aleatoria con distribución $N(\mu, \sigma)$. Para generar números aleatorios de dicha distribución, se generan para la $N(0, 1)$ y se pasan a la $N(\mu, \sigma)$ mediante la transformación $a_i = \mu + z_i\sigma$.

Generación de valores aleatorios de las distribuciones circulares

La generación de estos valores es bien sencilla. Para la uniforme circular se genera un valor aleatorio u de la distribución $U[0,1]$ y el valor de la distribución Uniforme circular es $2\pi u$.

Para la Normal Circular se genera un valor aleatorio z de la distribución $N(\mu, \sigma)$ para $\sigma = (-2\log\rho)^{\frac{1}{2}}$ y el valor de la distribución Normal circular (μ, ρ) es $z[mod\ 2\pi]$.

Método de Box-Muller

Otro método muy utilizado para generar valores de la distribución $N(0, 1)$ es el método de Box-Muller, que consiste en:

— Se generan dos valores u_1 y u_2 procedentes de una distribución $U(0, 1)$.

— Se obtienen dos valores de la distribución $N(0, 1)$:

$$z_1 = (-2\ ln\ u_1)^{1/2}\ cos\ 2\pi u_2$$
$$z_2 = (-2\ ln\ u_1)^{1/2}\ sen\ 2\pi u_2$$

— Se repite el proceso hasta obtener el número de valores normales deseados.

7.2. Método bootstrap para estimar errores cuadráticos medios

El bootstrap (o estimación autosuficiente) es un método de remuestreo que permite la estimación del error cuadrático medio con poco análisis, pero que, sin embargo, necesita un importante esfuerzo computacional, de manera que interviene la simulación para facilitar el cálculo. Además, este método es interesante en el análisis de resultados de la simulación.

Se considera una muestra $A_1,...,A_n$ independiente procedente de una variable aleatoria A con función de distribución de probabilidad F, que depende de un parámetro desconocido θ, que se pretende estimar mediante una función dependiente de la muestra $h(a_1,...,a_n)$. Para estudiar su bondad como estimador, se calcula su error cuadrático medio según:

$$ECM(F) = E_F[(h(a_1,...,a_n) - \theta_F)^2]$$

Si la distribución F fuera conocida, se podría calcular sin problemas el error cuadrático medio, pero dicha distribución es desconocida. Pero cuando el tamaño de la

muestra n es grande, la función de distribución F se puede aproximar mediante su función de distribución empírica, definida como:

$$F_n(a) = \frac{número\ de\ valores\ muestrales \leq a}{n}$$

Entonces, el parámetro $\theta_{F_n} = \theta_n$ está próximo a θ_F y el $ECM(F)$ se aproxima mediante la aproximación bootstrap del error cuadrático medio, dado por:

$$ECM(F_n) = E_{F_n}\left[(h(a_1^*,...,a_n^*) - \theta_n)^2\right]$$

para $(a_1^*,...,a_n^*)$ una muestra aleatoria simple de F_n (ver apartado 8, dedicado al muestreo).

Debido a la construcción de F_n, el vector $(a_1^*,...,a_n^*)$ tiene la misma probabilidad de tomar cualquiera de los n^n valores posibles $(a_{i_1},...,a_{i_n})$ para $i_j \in \{1,...,n\}$, $j=1,...,n$, por lo que:

$$ECM(F_n) = \sum_{i_n}\cdots\sum_{i_1}\frac{\left(h(a_{i_1},...,a_{i_n}) - \theta_n\right)^2}{n^n}$$

lo cual necesita la suma de n^n términos, que es imposible para n grande.

Pero se puede calcular el promedio de un gran número de términos mediante simulación, generando n variables aleatorias independientes $(a_{1_i}^*,..., a_{n_i}^*)$ cada una con distribución F_n, y después calcular:

$$b_1 = \left[h(a_{1_i}^*,...,a_{n_i}^*) - \theta_n\right]^2$$

Se repite el proceso, hasta obtener las variables $b_1,...,b_r$ ($r = 100$ se suele considerar suficiente para obtener una buena aproximación). Entonces, se puede estimar el error cuadrático medio mediante el promedio de las variables b:

$$\hat{ECM}(F_n) = \frac{\sum_{i=1}^{r} b_i}{r}$$

8. MUESTREO

La recogida de datos es un punto importante de cualquier trabajo a realizar y la Teoría de Muestras se encarga de dar una metodología para guiar los problemas de recogida de información. Tiene dos aspectos fundamentales: a) un proceso de selección, mediante el cual se incluyen en la muestra unos miembros de la población, y b) un proceso de estimación para calcular ciertas características de la población. Como no se estudia toda la población, se comete un riesgo al pasar de lo particular a lo general. Si N es el tamaño de la población y n es el tamaño de la muestra, $f = n / N$ se denomina fracción de muestreo.

Se analizarán aquí distintos métodos de muestreo y, en cada uno de ellos, se procederá al estudio de tres tipos de parámetros: el total poblacional de una característica (A), su media poblacional ($\overline{A}$) y la proporción de la población que está en un grupo determinado (π). La desviación típica muestral será S.

Es de interés conocer el intervalo de confianza para cualquier parámetro C bajo estudio. Este intervalo se obtiene a partir del estimador del parámetro más menos la desviación típica poblacional multiplicada por una constante k, la cual indica el nivel de confianza del intervalo. La constante k vale $z_{\alpha/2}$ si $n > 60$ y valdrá $t_{n-1,\alpha/2}$ si $n < 60$.

También se va a tratar el tamaño muestral a utilizar en el muestreo, dependiendo del parámetro elegido, para un error máximo admisible ε y un nivel de confianza $1-\alpha$. En su cálculo se utilizarán los valores poblacionales de la varianza (S^2) y la proporción (π), que habrá que conjeturar a partir del conocimiento que se tenga de la población en una fecha anterior, o de poblaciones parecidas a ésta o bien de una muestra piloto previamente seleccionada.

El muestreo es una técnica de gran importancia en la Estadística y de gran utilidad práctica, existen numerosos y extensos manuales dedicados a esta parte de la Estadística, entre ellos destacar Azorín (1994), Pérez (1999) y Cid (1999) desde el punto de vista teórico y Rueda (1992) desde el punto de vista práctico, pues incluye bastantes problemas resueltos.

Además, en este apartado se van a presentar los planes de muestreo, que no son más que métodos estandarizados que han sido adoptados por la industria en el control de la aceptación de suministros por lotes, es decir, para decidir si se acepta o no la mercancía que se pretende adquirir de un suministrador. Desde esta perspectiva también existen numerosos manuales, entre ellos destacar el ya clásico tratado de Hansen (1990).

8.1. Muestreo con probabilidades iguales

Es el método base de los muestreos, puesto que los conceptos y fórmulas importantes de los otros tipos de muestreo se derivan de éste y en él todas las unidades tienen la misma probabilidad de ser elegidas en cada selección.

En el muestreo de poblaciones finitas, cuando la muestra se obtiene unidad a unidad sin reposición de éstas a la población después de cada selección, se tiene el muestreo aleatorio simple (sin reemplazamiento). En él, todas las unidades tienen la misma probabilidad de selección, todas las muestras son equiprobables y las muestras con los mismos elementos colocados en orden distinto se consideran iguales, por lo que son imposibles las muestras con elementos distintos.

Cuando la unidad seleccionada se repone en la población, se tiene el muestreo con reemplazamiento, de manera que son posibles las muestras con elementos repetidos.

8.1.1. MUESTREO ALEATORIO SIMPLE

Dada una población finita, se obtiene una muestra unidad a unidad y sin reposición. En este caso todas las posibles muestras tienen la misma probabilidad de ser elegidas. Los intervalos de confianza para los parámetros considerados son los que se presentan en la Tabla A.12.

Tabla A.12 - Intervalos de confianza y tamaños para totales, medias y proporciones en muestreos sin reemplazamiento

	Intervalo de confianza	Tamaño muestra
Total	$\left(N\bar{a} \pm k\frac{Ns}{\sqrt{n}}\sqrt{1-f}\right)$	$n=\frac{n_0}{1+\frac{n_0}{N}}$ donde: $n_0=\frac{N^2k^2S^2}{\varepsilon^2}$
Media	$\left(\bar{a} \pm k\frac{s}{\sqrt{n}}\sqrt{1-f}\right)$	$n=\frac{n_0}{1+\frac{n_0}{N}}$ donde: $n_0=\frac{k^2S^2}{\varepsilon^2}$
Proporción	$\left(\hat{\pi} \pm k\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n-1}(1-f)}\right)$	$n=\frac{\frac{N}{N-1}n_0}{1+\frac{n_0}{N-1}}$ donde: $n_0=\frac{k^2\pi(1-\pi)}{\varepsilon^2}$

8.1.2. MUESTREO CON REEMPLAZAMIENTO

Se parte de una población finita de la que se obtiene una muestra, unidad a unidad, pero con reposición de cada unidad a la población después de cada extracción. En este caso se dice que el muestreo es con reemplazamiento. La Tabla A.13 presenta las fórmulas que permiten determinar el tamaño de muestreo y los intervalos de confianza para los totales, medias y proporciones estimadas.

Tabla A.13 - Intervalos de confianza y tamaños para totales, medias y proporciones en muestreos con reemplazamiento

	Intervalo de confianza	Tamaño muestra
Total	$\left(N\bar{a} \pm k\frac{Ns}{\sqrt{n}}\right)$	$n = \frac{N^2 k^2 S^2}{\varepsilon^2}$
Media	$\left(\bar{a} \pm k\frac{s}{\sqrt{n}}\right)$	$n = \frac{k^2 S^2}{\varepsilon^2}$
Proporción	$\left(\hat{\pi} \pm k\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n-1}}\right)$	$n = \frac{k^2 \pi(1-\pi)}{\varepsilon^2}$

8.2. Muestreo estratificado

Este método es adecuado cuando se desea asegurar la presencia de ciertas clases en el muestreo. Supone un cierto conocimiento previo de la estructura de la población.

Se divide la población heterogénea de N unidades en L subpoblaciones lo más homogéneas posibles, llamadas estratos, cada cual de tamaño N_h con $h = 1, ..., L$. Los estratos no se superponen y, en su conjunto, forman la totalidad de la población. En cada estrato, de manera independiente[1], se extrae una muestra de tamaño n_h.

Cada estrato tiene un peso o ponderación dado por $W_h=N_h/N$ y una fracción de muestreo dada por $f_h=n_h/N_h$, y sus respectivas medias poblacionales, medias muestrales, varianzas poblacionales y varianzas muestrales.

Este método se utiliza porque, en ocasiones, puede interesar disponer de información de algunas subdivisiones de la población, o bien, si la población es heterogénea, se puede subdividir en subpoblaciones lo más homogéneas posibles, de manera que en estos estratos se puedan obtener estimadores precisos de los parámetros de los estratos, para combinarse en un estimador de toda la población.

[1] Esta independencia permite la aplicación simultánea de métodos de muestreo diferentes en cada estrato.

En este caso los tamaños de las muestras de los estratos son elegidos por el diseñador del muestreo, de manera que la suma de todos los tamaños muestrales sea n. Las cuatro formas más usuales de repartir la muestra son:

1. Afijación uniforme: Todos los n_h son iguales: $n_h = n/L$.

2. Afijación proporcional: El valor de cada n_h es proporcional al tamaño del estrato, por lo que todas las unidades tienen la misma probabilidad de ser seleccionadas en la muestra: $n_h = N_h \cdot cte$.

3. Afijación de mínima varianza: Los n_h se eligen de manera que se minimice la varianza para n fijo, esto se consigue con el siguiente valor:

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} n$$

4. Afijación óptima: Los n_h se eligen de manera que se minimice la varianza para un coste fijo. En este caso es necesario definir una función de coste, por ejemplo:

$$C = C_0 + \sum_{h=1}^{L} n_h C_h$$

donde C_h es el coste de elegir una unidad en el estrato h, en este caso se tiene:

$$n_h = \frac{N_h S_h \frac{1}{\sqrt{C_h}}}{\sum_{h=1}^{L} N_h S_h \frac{1}{\sqrt{C_h}}} \qquad n = \frac{(C - C_0) W_h S_h \frac{1}{\sqrt{C_h}}}{\sum_{h=1}^{L} W_h S_h \sqrt{C_h}} n$$

Si se dispone de una sola variable, la mejor característica para la construcción de los estratos es la distribución de frecuencias de la variable estudiada, y para determinar el número de estratos se utiliza la expresión:

$$L = 2n \frac{C_u}{C_e}$$

donde:

C_u Coste por unidad.
C_e Coste por estrato.

La Tabla A.14 presenta las fórmulas que permiten determinar el tamaño de muestreo y los intervalos de confianza para los totales, medias y proporciones estimadas.

Tabla A.14 - Intervalos de confianza y tamaños para totales, medias y proporciones en muestreos estratificados

	Intervalo de confianza *	Tamaño de muestra (**)
Total	$\left(\sum_{h=1}^{L} N_h \bar{a}_h \pm z_{\alpha/2}\sqrt{\sum_{h=1}^{L} N_h(N_h - n_h)\frac{s_h^2}{n_h}}\right)$	$n = \frac{\sum_{h=1}^{L} \frac{N_h^2 S_h^2}{w_h}}{\frac{\varepsilon^2}{k^2} + \sum_{h=1}^{L} N_h S_h^2}$
Media	$\left(\frac{1}{N}\sum_{h=1}^{L} N_h \bar{a}_h \pm z_{\alpha/2}\sqrt{\frac{1}{N^2}\sum_{h=1}^{L} N_h(N_h - n_h)\frac{s_h^2}{n_h}}\right)$	$n = \frac{\sum_{h=1}^{L} \frac{W_h^2 S_h^2}{w_h}}{\frac{\varepsilon^2}{k^2} + \sum_{h=1}^{L} W_h S_h^2}$
Proporción	$\left(\frac{1}{N}\sum_{h=1}^{L} N_h \hat{\pi}_h \pm z_{\alpha/2}\sqrt{\sum_{h=1}^{L} W_h^2 (1 - f_h)\frac{\hat{\pi}_h(1-\hat{\pi}_h)}{n_h - 1}}\right)$	$n = \frac{\sum_{h=1}^{L} W_h^2 \frac{N_h}{N_h - 1}\frac{\pi_h(1-\pi_h)}{w_h}}{\frac{\varepsilon^2}{k^2} + \frac{1}{N^2}\sum_{h=1}^{L} \frac{N_h}{N_h - 1}\pi_h(1-\pi_h)}$

NOTA: (*) Si el tamaño de la muestra es suficientemente grande; (**) Se considera que $w_h = n_h/n$.

8.3. Muestreo por conglomerados

Hasta ahora, las unidades de muestreo eran las mismas que constituían el objeto de estudio, pero en este caso las unidades de muestreo comprenden dos o más unidades últimas, de manera que cada unidad de muestreo constituye un conglomerado de unidades últimas. Este método se utiliza profusamente porque en los muestreos anteriormente analizados es necesario disponer de una lista de todos los elementos de la población y, en la práctica, no se suele disponer de dicha lista. Entonces, es preferible la división previa de la población en conglomerados, de los cuales se selecciona cierto número, para lo cual sólo se necesita disponer de la lista de los conglomerados. La muestra, una vez seleccionados algunos conglomerados, está formada por todas las unidades que componen el conglomerado. Por tanto, para que la muestra sea representativa, los elementos que componen el conglomerado deben ser muy heterogéneos entre sí, aspecto de gran diferencia respecto al método estratificado. Una práctica común es seleccionar y medir una muestra en cada conglomerado seleccionado (muestreo bietápico), que puede extenderse a más etapas (muestreo polietápico). Existe gran diversidad de situaciones en el muestreo por conglomerados: en algunos casos los conglomerados son conocidos y de tamaños iguales, en otros de tamaños desiguales y, frecuentemente son desconocidos los tamaños.

La notación que se va a utilizar es la siguiente:

N	Número de conglomerados de la población.
n	Número de conglomerados seleccionados en la muestra.
M_i	Número de elementos en el conglomerado i.
a_{ij}	Valor de la variable en la unidad j del conglomerado i.
B_i	Número de individuos en el conglomerado i con la característica deseada.
π_i	Proporción de individuos en el conglomerado i con la característica deseada.
π	Proporción poblacional de individuos con la característica deseada.
$\overline{m} = \frac{1}{n}\sum_{i=1}^{n} M_i$	Tamaño promedio del conglomerado en la muestra.
$M = \sum_{i=1}^{N} M_i$	Número de elementos de la población.
$\overline{M} = \frac{M}{N}$	Tamaño promedio del conglomerado en la población.
$A_i = \sum_{j=1}^{M_i} A_{ij}$	Total de la variable en el conglomerado i.
$A = \sum_{i=1}^{N} A_i$	Total poblacional de la variable.
$\overline{A}_i = \sum_{j=1}^{M_i} \frac{A_{ij}}{M_i}$	Media de la variable en el conglomerado i.
$\overline{A} = \frac{1}{M} A$	Media poblacional de la variable.
$S_i^2 = \frac{1}{M_i - 1}\sum_{j=1}^{M_i} (a_{ij} - \overline{A}_i)^2$	Varianza en el conglomerado i.
$S^2 = \frac{1}{M-1}\sum_{i=1}^{N}\sum_{j=1}^{M_i} (a_{ij} - \overline{A})^2$	Varianza poblacional.
$S_b^2 = \frac{1}{N-1}\sum_{i=1}^{N}\sum_{j=1}^{M_i} (\overline{A}_i - \overline{A})^2$	Varianza entre las medias de los conglomerados.

Se pueden distinguir dos tipos de muestreo por conglomerados:

— Unietápico, en el que se seleccionan los conglomerados recogiendo la información del 100% de las unidades elementales contenidas en cada conglomerado.

— Bietápico, en el que una vez seleccionados los conglomerados, se realiza una segunda selección sobre los elementos pertenecientes a cada uno de ellos y la muestra total está formada por las muestras obtenidas en cada conglomerado.

8.3.1. MUESTREO POR CONGLOMERADOS EN UNA ETAPA

En el muestreo por conglomerados en una etapa se van a plantear los siguientes casos:

— Conglomerados del mismo tamaño, con probabilidades iguales y sin reemplazamiento, que es el caso más general.

— Conglomerados de distinto tamaño, con probabilidades iguales y sin reemplazamiento, que se utilizan cuando no es posible decidir cuál es el tamaño de los conglomerados y éstos tienen distinto número de unidades o no es conveniente tomarlos todos del mismo tamaño.

— Conglomerados con probabilidades desiguales y con reemplazamiento, que se utilizan cuando las probabilidades de los conglomerados son distintas.

— Conglomerados combinados con estratificación, que se utilizan cuando los conglomerados no tienen tamaño reducido ni medias semejantes.

Conglomerados del mismo tamaño con probabilidades iguales y sin reemplazamiento

En este caso, los estimadores de la media y la proporción poblacional son:

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} \bar{A}_i \qquad \hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} \pi_i$$

Entonces, los intervalos de confianza, si el tamaño de la muestra es suficientemente grande, para los parámetros considerados, son los que aparecen en la Tabla A.15.

Tabla A.15 - Intervalos de confianza para un Muestreo por conglomerados del mismo tamaño, con probabilidades iguales y sin reemplazamiento

	Intervalo de confianza
Total	$\left(N\bar{M}\sum_{i=1}^{n} \bar{A}_i \pm z_{\alpha/2} \sqrt{N^2 \bar{M}^2 \frac{1-f}{n\bar{M}^2} \frac{\sum_{i=1}^{n} A_i^2 - n\bar{M}^2\bar{a}^2}{n-1}} \right)$
Media	$\left(\frac{1}{n}\sum_{i=1}^{n} \bar{A}_i \pm z_{\alpha/2} \sqrt{\frac{1-f}{n\bar{M}^2} \frac{\sum_{i=1}^{n} A_i^2 - n\bar{M}^2\bar{a}^2}{n-1}} \right)$
Proporción	$\left(\frac{1}{n}\sum_{i=1}^{n} \pi_i \pm z_{\alpha/2} \sqrt{\frac{1-f}{n} \frac{\sum_{i=1}^{n} \pi_i^2 - n\hat{\pi}^2}{n-1}} \right)$

Conglomerados de distinto tamaño con probabilidades iguales y sin reemplazo

En el caso de conglomerados de distinto tamaño, se tiene que los estimadores de la media y proporción poblacionales son:

$$\bar{a} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} M_i} \qquad \hat{\pi} = \frac{\sum_{i=1}^{n} B_i}{\sum_{i=1}^{n} M_i}$$

Entonces, los intervalos de confianza, si el tamaño de la muestra es suficientemente grande, para los parámetros considerados son los que se presentan en la Tabla A.16.

Tabla A.16 - Intervalos de confianza para un Muestreo por conglomerados de distinto tamaño, con probabilidades iguales y sin reemplazamiento

	Intervalo de confianza
Total	Tamaño población conocido $\left(M\bar{a} \pm z_{\alpha/2} M \sqrt{\frac{N-n}{Nn\bar{M}^2} \frac{1}{n-1} \sum_{i=1}^{n} (\bar{A}_i - \bar{a}M_i)^2} \right)$ Tamaño población desconocido $\left(N \frac{\sum_{i=1}^{n} A_i}{n} \pm z_{\alpha/2} \sqrt{N^2 \frac{N-n}{Nn} \frac{1}{n-1} \left(\sum_{i=1}^{n} A_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} A_i \right)^2 \right)} \right)$
Media	$\left(\bar{a} \pm z_{\alpha/2} \sqrt{\frac{N-n}{Nn\bar{M}^2} \frac{1}{n-1} \sum_{i=1}^{n} (\bar{A}_i - \bar{a}M_i)^2} \right)$
Proporción	$\left(\frac{1}{n} \sum_{i=1}^{n} \pi_i \pm z_{\alpha/2} \sqrt{\frac{N-n}{Nn\bar{M}^2} \frac{1}{n-1} \sum_{i=1}^{n} (B_i - \hat{\pi}M_i)^2} \right)$

Extracción de los conglomerados con probabilidades desiguales

En este caso se extrae una muestra de n conglomerados con reemplazo, donde en cada extracción el conglomerado i se extrae con probabilidad p_i, cumpliéndose:

$$\sum_{i=1}^{n} p_i = 1$$

Si los estimadores del total, media y proporción poblacional son:

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{p_i}$$

$$\bar{a} = \frac{1}{M}\sum_{i=1}^{n}\frac{A_i}{np_i}$$

$$\hat{\pi} = \frac{1}{M}\sum_{i=1}^{n}\frac{B_i}{np_i}$$

entonces, los intervalos de confianza, si el tamaño de la muestra es suficientemente grande, para los parámetros considerados son los que presenta la Tabla A.17.

Tabla A.17 - Intervalos de confianza para un Muestreo por conglomerados combinado con estratificación

	Intervalo de confianza
Total	$\left(\hat{A} \pm z_{\alpha/2}\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{A_i}{p_i} - \hat{A}\right)^2}\right)$
Media	$\left(\bar{a} \pm z_{\alpha/2}\frac{1}{M}\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{A_i}{p_i} - \hat{A}\right)^2}\right)$
Proporción	$\left(\hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{1}{M^2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{B_i}{p_i} - M\hat{\pi}\right)^2}\right)$

Muestreo por conglomerados combinado con estratificación

En los casos estudiados son deseables los conglomerados de tamaño reducido y de medias semejantes, lo cual es difícil de respetar en la práctica. Por ello se estratifican los conglomerados, incluyendo en un mismo estrato los más homogéneos. Dentro de cada estrato, el muestreo por conglomerados da buenas estimaciones. Se consideran L estratos, siendo los estimadores de la media y proporción poblacional:

$$\bar{a}_{RC} = \frac{\sum_{k=1}^{L} N_K \bar{a}_{kt}}{\sum_{k=1}^{L} N_K \bar{m}_k} \qquad \hat{\pi}_{RC} = \frac{\sum_{k=1}^{L} N_k \bar{B}_{kt}}{\sum_{k=1}^{L} N_k \bar{m}_k}$$

donde:

$\bar{a}_{kl} = \dfrac{\sum_{i=1}^{n_k} A_{ki}}{n_k}$ Media de los totales de los conglomerados en la muestra para el estrato k

$\bar{m}_k = \dfrac{\sum_{i=1}^{n_k} M_{ki}}{n_k}$ Media de los tamaños de los conglomerados en el estrato k

$\bar{B}_{kl} = \dfrac{\sum_{i=1}^{n_k} B_{kl}}{n_k}$ Media de individuos por conglomerado con la característica deseada para el estrato k.

Los intervalos de confianza, si el tamaño de la muestra es suficientemente grande, son los que se presentan en la Tabla A.18.

Tabla A.18 - Intervalos de confianza para un Muestreo por conglomerados combinado con estratificación

	Intervalo de confianza
Total	$\left(M\bar{a}_{RC} \pm z_{\alpha/2} \sqrt{\sum_{k=1}^{L} \left(\frac{N_k (N_k - n_k)}{n_k (n_k - 1)} \sum_{i=1}^{n_k} \left((A_{ki} - \bar{a}_{kl}) - \bar{a}_{RC} (M_{ki} - \bar{m}_k) \right)^2 \right)} \right)$
Media	$\left(\bar{a}_{RC} \pm z_{\alpha/2} \frac{1}{M} \sqrt{\sum_{k=1}^{L} \left(\frac{N_k (N_k - n_k)}{n_k (n_k - 1)} \sum_{i=1}^{n_k} \left((A_{ki} - \bar{a}_{kl}) - \bar{a}_{RC} (M_{ki} - \bar{m}_k) \right)^2 \right)} \right)$
Proporción	$\left(\hat{\pi}_{RC} \pm z_{\alpha/2} \frac{1}{M} \sqrt{\sum_{k=1}^{L} \left(\frac{N_k (N_k - n_k)}{n_k (n_k - 1)} \sum_{i=1}^{n_k} \left((B_{ki} - \bar{B}_{kl}) - \hat{\pi}_{RC} (M_{ki} - \bar{m}_k) \right)^2 \right)} \right)$

8.3.2. Muestreo bietápico o por conglomerados con submuestreo

En este caso se seleccionan en primer lugar n conglomerados, y dentro de cada conglomerado se selecciona una muestra de tamaño m_i. La media de cada conglomerado es:

$$\bar{a}_i = \frac{1}{m_i} \sum_{i=1}^{m_i} a_{ij}$$

En el caso de hacer una selección con probabilidades iguales en cada etapa, hay dos opciones: a) que los tamaños de los conglomerados sean iguales, b) que los tamaños no sean iguales.

Si son iguales, se tiene que $M_i = \overline{M}$ y $m_i = \overline{m}$, y los estimadores de la media y proporción poblacional son:

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} \bar{a}_i \qquad \hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} \pi_i$$

donde:

$\pi_i = \frac{1}{m_i}\sum_{i=1}^{m_i} B_{ij}$ Proporción muestral de cada conglomerado.

Entonces, los intervalos de confianza, si el tamaño de la muestra es suficientemente grande, para los parámetros considerados son los que presenta la Tabla A.19.

Tabla A.19 - Intervalos de confianza para un Muestreo bietápico por conglomerados con submuestreo con probabilidades iguales en cada etapa y tamaños de conglomerados iguales

	Intervalo de confianza
Total	$\left(M\bar{a} \pm z_{\alpha/2} M \sqrt{\left(1-\frac{N-n}{N}\right)\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\bar{a}_i-\bar{a})^2}{n} + \frac{N-n}{N}\left(1-\frac{\overline{M}-\overline{m}}{\overline{M}}\right)\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{\overline{m}}\frac{(a_{ij}-\bar{a}_i)^2}{\overline{m}-1}}{n\overline{m}}} \right)$
Media	$\left(\bar{a} \pm z_{\alpha/2} \sqrt{\left(1-\frac{N-n}{N}\right)\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\bar{a}_i-\bar{a})^2}{n} + \frac{N-n}{N}\left(1-\frac{\overline{M}-\overline{m}}{\overline{M}}\right)\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{\overline{m}}\frac{(a_{ij}-\bar{a}_i)^2}{\overline{m}-1}}{n\overline{m}}} \right)$
Proporción	$\left(\hat{\pi} \pm z_{\alpha/2} \sqrt{\frac{1-\frac{N-n}{N}}{n}\sum_{i=1}^{n}\frac{(\pi_i-\hat{\pi})^2}{n-1} + \frac{\frac{N-n}{N}\left(1-\frac{\overline{M}-\overline{m}}{\overline{M}}\right)}{n^2(\overline{m}-1)}\sum_{i=1}^{n}\pi_i(1-\pi_i)} \right)$

Si los tamaños de los conglomerados son desiguales, para los casos de media y proporción cabe distinguirse si se conoce o no el tamaño de la población. Para el caso de tamaño conocido, los estimadores de la media y proporción poblacional son:

$$\bar{a} = \frac{N}{M}\sum_{i=1}^{n}\frac{M_i\bar{a}_i}{n} \qquad \hat{\pi} = \frac{N}{M}\sum_{i=1}^{n}\frac{M_i\pi_i}{n}$$

Si el tamaño es desconocido, ahora se utilizan los siguientes estimadores de la media y proporción poblacional:

$$\bar{a}_t = \frac{\sum_{i=1}^{n} M_i \bar{a}_i}{\sum_{i=1}^{n} M_i} \qquad \hat{\pi} = \frac{\sum_{i=1}^{n} M_i \pi_i}{\sum_{i=1}^{n} M_i}$$

Los intervalos de confianza, si el tamaño de la muestra es suficientemente grande, para los parámetros son los que se recogen en la Tabla A.20.

Tabla A.20 - Intervalos de confianza para un Muestreo bietápico por conglomerados con submuestreo con probabilidades iguales en cada etapa y tamaños de conglomerados desiguales

	Intervalo de confianza
Total	$\left(N\sum_{i=1}^{n} \frac{M_i \bar{a}_i}{n} \pm z_{\alpha/2} \sqrt{\frac{N-n}{N}\frac{N^2}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{a}_i - \overline{M\bar{a}}\right)^2 + \frac{N}{n}\sum_{i=1}^{n} M_i^2 \frac{M_i - m_i}{M_i}\frac{\frac{1}{m_i-1}\sum_{j=1}^{m_i}(a_{ij}-\bar{a}_i)^2}{m_i}} \right)$
Media	Conocido el tamaño de la población $\left(\bar{a} \pm z_{\alpha/2} \sqrt{\frac{N-n}{N}\frac{1}{n\bar{M}^2}\frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\bar{a}_i - \overline{M\bar{a}}\right)^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2 \frac{M_i - m_i}{M_i}\frac{\frac{1}{m_i-1}\sum_{j=1}^{m_i}(a_{ij}-\bar{a}_i)^2}{m_i}} \right)$ Desconocido el tamaño de la población $\left(\bar{a}_t \pm z_{\alpha/2} \sqrt{\frac{N-n}{N}\frac{1}{n\bar{M}^2}\frac{1}{n-1}\sum_{i=1}^{n} M_i^2\left(\bar{a}_i - \bar{a}_t\right)^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2 \frac{M_i - m_i}{M_i}\frac{\frac{1}{m_i-1}\sum_{j=1}^{m_i}(a_{ij}-\bar{a}_i)^2}{m_i}} \right)$
Proporción	Conocido el tamaño de la población $\left(\hat{\pi} \pm z_{\alpha/2} \sqrt{\frac{N-n}{N}\frac{1}{n\bar{M}^2}\frac{1}{n-1}\sum_{i=1}^{n}\left(M_i\pi_i - \overline{M\hat{\pi}}\right)^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} \frac{M_i - m_i}{M_i}\frac{\pi_i(1-\pi_i)}{m_i - 1}} \right)$ Desconocido el tamaño de la población $\left(\hat{\pi} \pm z_{\alpha/2} \sqrt{\frac{N-n}{N}\frac{1}{n\bar{M}^2}\frac{1}{n-1}\sum_{i=1}^{n} M_i^2\left(\pi_i - \hat{\pi}\right)^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} \frac{M_i - m_i}{M_i}\frac{\pi_i(1-\pi_i)}{m_i - 1}} \right)$

Selección de las unidades con probabilidades desiguales en la primera etapa y con probabilidades iguales en la segunda

Se seleccionan las unidades primarias con probabilidades desiguales y con reemplazo. La submuestra de m_i unidades se selecciona en cada conglomerado con un muestreo aleatorio simple. Sea p_i la probabilidad de seleccionar el conglomerado i, cumpliéndose que:

$$\sum_{i=1}^{N} p_i = 1$$

Los estimadores del total, media y proporción poblacional son:

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i\bar{a}_i}{p_i} \qquad \bar{a} = \frac{1}{nM}\sum_{i=1}^{n}\frac{M_i\bar{a}_i}{p_i} \qquad \hat{\pi} = \frac{1}{nM}\sum_{i=1}^{n}M_i\frac{\pi_i}{p_i}$$

Entonces, los intervalos de confianza, si el tamaño de la muestra es suficientemente grande, para los parámetros considerados son los que se presentan en la Tabla A.21.

Tabla A.21 - Intervalos de confianza para un Muestreo bietápico con probabilidades desiguales en la primera etapa y con probabilidades iguales en la segunda

	Intervalo de confianza
Total	$\left(\hat{A} \pm z_{\alpha/2}\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{M_i\bar{a}_i}{p_i} - \hat{A}\right)^2}\right)$
Media	$\left(\bar{a} \pm z_{\alpha/2}\sqrt{\frac{1}{n(n-1)M^2}\sum_{i=1}^{n}\left(\frac{M_i\bar{a}_i}{p_i} - \hat{A}\right)^2}\right)$
Proporción	$\left(\hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{1}{n(n-1)M^2}\sum_{i=1}^{n}\left(\frac{M_i\pi_i}{p_i} - \hat{\pi}\right)^2}\right)$

8.4. Planes de muestreo

En este apartado se van a presentar las ideas básicas de los muestreos de aceptación por atributos, lote a lote, y los procedimientos de aceptación basados en el Nivel de Calidad Aceptable (*NCA* o *AQL*, en inglés). Los planes de muestreo no son mas que métodos estandarizados de muestreo que han sido adoptados por la industria

para el control en la aceptación de lotes de mercancía suministrada. Por ejemplo, una institución cartográfica puede aplicar estos métodos para aceptar/rechazar un trabajo de imprenta, de restitución, de escaneo de fotogramas, etc.

Los muestreos de aceptación por atributos, lote a lote, se utilizan cuando se revisa de manera estática un número múltiple de unidades. Los planes de muestreo para la aceptación lote a lote consisten en la elección de una muestra de tamaño n y en determinar una cantidad de aceptación Ac, que asegure la calidad que se pretende con cierto nivel de confianza. La regla de decisión es que se acepta el lote si hay Ac o menos unidades defectuosas en la muestra tomada del mismo.

Las bases analíticas del método están en la distribución de Poisson y sólo hace falta considerar un número de unidades defectuosas en un lote de tamaño dado.

Para conocer los riesgos de productor y usuario que utilizan este plan de muestreo se introduce la curva operativa (curva *OC* en inglés), que permite determinar la probabilidad de aceptación de un lote cualquiera. Su construcción es fácil, basta situar en una tabla los siguientes valores:

— *Pa*: Es la probabilidad de aceptación del lote. Los valores más comunes son: 0,98, 0,95, 0,70, 0,50, 0,20, 0,05 y 0,02.

— np': Es el número de unidades defectuosas.

— p': Es la proporción de unidades defectuosas.

— $p'Pa$: Es la calidad media de salida (*CMS* o *AOQ* en inglés).

La determinación de todos ellos se puede realizar por medio de valores tabulados. Así, la Tabla 15 del Apéndice A permite que, a partir de un valor de Ac considerado (c en la tabla), al llegar al Pa buscado se obtenga el valor de np' asociado, del que, conocido n, se podrá calcular el valor de p'. El valor p' así obtenido es la abscisa que permite obtener la ordenada, Pa, en la curva operativa *OC*, mientras que el producto $p'Pa$ es la ordenada en la curva *CMS*.

Los puntos más importantes de la curva operativa y la curva *CMS* son:

— Riesgo del productor (α): Probabilidad de que se le rechace al fabricante o productor un producto correcto. Suele estar establecida en $Pa = 0,95$ ($\alpha = 0,05$). El punto viene dado por p'_{α}.

— Riesgo del consumidor (β): Probabilidad de que el consumidor o usuario acepte un producto defectuoso tomándolo como bueno. Suele estar establecida en $Pa = 0,10$ ($\beta = 0,10$). El punto viene dado por p'_{β}.

— Límite de calidad media de salida (*LCMS* o *AOQL* en inglés): Es el promedio máximo de calidad de salida en la aplicación del plan. El punto viene dado por p'_{LCMS}.

— Punto de indiferencia: Marca la calidad de indiferencia, es decir, donde el fabricante y consumidor tienen los mismos riesgos. El punto viene dado por $p'_{0,5}$.

También se puede realizar un plan de muestreo en el que no estén determinados ni el tamaño de la muestra ni la cantidad para la aceptación a partir de los puntos α y β de la curva operativa. Si se establece el riesgo del productor de manera que los lotes analizados de calidad defectuosa p_1 se rechacen el 100α % de los casos, y si se establece el riesgo del consumidor de manera que los lotes defectuosos de calidad defectuosa p_2 no se acepten más del 100β % de los casos, se puede realizar un plan de muestreo de atributos. En este caso la probabilidad de aceptación es $1-\alpha$ y la Tabla 16 del Apéndice A da los valores *np* para obtener la probabilidad de aceptación para distintas cantidades críticas *Ac*. Para diversas cantidades críticas se calcula el valor:

$$R = \frac{np \text{ para la probabilidad de aceptación } \beta}{np \text{ para la probabilidad de aceptación } \alpha}$$

mientras que el cociente deseado es:

$$R = \frac{p_2}{p_1}$$

Se utiliza el valor de *Ac* que da un *R* más próximo al buscado. El tamaño muestral en el punto de riesgo del producto es:

$$n = \frac{np \text{ para la probabilidad de aceptación } \alpha}{p_1}$$

el tamaño muestral en el punto de riesgo del consumidor es:

$$n = \frac{np \text{ para la probabilidad de aceptación } \beta}{p_2}$$

y se acepta un tamaño muestral promedio de los dos. Una vez obtenidos *n* y *Ac*, se tiene el plan de muestreo buscado y se pueden representar sus correspondientes curvas.

En general, cuanto mayor sea la muestra, menor será el riesgo para el consumidor, mientras que permitir más elementos defectuosos da menor riesgo de rechazo para el fabricante.

8.4.1. PLANES DE ACEPTACIÓN POR MUESTREO PARA ATRIBUTOS

Realmente, la elección de un plan de muestreo es una decisión para adoptar una curva operativa, pues sus tablas están diseñadas para concordar en un punto decisivo de la curva operativa y para aproximarse lo máximo posible en los demás puntos. En cualquier curva operativa hay tres puntos importantes y se pueden crear tablas independientes para destacar cada uno de dichos puntos:

— Primer punto (q_1), en el que la probabilidad de aceptación o error tipo II es $1-\alpha$, donde α es generalmente 0,05. La calidad q_1 se puede designar como nivel de calidad aceptable y todos los procedimientos que destaquen este punto se dice que están basados en *NCA*.

— Segundo punto (q_2), en el que la probabilidad de aceptación es 0,5. La calidad q_2 se puede definir como nivel de calidad indiferente y los procedimientos que destaquen este punto se dice que se basan en indiferencia de calidad.

— Tercer punto (q_3), en el que la probabilidad de aceptación es 0,1. La calidad q_3 se le llama calidad límite, que es el porcentaje máximo de defectuosos que se pueden admitir en un lote y los procedimientos que destaquen este punto se dice que se basan en *CL* (calidad límite).

Procedimientos de aceptación basados en *NCA*

El *NCA* o *AQL* es el Nivel de Calidad Aceptable que el consumidor considera y cuando un proceso de aceptación se basa en él, el resultado es una clara decisión de aceptación o rechazo del lote, sin inspección rectificadora. Dada su comodidad y robustez, estos métodos se han convertido en la norma para procedimientos de inspección en todos los sectores industriales. Los procedimientos basados en el NCA se pueden utilizar en inspecciones de aceptación por atributos, generalmente, lote a lote.

En primer lugar se debe indicar el nivel de calidad aceptable, de manera que el productor no tenga un riesgo demasiado alto de rechazo de productos correctos, y la proporción de productos recibidos que se van a inspeccionar. Seguidamente, se debe especificar de forma clara el concepto de elemento defectuoso y, finalmente, la cantidad relativa de inspección (si el lote tiene N elementos y se revisan n, la fracción de productos revisada es n/N). Para un valor constante del riesgo del productor, la

fracción disminuye si *N* aumenta; si *N* es constante, la fracción crece a medida que disminuye el riesgo del consumidor. Además, otro factor que afecta a la cantidad relativa de inspección es el coste de dicha inspección.

La Norma UNE 66020-1:2000 (equivalente a la ISO 2859-1:1999) establece los procedimientos de muestreo para la inspección por atributos, lote a lote, en función del NCA. Estas normas se basan en la MIL-STD-105E, desarrollada por el ejército de los EE.UU. durante la Segunda Guerra Mundial y ampliada y desarrollada posteriormente para el mundo civil, es la pionera en el ámbito del control estadístico y está basada en el uso del *NCA*, que lo define como el valor nominal, expresado en porcentaje de elementos defectuosos especificado para determinado grupo de defectos, es decir, el nivel de calidad en el que el fabricante asume un riesgo comparativamente pequeño de que su producto de calidad igual o mejor que el *NCA* sea rechazado.

El nivel de inspección indica la cantidad relativa de inspección realizada. Distintos niveles de inspección dan, más o menos, la misma protección al fabricante si el producto es de calidad aceptable, pero distintos niveles de protección al consumidor. La norma da siete niveles de inspección: los niveles de inspección general I, II y III y los niveles de inspección especial S-1, S-2, S-3 y S-4. En la mayor parte de los casos se utilizan los niveles de inspección general, mientras que los niveles de inspección especial se utilizan en casos muy excepcionales (en los que es muy caro realizar una inspección). La cantidad relativa de inspección aumenta en el siguiente orden: S-1, S-2, S-3, S-4, I, II y III, que da el máximo de inspección relativa. Si se considera el nivel II como la norma, la cantidad de inspección que dan los tres niveles es la que indica la Tabla A.22.

Tabla A.22. Comparación de niveles de inspección

Nivel de inspección general	Cantidad relativa de inspección
I	La mitad de la norma
II	La norma
III	El doble de la norma

El nivel de inspección habitual para un producto nuevo es el II o normal, en el cual se aceptan casi todos los lotes presentados con calidad igual que el *NCA*, además de un porcentaje alto de los que la tengan algo inferior. Se recurre al nivel III o riguroso cuando hay evidencias de falta de calidad, cuando la calidad media es suficientemente peor que el *NCA* como para aceptar que no es debido a una fluctuación muestral, y se realiza sin aumentar el tamaño de la muestra; y el nivel I o reducido se utiliza cuando hay evidencias del buen comportamiento. El paso a una inspección rigurosa mejora el riesgo del consumidor, mientras que el paso a una inspección reducida hace que éste aumente de manera considerable. Los criterios de paso entre niveles se presentan en la Tabla A.23.

Tabla A.23. Criterios de pasos entre niveles de inspección		
De	**a**	**Criterio**
I	II	Se obtiene un rechazo
II	III	2 lotes de 5 consecutivos rechazados
III	II	5 lotes consecutivos sin rechazo
II	I	10 lotes consecutivos sin rechazo

El uso de la norma se realiza mediante tablas. En primer lugar se utiliza la Tabla 17 del Apéndice A que, en función de los niveles de inspección y del tamaño del lote da el código de entrada a otra tabla propia para cada nivel de inspección: Tabla 18 para la inspección normal, Tabla 19 para la inspección rigurosa y Tabla 20 para inspección reducida. De esta segunda tabla se obtiene el tamaño de la muestra y los valores de los números de aceptación (máximo número de defectos admitidos en la muestra) y el de rechazo (número de defectos que si se alcanza suspende la inspección y se rechaza el lote), para cada *NCA*.

Para conservar un historial sobre la calidad de un producto hay que registrar todos los datos del muestreo realizado. La media estimada del proceso es el porcentaje medio de elementos defectuosos, la cual es útil para conocer la calidad del fabricante para poder tomar decisiones relativas a cambios en el proceso de inspección. La verdadera media del proceso, en general, es desconocida, pero se puede estimar a partir de los resultados medios de los lotes anteriores y su precisión depende tanto del promedio como de la cantidad total de unidades incluidas en el cálculo de la media. Cuando los lotes revisados tienen tamaños muy distintos, la media está ponderada por el tamaño del lote:

$$Media\ del\ Proceso\ para\ K\ lotes = \frac{100(p_1 N_1 + \ldots + p_k N_k)}{N_1 + \ldots + N_k}$$

donde:

$$p_i = \frac{n^o\ defectuosos\ en\ el\ i\text{-}ésimo\ lote\ revisado}{n^o\ unidades\ en\ el\ i\text{-}ésimo\ lote\ revisado}$$

Cuando los lotes revisados tienen tamaños similares:

$$Media\ del\ Proceso\ para\ K\ lotes = \frac{100(p_1 + \ldots + p_k)}{k}$$

Procedimientos de aceptación de calidad indiferente: El procedimiento Philips

El punto de calidad indiferente de la curva operativa da igual probabilidad de aceptación y de rechazo, por lo que el riesgo está dividido entre el productor y el consumidor.

El procedimiento Philips, desarrollado por la empresa Philips, es un ejemplo del concepto de riesgos divididos y utiliza la Tabla 21 del Apéndice A. Se usa un muestreo único para partidas de 1000 o menos artículos y muestreo doble para partidas mayores, teniendo la segunda muestra un tamaño del doble de la primera.

Procedimientos de aceptación basados en *CL:* El procedimiento Dodge-Romig

Para que los métodos estén protegidos respecto de porcentajes de defectuosos elevados, se establece el concepto de calidad límite (*CL*), con el que se indica un porcentaje máximo de defectuosos que se pueden admitir en un lote aislado, de manera que no se admitan lotes con calidad inferior al mismo. El procedimiento de Dodge-Romig utiliza el criterio *CL* para minimizar la inspección media total, el número medio de unidades por lote. El procedimiento está recogido en tablas de dos tipos:

— Tablas de *CL* (Tablas 22 y 23 del Apéndice A): Existen tablas normalizadas de *CL* de valores *0,5*, *1*, *2*, *3*, *4*, *5*, *6*, *7* y *10* por *100* tanto para muestreos simples como dobles. En las tablas se entra con el tamaño del lote y con una estimación de la media del proceso (porcentaje medio de defectuosos), obteniéndose el tamaño de la muestra, el *Ac* y el *LCMS* expresado en porcentaje.

— Tablas de *LCMS* (Tablas 24 y 25 del Apéndice A): Igual que en el caso anterior, existen tablas normalizadas para los valores de *0,1*, *0,25*, *1*, *1,5*, *2*, *2,5*, *3*, *4*, *7* y *10* por *100* tanto para muestreos simples como dobles. Se entra en las tablas con el tamaño del lote *N*, con la media del proceso y la *CLMS* y se obtiene el tamaño de la muestra y el valor de *CL*.

Para obtener un valor estimado de la calidad media del proceso, los datos históricos del mismo dan una excelente base para su cálculo. Pero para procesos nuevos hay dos formas de obtener buenos estimadores:

— Se realiza un gráfico de control de la proporción de elementos defectuosos durante un periodo base de 20 ó 30 lotes. Si el gráfico está bajo control, la línea central indica la calidad media del proceso y se puede usar como base para el método. Esto sólo lo puede usar el fabricante o un consumidor con información sobre la producción.

— Se usa una suposición previa de la calidad media del proceso para realizar un plan de muestreo tentativo. El *NCA* da un buen punto de partida. Este plan tentativo se usa para un número de lotes preestablecido. Los resultados de estas inspecciones de muestras dan un mejor valor estimado de la calidad media del proceso, pero hay que tener cuidado de usar solo los datos de las primeras muestras. Este método lo puede usar un consumidor que no conoce el comportamiento de la calidad que da un proveedor.

9. ANÁLISIS UNIVARIANTE DE SERIES TEMPORALES

Este tipo de análisis tiene gran importancia en los análisis de la evolución de los índices en los mercados de valores (bolsas), de los precios de productos, del nivel de paro, o en la evolución temporal de los índices de calidad en el producto que sale de un proceso, etc.

En este apartado se va a presentar cómo construir un modelo para explicar la estructura temporal y prever la evolución de una variable observada a lo largo del tiempo o de un eje que la organice[2]. Los datos estarán disponibles en intervalos regulares de tiempo y se desea utilizar la posible inercia en el comportamiento de la serie para prever su evolución futura.

El marco teórico para estudiar variables dinámicas data de las primeras décadas del siglo XX en las que se desarrolló toda una teoría econométrica para seguir los precios y la evolución de los mercados de valores. Un manual clásico en este ámbito es el trabajo de Pindyck y Rubinfeld (1980), también son de interés las monografías de los autores españoles Peña (1989) y Uriel (1995).

El modelo matemático para la representación de fenómenos que evolucionan en el tiempo con un comportamiento no determinístico es un proceso estocástico. El valor observado de la serie en el instante t es una extracción al azar de una variable aleatoria definida en dicho instante; por tanto, una serie de n datos será una muestra de un vector de n variables aleatorias ordenadas en el tiempo $(z_1, \ldots, z_t, \ldots, z_n)$ y se denomina proceso estocástico al conjunto de estas variables $\{z_t\}$, $t = 1, \ldots, n$, y la serie observada se considera una realización o trayectoria del proceso. Se dirá que se conoce la estructura probabilística de un proceso estocástico cuando se conozca la distribución conjunta de las n variables aleatorias $\{z_t\}$. De esta forma se definen:

[2] Las series temporales facilitan analizar la evolución de una característica en el tiempo, pero el eje organizativo también puede ser el espacio, lo que permite aplicarlas a algunos procesos de interés en cartografía, como pueden ser las digitalizaciones, el control posicional por elementos lineales, etc.

— La función de medias del proceso: Una función del tiempo que da las medias de las distribuciones marginales z_t para cada instante: $\mu_t = E[z_t]$. Si todas las variables tienen la misma media, se dice que el proceso es estable en la media.

— La función de varianzas del proceso: Da las varianzas en cada instante de tiempo: $\sigma_t^2 = Var(z_t)$. Si todas las variables tienen la misma varianza, se dice que el proceso es estable en la varianza.

— La función de autocovarianzas del proceso: Da las covarianzas en dos instantes cualesquiera: $Cov(z_t, z_{t+j}) = E[(z_t - \mu_t)(z_{t+j} - \mu_{t+j})]$.

— La función de autocorrelación: Es la estandarización de la función de covarianzas, su formulación es:

$$\rho_{t,t+j} = \frac{Cov(z_t, z_{t+j})}{\sigma_t \sigma_{t+j}}$$

Estas dos últimas funciones dependen de dos parámetros (t,j), donde t es el instante inicial y j es el intervalo entre observaciones. Una condición de estabilidad es que la dependencia entre dos observaciones sólo lo sea del intervalo entre ellas y no del origen, es decir que:

$$Cov(z_1, z_{1+k}) = Cov(z_2, z_{2+k}) = \gamma_k$$

y que las autocovarianzas dependan sólo de la distancia entre las variables.

Un proceso estocástico es estacionario en sentido débil si existen, y son estables la media, la varianza y las covarianzas. En un proceso estacionario la función de autocorrelación es $\rho_k = \gamma_k / \gamma_0$, teniendo en cuenta que $\gamma_0 = \sigma^2$, se verifica que $\rho_{-k} = \rho_k$.

En un proceso estacionario en sentido débil, se denomina función de autocorrelación simple, o correlograma, a la representación de los coeficientes de autocorrelación en función del retardo (k).

Si la dependencia entre observaciones tiende a cero al aumentar el retardo, se dice que el proceso es ergódico, lo cual es necesario para poder estimar las características del proceso a partir de una única realización. A partir de este momento, se utilizarán procesos estacionarios ergódicos, es decir, aquellos en los que se han eliminado, si existen, los términos sinusoidales deterministas.

La estacionariedad en sentido débil no garantiza la estabilidad completa del proceso, pues la distribución de las variables z_t puede cambiar en el tiempo. Entonces, se dice que un proceso estocástico $\{z_t, t \in T\}$ es estacionario estricto si para cualquier

conjunto $\{t_1, ..., t_n\}$ y cualquier h entero, las distribuciones conjuntas de $\{z_{t_1+h},...,z_{t_n+h}\}$ y $\{z_{t_1},...,z_{t_n}\}$ coinciden.

La estacionariedad estricta implica estacionariedad débil, pero el recíproco sólo se cumple cuando las n variables del proceso siguen una Normal n-variante. Una propiedad importante de los procesos estacionarios es que tienen incrementos estacionarios, es decir, si z_t es estacionario, $w_t = z_t - z_{t-1}$ también es estacionario.

Un proceso estacionario importante es el proceso de ruido, que está definido por:

$$E[z_t] = 0,\ Var(z_t) = \sigma^2\ ;\ Cov(z_t, z_{t-k}) = 0\ (k = \pm 1, \pm 2, ...)$$

Si todas las variables tienen distribución Normal, la falta de correlación garantiza la independencia y el proceso resultante se llama ruido blanco.

La gran mayoría de los procesos que se observan no son estacionarios y su nivel medio varía con el tiempo. Pero es frecuente que el proceso se convierta en estacionario al diferenciarlo:

— Se llama proceso en primera diferencia del proceso $\{z_t\}$ a uno nuevo $\{w_t\}$ obtenido como $w_t = z_t - z_{t-1}$.

— Se llama proceso en segunda diferencia de $\{z_t\}$ a un nuevo proceso obtenido como $w_t - w_{t-1}$.

— Un proceso es integrado de orden h cuando al diferenciarlo h veces se convierte en estacionario.

Los procesos que son suma de una tendencia polinómica y un proceso estacionario son integrados de orden el grado del polinomio.

9.1. Procesos autorregresivos (AR)

Los proceso autoregresivos (AR) son una clase importante de procesos estacionarios. Los procesos AR resultan de imponer una dependencia lineal entre las variables del proceso.

Un proceso es autorregresivo de orden p si:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + ... + \phi_p \tilde{z}_{t-p} + a_t$$

donde:

$\tilde{z}_t = z_t - \mu$

a_t Proceso de ruido blanco independiente de z_t.

Para facilitar el manejo, se define el operador retardo B como:

$$\text{B:} \begin{cases} Bz_t = z_{t-1} \\ B^k z_t = z_{t-k} \\ B\,cte = cte \end{cases}$$

con lo que la ecuación del proceso $AR(p)$ se puede escribir ahora:

$$\tilde{z}_t = (\phi_1 B + ... + \phi_p B^p)\tilde{z}_t + a_t$$

es decir:

$$(1 - \phi_1 B + ... + \phi_p B^p)\tilde{z}_t = \Phi_p(B) = a_t$$

donde:

$\Phi_p(B) = 0$ Ecuación característica del proceso, que tiene p raíces $(G_1^{-1},...,G_p^{-1})$.

El proceso es estacionario si $|G_i|<1$ para todo $i = 1,. ..,p$.

La función de autocorrelación se obtiene a partir de las ecuaciones de Yule-Walker:

$$\begin{cases} \rho_1 = \phi_1 + \phi_2\rho_1 + ... + \phi_p\rho_{p-1} \\ \rho_2 = \phi_1\rho_1 + \phi_2 + ... + \phi_p\rho_{p-2} \\ \vdots \\ \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + ... + \phi_p \end{cases}$$

Determinar el orden de un proceso autorregresivo a partir de su función de autocorrelación es difícil, puesto que, en general, es una mezcla de decrecimientos exponenciales y sinusoidales que se amortiguan al avanzar el retardo, y no presenta rasgos fácilmente identificables con el orden del proceso. Para resolver este problema se introduce la función de autocorrelación parcial.

En un $AR(1)$, conocido z_{t-1}, el valor de z_{t-2} es irrelevante para conocer z_t, mientras que en un $AR(2)$ sí existe influencia. En general, un $AR(p)$ presenta efectos directos de observaciones separadas por $1, 2, ...p$ retardos y los efectos directos de las $p + 1$,... son nulos.

Se define el coeficiente de autocorrelación parcial de orden k como una medida de la relación lineal entre observaciones separadas k periodos con independencia de los valores intermedios.

En el proceso $AR(p)$ los p primeros coeficientes de autocorrelación parcial son distintos de cero y, por tanto, indican el orden del proceso AR.

9.2. Procesos de media móvil (MA)

Antes de presentar los procesos de media móvil (MA) conviene comentar lo que se denomina descomposición de Wold.

Todo proceso estocástico débilmente estacionario de media cero, z_t, que no contenga componentes deterministas, se puede escribir como una función lineal de variables aleatorias incorreladas a_t, es decir:

$$\tilde{z}_t = \sum_{i=0}^{\infty} \psi_i a_{t-i} = \Psi(B) a_t \text{ con } \psi_0 = 1$$

donde:

$E[a_t] = 0$

$Var(a_t) = \sigma^2$

$E[a_t a_{t-k}] = 0 \ (k > 1)$

A partir de este momento se va a suponer que $\tilde{z}_t$ sigue una distribución Normal, por lo que la estacionariedad débil coincide con la estricta, las variables a_t son Normales y el proceso $\{a_t\}$ de variables independientes de media cero y varianza constante es un proceso de ruido blanco.

Bajo éstas circunstancias, la varianza de z_t vale:

$$Var(z_t) = \gamma_0 = \sigma^2 \sum_{i=1}^{\infty} \psi_i^2$$

Y para que el proceso sea estacionario la serie $\{\psi_i^2\}$ debe ser convergente. Se entiende pues que los procesos AR estudiados son casos particulares de la representación de Wold.

Se define el operador inverso de $(1 - \phi B)$ cuando $|\phi| < 1$ como:

$$(1-\phi B)^{-1} = 1+\phi B+\phi^2 B^2+...$$

Se dice que un proceso es invertible cuando el efecto del pasado decrece con el tiempo.

Un proceso de media móvil de orden q tiene los q primeros ψ_i no nulos y siempre es estacionario. Viene dado por:

$$\tilde{z}_t = (1-\theta_1 B-...-\theta_q B^q)a_t = \Theta_q(B)a_t$$

El proceso es invertible si las raíces de $\Theta_q(B) = 0$ son, en módulo, mayores que uno.

La función de autocovarianzas vale:

$$\gamma_k = \begin{cases} \sigma^2(1+\theta_1^2+...+\theta_q^2) & si\ k=0 \\ \sigma^2(-\theta_k+\theta_1\theta_{k+1}+\theta_{q-k}\theta_q) & si\ k=1,...,q \\ 0 & si\ k>q \end{cases},$$

por lo que tiene los q primeros coeficientes distintos de cero. La función de autocorrelación vale:

$$\rho_k = \begin{cases} 1 & si\ k=0 \\ \dfrac{-\theta_k+\theta_1\theta_{k+1}+\theta_{q-k}\theta_q}{1+\theta_1^2+...+\theta_q^2} & si\ k=1,...,q \\ 0 & si\ k>q \end{cases}$$

La función de autocorrelación parcial de un proceso $MA(q)$ tiene la estructura de la función de autocorrelación simple de un proceso $AR(q)$. Un proceso $AR(1)$ equivale a un proceso $MA(\infty)$.

9.3. Procesos mixtos autorregresivos de medias móviles (ARMA)

Los procesos *AR* y *MA* son aproximaciones al proceso lineal general $MA(\infty)$ desde puntos de vista complementarios, pues los *AR* imponen restricciones sobre las pautas de decrecimiento de los coeficientes y los *MA* suponen un número de términos finito. Los procesos *ARMA* intentan combinar estas ventajas y permiten representar

utilizando pocos parámetros procesos cuyos q primeros coeficientes son cualesquiera, y los siguientes decrecen según leyes simples.

El proceso $ARMA(p,q)$ viene dado por la expresión:

$$(1-\phi_1 B-\ldots-\phi_p B^p)\tilde{z}_t=(1-\theta_1 B-\ldots-\theta_q B^q)a_t$$

es decir:

$$\Phi_p(B)\tilde{z}_t=\Theta_q(B)a_t.$$

El proceso es estacionario si las raíces de $\Phi_p(B)=0$ están fuera del círculo unidad, e invertible si lo están las de $\Theta_q(B)=0$.

La función de autocorrelación tiene el siguiente comportamiento: los primeros $p-q+1$ valores tienen cualquier estructura, decreciendo a partir del valor p - q como una mezcla de exponenciales y sinusoidales, determinada por la parte autorregresiva. En la función de autocorrelación parcial se tienen valores iniciales dependientes del orden del AR, seguidos del decrecimiento debido a la parte MA.

9.4. Procesos ARIMA

Hay procesos no estacionarios que mediante diferenciaciones sucesivas se convierten en estacionarios. Entonces, se define el operador diferencia como $\nabla=1-B$, es decir, $\nabla A_t=A_t-A_{t-1}$.

Entonces, se define el modelo $ARIMA(p,d,q)$ como:

$$\Phi_p(B)\nabla^d\tilde{z}_t=\Theta_q(B)a_t$$

donde d es el número de diferencias a considerar.

Si el proceso ya es estacionario, se toma $d=0$, con lo que se tiene el modelo $ARMA(p,q)$.

9.5. Método de Box-Jenkins para el análisis de Series Temporales

Una serie temporal es la realización de un proceso estocástico y hay que buscar un modelo dentro de la clase $ARIMA$ que represente satisfactoriamente el

comportamiento de la serie. Para ello hay dos herramientas fundamentales, la función de autocorrelación y la función de autocorrelación parcial.

El procedimiento consiste en una secuencia iterativa en la que cada paso tiene tres fases: identificación, estimación y diagnóstico.

9.5.1. IDENTIFICACIÓN DEL MODELO ARIMA

En los apartados anteriores se han presentado las propiedades teóricas de los procesos *ARIMA*; ahora se va a ver cómo se ajustan estos modelos a una serie real observada. El primer paso va a ser identificar la estructura del modelo, lo que requiere:

— Decidir las transformaciones a aplicar para convertir el proceso subyacente en estacionario.

— Determinar un modelo para el proceso estacionario, es decir, obtener p y q.

Identificación de la estructura no estacionaria

Si se tiene una serie observada que es una realización del proceso estudiado, se tiene que determinar si es necesario transformar la serie para que tenga varianza constante y determinar el número de diferencias que se deben aplicar para que la media sea constante.

Determinación de la transformación Box-Cox

Muchas series presentan una variabilidad creciente con el nivel de la serie. Si existe una relación del tipo $\sigma_i = k\mu_i^{\alpha}$ entre las medias de un conjunto de variables aleatorias y sus desviaciones típicas; al transformar las variables mediante la familia Box-Cox con parámetro $\lambda = 1-\alpha$, las variables transformadas tienen varianza constante.

La transformación es la siguiente:

$$Z_t^{(\lambda)} = \begin{cases} \dfrac{Z_t - 1}{\lambda} & \lambda \neq 0 \\ \ln(Z_t) & \lambda = 0 \end{cases}$$

Para calcular λ se divide la serie en periodos y en cada uno se calcula la media y la varianza. Entonces, α se obtiene realizando regresión lineal entre $ln(\sigma_i)$ y $ln(\mu_i)$, esto es:

$$ln(\sigma_i) = ln(k) + \alpha\, ln(\mu_i)$$
$$\lambda = 1-\alpha$$

Determinación del orden de diferenciación regular

Una vez estabilizada la varianza, hay que estabilizar la media. Si la serie tiene tendencia o muestra cambios de nivel en la media, hay que diferenciarla para transformarla en estacionaria. Cuando la decisión no esté clara por el gráfico de la serie, hay que estudiar la función de autocorrelación de la serie. Si la serie no es estacionaria, el correlograma tiene un decrecimiento casi lineal, por lo que hay que diferenciar; mientras que si se diferencia una serie estable, la serie resultante también es estable, pero más regular.

La identificación de la estructura ARMA

La identificación de los valores p y q del modelo *ARMA* se realiza calculando las funciones de autocorrelación y autocorrelación parcial estimadas del proceso y comparándolas con las funciones teóricas estudiadas para cada modelo.

Si w_t es la serie diferenciada, los coeficientes de autocorrelación se calculan mediante la expresión:

$$r_k = \frac{\sum_{t=d+sD+1}^{n}(w_t - \overline{w})(w_{t+k} - \overline{w})}{\sum_{t=d+sD+1}^{n}(w_t - \overline{w})^2} \quad \text{para } k=1, 2...$$

donde se ha tenido en cuenta que al diferenciar se pierden d observaciones.

Para juzgar cuando un coeficiente r_k es distinto de cero, se necesita su error estándar, cuya determinación depende de la estructura del proceso. Una solución es tomar como error estándar $1/(n^{1/2})$, que es aproximadamente el error estándar de un coeficiente de correlación entre variables independientes. Si todos los coeficientes teóricos de autocorrelación fueran nulos, las desviaciones típicas de estimación serían aproximadamente $1/(n^{1/2})$; por lo que se pueden trazar bandas de confianza a $\pm 2/(n^{1/2})$ y considerar como significativos en una primera aproximación a los coeficientes fuera de dichas bandas.

Otro método consiste en considerar que las primeras h autocorrelaciones teóricas son distintas de cero, pero las siguientes son cero, entonces:

$$Var(r_k) = \frac{1}{n}\left(1 + 2\sum_{j=1}^{h} \rho_j^2\right) \text{ para } k > h$$

Se puede construir un intervalo de confianza para r_k condicionado a que los coeficientes $\rho_1, \ldots, \rho_{k-1}$ son no nulos y los siguientes son nulos. El procedimiento es totalmente análogo al anterior.

Los coeficientes de autocorrelación parcial estimados se obtienen con las regresiones $\tilde{w}_t = \alpha_{k1}\tilde{w}_{t-1} + \ldots + \alpha_{kk}\tilde{w}_{t-k}$, donde $\tilde{w}_t = w_t - \overline{w}$. La serie α_{kk} ($k = 1, 2, \ldots$) es la función de autocorrelación parcial estimada. En los gráficos de la función de autocorrelación parcial se utilizan los límites asintóticos $\pm 2/(n^{1/2})$.

Además de los órdenes p y q, se debe comprobar si la serie estacionaria tiene media distinta de cero. Para ello, se calcula la media muestral:

$$\overline{w} = \frac{\sum_{i=1}^{T} w_i}{T}$$

donde:

T es el número de sumandos (generalmente $T = n\text{-}d$)

y

$$s(\overline{w}) \cong \frac{s_w}{\sqrt{T}}(1 + 2r_1 + \ldots + 2r_k)^{1/2}$$

es su desviación típica, con s_w la desviación típica de w_t y suponiendo que los k primeros coeficientes de autocorrelación son significativos. Si $\overline{w} \geq 2s(\overline{w})$ se admite que la media es distinta de cero y se incluye como un parámetro a estimar; en caso contrario, se supone que $E[w_t]=0$.

En la Tabla A.24 se resumen las principales características de las funciones de autocorrelación y autocorrelación parcial en cada modelo estudiado.

Tabla A.24 - Resumen de las principales características de las funciones de autocorrelación y autocorrelación parical de los modelos *AR*, *MA* y *ARMA*.

	AR(p)	***MA(q)***	***ARMA(p,q)***
FAC	Decrece exponencialmente a cero	q valores no nulos y nulos los demás	Decrece exponencialmente a cero después de unos valores
FACP	p valores no nulos y nulos los demás	Decrece exponencialmente a cero	Decrece exponencialmente a cero después de unos valores

9.5.2. La estimación del modelo ARMA

Se dispone de una serie estacionaria de T elementos $\{w_t\}$ $(t = 1,...,T)$ para estimar los parámetros de un modelo *ARMA* concreto.

La estimación exacta de un $AR(p)$ se realiza con un algoritmo de optimización no lineal, siendo el más frecuente una modificación del algoritmo de Gauss-Newton realizada por Marquardt (1946). El algoritmo es el siguiente:

1. Se obtiene un valor inicial $\hat{\beta}_0 = (\hat{c}_0, \hat{\phi}_{10}, ..., \hat{\phi}_{p0})$ de los parámetros, que es el estimador condicionado, por mínimos cuadrados:

   Si se escribe el modelo $AR(p)$ como $w_t = v_t' \beta + a_t$ para $t = p+1,...,T$, donde:

   $$v_t' = (1, w_{t-1}, ..., w_{t-p}); \; \beta' = (c, \phi_1, ..., \phi_p)$$

   entonces:

   $$\hat{\beta} = \left(\sum_{t=p+1}^{T} v_t v_t' \right)^{-1} \left(\sum v_t w_t \right)$$

2. Se calculan los residuos del modelo para el valor supuesto de los parámetros mediante: $a_t(\hat{\beta}_0) = w_t - \hat{c}_0 - \hat{\phi}_{10} w_{t-1} - ... - \hat{\phi}_{p0} w_{t-p}$.

3. Se calculan las derivadas primeras de los residuos respecto al vector de parámetros, que se puede hacer numéricamente definiendo:

   $$\frac{\partial a_t(\hat{\beta}_0)}{\partial \beta_i} \cong \frac{a_t(\hat{\beta}_0 + 1_i \varepsilon) - a_t(\hat{\beta}_0)}{\varepsilon} \text{ para } i = 1,..., p+1$$

   donde 1_i es un vector con un uno en la componente i y cero en el resto; ε es un número pequeño (0,001 ó similar). El cálculo de derivadas requiere repetir $(p+1)$ veces el 2º paso, modificando cada vez ligeramente un componente de $\hat{\beta}_0$.

4. Estimar los coeficientes de una regresión entre los residuos calculados en el paso 2º como la variable dependiente y un vector de $(p+1)$ variables explicativas V_t, cuyas componentes son las derivadas parciales de los residuos calculadas en el paso 3 y cambiadas de signo, es decir:

$$v_{it} = -\frac{\partial a_t(\hat{\beta}_0)}{\partial \beta_i}$$

5. Sea b el vector de parámetros estimado con la regresión anterior que mide el efecto lineal sobre los residuos de cambios en el vector de parámetros. Se define un nuevo estimador $\hat{\beta}_1$ con: $\hat{\beta}_1 = \hat{\beta}_0 + b$ y se vuelve al paso 2, tomando $\hat{\beta}_1$ como estimador inicial. Se repite el proceso hasta obtener convergencia, definida por $|\beta_{i-1} - \beta_i| < \alpha$ y $|\partial L(\beta_i)/\partial\beta| < \gamma$, para todos sus componentes, siendo α y γ constantes pequeñas y

$$L(\phi \mid w_p,...,w_1) = -\frac{(T-p)}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+11}^{T} a_t^{\,2}$$

la función de verosimilitud condicionada a los p valores iniciales, que ha sido maximizada.

Estimación de procesos MA y mixtos

Es más complicada que en caso anterior, ya que las ecuaciones no van a ser siempre lineales en los parámetros. Se busca, mediante algún algoritmo no lineal, el valor paramétrico que maximiza la función de verosimilitud condicional.

La estimación de los modelos $ARMA(p,q)$ se realiza siguiendo los mismos principios, a partir de la función de verosimilitud condicional:

$$L(\beta \mid a_q w_p) = -\frac{(T-p)}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+11}^{T} a_t^{\,2}(\beta \mid a_q w_p)$$

donde β es el vector de parámetros, a_q y w_p son los vectores de perturbaciones y valores observados iniciales a los que condicionamos la estimación. Los residuos se calculan recursivamente mediante:

$$\hat{a}_t = w_t - \hat{\phi}_1 w_{t-1} - ... - \hat{\phi}_p w_{t-p} + \hat{\theta}_1 \hat{a}_{t-1} + ... + \hat{\theta}_q \hat{a}_{t-q} \quad \text{para } t = p+1,..., T$$

y se supone que los primeros residuos son cero. La maximización de la función de verosimilitud condicional se realiza por el algoritmo de Marquardt.

9.5.3. DIAGNOSIS DEL MODELO

La diagnosis del modelo se realiza para comprobar que ciertas hipótesis básicas realizadas respecto a los residuos son ciertas, recordemos que éstas son: media cero, varianza constante, falta de correlación para cualquier retardo y distribución Normal.

El primer análisis a realizar es para ver si los residuos estimados están incorrelados. Para ello, se calcula su función de autocorrelación, mediante:

$$\hat{r}_a = \frac{\sum_{t=1}^{T-k}(\hat{a}_t - \bar{a})(\hat{a}_{t+k} - \bar{a})}{\sum_{t=1}^{T}(\hat{a}_t - \bar{a})^2}$$

donde $\bar{a}$ es la media de los T residuos. El procedimiento habitual es comprobar si todos los coeficientes obtenidos están dentro de los límites de confianza de $\pm 2/(T^{1/2})$. Pero estos límites sobreestiman la varianza en los retardos iniciales, por lo que un valor próximo a los límites de confianza en un retardo inicial se debe considerar como un indicio claro de que el modelo es inadecuado. Por ello, se realizan contrastes globales de que los primeros h coeficientes son cero.

Uno de ellos es el contraste de Ljung-Box, que dice que si los residuos son realmente ruido blanco, el estadístico:

$$Q(h) = T(T+2)\sum_{j=1}^{h}\frac{\hat{r}_a^2(j)}{T-j}$$

sigue una distribución χ^2 con grados de libertad igual al número de coeficientes en la suma (h) menos el número de parámetros estimados ($p+q$). Entonces, un modelo es inadecuado si $P[\chi^2(h\text{-}p\text{-}q)) > Q(h)]$ es pequeña.

La Normalidad de los residuos se comprueba con un contraste de bondad de ajuste como los presentados anteriormente en este anejo.

A continuación se realiza el contraste de que la media de los residuos es cero. Para ello, se utiliza el contraste paramétrico de la media para varianza poblacional desconocida, donde la H_0 es que la media es cero, frente a la hipótesis alternativa de que es distinta de cero.

La estabilidad de la varianza se comprueba estudiando el gráfico de los residuos a lo largo del tiempo. En caso de duda se puede dividir el intervalo muestral en tres o cuatro partes y se aplica un contraste de homogeneidad de varianzas entre las partes.

Reformulación

Si se ha estimado el modelo $\Phi(B)z_t = \Theta(B)\hat{a}_t$ y al analizar las funciones de autocorrelación de los residuos estimados se concluye que siguen el modelo $\alpha(B)\hat{a}_t = \gamma(B)u_t$, se tiene para la serie el modelo $\alpha(B)\Phi(B)z_t = \Theta(B)\gamma(B)u_t$.

Estimando este modelo se obtienen nuevos residuos, repitiéndose todo el proceso.

Sobreajuste

Para comprobar si el modelo es adecuado es conveniente utilizar la técnica del sobreajuste, que consiste en estimar un modelo de orden mayor que el analizado y comprobar si se obtienen coeficientes estimados significativos. Así se pueden mejorar un poco las predicciones. En general, si se ha ajustado un *ARIMA*(p,d,q), el sobreajuste se aplica estimando los modelos *ARIMA*($p+1$,d,q) y *ARIMA*(p,d,$q+1$) y comprobando si los parámetros adicionales son significativos. Nunca conviene ampliar el orden de las partes *AR* y *MA* a la vez, ya que se pueden compensar los efectos.

APÉNDICE A

TABLAS ESTADÍSTICAS

Tabla 1

Función de distribución Binomial

		P										
n	A	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0,9801	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4225	0,3600	0,3025	0,2500
	1	0,9999	0,9975	0,9900	0,9775	0,9600	0,9375	0,9100	0,8775	0,8400	0,7975	0,7500
	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
3	0	0,9703	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2746	0,2160	0,1664	0,1250
	1	0,9997	0,9928	0,9720	0,9392	0,8960	0,8438	0,7840	0,7183	0,6480	0,5748	0,5000
	2	1,0000	0,9999	0,9990	0,9966	0,9920	0,9844	0,9730	0,9571	0,9360	0,9089	0,8750
	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4	0	0,9606	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1785	0,1296	0,0915	0,0625
	1	0,9994	0,9860	0,9477	0,8905	0,8192	0,7383	0,6517	0,5630	0,4752	0,3910	0,3125
	2	1,0000	0,9995	0,9963	0,9880	0,9728	0,9492	0,9163	0,8735	0,8208	0,7585	0,6875
	3	1,0000	1,0000	0,9999	0,9995	0,9984	0,9961	0,9919	0,9850	0,9744	0,9590	0,9375
	4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0	0,9510	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1160	0,0778	0,0503	0,0313
	1	0,9990	0,9774	0,9185	0,8352	0,7373	0,6328	0,5282	0,4284	0,3370	0,2562	0,1875
	2	1,0000	0,9988	0,9914	0,9734	0,9421	0,8965	0,8369	0,7648	0,6826	0,5931	0,5000
	3	1,0000	1,0000	0,9995	0,9978	0,9933	0,9844	0,9692	0,9460	0,9130	0,8688	0,8125
	4	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990	0,9976	0,9947	0,9898	0,9815	0,9688
	5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
6	0	0,9415	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0754	0,0467	0,0277	0,0156
	1	0,9985	0,9672	0,8857	0,7765	0,6554	0,5339	0,4202	0,3191	0,2333	0,1636	0,1094
	2	1,0000	0,9978	0,9842	0,9527	0,9011	0,8306	0,7443	0,6471	0,5443	0,4415	0,3438
	3	1,0000	0,9999	0,9987	0,9941	0,9830	0,9624	0,9295	0,8826	0,8208	0,7447	0,6563
	4	1,0000	1,0000	0,9999	0,9996	0,9984	0,9954	0,9891	0,9777	0,9590	0,9308	0,8906
	5	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9993	0,9982	0,9959	0,9917	0,9844
	6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
7	0	0,9321	0,6983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0490	0,0280	0,0152	0,0078
	1	0,9980	0,9556	0,8503	0,7166	0,5767	0,4449	0,3294	0,2338	0,1586	0,1024	0,0625
	2	1,0000	0,9962	0,9743	0,9262	0,8520	0,7564	0,6471	0,5323	0,4199	0,3164	0,2266
	3	1,0000	0,9998	0,9973	0,9879	0,9667	0,9294	0,8740	0,8002	0,7102	0,6083	0,5000
	4	1,0000	1,0000	0,9998	0,9988	0,9953	0,9871	0,9712	0,9444	0,9037	0,8471	0,7734
	5	1,0000	1,0000	1,0000	0,9999	0,9996	0,9987	0,9962	0,9910	0,9812	0,9643	0,9375
	6	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9994	0,9984	0,9963	0,9922
	7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
8	0	0,9227	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0319	0,0168	0,0084	0,0039
	1	0,9973	0,9428	0,8131	0,6572	0,5033	0,3671	0,2553	0,1691	0,1064	0,0632	0,0352
	2	0,9999	0,9942	0,9619	0,8948	0,7969	0,6785	0,5518	0,4278	0,3154	0,2201	0,1445
	3	1,0000	0,9996	0,9950	0,9786	0,9437	0,8862	0,8059	0,7064	0,5941	0,4770	0,3633
	4	1,0000	1,0000	0,9996	0,9971	0,9896	0,9727	0,9420	0,8939	0,8263	0,7396	0,6367
	5	1,0000	1,0000	1,0000	0,9998	0,9988	0,9958	0,9887	0,9747	0,9502	0,9115	0,8555
	6	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9987	0,9964	0,9915	0,9819	0,9648
	7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9993	0,9983	0,9961
	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
9	0	0,9135	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0207	0,0101	0,0046	0,0020
	1	0,9966	0,9288	0,7748	0,5995	0,4362	0,3003	0,1960	0,1211	0,0705	0,0385	0,0195
	2	0,9999	0,9916	0,9470	0,8591	0,7382	0,6007	0,4628	0,3373	0,2318	0,1495	0,0898
	3	1,0000	0,9994	0,9917	0,9661	0,9144	0,8343	0,7297	0,6089	0,4826	0,3614	0,2539
	4	1,0000	1,0000	0,9991	0,9944	0,9804	0,9511	0,9012	0,8283	0,7334	0,6214	0,5000
	5	1,0000	1,0000	0,9999	0,9994	0,9969	0,9900	0,9747	0,9464	0,9006	0,8342	0,7461
	6	1,0000	1,0000	1,0000	1,0000	0,9997	0,9987	0,9957	0,9888	0,9750	0,9502	0,9102
	7	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9986	0,9962	0,9909	0,9805
	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	0,9980
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
10	0	0,9044	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
	1	0,9957	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,0860	0,0464	0,0233	0,0107
	2	0,9999	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2616	0,1673	0,0996	0,0547
	3	1,0000	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5138	0,3823	0,2660	0,1719
	4	1,0000	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7515	0,6331	0,5044	0,3770
	5	1,0000	1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9051	0,8338	0,7384	0,6230
	6	1,0000	1,0000	1,0000	0,9999	0,9991	0,9965	0,9894	0,9740	0,9452	0,8980	0,8281
	7	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9952	0,9877	0,9726	0,9453
	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9983	0,9955	0,9893
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
	10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
11	0	0,8953	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0088	0,0036	0,0014	0,0005
	1	0,9948	0,8981	0,6974	0,4922	0,3221	0,1971	0,1130	0,0606	0,0302	0,0139	0,0059
	2	0,9998	0,9848	0,9104	0,7788	0,6174	0,4552	0,3127	0,2001	0,1189	0,0652	0,0327
	3	1,0000	0,9984	0,9815	0,9306	0,8389	0,7133	0,5696	0,4256	0,2963	0,1911	0,1133
	4	1,0000	0,9999	0,9972	0,9841	0,9496	0,8854	0,7897	0,6683	0,5328	0,3971	0,2744
	5	1,0000	1,0000	0,9997	0,9973	0,9883	0,9657	0,9218	0,8513	0,7535	0,6331	0,5000
	6	1,0000	1,0000	1,0000	0,9997	0,9980	0,9924	0,9784	0,9499	0,9006	0,8262	0,7256
	7	1,0000	1,0000	1,0000	1,0000	0,9998	0,9988	0,9957	0,9878	0,9707	0,9390	0,8867
	8	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9994	0,9980	0,9941	0,9852	0,9673
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9993	0,9978	0,9941
	10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9995
	11	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Función de distribución Binomial

		P									
n	A	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
12	0	0.8864	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.9938	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9998	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	1.0000	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	1.0000	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	0	0.8775	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.9928	0.8646	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
	2	0.9997	0.9755	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
	3	1.0000	0.9969	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
	4	1.0000	0.9997	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334
	5	1.0000	1.0000	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
	6	1.0000	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
	8	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.8687	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.9916	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9997	0.9699	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	1.0000	0.9958	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	1.0000	0.9996	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	1.0000	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.8601	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.9904	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9996	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	1.0000	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	1.0000	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	1.0000	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	0.8515	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.9891	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003
	2	0.9995	0.9571	0.7893	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021
	3	1.0000	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106
	4	1.0000	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384
	5	1.0000	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051
	6	1.0000	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272
	7	1.0000	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018
	8	1.0000	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982
	9	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Función de distribución Binomial

		P										
n	d	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
17	0	0.8429	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.9877	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9994	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	1.0000	0.9912	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064
	4	1.0000	0.9988	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5	1.0000	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6	1.0000	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7	1.0000	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145
	8	1.0000	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.8345	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.9862	0.7735	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9993	0.9419	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	1.0000	0.9891	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038
	4	1.0000	0.9985	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154
	5	1.0000	0.9998	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6	1.0000	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1189
	7	1.0000	1.0000	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8	1.0000	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073
	9	1.0000	1.0000	1.0000	0.9999	0.9991	0.9916	0.9790	0.9403	0.8653	0.7473	0.5927
	10	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9519
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
19	0	0.8262	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.9847	0.7547	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000
	2	0.9991	0.9335	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004
	3	1.0000	0.9868	0.8850	0.6841	0.4551	0.2631	0.1332	0.0591	0.0230	0.0077	0.0022
	4	1.0000	0.9980	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5	1.0000	0.9998	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318
	6	1.0000	1.0000	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835
	7	1.0000	1.0000	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796
	8	1.0000	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9165
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.8179	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.9831	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9990	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	1.0000	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	1.0000	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5	1.0000	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Función de distribución Binomial

		P									
n	A	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
25	0	0.7778	0.2774	0.0718	0.0172	0.0038	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.9742	0.6424	0.2712	0.0931	0.0274	0.0070	0.0016	0.0003	0.0001	0.0000	0.0000
	2	0.9980	0.8729	0.5371	0.2537	0.0982	0.0321	0.0090	0.0021	0.0004	0.0001	0.0000
	3	0.9999	0.9659	0.7636	0.4711	0.2340	0.0962	0.0332	0.0097	0.0024	0.0005	0.0001
	4	1.0000	0.9928	0.9020	0.6821	0.4207	0.2137	0.0905	0.0320	0.0095	0.0023	0.0005
	5	1.0000	0.9988	0.9666	0.8385	0.6167	0.3783	0.1935	0.0826	0.0294	0.0086	0.0020
	6	1.0000	0.9998	0.9905	0.9305	0.7800	0.5611	0.3407	0.1734	0.0736	0.0258	0.0073
	7	1.0000	1.0000	0.9977	0.9745	0.8909	0.7265	0.5118	0.3061	0.1536	0.0639	0.0216
	8	1.0000	1.0000	0.9995	0.9920	0.9532	0.8506	0.6769	0.4668	0.2735	0.1340	0.0539
	9	1.0000	1.0000	0.9999	0.9979	0.9827	0.9287	0.8106	0.6303	0.4246	0.2424	0.1148
	10	1.0000	1.0000	1.0000	0.9995	0.9944	0.9703	0.9022	0.7712	0.5858	0.3843	0.2122
	11	1.0000	1.0000	1.0000	0.9999	0.9985	0.9893	0.9558	0.8746	0.7323	0.5426	0.3450
	12	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966	0.9825	0.9396	0.8462	0.6937	0.5000
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9940	0.9745	0.9222	0.8173	0.6550
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9982	0.9907	0.9656	0.9040	0.7878
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9971	0.9868	0.9560	0.8852
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9957	0.9826	0.9461
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9942	0.9784
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9927
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9980
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tabla 2

Función de distribución de Poisson

					λ					
A	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,9953	0,9825	0,9631	0,9384	0,9098	0,8781	0,8442	0,8088	0,7725	0,7358
2	0,9998	0,9989	0,9964	0,9921	0,9856	0,9769	0,9659	0,9526	0,9371	0,9197
3	1,0000	0,9999	0,9997	0,9992	0,9982	0,9966	0,9942	0,9909	0,9865	0,9810
4	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9986	0,9977	0,9963
5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9997	0,9994
6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

					λ					
A	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,6990	0,6626	0,6268	0,5918	0,5578	0,5249	0,4932	0,4628	0,4338	0,4060
2	0,9004	0,8795	0,8571	0,8335	0,8088	0,7834	0,7572	0,7306	0,7037	0,6767
3	0,9743	0,9662	0,9569	0,9463	0,9344	0,9212	0,9068	0,8913	0,8747	0,8571
4	0,9946	0,9923	0,9893	0,9857	0,9814	0,9763	0,9704	0,9636	0,9559	0,9473
5	0,9990	0,9985	0,9978	0,9968	0,9955	0,9940	0,9920	0,9896	0,9868	0,9834
6	0,9999	0,9997	0,9996	0,9994	0,9991	0,9987	0,9981	0,9974	0,9966	0,9955
7	1,0000	1,0000	0,9999	0,9999	0,9998	0,9997	0,9996	0,9994	0,9992	0,9989
8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9998	0,9998
9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

					λ					
A	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
1	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498
2	0,3796	0,3546	0,3309	0,3084	0,2873	0,2674	0,2487	0,2311	0,2146	0,1991
3	0,6496	0,6227	0,5960	0,5697	0,5438	0,5184	0,4936	0,4695	0,4460	0,4232
4	0,8386	0,8194	0,7993	0,7787	0,7576	0,7360	0,7141	0,6919	0,6696	0,6472
5	0,9379	0,9275	0,9163	0,9041	0,8912	0,8774	0,8629	0,8477	0,8318	0,8153
6	0,9796	0,9751	0,9700	0,9643	0,9580	0,9510	0,9433	0,9349	0,9258	0,9161
7	0,9941	0,9925	0,9906	0,9884	0,9858	0,9828	0,9794	0,9756	0,9713	0,9665
8	0,9985	0,9980	0,9974	0,9967	0,9958	0,9947	0,9934	0,9919	0,9901	0,9881
9	0,9997	0,9995	0,9994	0,9991	0,9989	0,9985	0,9981	0,9976	0,9969	0,9962
10	0,9999	0,9999	0,9999	0,9998	0,9997	0,9996	0,9995	0,9993	0,9991	0,9989
11	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997
12	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999

					λ					
A	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
1	0,0450	0,0408	0,0369	0,0334	0,0302	0,0273	0,0247	0,0224	0,0202	0,0183
2	0,1847	0,1712	0,1586	0,1468	0,1359	0,1257	0,1162	0,1074	0,0992	0,0916
3	0,4012	0,3799	0,3594	0,3397	0,3208	0,3027	0,2854	0,2689	0,2531	0,2381
4	0,6248	0,6025	0,5803	0,5584	0,5366	0,5152	0,4942	0,4735	0,4533	0,4335
5	0,7982	0,7806	0,7626	0,7442	0,7254	0,7064	0,6872	0,6678	0,6484	0,6288
6	0,9057	0,8946	0,8829	0,8705	0,8576	0,8441	0,8301	0,8156	0,8006	0,7851
7	0,9612	0,9554	0,9490	0,9421	0,9347	0,9267	0,9182	0,9091	0,8995	0,8893
8	0,9858	0,9832	0,9802	0,9769	0,9733	0,9692	0,9648	0,9599	0,9546	0,9489
9	0,9953	0,9943	0,9931	0,9917	0,9901	0,9883	0,9863	0,9840	0,9815	0,9786
10	0,9986	0,9982	0,9978	0,9973	0,9967	0,9960	0,9952	0,9942	0,9931	0,9919
11	0,9996	0,9995	0,9994	0,9992	0,9990	0,9987	0,9984	0,9981	0,9977	0,9972
12	0,9999	0,9999	0,9998	0,9998	0,9997	0,9996	0,9995	0,9994	0,9993	0,9991
13	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997
14	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999

					λ					
A	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	0,0166	0,0150	0,0136	0,0123	0,0111	0,0101	0,0091	0,0082	0,0074	0,0067
1	0,0845	0,0780	0,0719	0,0663	0,0611	0,0563	0,0518	0,0477	0,0439	0,0404
2	0,2238	0,2102	0,1974	0,1851	0,1736	0,1626	0,1523	0,1425	0,1333	0,1247
3	0,4142	0,3954	0,3772	0,3595	0,3423	0,3257	0,3097	0,2942	0,2793	0,2650
4	0,6093	0,5898	0,5704	0,5512	0,5321	0,5132	0,4946	0,4763	0,4582	0,4405
5	0,7693	0,7531	0,7367	0,7199	0,7029	0,6858	0,6664	0,6510	0,6335	0,6160
6	0,8787	0,8675	0,8558	0,8436	0,8311	0,8180	0,8046	0,7908	0,7767	0,7622
7	0,9427	0,9361	0,9290	0,9214	0,9134	0,9050	0,8960	0,8867	0,8769	0,8666
8	0,9755	0,9721	0,9683	0,9642	0,9597	0,9549	0,9497	0,9442	0,9382	0,9319
9	0,9905	0,9889	0,9871	0,9851	0,9829	0,9805	0,9778	0,9749	0,9717	0,9682
10	0,9966	0,9959	0,9952	0,9943	0,9933	0,9922	0,9910	0,9896	0,9880	0,9863
11	0,9989	0,9986	0,9983	0,9980	0,9976	0,9971	0,9966	0,9960	0,9953	0,9945
12	0,9997	0,9996	0,9995	0,9993	0,9992	0,9990	0,9988	0,9986	0,9983	0,9980
13	0,9999	0,9999	0,9998	0,9998	0,9997	0,9997	0,9996	0,9995	0,9994	0,9993
14	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998
15	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999

Función de distribución de Poisson

					λ					
A	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	0.0061	0.0055	0.0050	0.0045	0.0041	0.0037	0.0033	0-0030	0.0027	0.0025
1	0.0372	0.0342	0.0314	0.0289	0.0266	0.0244	0.0224	0.0206	0.0189	0.0174
2	0.1165	0.1088	0.1016	0.0948	0.0884	0.0824	0.07&8	0.0715	0.0666	0.0620
3	0.2513	0.2381	0.2254	0.2133	0.2017	0.1906	0.1801	0.1700	0.1604	0.1512
4	0.4231	0.4061	0.3895	0.3733	0.3575	0.3422	0.3272	0.3127	0.2987	0.2&51
5	0.5984	0.5809	0.5635	0.5461	0.5289	0.5119	0.4950	0.4783	0.4619	0.4457
6	0.7474	0.7324	0.7171	0.7017	0.6860	0.6703	0.6544	0.6384	0.6224	0.6063
7	0.8560	0.8449	0.8335	0.8217	0.8095	0.7970	0.7842	0.7710	0.7576	0.7440
8	0.9252	0.9181	0.9106	0.9027	0.8944	0.8857	0.8766	0.8672	0.8574	0.8472
9	0.9644	0.9603	0.9559	0.9512	0.9462	0.9409	0.9352	0.9292	0.9228	0.9161
10	0.9844	0.9823	0.9800	0.9775	0.9747	0.9718	0.9686	0.9651	0.9614	0.9574
11	0.9937	0.9927	0.9916	0.9904	0.9890	0.9875	0.9859	0.9841	0.9821	0.9799
12	0.9976	0.9972	0.9967	0.9962	0.9955	0.9949	0.9941	0.9932	0.9922	0.9912
13	0.9992	0.9990	0.9988	0.9986	0.9983	0.9980	0.9977	0.9973	0.9969	0.9964
14	0.9997	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991	0.9990	0.9988	0.9986
15	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9996	0.9996	0.9995
16	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

					λ					
A	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	0.0022	0.0020	0.0018	0.0017	0.0015	0.0014	0.0012	0.0011	0.0010	0.0009
1	0.0159	0.0146	0.0134	0.0123	0.0113	0.0103	0.0095	0.0087	0.0080	0.0073
2	0.0577	0.0536	0.0498	0.0463	0.0430	0.0400	0.0371	0.0344	0.0320	0.0296
3	0.1425	0.1342	0.1264	0.1189	0.1119	0.1052	0.0988	0.0928	0.0871	0.0818
4	0.2719	0.2592	0.2469	0.2351	0.2237	0.2127	0.2022	0.1920	0.1823	0.1730
5	0.4298	0.4141	0.3988	0.3837	0.3690	0.3547	0.3407	0.3270	0.3137	0.3007
6	0.5902	0.5742	0.5582	0.5423	0.5265	0.5108	0.4953	0.4799	0.4647	0.4497
7	0.7301	0.7160	0.7018	0.6873	0.6728	0.6581	0.6433	0.6285	0.6136	0.5987
8	0.8367	0.8259	0.8148	0.8033	0.7916	0.7796	0.7673	0.7548	0.7420	0.7291
9	0.9090	0.9016	0.8939	0.8858	0.8774	0.8686	0.8596	0.8502	0.8405	0.8305
10	0.9531	0.9486	0.9437	0.9386	0.9332	0.9274	0.9214	0.9151	0.9084	0.9015
11	0.9776	0.9750	0.9723	0.9693	0.9661	0.9627	0.9591	0.9552	0.9510	0.9467
12	0.9900	0.9887	0.9873	0.9857	0.9840	0.9821	0.9801	0.9779	0.9755	0.9730
13	0.9958	0.9952	0.9945	0.9937	0.9929	0.9920	0.9909	0.9898	0.9885	0.9872
14	0.9984	0.9981	0.9978	0.9974	0.9970	0.9966	0.9961	0.9956	0.9950	0.9943
15	0.9994	0.9993	0.9992	0.9990	0.9988	0.9986	0.9984	0.9982	0.9979	0.9976
16	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9994	0.9993	0.9992	0.9990
17	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996
18	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
19	1.0000	1.0000	1.0000	1.0000	1.0000	1 0000	1.0000	1.0000	1.0000	1.0000

					λ					
A	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	0.0008	0.0007	0.0007	0.0006	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003
1	0.0067	0.0061	0.0056	0.0051	0.0047	0.0043	0.0039	0.0036	0.0033	0.0030
2	0.0275	0.0255	0.0236	0.0219	0.0203	0.0188	0.0174	0.0161	0.0149	0.0138
3	0.0767	0.0719	0.0674	0.0632	0.0591	0.0554	0.0518	0.0485	0.0453	0.0424
4	0.1641	0.1555	0.1473	0.1395	0.1321	0.1249	0.1181	0.1117	0.1055	0.0996
5	0.2881	0.2759	0.2640	0.2526	0.2414	0.2307	0.2203	0.2103	0.2006	0.1912
6	0.4349	0.4204	0.4060	0.3920	0.3782	0.3646	0.3514	0.3384	0.3257	0.3134
7	0.5838	0.5689	0.5541	0.5393	0.5246	0.5100	0.4956	0.4812	0.4670	0.4530
8	0.7160	0.7027	0.6892	0.6757	0.6620	0.6482	0.6343	0.6204	0.6065	0.5926
9	0.8202	0.8097	0.7988	0.7877	0.7764	0.7649	0.7531	0.7411	0.7290	0.7166
10	0.8942	0.8867	0.8788	0.8707	0.8622	0.8535	0.8445	0.8352	0.8257	0.8159
11	0.9420	0.9371	0.9319	0.9265	0.9208	0.9148	0.9085	0.9020	0.8952	0.8881
12	0.9703	0.9673	0.9642	0.9609	0.9573	0.9536	0.9496	0.9454	0.9409	0.9362
13	0.9857	0.9841	0.9824	0.9805	0.9784	0.9762	0.9739	0.9714	0.9687	0.9658
14	0.9935	0.9927	0.9918	0.9908	0.9897	0.9886	0.9873	0.9859	0.9844	0.9827
15	0.9972	0.9969	0.9964	0.9959	0.9954	0.9948	0.9941	0.9934	0.9926	0.9918
16	0.9989	0.9987	0.9985	0.9983	0.9980	0.9978	0.9974	0.9971	0.9967	0.9963
17	0.9996	0.9995	0.9994	0.9993	0.9992	0.9991	0.9989	0.9988	0.9986	0.9984
18	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9994	0.9993
19	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Función de distribución de Poisson

					λ					
A	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
1	0.0028	0.0025	0.0023	0.0021	0.0019	0.0018	0.0016	0.0015	0.0014	0.0012
2	0.0127	0.0118	0.0109	0.0100	0.0093	0.0086	0.0079	0.0073	0.0068	0.0062
3	0.0396	0.0370	0.0346	0.0323	0.0301	0.0281	0.0262	0.0244	0.0228	0.0212
4	0.0941	0.0887	0.0837	0.0789	0.0744	0.0701	0.0660	0.0621	0.0584	0.0550
5	0.1823	0.1736	0.1653	0.1573	0.1496	0.1422	0.1352	0.1284	0.1219	0.1157
6	0.3013	0.2896	0.2781	0.2670	0.2562	0.2457	0.2355	0.2256	0.2160	0.2068
7	0.4391	0.4254	0.4119	0.3987	0.3856	0.3728	0.3602	0.3478	0.3357	0.3239
8	0.5786	0.5647	0.5508	0.5369	0.5231	0.5094	0.4958	0.4823	0.4689	0.4557
9	0.7041	0.6915	0.6788	0.6659	0.6530	0.6400	0.6269	0.6137	0.6006	0.5874
10	0.8058	0.7956	0.7850	0.7743	0.7634	0.7522	0.7409	0.7294	0.7178	0.7060
11	0.8807	0.8731	0.8652	0.8571	0.8487	0.8400	0.8311	0.8220	0.8126	0.8030
12	0.9313	0.9261	0.9207	0.9150	0.9091	0.9029	0.8965	0.8898	0.8829	0.8758
13	0.9628	0.9595	0.9561	0.9524	0.9486	0.9445	0.9403	0.9358	0.9311	0.9262
14	0.9810	0.9791	0.9771	0.9749	0.9726	0.9701	0.9675	0.9647	0.9617	0.9585
15	0.9908	0.9898	0.9887	0.9875	0.9862	0.9848	0.9832	0.9816	0.9798	0.9780
16	0.9958	0.9953	0.9947	0.9941	0.9934	0.9926	0.9918	0.9909	0.9899	0.9889
17	0.9982	0.9979	0.9977	0.9973	0.9970	0.9966	0.9962	0.9957	0.9952	0.9947
18	0.9992	0.9991	0.9990	0.9989	0.9987	0.9985	0.9983	0.9981	0.9978	0.9976
19	0.9997	0.9997	0.9996	0.9995	0.9995	0.9994	0.9993	0.9992	0.9991	0.9989
20	0.9999	0.9999	0.9998	0.9998	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996
21	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

					λ					
A	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10.0
0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
1	0.0011	0.0010	0.0009	0.0009	0.0008	0.0007	0.0007	0.0006	0.0005	0.0005
2	0.0058	0.0053	0.0049	0.0045	0.0042	0.0038	0.0035	0.0033	0.0030	0.0028
3	0.0198	0.0184	0.0172	0.0160	0.0149	0.0138	0.0129	0.0120	0.0111	0.0103
4	0.0517	0.0486	0.0456	0.0429	0.0403	0.0378	0.0355	0.0333	0.0312	0.0293
5	0.1098	0.1041	0.0987	0.0935	0.0885	0.0838	0.0793	0.0750	0.0710	0.0671
6	0.1978	0.1892	0.1808	0.1727	0.1650	0.1575	0.1502	0.1433	0.1366	0.1301
7	0.3123	0.3010	0.2900	0.2792	0.2687	0.2584	0.2485	0.2388	0.2294	0.2202
8	0.4426	0.4296	0.4168	0.4042	0.3918	0.3796	0.3676	0.3558	0.3442	0.3328
9	0.5742	0.5611	0.5480	0.5349	0.5218	0.5089	0.4960	0.4832	0.4705	0.4579
10	0.6941	0.6820	0.6699	0.6576	0.6453	0.6330	0.6205	0.6081	0.5956	0.5830
11	0.7932	0.7832	0.7730	0.7626	0.7570	0.7412	0.7303	0.7193	0.7081	0.6968
12	0.8684	0.8607	0.8529	0.8448	0.8364	0.8279	0.8191	0.8101	0.8009	0.7916
13	0.9210	0.9156	0.9100	0.9042	0.8981	0.8919	0.8853	0.8786	0.8716	0.8645
14	0.9552	0.9517	0.9480	0.9441	0.9400	0.9357	0.9312	0.9265	0.9216	0.9165
15	0.9760	0.9738	0.9715	0.9691	0.9665	0.9638	0.9609	0.9579	0.9546	0.9513
16	0.9878	0.9865	0.9852	0.9838	0.9823	0.9806	0.9789	0.9770	0.9751	0.9730
17	0.9941	0.9934	0.9927	0.9919	0.9911	0.9902	0.9892	0.9881	0.9870	0.9857
18	0.9973	0.9969	0.9966	0.9962	0.9957	0.9952	0.9947	0.9941	0.9935	0.9928
19	0.9988	0.9986	0.9985	0.9983	0.9980	0.9978	0.9975	0.9972	0.9969	0.9965
20	0.9995	0.9994	0.9993	0.9992	0.9991	0.9990	0.9989	0.9987	0.9986	0.9984
21	0.9998	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9995	0.9994	0.9993
22	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9998	0.9997	0.9997
23	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Función de distribución de Poisson

	λ								
A	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000
5	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003	0.0002	0.0001
6	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010	0.0005	0.0003
7	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029	0.0015	0.0008
8	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071	0.0039	0.0021
9	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154	0.0089	0.0050
10	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304	0.0183	0.0108
11	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549	0.0347	0.0214
12	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917	0.0606	0.0390
13	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426	0.0984	0.0661
14	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081	0.1497	0.1049
15	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867	0.2148	0.1565
16	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751	0.2920	0.2211
17	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686	0.3784	0.2970
18	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622	0.4695	0.3814
19	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509	0.5606	0.4703
20	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307	0.6472	0.5591
21	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991	0.7255	0.6437
22	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551	0.7931	0.7206
23	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989	0.8490	0.7875
24	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317	0.8933	0.8432
25	0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554	0.9269	0.8878
26	1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718	0.9514	0.9221
27	1.0000	0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827	0.9687	0.9475
28	1.0000	1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897	0.9805	0.9657
29	1.0000	1.0000	1.0000	0.9999	0.9996	0.9989	0.9973	0.9941	0.9882	0.9782
30	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986	0.9967	0.9930	0.9865
31	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9982	0.9960	0.9919
32	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9990	0.9978	0.9953
33	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995	0.9988	0.9973
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9985
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992
36	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996
37	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
38	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

Tabla 3

Función de distribución y de probabilidad Hipergeométrica

N	n	k	a	F(a)	P(a)
2	1	1	0	0,500000	0,500000
2	1	1	1	1,000000	0,500000
3	1	1	0	0,666667	0,666667
3	1	1	1	1,000000	0,333333
3	2	1	0	0,333333	0,333333
3	2	1	1	1,000000	0,666667
3	2	2	1	0,666667	0,666667
3	2	2	2	1,000000	0,333333
4	1	1	0	0,750000	0,760000
4	1	1	1	1,000000	0,250000
4	2	1	0	0,500000	0,500000
4	2	1	1	1,000000	0,500000
4	2	2	0	0,166667	0,166667
4	2	2	1	0,833333	0,666667
4	2	2	2	1,000000	0,166667
4	3	1	0	0,250000	0,250000
4	3	1	1	1,000000	0,750000
4	3	2	1	0,500000	0,500000
4	3	2	2	1,000000	0,500000
4	3	3	2	0,750000	0,750000
4	3	3	3	1,000000	0,250000
5	1	1	0	0,800000	0,800000
5	1	1	1	1,000000	0,200000
5	2	1	0	0,600000	0,600000
5	2	1	1	1,000000	0,400000
5	2	2	0	0,300000	0,300000
5	2	2	1	0,900000	0,600000
5	2	2	2	1,000000	0,100000
5	3	1	0	0,400000	0,400000
5	3	1	1	1,000000	0,600000
5	3	2	0	0,100000	0,100000
5	3	2	1	0,700000	0,600000
5	3	2	2	1,000000	0,300000
5	3	3	1	0,300000	0,300000
5	3	3	2	0,900000	0,600000
5	3	3	3	1,000000	0,100000
5	4	1	0	0,200000	0,200000
5	4	1	1	1,000000	0,800000
5	4	2	1	0,400000	0,400000
5	4	2	2	0,000000	0,600000
5	4	3	2	0,600000	0,600000
5	4	3	3	1,000000	0,400000
5	4	4	3	0,800000	0,800000
5	4	4	4	1,000000	0,200000
6	1	1	0	0,833333	0,833333
6	1	1	1	1,000000	0,166667
6	2	1	0	0,666667	0,666667
6	2	1	1	1,000000	0,333333
6	2	2	0	0,400000	0,400000
6	2	2	1	0,933333	0,533333
6	2	2	2	1,000000	0,066667
6	3	1	0	0,500000	0,500000
6	3	1	1	1,000000	0,500000
6	3	2	0	0,200000	0,200000
6	3	2	1	0,800000	0,600000
6	3	2	2	1,000000	0,200000
6	3	3	0	0,050000	0,050000
6	3	3	1	0,500000	0,450000
6	3	3	2	0,950000	0,450000
6	3	3	3	1,000000	0,050000
6	4	1	0	0,333333	0,333333
6	4	1	1	1,000000	0,666667
6	4	2	0	0,066667	0,066667
6	4	2	1	0,600000	0,533333
6	4	2	2	1,000000	0,400000
6	4	3	1	0,200000	0,200000
6	4	3	2	0,800000	0,600000
6	4	3	3	1,000000	0,200000
6	4	4	2	0,400000	0,400000
6	4	4	3	0,933333	0,533333
6	4	4	4	1,000000	0,066667
6	5	1	0	0,166667	0,166667
6	5	1	1	1,000000	0,833333
6	5	2	1	0,333333	0,333333
6	5	2	2	1,000000	0,666667
6	5	3	2	0,500000	0,500000
6	5	3	3	1,000000	0,500000
6	5	4	3	0,666667	0,666667
6	5	4	4	1,000000	0,333333
6	5	5	4	0,833333	0,833333
6	5	5	5	1,000000	0,166667
7	1	1	0	0,857143	0,857143
7	1	1	1	1,000000	0,142857
7	2	1	0	0,714286	0,714286
7	2	1	1	1,000000	0,285714
7	2	2	0	0,476190	0,476190
7	2	2	1	0,952381	0,476190
7	2	2	2	1,000000	0,047619
7	3	1	0	0,571429	0,571429
7	3	1	1	1,000000	0,428571
7	3	2	0	0,285714	0,285714
7	3	2	1	0,857143	0,571429
7	3	2	2	1,000000	0,142857
7	3	3	0	0,114286	0,114286
7	3	3	1	0,628571	0,514286
7	3	3	2	0,971428	0,342857
7	3	3	3	1,000000	0,028571
7	4	1	0	0,428571	0,428571
7	4	1	1	1,000000	0,571429
7	4	2	0	0,142857	0,142857
7	4	2	1	0,714286	0,571429
7	4	2	2	1,000000	0,285714
7	4	3	0	0,025571	0,028571
7	4	3	1	0,371429	0,342857
7	4	3	2	0,885714	0,514286
7	4	3	3	1,000000	0,114286
7	4	4	1	0,114286	0,114286
7	4	4	2	0,628571	0,514286
7	4	4	3	0,971428	0342857
7	4	4	4	1,000000	0,028571
7	5	1	0	0,285714	0,285714
7	5	1	1	1,000000	0,714286
7	5	2	0	0,047619	0,047619
7	5	2	1	0,523809	0,476190
7	5	2	2	1,000000	0,476190
7	5	3	1	0,142857	0,142857
7	5	3	2	0,714286	0,571429
7	5	3	3	1,000000	0,285714
7	5	4	2	0,285714	0,285714
7	5	4	3	0,857143	0,571429
7	5	4	4	1,000000	0,142857
7	5	5	3	0,476190	0,476190
7	5	5	4	0,952381	0,476190
7	5	5	5	1,000000	0,047619
7	6	1	0	0,142857	0,142857
7	6	1	1	1,000000	0,857143
7	6	2	1	0,285714	0,285714
7	6	2	2	1,000000	0,714286
7	6	3	2	0,428571	0,428571
7	6	3	3	1,000000	0,571429
7	6	4	3	0,571429	0,571429
7	6	4	4	1,000000	0,428571
7	6	5	4	0,714286	0,714286
7	6	5	5	1,000000	0,285714
7	6	6	5	0,857143	0,857143
7	6	6	6	1,000000	0,142857
8	1	1	0	0,875000	0,875000
8	1	1	1	1,000000	0,125000
8	2	1	0	0,750000	0,750000
8	2	1	1	1,000000	0,250000
8	2	2	0	0,535714	0,535714
8	2	2	1	0,964286	0,428571
8	2	2	2	1,000000	0,035714
8	3	1	0	0,625000	0,625000
8	3	1	1	1,000000	0,375000
8	3	2	0	0,357143	0,357143
8	3	2	1	0,892857	0,535714
8	3	2	2	1,000000	0,107143
8	3	3	0	0,178571	0,178571
8	3	3	1	0,714286	0,535714
8	3	3	2	0,982143	0,267857
8	3	3	3	1,000000	0,017857
8	4	1	0	0,500000	0,500000
8	4	1	1	1,000000	0,500000
8	4	2	0	0,214286	0,214286
8	4	2	1	0,785714	0,571429
8	4	2	2	1,000000	0,214286
8	4	3	0	0,071429	0,071429
8	4	3	1	0,500000	0,428571
8	4	3	2	0,928571	0,428571

Función de distribución y de probabilidad Hipergeométrica

N	n	k	a	F(a)	P(a)	N	n	k	a	F(a)	P(a)
8	4	3	3	1.000000	0.071429	9	5	4	3	0.960317	0.317460
8	4	4	0	0.014286	0.014286	9	5	4	4	1.000000	0.039683
8	4	4	1	0.242857	0.228571	9	5	5	1	0.039683	0.039683
8	4	4	2	0.757143	0.5142S6	9	5	5	2	0.357143	0.317460
8	4	4	3	0.985714	0.228571	9	5	5	3	0.833333	0.476190
8	4	4	4	1.000000	0.014286	9	5	5	4	0.992063	0.158730
8	5	1	0	0.375000	0.375000	9	5	5	5	1.000000	0.007936
8	5	1	1	1.000000	0.625000	9	6	1	0	0.333333	0.333333
8	5	2	0	0.107143	0.107143	9	6	1	1	1.000000	0.666667
8	5	2	1	0.642857	0.5357i4	9	6	2	0	0.083333	0.083333
8	5	2	2	1.000000	0.357143	9	6	2	1	0.583333	0.500000
8	5	3	0	0.017857	0.017857	9	6	2	2	1.000000	0.416667
8	5	3	1	0.285714	0.267857	9	6	3	0	0.011905	0.011905
8	5	3	2	0.821429	0.535714	9	6	3	1	0.226190	0.214286
8	5	3	3	1.000000	0.178571	9	6	3	2	0.761905	0.535714
8	5	4	1	0.071429	0.071429	9	6	3	3	1.000000	0.238095
8	5	4	2	0.500000	0.428571	9	6	4	1	0.047619	0.047619
8	5	4	3	0.928571	0.428571	9	6	4	2	0.404762	0.357143
8	5	4	4	1.000000	0.071429	9	6	4	3	0.880952	0.476190
8	5	5	2	0.178571	0.178571	9	6	4	4	1.000000	0.119048
8	5	5	3	0.714286	0.535714	9	6	5	2	0.119048	0.119048
8	5	5	4	0.982143	0.267857	9	6	5	3	0.595238	0.476190
8	5	5	5	1.000000	0.017857	9	6	5	4	0.952381	0.357143
8	6	1	0	0.250000	0.250000	9	6	5	5	1.000000	0.047619
8	6	1	1	1.000000	0.750000	9	6	6	3	0.238095	0.238095
8	6	2	0	0.035714	0.035714	9	6	6	4	0.773809	0.535714
8	6	2	1	0.464286	0.428571	9	6	6	5	0.988095	0.214286
8	6	2	2	1.000000	0.535714	9	6	6	6	1.000000	0.011905
8	6	3	1	0.107143	0.107143	9	7	1	0	0.222222	0.222222
8	6	3	2	0.642857	0.535714	9	7	1	1	1.000000	0.777778
8	6	3	3	1.000000	0.357143	9	7	2	0	0.027778	0.027778
8	6	4	2	0.214286	0.214286	9	7	2	1	0.416667	0.388889
8	6	4	3	0.785714	0.571429	9	7	2	2	1.000000	0.583333
8	6	4	4	1.000000	0.214286	9	7	3	1	0.083333	0.083333
8	6	5	3	0.357143	0.357143	9	4	4	1	0.357143	0.317460
8	6	5	4	0.892857	0.535714	9	4	4	2	0.833333	0.476190
8	6	5	5	1.000000	0.107143	9	4	4	3	0.992063	0.158730
8	6	6	4	0.535714	0.535714	9	4	4	4	1.000000	0.007936
8	6	6	5	0.964286	0.428571	9	5	1	0	0.444444	0.444444
8	6	6	6	1.000000	0.035714	9	5	1	1	1.000000	0.555556
8	7	1	0	0.125000	0.125000	9	5	2	0	0.166667	0.166667
8	7	1	1	1.000000	0.875000	9	5	2	1	0.722222	0.555556
8	7	2	1	0.250000	0.250000	9	5	2	2	1.000000	0.277778
8	7	2	2	1.000000	0.750000	9	5	3	0	0.047619	0.047619
8	7	3	2	0.375000	0.375000	9	7	6	6	1.000000	0.833333
8	7	3	3	1.000000	0.625000	9	7	7	5	0.583333	0.583333
8	7	4	3	0.500000	0.500000	9	7	7	6	0.972222	0.388889
8	7	4	4	1.000000	0.500000	9	7	7	7	1.000000	0.027778
8	7	5	4	0.625000	0.625000	9	8	1	0	0.111111	0.111111
8	7	5	5	1.000000	0.375000	9	8	1	1	1.000000	0.888889
8	7	6	5	0.750000	0.750000	9	8	2	1	0.222222	0.222222
8	7	6	6	1.000000	0.250000	9	8	2	2	1.000000	0.777778
8	7	7	6	0.875000	0.875000	9	8	3	2	0.333333	0.333333
8	7	7	7	1.000000	0.125000	9	8	3	3	1.000000	0.666667
9	1	1	0	0.888889	0.888889	9	8	4	3	0.444444	0.444444
9	1	1	1	1.000000	0.111111	9	8	4	4	1.000000	0.555556
9	2	1	0	0.777778	0.777778	9	8	5	4	0.555556	0.555556
9	2	1	1	1.000000	0.222222	9	8	5	5	1.000000	0.444444
9	2	2	0	0.583333	0.583333	9	8	6	5	0.666667	0.666667
9	2	2	1	0.972222	0.388889	9	8	6	6	1.000000	0.333333
9	2	2	2	1.000000	0.027778	9	8	7	6	0.777778	0.777778
9	3	1	0	0.666667	0.666667	9	8	7	7	1.000000	0.222222
9	3	1	1	1.000000	0.333333	9	8	8	7	0.888889	0.888889
9	3	2	0	0.416667	0.416667	9	8	8	8	1.000000	0.111111
9	3	2	1	0.916667	0.500000	9	7	3	2	0.583333	0.500000
9	3	2	2	1.000000	0.083333	9	7	3	3	1.000000	0.416667
9	3	3	0	0.238095	0.238095	9	7	4	2	0.166667	0.166667
9	3	3	1	0.773809	0.535714	9	7	4	3	0.722222	0.555556
9	3	3	2	0.988095	0.214286	9	7	4	4	1.000000	0.277778
9	3	3	3	1.000000	0.011905	9	7	5	3	0.277778	0.277778
9	4	1	0	0.555556	0.555556	9	7	5	4	0.833333	0.555556
9	4	1	1	1.000000	0.444444	9	7	5	5	1.000000	0.166667
9	4	2	0	0.277778	0.277778	9	7	6	4	0.416667	0.416667
9	4	2	1	0.833333	0.555556	9	7	6	5	0.916667	0.500000
9	4	2	2	1.000000	0.166667	10	5	1	0	0.500000	0.500000
9	4	3	0	0.119048	0.119048	10	5	1	1	1.000000	0.500000
9	4	3	1	0.595238	0.476190	10	5	2	0	0.222222	0.222222
9	4	3	2	0.952381	0.357143	10	5	2	1	0.777778	0.555556
9	4	3	3	1.000000	0.047619	10	5	2	2	1.000000	0.222222
9	4	4	0	0.039683	0.039683	10	5	3	0	0.083333	0.083333
9	5	3	1	0.404762	0.357143	10	5	3	1	0.500000	0.416667
9	5	3	2	0.880952	0.476190	10	5	3	2	0.916667	0.416667
9	5	3	3	1.000000	0.119048	10	5	3	3	1.000000	0.083333
9	5	4	0	0.007936	0.007936	10	5	4	0	0.023810	0.023810
9	5	4	1	0.166667	0.158730	10	5	4	1	0.261905	0.238095
9	5	4	2	0.642857	0.476190	10	5	4	2	0.738095	0.476190

Función de distribución y de probabilidad Hipergeométrica

N	*n*	*k*	*a*	*F(a)*	*P(a)*
10	5	4	3	0,976190	0,238095
10	5	4	4	1,000000	0,023810
10	5	5	0	0,003968	0,003968
10	5	5	1	0,103175	0,099206
10	5	5	2	0,500000	0,396825
10	5	5	3	0,896825	0,396825
10	5	5	4	0,996032	0,099206
10	5	5	5	1,000000	0,003968
10	1	1	0	0,900000	0,900000
10	1	1	1	1,000000	0,100000
10	2	1	0	0,800000	0,800000
10	2	1	1	1,000000	0,200000
10	2	2	0	0,622222	0,622222
10	2	2	1	0,977778	0,355556
10	2	2	2	1,000000	0,022222
10	3	1	0	0,700000	0,700000
10	3	1	1	1,000000	0,300000
10	3	2	0	0,466667	0,466667
10	3	2	1	0,933333	0,466667
10	3	2	2	1,000000	0,066667
10	3	3	0	0,291667	0,291667
10	3	3	1	0,816667	0,525000
10	3	3	2	0,991667	0,175000
10	3	3	3	1,000000	0,008333
10	4	1	0	0,600000	0,600000
10	4	1	1	1,000000	0,400000
10	4	2	0	0,333333	0,333333
10	4	2	1	0,866667	0,533333
10	4	2	2	1,000000	0,133333
10	4	3	0	0,166667	0,166667
10	4	3	1	0,666667	0,500000
10	4	3	2	0,966667	0,300000
10	4	3	3	1,000000	0,033333
10	4	4	0	0,071429	0,071429
10	4	4	1	0,452381	0,380952
10	4	4	2	0,880952	0,428571
10	4	4	3	0,995238	0,114286
10	4	4	4	1,000000	0,004762
10	6	1	0	0,400000	0,400000
10	6	1	1	1,000000	0,600000
10	6	2	0	0,133333	0,133333
10	6	2	1	0,666667	0,533333
10	6	2	2	1,000000	0,333333
10	6	3	0	0,033333	0,033333
10	6	3	1	0,333333	0,300000
10	6	3	2	0,833333	0,500000
10	6	3	3	1,000000	0,166667
10	6	4	0	0,004762	0,004762
10	6	4	1	0,119048	0,114286
10	6	4	2	0,547619	0,428571
10	6	4	3	0,928571	0,380952
10	6	4	4	1,000000	0,071429
10	6	5	1	0,023810	0,023810
10	6	5	2	0,261905	0,238095
10	6	5	3	0,738095	0,476190
10	6	5	4	0,976190	0,238095
10	6	5	5	1,000000	0,023810
10	6	6	2	0,071429	0,071429
10	6	6	3	0,452381	0,380952
10	6	6	4	0,880952	0,428571
10	6	6	5	0,995238	0,114286
10	6	6	6	1,000000	0,004762
10	7	1	0	0,300000	0,300000
10	7	1	1	1,000000	0,700000
10	7	2	0	0,066667	0,066667
10	7	2	1	0,533333	0,466667
10	7	2	2	1,000000	0,466667
10	7	3	0	0,008333	0,008333

Tabla 4

Función de distribución Normal (0, 1)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.5	0,0002	0.0002	0,0002	0,0002	0,0002	0,0002	0.0002	0,0002	0.0002	0.0002
-3.4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3.3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3.2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3.1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3.0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2.9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2.8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2.7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2.6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2.5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2.4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2.3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2.2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2.1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2.0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1.9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1.8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1.7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1.6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1.5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1.4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1.3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1.2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1.1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1.0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0.9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0.8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0.7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2206	0,2177	0,2148
-0.6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0.5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0.4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0.3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0.2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0.1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0.0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0.0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0.1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0.2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0.3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0.4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0.5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0.6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0.7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0.8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0.9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1.0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1.1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1.2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1.3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1.4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1.5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1.6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1.7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1.8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1.9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2.0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2.1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2.2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2.3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2.4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2.5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2.6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2.7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2.8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2.9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3.0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3.1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3.2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3.3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3.4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3.5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Tabla 5

Función de distribución de la Normal Circular

Probabilidad Acumulada	Radio en la normal circular	Distancia Znormal de probabilidad equivalente
0,1000	0,4590	-1,2815
0,2000	0,6680	-0,8416
0,3000	0,8446	-0,5244
0,4000	1,0108	-0,2534
0,5000	1,1774	0,0000
0,6000	1,3537	0,2533
0,7000	1,5518	0,5244
0,8000	1,7941	0,8416
0,9000	2,1460	1,2816
0,9500	2,4477	1,6448
0,9750	2,7162	1,9600
0,9990	3,7169	3,0902

Función de distribución de la Normal Circular

Radio en la normal circular	Probabilidad Acumulada	Distancia Znormal de probabilidad equivalente
0,00	0,00000	
0,10	0,00499	-2,57660
0,20	0,01980	-2,05786
0,30	0,04400	-1,70600
0,40	0,07689	-1,42634
0,50	0,11751	-1,18755
0,60	0,16474	-0,97517
0,70	0,21730	-0,78133
0,80	0,27386	-0,60118
0,90	0,33303	-0,43156
1,00	0,39348	-0,27026
1,10	0,45393	-0,11573
1,20	0,51326	0,03324
1,30	0,57045	0,17751
1,40	0,62469	0,31783
1,50	0,67535	0,45473
1,60	0,72196	0,58869
1,70	0,76426	0,72006
1,80	0,80210	0,84916
1,90	0,83553	0,97624
2,00	0,86467	1,10153
2,10	0,88975	1,22520
2,20	0,91108	1,34743
2,30	0,92900	1,46835
2,40	0,94387	1,58808
2,50	0,95606	1,70673
2,60	0,96595	1,82438
2,70	0,97388	1,94113
2,80	0,98016	2,05704
2,90	0,98508	2,17219
3,00	0,98889	2,28662
3,10	0,99181	2,40041
3,20	0,99402	2,51355
3,30	0,99568	2,62615
3,40	0,99691	2,73823
3,50	0,99781	2,84977
3,60	0,99847	2,96088
3,70	0,99894	3,07162
3,80	0,99927	3,18192
3,90	0,99950	3,29164
3,95	0,99959	3,34636
3,96	0,99961	3,35742
3,97	0,99962	3,36848
3,98	0,99964	3,37954
3,99	0,99965	3,39031
4,00	0,99966	3,40107

Tabla 6

Función de distribución χ^2

	P									
g.l.	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
1	0,00	0,00	0.00	0,00	0,02	2,71	3,84	5,02	6,64	7,90
2	0,01	0,02	0.05	0.10	0,21	4.60	5,99	7,38	9.22	10,59
3	0,07	0,11	0,22	0.35	0,58	6,25	7,82	9,36	11,32	12,82
4	0,21	0,30	0.48	0,71	1,06	7.78	9,49	11,15	13,28	14.82
5	0.41	0,55	0,83	1.15	1,61	9,24	11.07	12.84	15,09	16,76
6	0,67	0,87	1.24	1,63	2,20	10,65	12,60	14,46	16.81	18,55
7	0.99	1,24	1,69	2,17	2,83	12,02	14.07	16,02	18,47	20.27
8	1,34	1,64	2,18	2,73	3.49	13,36	15,51	17,55	20.08	21,94
9	1,73	2,09	2,70	3,32	4,17	14,69	16.93	19,03	21.65	23,56
10	2,15	2.55	3,24	3,94	4.86	15,99	18,31	20.50	23.19	25,15
11	2,60	3,05	3.81	4,57	5,58	17.28	19.68	21,93	24,75	26,71
12	3,06	3.57	4,40	5.22	6.30	18,55	21,03	23,35	26,25	28,25
13	3,56	4,10	5,01	5,89	7,04	19.81	22,37	24,75	27,72	29,88
14	4,07	4,65	5.62	6,57	7,79	21.07	23,69	26.13	29.17	31,38
15	4.59	5,23	6,26	7.26	8,55	22,31	25.00	27,50	30,61	32,86
16	5,14	5,81	6,90	7,96	9,31	23,55	26,30	28,86	32,03	34,32
17	5.69	6,40	7,56	8.67	10,08	24,77	27,59	30,20	33,43	35,77
18	6,25	7.00	8,23	9,39	10,86	25,99	28,88	31.54	34.83	37,21
19	6.82	7,63	8.90	10,11	11,65	27,21	30,15	32,87	36,22	38,63
20	7,42	8,25	9,59	10,85	12,44	28,42	31,42	34,18	37,59	40,05
21	8,02	8,89	10,28	11,59	13,24	29,62	32,68	35,49	38,96	41,45
22	8,62	9,53	10,98	12,34	14,04	30,82	33,93	36,79	40,31	42,84
23	9,25	10,19	11,69	13,09	14,85	32,01	35,18	38,09	41,66	44,23
24	9.87	10,85	12,40	13,84	15,66	33,20	36,42	39,38	43,00	45,60
25	10.50	11,51	13,11	14,61	16,47	34,38	37,66	40,66	44,34	46,97
26	11,13	12,19	13,84	15,38	17,29	35,57	38,89	41,94	45,66	48.33
27	11,79	12,87	14,57	16,15	18,11	36,74	40,12	43,21	46,99	49,69
28	12,44	13,55	15,30	16,92	18,94	37,92	41,34	44.47	48,30	51,04
29	13,09	14,24	16,04	17,70	19,77	39,09	42,56	45,74	49,61	52,38
30	13,77	14,94	16,78	18,49	20,60	40,26	43,78	46,99	50,91	53,71
35	17,16	18,49	20,56	22,46	24,79	46,06	49,81	53,22	57,36	60,31
40	20,67	22,14	24,42	26,51	29,06	51,80	55,75	59,34	63,71	66,80
45	24,28	25,88	28,36	30,61	33,36	57,50	61,65	65,41	69,98	73,20
50	27,96	29,68	32,35	34,76	37,69	63,16	67,50	71,42	76,17	79,52
60	35,50	37,46	40,47	43,19	46,46	74,39	79,08	83,30	88,40	91,98
70	43,25	45,42	48,75	51,74	55,33	85,52	90,53	95,03	100,44	104,24
80	51,14	53,52	57,15	60,39	64,28	96,57	101,88	106,63	112,34	116,35
90	59,17	61,74	65,64	69,13	73,29	107,56	113,14	118,14	124,13	128,32
100	67,30	70,05	74,22	77,93	82,36	118,49	124,34	129,56	135,82	140,19

Tabla 7

Función de distribución t-Student

g.l.	0.001	0.005	0.010	0.025	0.050	0.100	0.200	0.800	0.900	0.950	0.975	0.990	0.995	0.999
1	-318,30	-63,657	-31,821	-12,706	-6,314	-3,078	-1,376	1,376	3,078	6,314	12,706	31,821	63,657	318,309
2	-22,327	-9,925	-6,965	-4,303	-2,920	-1,886	-1,061	1,061	1,886	2,920	4,303	6,965	9,925	22,327
3	-10,215	-5,841	-4,541	-3,182	-2,353	-1,638	-0,978	0,978	1,638	2,353	3,182	4,541	5,841	10,215
4	-7,173	-4,604	-3,747	-2,776	-2,132	-1,533	-0,941	0,941	1,533	2,132	2,776	3,747	4,604	7,173
5	-5,893	-4,032	-3,365	-2,571	-2,015	-1,476	-0,920	0,920	1,476	2,015	2,571	3,365	4,032	5,893
6	-5,208	-3,707	-3,143	-2,447	-1,943	-1,440	-0,906	0,906	1,440	1,943	2,447	3,143	3,707	5,208
7	-4,785	-3,499	-2,998	-2,365	-1,895	-1,415	-0,896	0,896	1,415	1,895	2,365	2,998	3,499	4,785
8	-4,501	-3,355	-2,896	-2,306	-1,860	-1,397	-0,889	0,889	1,397	1,860	2,306	2,896	3,355	4,501
9	-4,297	-3,250	-2,821	-2,262	-1,833	-1,383	-0,883	0,883	1,383	1,833	2,262	2,821	3,250	4,297
10	-4,144	-3,169	-2,764	-2,228	-1,812	-1,372	-0,879	0,879	1,372	1,812	2,228	2,764	3,169	4,144
11	-4,025	-3,106	-2,718	-2,201	-1,796	-1,363	-0,876	0,876	1,363	1,796	2,201	2,718	3,106	4,025
12	-3,930	-3,055	-2,681	-2,179	-1,782	-1,356	-0,873	0,873	1,356	1,782	2,179	2,681	3,055	3,930
13	-3,852	-3,012	-2,650	-2,160	-1,771	-1,350	-0,870	0,870	1,350	1,771	2,160	2,650	3,012	3,852
14	-3,787	-2,977	-2,624	-2,145	-1,761	-1,345	-0,868	0,868	1,345	1,761	2,145	2,624	2,977	3,787
15	-3,733	-2,947	-2,602	-2,131	-1,753	-1,341	-0,866	0,866	1,341	1,753	2,131	2,602	2,947	3,733
16	-3,686	-2,921	-2,583	-2,120	-1,746	-1,337	-0,865	0,865	1,337	1,746	2,120	2,583	2,921	3,686
17	-3,646	-2,898	-2,567	-2,110	-1,740	-1,333	-0,863	0,863	1,333	1,740	2,110	2,567	2,898	3,646
18	-3,610	-2,878	-2,552	-2,101	-1,734	-1,330	-0,862	0,862	1,330	1,734	2,101	2,552	2,878	3,610
19	-3,579	-2,861	-2,539	-2,093	-1,729	-1,328	-0,861	0,861	1,328	1,729	2,093	2,539	2,861	3,579
20	-3,552	-2,845	-2,528	-2,086	-1,725	-1,325	-0,860	0,860	1,325	1,725	2,086	2,528	2,845	3,552
21	-3,527	-2,831	-2,518	-2,080	-1,721	-1,323	-0,859	0,859	1,323	1,721	2,080	2,518	2,831	3,527
22	-3,505	-2,819	-2,508	-2,074	-1,717	-1,321	-0,858	0,858	1,321	1,717	2,074	2,508	2,819	3,505
23	-3,485	-2,807	-2,500	-2,069	-1,714	-1,319	-0,858	0,858	1,319	1,714	2,069	2,500	2,807	3,485
24	-3,467	-2,797	-2,492	-2,064	-1,711	-1,318	-0,857	0,857	1,318	1,711	2,064	2,492	2,797	3,467
25	-3,450	-2,787	-2,485	-2,060	-1,708	-1,316	-0,856	0,856	1,316	1,708	2,060	2,485	2,787	3,450
26	-3,435	-2,779	-2,479	-2,056	-1,706	-1,315	-0,856	0,856	1,315	1,706	2,056	2,479	2,779	3,435
27	-3,421	-2,771	-2,473	-2,052	-1,703	-1,314	-0,855	0,855	1,314	1,703	2,052	2,473	2,771	3,421
28	-3,408	-2,763	-2,467	-2,048	-1,701	-1,313	-0,855	0,855	1,313	1,701	2,048	2,467	2,763	3,408
29	-3,396	-2,756	-2,462	-2,045	-1,699	-1,311	-0,854	0,854	1,311	1,699	2,045	2,462	2,756	3,396
30	-3,385	-2,750	-2,457	-2,042	-1,697	-1,310	-0,854	0,854	1,310	1,697	2,042	2,457	2,750	3,385
35	-3,340	-2,724	-2,438	-2,030	-1,690	-1,306	-0,852	0,852	1,306	1,690	2,030	2,438	2,724	3,340
40	-3,307	-2,704	-2,423	-2,021	-1,684	-1,303	-0,851	0,851	1,303	1,684	2,021	2,423	2,704	3,307
45	-3,281	-2,690	-2,412	-2,014	-1,679	-1,301	-0,850	0,850	1,301	1,679	2,014	2,412	2,690	3,281
50	-3,261	-2,678	-2,403	-2,009	-1,676	-1,299	-0,849	0,849	1,299	1,676	2,009	2,403	2,678	3,261
60	-3,232	-2,660	-2,390	-2,000	-1,671	-1,296	-0,848	0,848	1,296	1,671	2,000	2,390	2,660	3,232
70	-3,211	-2,648	-2,381	-1,994	-1,667	-1,294	-0,847	0,847	1,294	1,667	1,994	2,381	2,648	3,211
80	-3,195	-2,639	-2,374	-1,990	-1,664	-1,292	-0,846	0,846	1,292	1,664	1,990	2,374	2,639	3,195
90	-3,183	-2,632	-2,369	-1,987	-1,662	-1,291	-0,846	0,846	1,291	1,662	1,987	2,369	2,632	3,183
100	-3,174	-2,626	-2,364	-1,984	-1,660	-1,290	-0,845	0,845	1,290	1,660	1,984	2,364	2,626	3,174
200	-3,131	-2,601	-2,345	-1,972	-1,652	-1,286	-0,843	0,843	1,286	1,652	1,972	2,345	2,601	3,131
500	-3,107	-2,586	-2,334	-1,965	-1,648	-1,283	-0,842	0,842	1,283	1,648	1,965	2,334	2,586	3,107
1000	-3,098	-2,581	-2,330	-1,962	-1,646	-1,282	-0,842	0,842	1,282	1,646	1,962	2,330	2,581	3,098

Tabla 8

Función de distribución F de Snedecor

$P = 0.9$

Gl_1

gl_2	1	2	3	4	5	6	7	8	9	10
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94
7	3,59	3,26	3,07	2,96	2,88	2,83	2,79	2,75	2,72	2,70
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25
12	3,18	2,81	2,61	2,48	1,39	2,33	2,28	2,24	2,21	2,19
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82
35	2,85	2,46	2,25	2,11	2,02	1,95	1,90	1,85	1,82	1,79
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71
80	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71	1,68
100	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69	1,66
200	2,73	2,33	2,11	1,97	1,88	1,80	1,75	1,70	1,66	1,63
500	2,72	2,31	2,09	1,96	1,86	1,79	1,73	1,68	1,64	1,61
1000	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61

$P = 0.9$

gl_1

gl_2	11	12	15	20	25	30	40	50	100	1000
1	60,47	60,71	61,22	61,74	62,06	62,26	62,53	62,69	63,00	63,29
2	9,40	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,22	5,22	5,20	5,19	5,17	5,17	5,16	5,15	5,14	5,13
4	3,91	3,90	3,87	3,84	3,83	3,82	3,80	3,80	3,78	3,76
5	3,28	3,27	3,24	3,21	3,19	3,17	3,16	3,15	3,13	3,11
6	2,92	2,90	2,87	2,84	2,81	2,80	2,78	2,77	2,75	2,72
7	2,68	2,67	2,63	2,59	2,57	2,56	2,54	2,52	2,50	2,47
8	2,52	2,50	2,46	2,42	2,40	2,38	2,36	2,35	2,32	2,30
9	2,40	2,38	2,34	2,30	2,27	2,25	2,23	2,22	2,19	2,16
10	2,30	2,28	2,24	2,20	2,17	2,16	2,13	2,12	2,09	2,06
11	2,23	2,21	2,17	2,12	2,10	2,08	2,05	2,04	2,00	1,98
12	2,17	2,15	2,10	2,06	2,03	2,01	1,99	1,97	1,94	1,91
13	2,12	2,10	2,05	2,01	1,98	1,96	1,93	1,92	1,88	1,85
14	2,07	2,05	2,01	1,96	1,93	1,91	1,89	1,87	1,83	1,80
15	2,04	2,02	1,97	1,92	1,89	1,87	1,85	1,83	1,79	1,76
16	2,01	1,99	1,94	1,89	1,86	1,84	1,81	1,79	1,76	1,72
17	1,98	1,96	1,91	1,86	1,83	1,81	1,78	1,76	1,73	1,69
18	1,95	1,93	1,89	1,84	1,80	1,78	1,75	1,74	1,70	1,66
19	1,93	1,91	1,86	1,81	1,78	1,76	1,73	1,71	1,67	1,64
20	1,91	1,89	1,84	1,79	1,76	1,74	1,71	1,69	1,65	1,61
25	1,84	1,82	1,77	1,72	1,68	1,66	1,63	1,61	1,56	1,52
30	1,79	1,77	1,72	1,67	1,63	1,61	1,57	1,55	1,51	1,46
35	1,76	1,74	1,69	1,63	1,60	1,57	1,53	1,51	1,47	1,42
40	1,74	1,71	1,66	1,61	1,57	1,54	1,51	1,48	1,43	1,38
50	1,70	1,68	1,63	1,57	1,53	1,50	1,46	1,44	1,39	1,33
60	1,68	1,66	1,60	1,54	1,50	1,48	1,44	1,41	1,36	1,30
80	1,65	1,63	1,57	1,51	1,47	1,44	1,40	1,38	1,32	1,25
100	1,64	1,61	1,56	1,49	1,45	1,42	1,38	1,35	1,29	1,22
200	1,60	1,58	1,52	1,46	1,41	1,38	1,34	1,31	1,24	1,16
500	1,58	1,56	1,50	1,44	1,39	1,36	1,31	1,28	1,21	1,11
1000	1,58	1,55	1,49	1,43	1,38	1,35	1,30	1,27	1,20	1,08

Función de distribución F de Snedecor

$P = 0.9$

gl_1

gl_2	1	2	3	4	5	6	7	8	9	10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,97
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,73
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85
1000	3,85	3,01	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84

$P = 0.9$

gl_1

gl_2	11	12	15	20	25	30	40	50	100	1000
1	242,98	243,91	245,96	248,01	249,26	250,08	251,15	251,77	253,01	254,17
2	19,40	19,41	19,43	19,45	19,46	19,46	19,47	19,48	19,49	19,50
3	8,76	8,74	8,70	8,66	8,63	8,62	8,59	8,58	8,55	8,53
4	5,94	5,91	5,86	5,80	5,77	5,74	5,72	5,70	5,66	5,63
5	4,70	4,68	4,62	4,56	4,52	4,50	4,46	4,44	4,41	4,37
6	4,03	4,00	3,94	3,87	3,84	3,81	3,77	3,75	3,71	3,67
7	3,60	3,57	3,51	3,44	3,40	3,38	3,34	3,32	3,27	3,23
8	3,31	3,28	3,22	3,15	3,11	3,08	3,04	3,02	2,97	2,93
9	3,10	3,07	3,01	2,94	2,89	2,86	2,83	2,80	2,76	2,71
10	2,94	2,91	2,85	2,77	2,73	2,70	2,66	2,64	2,59	2,54
11	2,82	2,79	2,72	2,65	2,60	2,57	2,53	2,51	2,46	2,41
12	2,72	2,69	2,62	2,54	2,50	2,47	2,43	2,40	2,35	2,30
13	2,63	2,60	2,53	2,46	2,41	2,38	2,34	2,31	2,26	2,21
14	2,57	2,53	2,46	2,39	2,34	2,31	2,27	2,24	2,19	2,14
15	2,51	2,48	2,40	2,33	2,28	2,25	2,20	2,18	2,12	2,07
16	2,46	2,42	2,35	2,28	2,23	2,19	2,15	2,12	2,07	2,02
17	2,41	2,38	2,31	2,23	2,18	2,15	2,10	2,08	2,02	1,97
18	2,37	2,34	2,27	2,19	2,14	2,11	2,06	2,04	1,98	1,92
19	2,34	2,31	2,23	2,16	2,11	2,07	2,03	2,00	1,94	1,88
20	2,31	2,28	2,20	2,12	2,07	2,04	1,99	1,97	1,91	1,85
25	2,20	2,16	2,09	2,01	1,96	1,92	1,87	1,84	1,78	1,72
30	2,13	2,09	2,01	1,93	1,88	1,84	1,79	1,76	1,70	1,63
35	2,07	2,04	1,96	1,88	1,82	1,79	1,74	1,70	1,63	1,57
40	2,04	2,00	1,92	1,84	1,78	1,74	1,69	1,66	1,59	1,52
50	1,99	1,95	1,87	1,78	1,73	1,69	1,63	1,60	1,52	1,45
60	1,95	1,92	1,84	1,75	1,69	1,65	1,59	1,56	1,48	1,40
80	1,91	1,88	1,79	1,70	1,64	1,60	1,54	1,51	1,43	1,34
100	1,89	1,85	1,77	1,68	1,62	1,57	1,52	1,48	1,39	1,30
200	1,84	1,80	1,72	1,62	1,56	1,52	1,46	1,41	1,32	1,21
500	1,81	1,77	1,69	1,59	1,53	1,48	1,42	1,38	1,28	1,14
1000	1,80	1,76	1,68	1,58	1,52	1,47	1,41	1,36	1,26	1,11

Función de distribución F de Snedecor

$P = 0.975$

gl_2 \ gl_1	1	2	3	4	5	6	7	8	9	10
1	647.80	799.50	864.20	899.60	921.80	937.10	948.20	956.70	963.30	968.60
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.83	2.77
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16
1000	5.02	3.69	3.12	2.79	2.59	2.41	2.29	2.19	2.11	2.05

$P = 0.975$

gl_2 \ gl_1	12	15	20	24	30	40	60	120	1000
1	976.70	978.90	993.10	997.20	1001.00	1006.00	1010.00	1014.00	1018.00
2	39.41	39.43	39.45	39.45	39.46	39.47	39.48	39.49	39.50
3	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.77	2.67	2.56	2.50	2.44	2.39	2.32	2.26	1.19
19	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
1000	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

Función de distribución F de Snedecor

$P = 0.99$

gl_2 \ gl_1	1	2	3	4	5	6	7	8	9	10
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.50	27.34	27.22
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34

$P = 0.99$

gl_2 \ gl_1	11	12	15	20	25	30	40	50	100	1000
2	99.41	99.42	99.43	99.45	99.46	99.46	99.47	99.48	99.49	99.51
3	27.12	27.03	26.85	26.67	26.58	26.50	26.41	26.35	26.24	26.14
4	14.45	14.37	14.19	14.02	13.91	13.84	13.75	13.69	13.58	13.48
5	9.96	5.89	9.72	9.55	9.45	9.38	9.30	9.24	9.13	9.03
6	7.79	7.72	7.56	7.40	7.29	7.23	7.15	7.09	6.99	6.89
7	6.54	6.47	6.31	6.16	6.06	5.99	5.91	5.86	5.75	5.66
8	5.73	5.67	5.52	5.36	5.26	5.20	5.12	5.07	4.96	4.87
9	5.18	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.41	4.32
10	4.77	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.01	3.92
11	4.46	4.40	4.25	4.10	4.00	3.94	3.86	3.81	3.71	3.61
12	4.22	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.47	3.37
13	4.02	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.27	3.18
14	3.86	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.11	3.02
15	3.73	3.67	3.52	3.37	3.28	3.21	3.13	3.08	2.98	2.88
16	3.62	3.55	3.41	3.26	3.16	3.10	3.02	2.97	2.86	2.76
17	3.52	3.46	3.31	3.16	3.07	3.00	2.92	2.87	2.76	2.66
18	3.43	3.37	3.23	3.08	2.98	2.92	2.84	2.78	2.68	2.58
19	3.36	3.30	3.15	3.00	2.91	2.84	2.76	2.71	2.60	2.50
20	3.29	3.23	3.09	2.94	2.84	2.78	2.69	2.64	2.54	2.43
25	3.06	2.99	2.85	2.70	2.60	2.54	2.45	2.40	2.29	2.18
30	2.91	2.84	2.70	2.55	2.45	2.39	2.30	2.24	2.13	2.02
35	2.80	2.74	2.60	2.44	2.35	2.28	2.19	2.14	2.02	1.90
40	2.73	2.66	2.52	2.37	2.27	2.20	2.11	2.06	1.94	1.82
50	2.62	2.56	2.42	2.27	2.17	2.10	2.01	1.95	1.82	1.70
60	2.56	2.50	2.35	2.20	2.10	2.03	1.94	1.88	1.75	1.62
80	2.48	2.42	2.27	2.12	2.01	1.94	1.85	1.79	1.65	1.51
100	2.43	2.37	2.22	2.07	1.97	1.89	1.80	1.74	1.60	1.45
200	2.34	2.27	2.13	1.97	1.87	1.79	1.69	1.63	1.48	1.30
500	2.28	2.22	2.07	1.92	1.81	1.74	1.63	1.57	1.41	1.20
1000	2.27	2.20	2.06	1.90	1.79	1.72	1.61	1.54	1.38	1.16

Tabla 9

Prueba de Kolmogorov-Smirnov sobre Bondad de Ajuste

Nivel de significación α

n	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
1	0.90000	0.95000	0.97500	0.99000	0.99500	0.99750	0.99900	0.99950
2	0.68337	0.77639	0.84189	0.90000	0.92929	0.95000	0.96838	0.97764
3	0.56481	0.63604	0.70760	0.78456	0.82900	0.86428	0.90000	0.92065
4	0.49265	0.56522	0.62394	0.68887	0.73424	0.77639	0.82217	0.85047
5	0.44698	0.50945	0.56328	0.62718	0.66853	0.70543	0.75000	0.78137
6	0.41037	0.46799	0.51926	0.57741	0.61661	0.65287	0.69571	0.72479
7	0.38148	0.43607	0.48342	0.53844	0.57581	0.60975	0.65071	0.67930
8	0.35831	0.40962	0.45427	0.50654	0.54179	0.57429	0.61368	0.64098
9	0.33910	0.38746	0.43001	0.47960	0.51332	0.54443	0.58210	0.60846
10	0.32260	0.36866	0.40925	0.45562	0.48893	0.51872	0.55500	0.58042
11	0.30829	0.35242	0.39122	0.43670	0.46770	0.49539	0.53135	0.55588
12	0.29577	0.33815	0.37543	0.41918	0.44905	0.47672	0.51047	0.53422
13	0.28470	0.32549	0.36143	0.40362	0.43247	0.45921	0.49189	0.51490
14	0.27481	0.31417	0.34890	0.38970	0.41762	0.44352	0.47520	0.49753
15	0.26589	0.30397	0.33750	0.37713	0.40420	0.42934	0.45611	0.48182
16	0.25778	0.29472	0.32733	0.36571	0.39201	0.41644	0.44637	0.46750
17	0.25039	0.28627	0.31796	0.35528	0.38086	0.40464	0.43380	0.45540
18	0.24360	0.27851	0.30936	0.34569	0.37062	0.39380	0.42224	0.44234
19	0.23735	0.27136	0.30143	0.33685	0.36117	0.38379	0.41156	0.43119
20	0.23156	0.26473	0.29408	0.32866	0.35241	0.37451	0.40165	0.42085
21	0.22517	0.25858	0.28724	0.32104	0.34426	0.36588	0.39243	0.41122
22	0.22115	0.25283	0.28087	0.31394	0.33666	0.35782	0.38382	0.40223
23	0.21646	0.24746	0.2749tl	0.30728	0.32954	0.35027	0.37575	0.39380
24	0.21205	0.24242	0.26931	0.30104	0.32286	0.34318	0.36787	0.38588
25	0.20790	0.23768	0.26404	0.29518	0.31657	0.33651	0.36104	0.37743
26	0.20399	0.23320	0.25908	0.28962	0.30963	0.33022	0.35431	0.37139
27	0.20030	0.22898	0.25438	0.28438	0.30502	0.32425	0.34794	0.36473
28	0.19680	0.22497	0.24993	0.27942	0.29971	0.31862	0.34190	0.35842
29	0.19348	0.22117	0.24571	0.27471	0.29466	0.31327	0.33617	0.35242
30	0.19032	0.21756	0.24170	0.27023	0.28986	0.30818	0.33072	0.34672
31	0.18732	0.21412	0.23788	0.26596	0.28529	0.30333	0.32553	0.34129
32	0.18445	0.21085	0.23424	0.26189	0.28094	0.29870	0.32058	0.33611
33	0.18171	0.20771	0.23076	0.25801	0.27577	0.29428	0.31584	0.33115
34	0.17909	0.21472	0.22743	0.25429	0.27271	0.29005	0.31131	0.32641
35	0.17659	0.20185	0.22425	0.25073	0.26897	0.28600	0.30597	0.32187
36	0.17418	0.19910	0.22119	0.24732	0.26532	0.28211	0.30281	0.31751
37	0.17188	0.19646	0.21826	0.24404	0.26180	0.27838	0.29882	0.31333
38	0.16966	0.19392	0.21544	0.24089	0.25843	0.27483	0.29498	0.30931
39	0.16753	0.19148	0.21273	0.23785	0.25518	0.27135	0.29125	0.30544
40	0.16547	0.18913	0.21012	0.23494	0.25205	0.26803	0.28772	0.30171
41	0.16349	0.18687	0.20760	0.23213	0.24904	0.26482	0.28429	0.29811
42	0.16158	0.18468	0.20517	0.22941	0.24613	0.26173	0.28097	0.29465
43	0.15974	0.18257	0.20283	0.22679	0.24332	0.25875	0.27778	0.29130
44	0.15795	0.18051	0.20056	0.22426	0.24060	0.25587	0.27468	0.28806
45	0.15623	0.17856	0.19837	0.22181	0.23798	0.25308	0.27169	0.28493
46	0.15457	0.17665	0.19625	0.21944	0.23544	0.25038	0.26880	0.28190
47	0.15295	0.17481	0.19420	0.21715	0.23298	0.24776	0.26600	0.27896
48	0.15139	0.17301	0.19221	0.21493	0.23059	0.24523	0.26328	0.27611
49	0.14987	0.17128	0.19028	0.21281	0.22832	0.24281	0.26069	0.27339
50	0.14840	0.16959	0.18841	0.21068	0.22604	0.24039	0.25809	0.27067
$n > 50$	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.52/\sqrt{n}$	$1.63/\sqrt{n}$	$1.73/\sqrt{n}$	$1.85/\sqrt{n}$	$1.95/\sqrt{n}$

Tabla 10

Prueba Lilliefors sobre Normalidad

	Nivel de significación α				
n	0.20	0.15	0.10	0.05	0.01
4	0.300	0.319	0.352	0.381	0.417
5	0.285	0.299	0.315	0.337	0.405
6	0.265	0.277	0.294	0.319	0.364
7	0.247	0.258	0.276	0.300	0.348
8	0.233	0.244	0.261	0.285	0.331
9	0.223	0.233	0.249	0.271	0.311
10	0.215	0.224	0.239	0.258	0.294
11	0.206	0.217	0.230	0.249	0.284
12	0.199	0.212	0.223	0.242	0.275
13	0.190	0.202	0.214	0.234	0.268
14	0.183	0.194	0.207	0.227	0.261
15	0.177	0.187	0.201	0.220	0.257
16	0.173	0.182	0.195	0.213	0.250
17	0.169	0.177	0.189	0.206	0.245
18	0.166	0.173	0.184	0.200	0.239
19	0.163	0.169	0.179	0.195	0.235
20	0.160	0.166	0.174	0.190	0.231
21	0.157	0.162	0.170	0.186	0.226
22	0.154	0.158	0.167	0.182	0.220
23	0.150	0.154	0.164	0.179	0.213
24	0.146	0.151	0.161	0.176	0.207
25	0.142	0.147	0.158	0.173	0.200
26	0.139	0.14S	0.155	0.171	0.197
27	0.137	0.143	0.153	0.168	0.194
28	0.135	0.140	0.150	0.166	0.192
29	0.133	0.138	0.147	0.163	0.190
30	0.131	0.136	0.144	0.161	0.187
$n > 30$	$0.736/\sqrt{n}$	$0.768/\sqrt{n}$	$0.805/\sqrt{n}$	$0.886/\sqrt{n}$	$1.031/\sqrt{n}$

Tabla 11

Prueba de homogeneidad de varianzas de Cochran

Nivel de significación $\alpha = 0.05$

	N													
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**17**	**37**	**145**	**1000**
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
4	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2513	0.2001
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3529	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3505	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.7353	0.2032	0.1655	0.1308	0.1000

Nivel de significación $\alpha = 0.01$

	N													
k	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**17**	**37**	**145**	**1000**
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7112	0.6761	0.6410	0.6129	0.5897	0.5702	0.5536	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6937	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8828	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8376	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.3704	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1250
9	0.7544	0.5727	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1111
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2513	0.2704	0.2813	0.1811	0.1376	0.1000

Tabla 12

Test de Rachas

		n-								
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$n+=2$	α=0.05	0-4	0-5	0-5	0-5	0-5	0-5	2-3	2-5	2-5
	α=0.01	0-4	0-5	0-5	0-5	0-5	0-5	0-5	0-5	0-5
$n+=3$	α=0.05	0-5	0-6	0-6	2-7	2-7	2-7	2-7	2-7	3-7
	α=0.01	0-5	0-6	0-7	0-7	0-7	0-7	0-7	2-7	2-7
$n+=4$	α=0.05	0-5	0-6	2-7	2-8	3-8	3-8	3-9	3-9	3-9
	α=0.01	0-5	0-7	0-8	0-9	2-9	2-9	2-9	2-9	2-9
$n+=5$	α=0.05	0-5	2-7	2-8	3-8	3-9	3-9	3-10	4-10	4-10
	α=0.01	0-5	0-7	0-9	2-9	2-10	2-10	2-11	3-10	3-11
$n+=6$	α=0.05	0 5	2-7	3-8	3-9	3-10	4-10	4-11	4-11	5-11
	α=0.01	0 5	0-7	2-9	2-10	2-11	3-11	3-12	3-12	3-13
$n+=7$	α=0.05	0-5	2-7	3-8	3-9	4-10	4-11	4-12	5-12	5-12
	α=0.01	0-5	0-7	2-9	2-10	3-11	3-12	3-12	4-13	4-14
$n+=8$	α=0.05	2-5	2-7	3-9	3-10	4-11	4-12	5-12	5-13	6-13
	α=0.01	0-5	0-7	2-9	2-11	3-12	3-12	4-13	4-14	4-14
$n+=9$	α=0.05	2-5	2-7	3-9	4-10	4-11	5-12	5-13	6-13	6-14
	α=0.01	0-5	2-7	2-9	3-10	3-12	4-13	4-14	4-15	5-15
$n+=10$	α=0.05	2-5	3-7	3-9	4-10	5-11	5-12	6-13	6-14	6-15
	α=0.01	0-5	2-7	2-9	3-11	3-13	4-14	4-14	5-15	5-16
$n+=11$	α=0.05	2-5	3-7	3-9	4-11	5-12	5-13	6-14	6-14	7-15
	α=0.01	0-5	2-7	2-9	3-11	4-13	4-14	4-15	5-16	5-17
$n+=12$	α=0.05	2-5	3-7	4-9	4-11	5-12	6-13	6-14	7-15	7-16
	α=0.01	0-5	2-7	3-9	3-11	4-13	4-14	5-15	5-16	6-17
$n+=13$	α=0.05	2-5	3-7	4-9	4-11	5-12	6-13	6-14	7-15	8-16
	α=0.01	0-5	2-7	3-9	3-11	4-13	5-15	5-16	6-17	6-18
$n+=14$	α=0.05	2-5	3-7	4-9	5-11	5-12	6-13	7-15	7-16	8-16
	α=0.01	0-5	2-7	3-9	3-11	4-13	5-15	5-16	6-17	6-18
$n+=15$	α=0.05	2-5	3-7	4-9	5-11	6-13	6-14	7-15	8-16	8-17
	α=0.01	0-5	2-7	3-9	4-11	4-13	5-15	5-16	6-17	7-18
$n+=16$	α=0.05	2-5	3-7	4-9	5-11	6-13	6-14	7-15	8-16	8-17
	α=0.01	0-5	2-7	3-9	4-11	4-13	5-15	6-16	6-17	7-19
$n+=17$	α=0.05	2-5	3-7	4-9	5-11	6-13	7-14	7-15	8-16	9-17
	α=0.01	0-5	2-7	3-9	4-11	5-13	5-15	6-17	7-17	7-19
$n+=18$	α=0.05	2-5	3-7	4-9	5-11	6-13	7-14	8-15	8-17	9-18
	α=0.01	0-5	2-7	3-9	4-11	5-13	5-15	6-17	7-18	7-19
$n+=19$	α=0.05	2-5	3-7	4-9	5-11	6-13	7-14	8-15	9-17	9-18
	α=0.01	2-5	2-7	3-9	4-11	5-13	6-15	6-17	7-18	8-19
$n+=20$	α=0.05	2-5	3-7	4-9	5-11	6-13	7-14	8-16	9-17	9-18
	α=0.01	2-5	2-7	3-9	4-11	5-13	6-15	6-17	7-18	8-19

Test de Rachas

		n- 11	12	13	14	15	16	17	18	19	20
n+ = 2	α=0.05	2-5	2-5	2-5	2-5	2-5	2-5	2-5	2-5	2-5	2-5
	α=0.01	0-5	0-5	0-5	0-5	0-5	0-5	0-5	0-5	2-5	2-5
n+ = 3	α=0.05	3-7	3-7	3-7	3-7	3-7	3-7	3-7	3-7	3-7	3-7
	α=0.01	2-7	2-7	2-7	2-7	2-7	2-7	2-7	2-7	2-7	2-7
n+ = 4	α=0.05	3-9	4-9	4-9	4-9	4-9	4-9	4-9	4-9	4-9	4-9
	α=0.01	2-9	3-9	3-9	3-9	3-9	3-9	3-9	3-9	3-9	3-9
n+ = 5	α=0.05	4-11	4-11	4-11	5-11	5-11	5-11	5-11	5-11	5-11	5-11
	α=0.01	3-11	3-11	3-11	3-11	4-11	4-11	4-11	4-11	4-11	4-11
n+ = 6	α=0.05	5-12	5-12	5-12	5-12	6-13	6-13	6-13	6-13	6-13	6-13
	α=0.01	4-13	4-13	4-13	4-13	4-13	4-13	5-13	5-13	5-13	5-13
n+ = 7	α=0.05	5-13	6-13	6-13	6-13	6-14	6-14	7-14	7-14	7-14	7-14
	α=0.01	4-14	4-14	5-15	5-15	5-15	5-15	5-15	5-15	6-15	6-15
n+ = 8	α=0.05	6-14	6-14	6-14	7-15	7-15	7-15	7-15	8-15	8-15	8-16
	α=0.01	4-15	5-15	5-16	5-16	5-16	6-16	6-17	6-17	6-17	6-17
n+ = 9	α=0.05	6-14	7-15	7-15	7-16	8-16	8-16	8-16	8-17	9-17	9-17
	α=0.01	5-16	5-16	6-17	6-17	6-17	6-17	7-17	7-18	7-18	7-18
n+ = 10	α=0.05	7-15	7-16	8-16	8-16	8-17	8-17	9-17	9-18	9-18	9-18
	α=0.01	5-17	6-17	6-18	6-18	7-18	7-19	7-19	7-19	8-19	8-19
n+ = 11	α=0.05	8-17	9-17	9-18	9-18	10-19	10-19	10-20	10-20	11-20	7-16
	α=0.01	7-18	7-19	7-20	8-20	8-21	8-21	8-21	9-22	9-22	6-17
n+ = 12	α=0.05	9-17	8-16	8-17	8-17	9-18	9-18	9-18	10-19	10-19	10-19
	α=0.01	7-19	6-18	6-18	7-19	7-19	7-20	8-20	8-20	8-21	8-21
n+ = 13	α=0.05	9-18	8-17	9-18	9-19	10-19	10-20	10-20	11-20	11-21	11-21
	α=0.01	7-20	6-18	7-20	8-20	8-21	8-21	9-22	9-22	9-23	10-23
n+ = 14	α=0.05	9-18	8-17	9-19	10-19	10-20	11-20	11-21	11-21	12-22	12-22
	α=0.01	8-20	7-19	8-20	8-21	8-22	9-22	9-23	9-23	10-23	10-24
n+ = 15	α=0.05	10-19	9-18	10-19	10-20	11-20	11-21	11-21	12-22	12-22	12-23
	α=0.01	8-21	7-19	8-21	8-22	9-22	9-23	10-23	10-24	10-24	11-25
n+ = 16	α=0.05	10-19	9-18	10-20	11-20	11-21	11-22	12-22	12-22	12-23	12-23
	α=0.01	8-21	7-20	8-21	9-22	9-23	10-23	10-24	10-24	10-25	10-25
n+ = 17	α=0.05	10-20	9-18	10-20	11-21	11-21	12-22	12-23	13-23	13-24	13-24
	α=0.01	8-21	8-20	9-22	9-23	10-23	10-24	10-25	11-25	11-26	11-26
n+ = 18	α=0.05	10-20	10-19	11-20	11-21	12-22	12-22	13-23	13-24	14-25	14-25
	α=0.01	9-22	8-20	9-22	9-23	10-24	10-24	11-25	11-26	12-26	12-27
n+ = 19	α=0.05	11-20	10-19	11-21	10-19	12-22	12-23	13-24	14-25	14-25	14-26
	α=0.01	9-22	8-21	9-23	10-23	10-24	10-25	11-26	12-26	12-27	12-28
n+ = 20	α=0.05	7-16	10-19	11-21	12-22	12-23	12-23	13-24	14-25	14-26	15-26
	α=0.01	6-17	8-21	10-23	10-24	11-25	10-25	11-26	12-27	12-28	13-28

Tabla 13

Prueba de Kolmogorov-Smirnov
(muestras del mismo tamaño)

	$1-\alpha$				
n	0.80	0.90	0.95	0.98	0.99
3	2	2			
4	3	3	3		
5	3	3	4	4	4
6	3	4	4	5	5
7	4	4	5	5	5
8	4	4	5	5	6
9	4	5	5	6	6
10	4	5	6	6	7
11	5	5	6	7	7
12	5	5	6	7	7
13	5	6	6	7	8
14	5	6	7	7	8
15	5	6	7	8	8
16	6	6	7	8	9
17	6	7	7	8	9
18	6	7	8	9	9
19	6	7	8	9	9
20	6	7	8	9	10
21	6	7	8	9	10
22	7	8	8	10	10
23	7	8	9	10	10
24	7	8	9	10	11
25	7	8	9	10	11
26	7	8	9	10	11
27	7	8	9	11	11
28	8	9	10	11	11
29	8	9	10	11	12
30	8	9	10	11	12
31	8	9	10	11	12
32	8	9	10	12	12
33	8	9	10	12	12
34	8	10	11	12	13
35	8	10	11	12	13
36	9	10	11	12	13
37	9	10	11	12	13
38	9	10	11	13	14
39	9	10	12	13	14
40	9	10	12	13	14
$n > 40$	$1.52/\sqrt{n}$	$1.73/\sqrt{n}$	$1.92/\sqrt{n}$	$2.15/\sqrt{n}$	$2.30/\sqrt{n}$

Tabla 14

Prueba de Kolmogorov-Smirnov (muestras de distinto tamaño)

		$1-\alpha$				
n_1	n_2	0.80	0.90	0.95	0.98	0.99
1	9	17/18				
	10	9/10				
2	3	5/6				
	4	3/4				
	5	4/5	4/5			
	6	5/6	5/6			
	7	5/7	6/7			
	8	3/4	7/8	7/8		
	9	7/9	8/9	8/9		
	10	7/10	4/5	9/10		
3	4	3/4	3/4			
	5	2/3	4/5	4/5		
	6	2/3	2/3	5/6		
	7	2/3	5/7	6/7	6/7	
	8	5/8	3/4	3/4	7/8	
	9	2/3	2/3	7/9	8/9	8/9
	10	3/5	7/10	4/5	9/10	9/10
	12	7/12	2/3	3/4	5/6	11/12
4	5	3/5	1/4	4/5	4/5	
	6	7/12	2/3	3/4	5/6	5/6
	7	17/28	5/7	1/4	6/7	6/7
	8	5/8	5/8	3/4	7/8	7/8
	9	5/9	2/3	3/4	7/9	8/9
	10	11/20	13/20	7/10	4/5	4/5
	12	7/12	2/3	2/3	1/4	5/6
	16	9/16	5/8	11/16	1/4	13/16
5	6	1/5	2/3	2/3	5/6	5/6
	7	4/7	23/25	5/7	29/15	6/7
	8	11/20	5/8	27/40	4/5	4/5
	9	5/9	3/5	31/45	7/9	4/5
	10	1/2	3/5	7/10	7/10	4/5
	15	8/15	3/5	2/3	11/15	11/15
	20	1/2	11/20	3/5	7/10	3/4
6	7	23/42	4/7	29/42	5/7	5/6
	8	1/2	7/12	2/3	3/4	3/4
	9	1/2	5/9	2/3	13/18	7/9
	10	1/2	17/30	19/10	7/10	11/45
	12	1/2	7/12	7/12	2/3	3/4
	18	4/9	5/9	11/18	2/3	13/18
	24	11/24	1/2	7/12	5/8	2/3
7	8	27/56	33/56	5/8	41/56	3/4
	9	31/63	5/9	40/63	5/7	47/63
	10	13/70	39/70	43/70	7/10	5/7
	14	3/7	1/2	4/7	9/14	5/7
	28	3/7	13/28	15/28	17/28	9/14
8	8	4/9	13/24	5/8	2/3	3/4
	10	19/40	21/40	21/40	27/40	7/10
	12	11/24	1/2	7/12	5/8	2/3
	16	7/16	1/2	9/16	5/8	5/8
	32	13/32	7/16	1/2	9/16	19/32
9	10	7/15	1/2	26/45	2/3	31/45
	12	4/9	1/2	5/9	11/18	2/3
	15	19/45	22/45	8/15	3/5	29/45
	18	7/18	4/9	1/2	5/9	11/18
	36	13/36	5/12	17/36	19/36	5/9
10	15	2/5	7/15	1/2	17/30	19/30
	20	2/5	9/20	1/2	11/20	3/5
	40	7/20	2/5	9/20	1/2	
12	15	23/60	9/20	1/2	11/20	7/12
	16	3/8	7/16	23/48	13/24	7/12
	18	13/36	5/12	17/36	19/36	5/9
	20	11/30	5/12	7/15	31/60	17/30
15	20	7/20	2/5	13/30	29/60	31/60
16	20	27/80	31/80	17/40	19/40	41/80
		$1.07\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1.22\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1.36\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1.48\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1.63\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$

Tabla 15

Distribución acumulativa de Poisson

(probabilidad × 1000 de c o menos defectos cuando el número esperado es de $c' = np' = \lambda$)

	c								
c'	0	1	2	3	4	5	6	7	8
0.02	980	1000							
0.04	961	999	1000						
0.06	942	998	1000						
0.08	923	997	1000						
0.10	905	995	1000						
0.15	861	990	999	1000					
0.20	819	982	999	1000					
0.25	779	974	998	1000					
0.30	741	963	996	1000					
0.35	705	951	994	1000					
0.40	670	938	992	999	1000				
0.45	638	925	989	999	1000				
0.50	607	910	986	998	1000				
0.55	577	894	982	998	1000				
0.60	549	878	977	997	1000				
0.65	522	861	972	996	999	1000			
0.70	497	844	966	994	999	1000			
0.75	472	827	959	993	999	1000			
0.80	449	809	953	991	999	1000			
0.90	407	772	937	987	998	1000			
1.00	368	736	920	981	996	999	1000		
1.10	333	699	900	974	995	999	1000		
1.20	301	663	879	966	992	998	1000		
1.30	273	627	857	957	989	998	1000		
1.40	247	592	833	946	986	997	999	1000	
1.50	223	558	809	934	981	996	999	1000	
1.60	202	525	783	921	976	994	999	1000	
1.70	183	493	757	907	970	992	998	1000	
1.80	165	463	731	891	964	990	997	999	1000
1.90	150	434	704	875	956	987	997	999	1000
2.00	135	406	677	857	947	983	995	999	1000
2.10	122	380	650	839	938	980	994	999	1000
2.20	110	354	622	819	927	974	993	998	1000
2.30	100	331	596	799	916	970	991	997	999
2.40	091	308	570	779	904	964	988	997	999
2.50	082	287	544	758	891	958	986	996	999
2.60	074	267	518	736	877	951	983	995	999
2.70	067	249	494	714	863	943	979	993	998
2.80	061	231	469	692	848	935	976	992	998
2.90	055	215	446	670	832	926	971	990	997
3.00	050	199	423	647	815	916	966	988	996
3.10	045	185	401	625	798	906	961	986	995
3.20	041	171	380	603	781	895	955	983	994
3.30	037	159	359	580	763	883	949	980	993
3.40	033	147	340	558	744	871	942	977	992
3.50	030	136	321	537	725	858	935	973	991
3.60	027	126	303	515	706	844	927	969	988
3.70	025	116	285	494	687	830	918	965	986
3.80	022	107	269	473	668	816	909	960	984
3.90	020	099	253	453	648	801	899	955	981
4.00	018	092	238	433	629	785	889	949	979
4.10	017	085	224	414	609	769	879	943	976
4.20	015	078	210	395	590	753	867	936	972
4.30	014	072	197	377	570	737	856	929	968
4.40	012	066	185	359	551	720	844	921	964
4.50	011	061	174	342	532	703	831	913	960
4.60	010	056	163	326	513	686	818	905	955
4.70	009	052	152	310	495	668	805	896	950
4.80	008	048	143	294	476	651	791	887	944
4.90	007	044	133	279	458	634	777	877	938
5.00	007	040	125	265	440	616	762	867	932
5.10	006	037	116	251	423	598	747	856	925
5.20	006	034	109	238	406	581	732	845	918
5.30	005	031	102	225	390	563	717	833	911
5.40	004	029	095	213	373	546	702	822	903
5.50	004	027	088	202	358	529	686	809	894
5.60	004	024	082	191	342	512	670	797	886
5.70	003	022	077	180	327	495	654	784	877
5.80	003	021	072	170	313	478	638	771	867
5.90	003	019	067	160	299	462	622	758	857

Distribución acumulativa de Poisson

(probabilidad × 1000 de c o menos defectos cuando el número esperado es de $c' = np = \lambda$)

	c								
c'	0	1	2	3	4	5	6	7	8
6.00	002	017	062	151	285	446	606	744	847
6.10	002	016	058	143	272	430	590	730	837
6.20	002	015	054	134	259	414	574	716	826
6.30	002	013	050	126	247	399	558	702	815
6.40	002	012	046	119	235	384	542	687	803
6.50	002	011	043	112	224	369	527	673	792
6.60	001	010	040	105	213	355	511	658	780
6.70	001	009	037	039	202	341	495	643	767
6.80	001	009	034	093	192	327	480	628	755
6.90	001	008	032	087	182	314	465	614	742
7.00	001	007	030	082	173	301	450	599	729
7.20	001	006	025	072	156	276	420	569	703
7.40	001	005	022	063	140	253	392	539	676
7.60	000	004	019	055	125	231	365	510	648
7.80		004	016	048	112	210	338	481	620
8.00		003	014	042	100	191	313	453	593
8.20		003	012	037	089	174	290	425	565
8.40		002	010	032	079	157	267	399	537
8.60		002	009	028	070	142	246	373	509
8.80		001	007	024	062	128	226	348	482
9.00		001	006	021	055	116	207	324	456
9.20		001	005	018	049	104	189	301	430
9.40		001	005	016	043	093	173	279	404
9.60		001	004	014	038	084	157	258	380
9.80		001	003	012	033	075	143	239	356
10.00		001	003	010	029	067	130	220	333
10.20		000	002	009	026	060	118	203	311
10.40			002	008	023	053	107	186	290
10.60			002	007	020	048	097	171	269
10.80			001	006	017	042	087	157	250
11.00			001	005	015	037	079	143	232
11.20			001	004	013	033	071	131	215
11.40			001	004	012	029	064	119	198
11.60			001	003	010	026	057	108	183
11.80			001	003	009	023	051	099	169
12.00			001	002	008	020	046	089	155
12.20			000	002	007	018	041	081	143
12.40				002	006	016	037	073	131
12.60				001	005	014	033	066	120
12.80				001	004	012	029	060	109
11.00				001	004	011	026	054	100
13.20				001	003	009	023	049	091
13.40				001	003	008	020	044	083
13.60				001	002	007	017	039	075
13.80				001	002	006	016	035	068
14.00				000	002	006	014	032	062
14.20					002	005	013	028	056
14.40					001	004	011	025	051
14.60					001	004	010	023	046
14.80					001	003	009	020	042
15.00					001	003	008	018	037

Tabla 16

Valores medios para obtener las probabilidades acumulativas de Poisson deseadas

c	0.99	0.95	0.90	0.50	0.10	0.05
0	0.010	0.051	0.105	0.693	2.303	2.996
1	0.149	0.355	0.532	1.678	3.890	4.744
2	0.436	0.818	1.102	2.674	5.322	6.296
3	0.823	1.366	1.745	3.672	6.681	7.754
4	1.279	1.970	2.433	4.671	7.994	9.154
5	1.785	2.613	3.152	5.670	9.275	10.513
6	2.330	3.286	3.895	6.670	10.532	11.842
7	2.906	3.981	4.656	7.669	11.771	13.148
8	3.507	4.695	5.432	8.669	12.995	14.434
9	4.130	5.426	6.221	9.669	14.206	15.705
10	4.771	6.169	7.021	10.668	15.407	16.962
11	5.428	6.924	7.829	11.668	16.598	18.208
12	6.099	7.690	8.646	12.668	17.782	19.442
13	6.782	8.464	9.470	13.668	18.958	20.668
14	7.477	9.246	10.300	14.668	20.128	21.886
15	8.181	10.035	11.135	15.668	21.292	23.098
16	8.895	10.831	11.976	16.668	22.452	24.302
17	9.616	11.633	12.822	17.668	23.606	25.500
18	10.346	12.442	13.672	18.668	24.756	26.692
19	11.082	13.254	14.525	19.668	25.902	27.879
20	11.825	14.072	15.383	20.668	27.045	29.062
21	12.574	14.894	16.244	21.668	28.184	30.241
22	13.329	15.719	17.108	22.668	29.320	31.416
23	14.088	16.548	17.975	23.668	30.453	32.586
24	14.853	17.382	18.844	24.668	31.584	33.752
25	15.623	18.218	19.717	25.667	32.711	34.916
30	19.532	22.444	24.113	30.667	38.315	40.690
35	23.525	26.731	28.556	35.667	43.872	46.404
40	27.587	31.066	33.038	40.667	49.390	52.069
45	31.704	35.441	37.550	45.667	54.878	57.695
50	35.867	39.849	42.089	50.667	60.339	63.287

Tabla 17

Muestreo por atributos

Letras de código para tamaño de muestra

UNE 66020, ISO 2859, MIL-STD-105E

Tamaño del lote o grupo	Niveles especiales de inspección				Niveles de inspección general		
	S-1	S-2	S-3	S-4	I	II	III
2-8	A	A	A	A	A	A	B
9-15	A	A	A	A	A	B	C
16-25	A	A	B	B	B	C	D
26-50	A	B	B	C	C	D	E
51-90	B	B	C	C	C	E	F
91-150	B	B	C	D	D	F	G
151-280	B	C	D	E	E	G	H
281-500	B	C	D	E	F	H	J
501-1.200	C	C	E	F	G	J	K
1.201-3.200	C	D	E	G	H	K	L
3.201 -10.000	C	D	F	G	J	L	M
10.001-35.000	C	D	F	H	K	M	N
35.001-150.000	D	E	G	J	L	N	P
150.001-500.000	D	E	G	J	M	P	Q
> 500.001	D	E	H	K	N	Q	R

Tabla 18

Muestreo por atributos
Tabla maestra para inspección normal
Muestreo simple
UNE 66020, ISO 2859, MIL-STD-105E

Letra-código de tamaño muestral	Tamaño de la muestra	Niveles de Calidad Aceptable (inspección normal)																							
		0.01		0.015		0.025		0.04		0.065		0.10		0.15		0.25		0.40		0.65		1.0		1.5		2.5	
		A	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*
A	2																										
B	3																									0	1
C	5																							0	1		
D	8																					0	1				
E	13																			0	1					1	2
F	20																	0	1					1	2	2	3
G	32															0	1					1	2	2	3	3	4
H	50													0	1					1	2	2	3	3	4	5	6
J	80																	1	2	2	3	3	4	5	6	7	8
K	125											0	1			1	2	2	3	3	4	5	6	7	8	10	11
L	200									0	1					1	2	2	3	3	4	5	6	7	8	10	11
M	315							0	1					1	2	2	3	3	4	5	6	7	8	10	11	14	15
N	500					0	1					1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22
P	800			0	1					1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22		
Q	1250	0	1					1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22				
R	2000					1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22						

Letra-código de tamaño muestral	Tamaño de la muestra	Niveles de Calidad Aceptable (inspección normal)																							
		4.0		6.5		10		15		25		40		65		100		150		250		400		650		1000	
		A	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*
A	2			0	1					1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22	30	31
B	3	0	1					1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22	30	31	44	45
C	5					1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22	30	31	44	45		
D	8			1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22	30	31	44	45				
E	13	1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22	30	31	44	45						
F	20	2	3	3	4	5	6	7	8	10	11	14	15	21	22												
G	32	3	4	5	6	7	8	10	11	14	15	21	22														
H	50	5	6	7	8	10	11	14	15	21	22																
J	80	7	8	10	11	14	15	21	22																		
K	125	10	11	14	15	21	22																				
L	200	14	15	21	22																						
M	315	21	22																								
N	500																										
P	800																										
Q	1250																										
R	2000																										

Tabla 19

Muestreo por atributos
Tabla maestra para inspección rigurosa
Muestreo simple
UNE 66020, ISO 2859, MIL-STD-105E

Letra-código de tamaño muestral	Tamaño de la muestra	Niveles de calidad aceptable (inspección rigurosa)																					
		0.01		0.015		0.025		0.04		0.065		0.10		0.15		0.25		0.40		0.65		1.0		1.5		2.5	
		A	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*
A	2																										
B	3																										
C	5																										
D	8																									0	1
E	13																							0	1		
F	20																					0	1				
G	32																			0	1					1	2
H	50																	0	1					1	2	2	3
J	80															0	1					1	2	2	3	3	4
K	125													0	1					1	2	2	3	3	4	5	6
L	200											0	1					1	2	2	3	3	4	5	6	8	9
M	315									0	1					1	2	2	3	3	4	5	6	8	9	12	13
N	500							0	1					1	2	2	3	3	4	5	6	8	9	12	13	18	19
P	800					0	1					1	2	2	3	3	4	5	6	8	9	12	13	18	19		
Q	1250			0	1					1	2	2	3	3	4	5	6	8	9	12	13	18	19				
R	2000	0	1					1	2	2	3	3	4	5	6	8	9	12	13	18	19						
S	3150					1	2																				

Letra-código de tamaño muestral	Tamaño de la muestra	Niveles de Calidad Aceptable (inspección rigurosa)																					
		4.0		6.5		10		15		25		40		65		100		150		250		400		650		1000	
		A	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*
A	2											1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28
B	3			0	1					1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42
C	5	0	1					1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42		
D	8					1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42				
E	13			1	2	2	3	3	4	5	6	8	9	12	13	18	19	27	28	41	42						
F	20	1	2	2	3	3	4	5	6	8	9	12	13	18	19												
G	32	2	3	3	4	5	6	8	9	12	13	18	19														
H	50	3	4	5	6	8	9	12	13	18	19																
J	80	5	6	8	9	12	13	18	19																		
K	125	8	9	11	13	18	19																				
L	200	12	13	18	19																						
M	315	18	19																								
N	500																										
P	800																										
Q	1250																										
R	2000																										
S	3150																										

Tabla 20

Muestreo por atributos
Tabla maestra para inspección reducida
Muestreo simple
UNE 66020, ISO 2859, MIL-STD-105E

Letra-código de tamaño muestral	Tamaño de la muestra	Niveles de Calidad Aceptable (inspección reducida)																							
		0.01		0.015		0.025		0.04		0.065		0.10		0.15		0.25		0.40		0.65		1.0		1.5		2.5	
		A	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*
A	2																										
B	2																										
C	2																									0	1
D	3																							0	1		
E	5																					0	1				
F	8																			0	1					0	2
G	13																	0	1					0	2	1	3
H	20															0	1					0	2	1	3	1	4
J	32													0	1					0	2	L	3	1	4	2	5
K	50											0	1					0	2	1	3	1	4	2	5	3	6
L	80									0	1					0	2	1	3	1	4	2	5	3	6	5	8
M	125							0	1					0	2	1	3	1	4	2	5	3	6	5	8	7	10
N	200					0	1					0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13
P	315			0	1					0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13		
Q	500	0	1					0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13				
R	800					0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13						

Letra-código de tamaño muestral	Tamaño de la muestra	Niveles de Calidad Aceptable (inspección reducida)																							
		4.0		6.5		10		15		25		40		65		100		150		250		400		650		1000	
		A	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*	*A*	*Re*
A	2			0	1				1	1	2	2	3	3	4	5	6	7	8	10	11	14	15	21	22	30	31
B	2	0	1					0	2	1	3	2	4	3	5	5	6	7	8	10	11	14	15	21	22	30	31
C	2					0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13	14	17	21	24		
D	3			0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13	14	17	21	24				
E	5	0	2	1	3	1	4	2	5	3	6	5	8	7	10	10	13	14	17	21	24						
F	8	1	3	1	4	2	5	3	6	5	8	7	10	10	13												
G	13	1	4	2	5	3	6	5	8	7	10	10	13														
H	20	2	5	3	6	5	8	7	10	10	13																
J	32	3	6	5	8	7	10	10	13																		
K	50	5	8	7	10	10	13																				
L	80	7	10	10	13																						
M	125	10	13																								
N	200																										
P	315																										
Q	500																										
R	800																										

Tabla 21

Sistema Philips de muestreo estándar

Tamaño del lote	Punto de control	0.25%			0.5%			1%			2%		
		n	c		n	c		n	c		n	c	
Muestreo único	20-50	A			A			A			30	0	
	51-100	A			A			60	0		30	0	
	101-200	A			100	0		60	0		35	0	
	201-500	175	0		100	0		135	1		75	1	
	501-1.000	225	0		225	1		150	1		85	1	
Muestreo doble		n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2
	1.001-2.000	330	0	1	150	0	1	110	0	2	55	0	2
	2.001-5.000	425	0	2	200	0	2	135	0	3	70	0	3
	5.001-10.000	525	0	3	260	0	3	220	1	5	110	1	5
	10.000-20.000	875	1	5	440	1	5	380	2	10	190	2	10
	20.000-50.000	1.500	2	10	750	2	10	540	3	15	270	3	15
	50.000-100.000	2.200	3	15	1.100	3	15	700	4	20	390	4	20

Tamaño del lote	Punto de control	3 %			5 %			7 %			10 %		
		n	c		n	c		n	c		n	c	
Muestreo único	20-50	20	0		13	0		10	0		7	0	
	51-100	20	0		13	0		10	0		7	0	
	101-200	55	1		35	1		25	1		17	1	
	201-500	55	1		35	1		40	2		25	2	
	501-1.000	85	2		55	2		55	3		35	3	
Muestreo doble		n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2	n_1	c_1	c_2
	1.001-2.000	45	0	3	25	0	3	30	1	5	22	1	5
	2.001-5.000	70	1	5	45	1	5	55	2	10	40	2	10
	5.001-10.000	125	2	10	75	2	10	75	3	15	55	3	15
	10.000-20.000	180	3	15	110	3	15	100	4	20	70	4	20
	20.000-50.000	240	4	20	140	4	20	120	5	25	85	5	25
	50.000-100.000	290	5	25	175	5	25	146	6	30	105	6	30

Tabla 22

Método Dodge-Romig

Tolerancia de lote con muestreo simple

% de media del proceso	0 - 0.05			0.06 - 0.50			0.51 - 1.00			1.01 - 1.50			1.51 - 2.00			2.01 - 2.50		
Tamaño del lote	n	Ac	LCMS	n	Ac	LCMS	n	Ac	LCMS	N	Ac	LCMS	n	Ac	LCMS	n	Ac	LCMS
1-30	todo	0	0	Todo	0	0	Todo	0	0	Todo	0	0	Todo	0	0	Todo	0	0
31-50	30	0	0.49	30	0	0.49	30	0	0.49	30	0	0.49	30	0	0.49	30	0	0.49
51-100	37	0	0.63	37	0	0.63	37	0	0.63	37	0	0.63	37	0	0.63	37	0	0.63
101-200	40	0	0.74	40	0	0.74	40	0	0.74	40	0	0.74	40	0	0.74	40	0	0.74
201-300	43	0	0.74	43	0	0.74	70	1	0.92	70	1	0.92	95	2	0.99	95	2	0.99
301-400	44	0	0.74	44	0	0.74	70	1	0.99	100	2	1.0	120	3	1.1	145	4	1.1
401-500	45	0	0.75	75	1	0.95	100	2	1.1	100	2	1.1	125	3	1.2	150	4	1.2
501-600	45	0	0.76	75	1	0.98	100	2	1.1	125	3	1.2	150	4	1.3	175	5	1.3
601-800	45	0	0.77	75	1	1.0	100	2	1.2	130	3	1.2	175	5	14	200	6	1.4
801-1000	45	0	0.78	75	1	1.0	105	2	1.2	155	4	1.4	180	5	1.4	225	7	1.5
1001-2000	45	0	0.80	75	1	1.0	130	3	1.4	180	5	1.6	230	7	1.7	280	9	1.8
2001-3000	75	1	1.1	105	2	1.3	135	3	1.4	210	6	1.7	280	9	1.9	370	13	2.1
3001-4000	75	1	1.1	105	2	1.3	160	4	1.5	210	6	1.7	305	10	2.0	420	15	2.2
4001-5000	75	1	1.1	105	2	1.3	160	4	1.5	235	7	1.8	330	11	2.0	440	10	2.2
5001-7000	75	1	1.1	105	2	1.3	185	5	1.7	260	8	1.9	350	12	2.2	490	18	2.4
7001-10000	75	1	1.1	105	2	1.3	185	5	1.7	260	8	1.9	380	13	2 2	535	20	2.5
10001-20000	75	1	1.1	135	3	1.4	210	6	1.8	285	9	2.0	425	15	2.3	610	23	2.6
20001-50000	75	1	1.1	135	3	1.4	235	7	1.9	305	10	2.1	470	17	2.4	700	27	2.7
50001-100000	75	1	1.1	160	4	1.6	235	7	1.9	355	12	2.2	515	19	2.5	770	30	2.8

Porcentaje de tolerancia por lote de elementos defectuosos = 5 %

Riesgo del consumidor = 0.10

Tabla 23

Método Dodge-Romig

Tolerancia de lote con muestreo doble

% de media del proceso	0 - 0.05			0.06 - 0.50			0.51 - 1.00		
Tamaño del lote	Prueba 1 n_1 c_1	Prueba 2 n_2 n_1+n_2 c_2	LCMS en %	Prueba 1 n_1 c_1	Prueba 2 n_2 n_1+n_2 c_2	LCMS en %	Prueba 1 n_1 c_1	Prueba 2 n_2 n_1+n_2 c_2	LCMS en %
1-30	Todo 0		0	Todo 0		0	Todo 0		0
31-50	30 0		0,49	30 0		0,49	30 0		0,49
51-75	38 0		0.59	38 0		0,59	38 0		0,59
76-100	44 0	21 65 1	0.64	44 0	21 65 1	0,64	44 0	21 65 1	0,64
101-200	49 0	26 75 1	0.84	49 0	26 75 1	0,84	49 0	26 75 1	0,84
201-300	50 0	30 80 1	0.91	50 0	30 80 1	0,91	50 0	55 105 2	1,0
301-400	55 0	30 85 1	0.92	55 0	55 110 2	1,1	55 0	55 110 2	1,1
401-500	55 0	30 85 1	0.93	55 0	55 110 2	1,1	55 0	80 135 3	1,2
501-600	55 0	30 85 1	0.94	55 0	60 115 2	1,1	55 0	85 140 3	1,2
601 -800	55 0	35 90 1	0.95	55 0	65 120 2	1,1	55 0	85 140 3	1,3
801-1000	55 0	35 90 1	0.96	55 0	65 120 2	1,1	55 0	115 170 4	1,4
1001-2000	55 0	35 90 1	0.98	55 0	95 150 3	1,3	55 0	120 175 4	1,4
2001-3000	55 0	65 120 2	1.2	55 0	95 150 3	1,3	55 0	150 205 5	1,5
3001-4000	55 0	65 120 2	1.2	55 0	85 150 3	1,3	90 1	140 230 6	1,6
4001-5000	55 0	65 320 2	1.2	55 0	95 150 3	1,4	90 1	165 255 7	1,8
5001-7000	55 0	65 120 2	1.2	55 0	95 150 3	1,4	90 1	165 255 7	1,8
7001-10000	55 0	65 120 2	1.2	55 0	120 175 4	3,5	90 1	190 280 8	1,9
10001-20000	55 0	65 120 2	1.2	55 0	120 175 4	1,5	90 1	190 280 8	1,9
20001-50000	55 0	65 120 2	1.2	55 0	150 205 5	1,7	90 1	215 305 9	2,0
50001-100000	55 0	65 120 2	1.2	55 0	150 205 5	1,7	90 1	240 330 10	2.1

Porcentaje de tolerancia por lote de **elementos** defectuosos = 5 %

Riesgo del consumidor = 0.10

% de media del proceso	1.01-1.50			1.51-2.00			2.01-2.50		
Tamaño del lote	Prueba 1 n_1 c_1	Prueba 2 n_2 n_1+n_2 c_2	LCMS en %	Prueba 1 n_1 c_1	Prueba 2 n_2 n_1+n_2 c_2	LCMS en %	Prueba 1 n_1 c_1	Prueba 2 n_2 n_1+n_2 c_2	LCMS en %
1-30	Todo 0		0	Todo 0		0	Todo 0		0
31-50	30 0		0,49	30 0		0.49	30 0		0.49
51-75	38 0		0,59	38 0		0.59	38 0		0.59
76-100	44 0	21 65 1	0.64	44 0	21 65 1	0.64	44 0	21 65 1	0.64
101-200	49 0	51 100 2	0.91	49 0	51 100 2	0.91	49 0	51 100 2	0.91
201-300	50 0	55 105 2	1.0	50 0	80 130 3	1.1	50 0	100 150 4	1.1
301-400	55 0	80 135 3	1.1	55 0	100 155 4	1.2	85 1	105 190 6	1.3
401-500	55 0	105 160 4	1.3	85 1	120 205 6	1.4	85 1	140 225 7	1.4
501-600	55 0	110 165 4	1.3	85 1	145 230 7	1.4	85 1	165 250 8	1.5
601 -800	90 1	125 215 6	1.5	90 1	170 260 8	1.5	120 2	185 305 10	1.6
801-1000	90 1	150 240 7	1.5	90 1	200 290 9	1.6	120 2	210 330 11	1.7
1001-2000	90 1	185 275 8	1.7	120 2	225 345 11	1.9	175 4	260 435 15	2.0
2001-3000	120 2	180 300 9	1.9	150 3	270 420 14	2.1	205 5	375 580 21	2.3
3001-4000	120 2	210 330 10	2.0	150 3	295 445 15	2.3	230 6	420 650 24	2.4
4001-5000	120 2	255 375 12	2.1	150 3	345 495 17	2.3	255 7	445 700 26	2.5
5001-7000	120 2	260 380 12	2.1	150 3	370 520 18	2.3	255 7	495 750 28	2.6
7001-10000	120 2	285 405 13	2.1	175 4	370 545 19	2.4	280 8	540 820 31	2.7
10001-20000	120 2	310 430 14	2.2	175 4	420 595 21	2.4	280 8	660 940 36	2.8
20001-50000	120 2	335 455 15	2.2	205 5	485 690 25	2.5	305 9	745 1050 41	2.9
50001-100000	120 2	360 480 16	2.3	205 5	555 700 28	2.6	330 10	810 1140 45	3.0

Porcentaje de tolerancia por lote de elementos **defectuosos** = 5 %

Riesgo del consumidor = 0.10

Tabla 24

Método Dodge-Romig

LCMS en muestreo simple

% de media del proceso	0 - 0.04			0.05 - 0.40			0.41 - 0.80			0.81 - 1.20			1.21 - 1.60			1.61 - 2.00		
Tamaño del lote	*n*	*c*	$100\ p_{0.1}$	*n*	*c*	$100\ p_{0.1}$	*n*	*c*	$100\ p_{0.10}$	*N*	*c*	$100\ p_{0.10}$	*n*	*c*	$100\ p_{0.1}$	*n*	*c*	$100\ p_{0.1}$
1-15	todo	0		todo	0		todo	0		Todo	0		todo	0		todo	0	
16-50	14	0	13.6	14	0	13.6	14	0	13.6	14	0	13.6	14	0	13.6	14	0	13.6
51-100	16	0	12.4	36	0	12.4	16	0	12.4	16	0	12.4	16	0	12.4	16	0	12.4
101-200	17	0	12.2	17	0	12.2	17	0	12.2	17	0	12.2	35	1	10.5	35	1	10.5
201-300	17	0	12.3	17	0	12.3	17	0	12.3	37	1	10.2	37	1	10.2	37	1	10.2
301-400	18	0	11.8	18	0	11.8	38	1	10.0	38	1	10.0	38	1	10.0	60	2	8.5
401-500	18	0	11.9	18	0	119	39	1	9.8	39	1	9.8	60	2	8.6	60	2	8.6
501-600	18	0	11.9	18	0	11.9	39	1	9.8	39	1	9.8	60	2	8.6	60	2	8.6
601-800	18	0	11.9	40	1	9.6	40	1	9.6	65	2	8.0	65	2	8.0	85	3	7.5
801-1000	18	0	12.0	40	1	9.6	40	1	9.6	65	2	8.1	65	2	S.1	90	3	7.4
1001-2000	18	0	12.0	41	1	9.4	65	2	8.2	65	2	8.2	95	3	7.0	120	4	6.5
2001-3000	18	0	12.0	41	1	9.4	65	2	8.2	95	3	7.0	120	4	6.5	180	6	5.8
3001-4000	Í8	0	12.0	42	1	9.3	65	2	8.2	95	3	7.0	155	5	6.0	210	7	5.5
4001-5000	18	0	12.0	42	1	9.3	70	2	7.5	125	4	6.4	155	5	6.0	245	8	5.3
5001-7000	18	0	12.0	42	1	9.3	95	3	7.0	125	4	6.4	185	6	5.6	280	9	5.1
7001-10000	42	1	9.3	70	2	7.5	95	3	7.0	155	5	6.0	220	7	5.4	350	11	4.8
10001-20000	42	1	9.3	70	2	7.6	95	3	7.0	190	6	5.6	290	9	4.9	460	14	4.4
20001-50000	42	1	9.3	70	2	7.6	125	4	6.4	220	7	5.4	395	12	4.5	720	21	3.9
50001-100000	42	1	9.3	95	3	7.0	160	5	5.9	290	9	4.9	505	15	4.2	955	27	3.7
Límite de calidad media de salida 2%																		

Tabla 25

Método Dodge-Romig

LCMS en muestreo doble

% de media del proceso	0 - 0.04							0.05 - 0.40							0.41 - 0.80						
Tamaño del lote	Prueba 1		Prueba 2				$100\,p_{0.1}$	Prueba 1		Prueba 2				$100\,p_{0.1}$	Prueba 1		Prueba 2				$100\,p_{0.1}$
	n_1	c_1	n_2	n_1+n_2	c_2			n_1	c_1	n_2	n_1+n_2	c_2			n_1	c_1	n_2	n_1+n_2	c_2		
1-15	todo	0						todo	0						todo	0					
16-50	14	0					13.6	14	0					13.6	14	0					13.6
51-100	21	0	12	33	1		11.7	21	0	12	33	1		11.7	21	0	12	33	1		11.7
101-200	24	0	13	37	1		11.0	24	0	13	37	1		11.0	24	0	13	37	1		11.0
201-300	26	0	15	41	1		10.4	26	0	15	41	1		10.4	29	0	31	60	2		9.1
301-400	26	0	16	42	1		10.3	26	0	16	42	1		10.3	30	0	35	65	2		9.0
401-500	27	0	16	43	1		10.3	30	0	35	65	2		9.0	30	0	35	65	2		9.0
501-600	27	0	16	43	1		10.3	31	0	34	65	2		8.9	35	0	55	90	3		7.9
601-800	27	0	17	44	1		10.2	31	0	39	70	2		8.8	35	0	60	95	3		7.7
801-1000	27	0	17	44	1		10.2	32	0	38	70	2		8.7	36	0	59	95	3		7.6
1001-2000	33	0	37	70	2		8.5	33	0	37	70	2		8.5	37	0	63	100	3		7.5
2001-3000	34	0	41	75	2		8.2	34	0	41	75	2		8.2	41	0	84	125	4		7.0
3001-4000	34	0	41	75	2		8.2	38	0	62	100	3		7.3	41	0	89	130	4		6.9
4001-5000	34	0	41	75	2		8.2	38	0	62	100	3		7.3	42	0	88	130	4		6.9
5001-7000	35	0	40	75	2		8.1	38	0	62	100	3		7.3	44	0	116	160	5		6.4
7001-10000	35	0	40	75	2		8.1	38	0	62	100	3		7.3	45	0	115	160	5		6.3
10001-20000	35	0	40	75	2		8.1	39	0	66	105	3		7.2	45	0	115	160	5		6.3
20001-50000	35	0	40	75	2		8.1	43	0	92	135	4		6.6	47	0	148	195	6		6.0
50001-100000	35	0	45	80	2		8.0	43	0	92	135	4		6.6	85	1	185	270	8		5.2
Límite de calidad media de salida 2%																					

% de media del proceso	0.81 - 1.20							1.21 - 1.60							1.61 - 2.00						
Tamaño del lote	Prueba 1		Prueba 2				$100\,p_{0.1}$	Prueba 1		Prueba 2				$100\,p_{0.1}$	Prueba 1		Prueba 2				$100\,p_{0.1}$
	n_1	c_1	n_2	n_1+n_2	c_2			n_1	c_1	n_2	n_1+n_2	c_2			n_1	c_1	n_2	n_1+n_2	c_2		
1-15	todo	0						todo	0						todo	0					
16-50	14	0					13.6	14	0					13.6	14	0					13.6
51-100	21	0	12	33	1		11.7	21	0	12	33	1		11.7	23	0	23	46	2		10.9
101-200	27	0	28	55	2		9.6	27	0	28	55	2		9.6	27	0	28	55	2		9.6
201-300	29	0	31	60	2		9.1	32	0	48	80	3		8.4	32	0	48	80	3		8.4
301-400	33	0	52	55	3		8.2	33	0	52	85	3		8.2	36	0	69	105	4		7.6
401-500	34	0	56	90	3		7.9	36	0	74	110	4		7.5	60	1	90	150	6		7.0
501-600	35	0	55	90	3		7.9	37	0	78	115	4		7.4	65	1	95	160	6		6.8
601-800	38	0	82	120	4		7.3	38	0	82	120	4		7.3	70	1	120	190	7		6.4
801-1000	38	0	87	125	4		7.2	70	1	100	170	6		6.5	70	1	145	215	8		6.2
1001-2000	43	0	112	155	5		6.5	80	1	160	240	8		5.8	110	2	205	315	11		5.5
2001-3000	75	1	115	190	6		6.1	115	2	195	310	10		5.3	160	3	310	470	15		4.7
3001-4000	80	1	140	220	7		5.8	120	2	255	375	12		5.0	235	5	415	650	20		4.3
4001-5000	80	1	175	255	8		5.5	125	2	285	410	13		4.9	275	6	475	750	23		4.2
5001-7000	85	1	205	290	9		5.3	125	2	320	445	14		4.8	280	6	575	855	26		4.1
7001-10000	85	1	210	295	9		5.2	165	3	335	500	15		4.5	320	7	645	965	29		4.0
10001-20000	90	1	260	350	11		5.1	170	3	425	595	18		4.4	395	9	835	1230	37		3.9
20001-50000	130	2	300	430	13		4.7	205	4	515	720	22		4.3	480	11	1090	1570	46		3.7
50001-100000	135	2	345	480	14		4.5	250	5	615	865	26		4.1	580	13	1460	2040	58		3.5
Límite de calidad media de salida 2%																					

APÉNDICE B

GLOSARIO

Acción correctora: Acción emprendida para eliminar las causas de una no conformidad o defecto u otra situación no deseable existente con el propósito de evitar que vuelva a ocurrir.

Acción preventiva: Aquella que se realiza con antelación a la aparición de una no conformidad.

Actualización: Operación para corregir o reemplazar datos en un conjunto de datos.

Aleatorio: Valor o suceso fortuito.

Análisis cluster: Técnica que se utiliza para clasificar un único conjunto de datos en una serie de grupos que no están definidos previamente.

Análisis de varianza (ANOVA, *Analysis of Variance*): Técnica estadística de análisis de grupos de datos basada en el análisis de la varianza en y entre los grupos de datos para determinar si se pueden considerar un mismo grupo general.

AR (*AutoRegresive*): Proceso o modelo autorregresivo, en el que existe correlación entre sus valores.

ARIMA (*AutoRegresive Integrated Moving Average*): Técnica estadística de análisis de series de datos no estacionarios homogéneos, basada en procesos o modelos autorregresivos integrados de media móvil.

ARMA (*AutoRegresive Moving Average*): Técnica estadística de análisis de series de datos basada en procesos o modelos autorregresivos de media móvil.

Aseguramiento de la calidad: Conjunto demostrable de acciones planificadas, implantadas dentro del sistema de la calidad para proporcionar la confianza adecuada de que una entidad cumplirá con los requisitos de la calidad.

Atributo: Representación de una propiedad, cualidad o rasgo esencial de un objeto o entidad.

Atributos: (ver *dato cualitativo*).

Auditor: Persona cualificada y autorizada para realizar la evaluación de sistemas de gestión de la calidad.

Auditoría de la calidad: Proceso sistemático, independiente y documentado para obtener evidencias y determinar si las actividades y los resultados relativos a la calidad cumplen las disposiciones establecidas, si éstas disposiciones están implantadas de forma efectiva y son adecuadas para alcanzar los objetivos.

Autoevaluación: En el ámbito del modelo EFQM (*European Foundation for Quality Management*) es un proceso interno de re-

flexión participativa que concluye con la elaboración de un informe donde se sustancian las valoraciones y propuestas de mejora.

Banda de error: Zona de incertidumbre alrededor de un elemento lineal. También se denomina banda épsilon.

Base de Datos Geográficos (BDG): Conjunto o colección identificable de datos geográficos.

Binomial: Distribución estadística en la que los sucesos sólo pueden tomar dos estados p ó q.

Bootstrap: Método de remuestreo para estimar el error medio cuadrático, en el que interviene la simulación. También se le denomina "estimación autosuficiente".

Calidad: Conjunto de características de una entidad que le confieren su aptitud para satisfacer las necesidades explícitas y las implícitas.

Calidad de procesos: Se refiere a la gestión de la calidad.

Calidad de producto: (ver *calidad*).

Calidad de rechazo: En la aceptación de productos, el nivel de defectuoso que lleva a rechazar un lote (calidad límite, CL).

Calidad técnica: Calidad que se obtiene en la realización de un producto o servicio.

Calidad teórica: Calidad del diseño de un producto o servicio.

Calidad total: Modelo de gestión basado en la organización, los procesos y personas tal que se orientan a la misión de la propia empresa para obtener niveles de calidad sostenidos.

Capacidad de un proceso: Puede referirse a la capacidad de un proceso completo o de una máquina. Se refiere a la variación natural de los valores obtenidos sobre una característica con dicho proceso. Los índices de capacidad expresan la probabilidad de que el proceso satisfaga las tolerancias marcadas.

Carácter: Es una propiedad de los elementos de la población, por ejemplo, las coordenadas x, y, z de una población de vértices geodésicos.

Categoría: (ver *modalidad*).

Celda: También denominado tesela o píxel. Es la primitiva geométrica bidimensional que discretiza la información geográfica espacial en el modelo ráster.

Certificación: Proceso llevado a cabo por una entidad reconocida como independiente de las partes interesadas, en la que se manifiesta por escrito que se dispone de la confianza adecuada de que un producto, proceso o servicio debidamente identificado, es conforme con una norma específica.

Certificado de conformidad: Documento emitido por una entidad de certificación que indica que existe la seguridad requerida de que un producto, servicio o proceso especificado es conforme a una norma específica.

CL: Calidad límite. Es el máximo porcentaje de defectos que se está dispuesto a admitir en un lote aislado (en inglés se denomina *Limiting Quality Level, LQL*.

Cliente: Organización o persona que recibe un producto. Cliente interno se refiere a clientes dentro de la propia organización.

CMS (AOQ, *Average Output Quality*): Calidad media de salida. Determina la calidad de todos los lotes resultantes de la inspección, tanto de los aceptados como de los rechazados después de haber sido inspeccionados al 100% y se hayan sustituido todos los elementos defectuosos.

Codificación: Procedimiento de conversión de la información en un formato especificado para su procesado.

Compleción: Hecho de estar completo o pleno. Parámetro de calidad que describe la presencia y ausencia de instancias de entidades, instancias de relaciones e instancias de atributos.

Confiabilidad: Probabilidad de que un producto determinado realice las funciones para las

que fue creado de manera satisfactoria durante un tiempo predeterminado y bajo ciertas condiciones de uso.

Confianza: Elemento de metacalidad que describe la corrección de la información de calidad.

Conformidad: Cumplimento de requisitos especificados.

Conglomerado: Unión de unidades últimas para realizar un muestreo.

Consistencia geométrica: Consistencia lógica relativa sólo a las especificaciones de los aspectos geométricos de la información geográfica.

Consistencia lógica: Grado de corrección o conformidad de un conjunto de datos geográficos con respecto a la estructura interna descrita por sus especificaciones.

Consistencia semántica: Consistencia lógica relativa sólo a las especificaciones de los aspectos semánticos de la información geográfica.

Consistencia topológica: Consistencia lógica relativa sólo a las especificaciones de los aspectos geométricos y topológicos de la información geográfica.

Control de calidad: Las técnicas y actividades de carácter operativo utilizadas para cumplir los requisitos para la calidad.

Curva Característica Operativa: Es la función que, para un plan de muestreo determinado, relaciona la probabilidad de aceptación con el porcentaje de defectuoso del lote.

Dato continuo: Son aquellos que pueden tomar cualquier valor dentro de un intervalo.

Dato cualitativo: Son aquellos que se refieren a una cualidad de la población estudiada. No pueden ser medidos numéricamente.

Dato cuantitativo: Son aquellos que pueden ser medidos numéricamente.

Dato discreto: Son aquellos que sólo toman valores numéricos aislados.

Disponibilidad: Relación entre fiabilidad y mantenibilidad de un equipo.

Distancia de Cook: Medida resumen de la influencia de una única observación basado en el cambio total del resto de los residuos cuando se elimina dicha observación del proceso de estimación del modelo de regresión. Grandes valores (generalmente mayores que uno) indican una influencia sustancial de la observación que afecta a los coeficientes de regresión estimados.

Distancia: Función con la que se mide la separación entre puntos.

Dominio: Conjunto de valores permitidos que puede tomar un atributo o una función.

Edición: procedimiento para: a) eliminación de errores en BDG, o para b) la codificación de los atributos de los datos geográficos.

Efecto de abstracción: Distorsión de la información de calidad debida a especificaciones imperfectas y poco precisas.

Eficacia: Cumplimiento de los objetivos marcados.

Eficiencia: relación entre los resultados y los recursos utilizados.

Elemento de calidad: Ítem de información que describe la calidad de un conjunto de datos geográficos.

Enfoque: Manera de dirigir la atención o el interés hacia un asunto o problema desde unos supuestos previos, para tratar de resolverlo acertadamente.

Entidad: Fenómeno real localizado sobre la superficie terrestre que no puede ser dividido en fenómenos del mismo tipo. Cada clase de información definida por unas propiedades comunes.

Error: Separación respecto a un valor teórico o real de una medida o atributo.

Error accidental: También denominado aleatorio dado que no responde a una ley fija. Se corresponden con defectos que se compensan cuando se realiza un numero elevado de mediciones.

Error grosero: Equivocación que se puede cometer en una medida o transcripción de un dato.

Error sistemático: El debido a la existencia de una causa permanente que introduce un sesgo o desviación que se puede eliminar si se conoce.

Escáner: Dispositivo para convertir automáticamente informaciones en soportes analógicos (por ejemplo, mapas en papel o fotografías) a informaciones digitales en modo ráster.

Especificación técnica: Documento que define las características requeridas en un producto (por ejemplo, niveles de calidad, uso específico, seguridad, dimensiones...).

Esperanza matemática: Valor promedio de una variable aleatoria.

Estadístico muestral: Función de la variable aleatoria que no depende de ningún parámetro desconocido.

Estimador: Estadístico muestral que corresponde a un parámetro desconocido.

Estratificación: Proceso por el que un conjunto mayor se divide, según cierta propiedad, en subconjuntos menores para su análisis.

Evidencia objetiva: Información cuya veracidad se puede demostrar con hechos obtenidos a través de la observación, la medición, el ensayo u otros medios.

Exactitud: Grado de concordancia entre el resultado de un ensayo y el valor de referencia aceptado. La exactitud se refiere a la calidad de un resultado y se diferencia del concepto de precisión en que es relativa a la calidad de la operación más que al propio resultado.

Exactitud posicional: Grado de exactitud/fiabilidad del procedimiento de adquisición de los atributos posicionales del dato geográfico. Parámetro de calidad que describe la exactitud (X, Y y Z) de las posiciones geográficas en un conjunto de datos geográficos.

Exactitud temática: También exactitud semántica. Grado de exactitud/fiabilidad del procedimiento de adquisición de los atributos temáticos del dato geográfico. Parámetro de la calidad que describe la exactitud de los aspectos temáticos/semánticos de un conjunto de datos geográficos.

Exactitud temporal: Parámetro de calidad que describe la exactitud de los aspectos temporales de los datos geográficos.

Experimento aleatorio: Experimento del cual se desconoce su resultado cada vez que se realiza.

Fiabilidad: Elementos de la metacalidad que describen la probabilidad de que una muestra de un conjunto de datos geográficos usada para la evaluación de la calidad, sea representativa del conjunto de datos total.

Frecuencia de Nyquist: Según la teoría de procesado de señales es la frecuencia mínima necesaria para muestrear una señal. La frecuencia de muestreo debe ser el doble de la que se desea caracterizar.

Gestión de la Calidad: Actividades coordinadas para dirigir y controlar una organización en lo relativo a la calidad.

Gráfico de control: Medios gráficos que permiten controlar procesos mediante la representación de la evolución de características medibles que poseen los productos generados en dicho proceso.

Herramientas avanzadas: En el ámbito de la mejora continua de la calidad, conjunto formado por siete herramientas de mayor complejidad que las básicas y que se orientan a niveles decisionales más altos. Las herramientas son: diagrama de afinidad, de relaciones, matricial, de árbol, de proceso de decisión, diagramas de actividades y las matrices de priorización.

Herramientas básicas: En el ámbito de la mejora continua de la calidad, conjunto formado por siete herramientas de fácil aplicación en la mejora: diagrama de Pareto, diagrama de Ishikawa, diagrama de dispersión, histograma, estratificación, hoja de verificación y gráficos de control.

Histograma: Distribución de las frecuencias de ocurrencia de un suceso en función de unas clases o valores del mismo.

Homologación: Acto administrativo, mediante el cual una entidad de derecho público, atestigua que un producto o servicio cumple la reglamentación técnica que le es de aplicación.

Incertidumbre: Falta de certeza, nivel de error.

Indicador de calidad: Para un conjunto de datos geográficos dado es el conjunto de medidas que describe un parámetro de calidad.

Individuo: Es cada uno de los elementos de la población.

Inferencia: Parte de la Estadística que toma decisiones sobre una población a partir de una muestra seleccionada de ella.

Información geográfica: Información que concierne a fenómenos directa o indirectamente asociados con una ubicación relativa a la Tierra.

Inspección: Actividades tales como medir, examinar, ensayar o comparar con un patrón una o varias características y confrontar los resultados con requisitos especificados para así establecer si se logra la conformidad para cada una de estas características.

Instancia: Valor individual.

Instrucciones de trabajo: Conjunto de todos los documentos esenciales para la operatividad del sistema de gestión de la calidad (instrucciones operativas, manuales de uso de los equipos, manuales del usuario, catálogos, normas de actuación, códigos de conducta, etc.).

Intervalo de confianza: Conjunto de valores entre los que se encuentra el verdadero valor del parámetro estimado con una probabilidad predeterminada.

Kappa: En la evaluación de la exactitud temática, índice global que permite valorar la bondad del ajuste entre columnas y filas en una matriz de confusión.

Linaje: Registro de las características, de las fuentes de datos, de las operaciones, etc., que han creado el producto final, con el fin de poder reconstruir la historia de un producto o servicio.

Lista de verificación: Documento de trabajo que describe los aspectos del sistema de calidad y que puede ser investigado por el auditor y con el cual éste, verifica que se cumplen los objetivos y alcance de la auditoría. También se le conoce como "*check list*".

Lote: Conjunto de productos obtenidos de un proceso homogéneo.

MA (*Moving Average*): Proceso o modelo de medias móviles.

Magnitud: Atributo de un fenómeno u objeto que puede ser distinguido cualitativamente y determinado cuantitativamente.

MANOVA: Análisis de la Varianza Multivariante.

Mantenibilidad: Probabilidad de reparación en un tiempo dado.

Manual de calidad: Documento que describe el sistema de calidad de la organización y en el que se define la política de calidad, la estructura organizativa y las políticas sobre aspectos específicos del sistema de calidad.

Manual de procedimientos: Documento complementario al manual de calidad que describe las actividades que se realizan en la organización, especialmente aquellas que afectan al sistema de calidad.

Marca de conformidad: Marca registrada emitida según las reglas de un sistema de certi-

ficación que indica que un producto, proceso o servicio es conforme a una norma específica.

Matriz de confusión: También denominada matriz de error. Tabla de contingencia que se utiliza para la cuantificación de la exactitud temática en cartografía. En ella se presentan tanto los elementos correctamente clasificados como aquellos erróneamente atribuidos a una clase.

Medición: Operación cuya finalidad es determinar el valor de una magnitud.

Medida de calidad: Definición de un ensayo específico a realizar sobre un conjunto de datos geográficos que incluye algoritmos y el tipo del valor o conjunto de valores a obtener.

Mejora continua: Actividad recurrente para aumentar la capacidad para cumplir con los requisitos.

Mejora de la calidad: Parte de la gestión de la calidad orientada a aumentar la capacidad de cumplir con los requisitos de la calidad

Mensurando: Magnitud sometida a medición.

Metacalidad: Datos que informan sobre los datos de calidad de un conjunto de datos geográficos.

Metadatos: Datos sobre un conjunto o conjuntos de datos geográficos.

Métrica: Medida en un conjunto.

Modalidad: Es cada uno de los valores que puede presentar cada carácter, por ejemplo, cada uno de los valores que puede tomar una coordenada.

Modelo de calidad: Modelo conceptual de los aspectos de calidad de los datos geográficos.

Modelo de datos: Representación de la organización lógica de los datos de una base de datos mediante la definición de entidades, atributos y relaciones entre los atributos. En los SIG se vienen utilizando comúnmente los modelos ráster y vectorial.

Modelo de excelencia: Modelo de gestión de las organizaciones con mayores exigencias que los sistema de gestión de la calidad.

Módulo: El módulo algebraico de orden m de un valor entero positivo d es el valor entero c entre 0 y $m-1$ tal que $d=k\cdot m+c$ para algún valor entero positivo de k.

Momento de una variable: Valores construidos a partir de la distribución de frecuencias o de probabilidad de la variable que resumen la información en relación a alguna propiedad de ella.

Muestra: Es un subconjunto pequeño de la población, que debe ser representativo de toda ella, por ejemplo, los puntos de control empleados para evaluar la exactitud posicional.

Multinomial: Distribución estadística en la que se consideran diversos estados posibles. La distribución binomial es un caso particular de ésta.

Nivel de calidad Aceptable (NCA): En la aceptación de productos, es el máximo número de defectuosos que se admite como promedio de los porcentajes de defectuoso que aparecen en los lotes (en inglés, *Acceptable Quality Level, AQL*).

No conformidad: Fallo en el cumplimiento de un requisito especificado en una norma o en los procedimientos de calidad.

Norma: Documento público, no obligatorio, consensuado entre las partes interesadas y aprobada por un organismo de Normalización, que recoge conocimientos y recomendaciones científicotécnicas y surgidas de la experiencia orientado a la simplificación, unificación y clasificación de los procesos y productos dentro de un ámbito territorial determinado.

Norma Europea: Especificación técnica que se desarrolla en base al consenso de los organismos de normalización europea, miembros del CEN (Comité Europeo de Normalización), CENELEC (Comité Europeo de Normalización Electrotécnica), ó ETSI (European Telecommunications Standards Institute, Instituto Europeo de Normas de Telecomunicaciones) que

adopta el estatus de norma nacional por la publicación del texto o mediante su ratificación en un periodo no superior a seis meses.

Normalización: Conjunto de actividades consistentes en la elaboración, difusión y aplicación de normas como forma de establecer soluciones productivas a situaciones repetitivas.

Números aleatorios: Una secuencia infinita de números es aleatoria si cualquier secuencia finita seleccionada previamente a su diseño es igualmente factible de estar en ella.

Números seudoaleatorios: Números que tienen carácter aleatorio, pero que han sido obtenidos mediante fórmulas recursivas, aritméticas, etc.

Organización: Compañía, corporación, firma, empresa o institución, o parte de las mismas constituidas como sociedad o no, pública o privada, que tiene sus propias funciones y su propia administración.

Outlier: Dato que se considera extraño, atípico o aberrante, al presentar una probabilidad de ocurrencia muy pequeña

Parámetro: Característica de la variable aleatoria que forma parte de la definición de su función de probabilidad.

Plan de calidad: Documento que enuncia las prácticas, los recursos y la secuencia de las actividades relacionados con la calidad, que son específicas a un producto, un proyecto o un contrato en particular.

Plan de muestreo: Conjunto de características que definen una aceptación por muestreos: tamaño del lote, calidad considerada, tamaño de la muestra, número de aceptación, número de rechazo, etc.

Planificación de la calidad: Actividades que establecen los objetivos de calidad, así como los requisitos para la aplicación de los elementos del sistema de calidad.

Población: Es el conjunto de elementos objeto de estudio, por ejemplo el conjunto de puntos de un mapa.

Política de calidad: Directrices y objetivos generales de una organización con respecto a la calidad, expresadas de manera formal, y que es aprobada y revisada periódicamente por la alta dirección de la organización.

Precisión: Calidad asociada al grado de refinamiento de los instrumentos de medida que determina el grado de uniformidad o coherencia de sucesivas mediciones del mismo fenómeno. La precisión también se refiere al número de cifras decimales con las que se aproxima una medición.

Prestación del servicio: Aquellas actividades del proveedor para suministrar el servicio.

Procedimiento: Forma especificada de desarrollar una actividad o un proceso.

Proceso: Conjunto de recursos y actividades interrelacionados que transforma entradas en salidas o productos.

Producto: Resultado de las actividades o procesos.

Punto de Riesgo Compartido: En una curva característica operativa es el punto en el que los riesgos del productor y usuario valen lo mismo (50%).

Racha: En una serie, el conjunto de elementos que comparten una característica.

Rango o recorrido: Se utiliza como estimación de la variación en una muestra, se determina con la amplitud del intervalo que marcan el valor máximo y mínimo.

Registro: Documento que suministra evidencia objetiva de las actividades efectuadas o de los resultados alcanzados.

Repetitividad: Grado de concordancia entre los resultados de mediciones sucesivas sobre un mismo mensurando.

Representatividad: Índice que expresa si la información contenida en una BDG es efectivamente suficiente para describir una cierta área geográfica (equivalente a la compleción).

Reproceso: Acciones desarrolladas sobre un producto no conforme para que cumpla con los requisitos.

Reproducibilidad: Grado de concordancia entre los resultados de las mediciones de un mismo mensurando realizadas bajo diferentes condiciones de medida.

Requisito: Necesidad o expectativa establecida, implícita u obligatoria.

Resolución: Dimensión del más pequeño elemento representado en un mapa, o de la celda ráster (resolución espacial). Lapso temporal en el paso de un satélite de observación de la Tierra (resolución temporal).

Revisión del contrato: Actividades sistemáticas efectuadas por el proveedor antes de firmar el contrato, que le permiten asegurarse de que los requisitos de calidad están definidos en forma adecuada, sin ambigüedad, documentados y que son realizables.

Revisión del diseño: Examen documentado, completo y sistemático de un diseño con el propósito de evaluar su capacidad para cumplir los requisitos de calidad, identificar problemas, si los hay, y proponer el desarrollo de soluciones.

Revisión por la Dirección: Evaluación formal, por parte de la alta dirección, acerca de la situación actual y de la adecuación del sistema de calidad en relación con la política de la calidad y sus objetivos.

Riesgo del Productor: En la aceptación de productos, probabilidad de que se le rechace un trabajo bien hecho.

Riesgo del Usuario: En la aceptación de productos, probabilidad de que acepte por bueno un trabajo cuya calidad es inferior a la establecida.

Serie temporal: Modelo estadístico que estudia la evolución de una variable en el tiempo.

Simulación: Programa de ordenador que incorpora un modelo de un sistema tal que permite estudiar el sistema y obtener conclusiones para la toma de decisiones.

Sistema: Conjunto de elementos mutuamente relacionados o que interactúan.

Sistema de Gestión de la Calidad (SGC): Parte del sistema de gestión de la organización enfocada al logro de resultados en relación con los objetivos de la calidad.

Sistematismo: Presencia de una perturbación constante en un conjunto de datos. Puede provocar la aparición de errores sistemáticos.

Suceso aleatorio: Cualquier posible resultado de un experimento aleatorio.

Terreno nominal: Concepción del mundo real derivada de las especificaciones de un conjunto de datos geográficos. El terreno nominal constituye el conjunto de datos geográficos ideal con el que se compara el conjunto de datos geográficos real para evaluar su calidad.

Tesauro: Diccionario o catálogo.

Tipificación: En Estadística, transformación de una variable en otra con media cero y varianza uno.

Topología: Estudio de las propiedades de los objetos geométricos que no dependen de las propiedades métricas (por ejemplo la adyacencia).

Trazabilidad: Capacidad para rastrear la historia, la aplicación o la localización de una entidad de interés mediante identificaciones registradas.

Universo del discurso: Conjunto de entidades percibidas de la realidad que sirve como marco de referencia para la obtención de un conjunto de datos. También se le denomina "universo abstracto" o "terreno nominal".

Variable: Se utiliza esta denominación para referirse al carácter o caracteres estudiados, por ejemplo, la x y la y en un control posicional planimétrico.

Variable aleatoria: Función que hace corresponder un número a cada suceso elemental de un experimento aleatorio, con la que se trabaja en Probabilidad.

Variable continua: Variable que utiliza datos continuos.

Variable discreta: Variable que utiliza datos discretos.

Variograma: Expresión de la correlación de la diferencia de valores para dos posiciones en función de la distancia entre las mismas. Para algunos autores el término es equivalente a semivariograma.

Verificación: Confirmación mediante examen y aporte de evidencia objetiva de que se han cumplido los requisitos especificados.

APÉNDICE C

REFERENCIAS

Álvarez, F. (1994). *Estadística Aplicada.* Mignon Libreros.

Ariza, F.J. (1997). Uso de la simulación en el control de errores en la rasterización. *Mapping*, 39: 24-34.

Ariza, F.J. (2002). *Calidad en la Producción Cartográfica*. Ra-Ma. Madrid.

Ariza, F.J.; Pinilla, C. (1996). Control de calidad del proceso de clasificación de imágenes de satélite. *Mapping*, 34: 74-86.

ASCI (1983). *Map uses, scales and accuracies for engineering and associated purposes.* American Society of Civil Engineers, Committee on Cartographic Surveying, Surveing and Mapping Division, New York.

ASP (American Society of Photogrammetry) (1985). Accuracy Specification for Large-Scale Line Maps. *Photogrammetric Engineering & Remote Sensing*, 51(2): 195-199.

ASPRS (American Society of Photogrammetry and Remote Sensing) (1990). Accuracy standards for large scale maps. *Photogrammetric Engineering and Remote Sensing*, 56(7): 1068-1070.

Atkinson, A. (1998). *Control de calidad posicional por GPS de las hojas 926 I-1 y I-2 del Mapa Topográfico de Andalucía.* Proyecto Fin de Carrera. Escuela Politécnica Superior. Universidad de Jaén.

Atkinson, A.; Ariza, F.J. (2000). Control de calidad posicional en cartografía mediante simulación y técnicas GPS. *Actas del XII Congreso Internacional de Ingeniería Gráfica*. Valladolid.

Atkinson, A.; García, J.L.; Ariza, F.J. (2001a). Los diferentes test para el control de calidad posicional en cartografía. *Actas del XIII Congreso Internacional de Ingeniería Gráfica*. Universidad de Extremadura.

Atkinson, A.; García, J.L.; Ariza, F.J. (2001b). Análisis comparativo de los diferentes test para el control de calidad posicional en cartografía empleando procesos de simulación. *Actas del XIII Congreso Internacional de Ingeniería Gráfica*. Universidad de Extremadura.

Azorín, F.; Sánchez-Crespo, J.L. (1994). *Métodos y Aplicaciones del Muestreo.* Alianza Universidad Textos.

Baltsavias, E.P. (1994). The Agfa Horizon DTP Scanner-Characteristics, Testing and Evaluation. ISPRS Commission I Symposium, Como, Italia. *IAPRS*, 30 (Part 1): 171-179.

Baltsavias, E.P.; Kaeser, C. (1998). Evaluation and Testing of the Zeiss SCAI Roll Film Scanner. ISPRS Commission I Symposium, Bangalore, India. *IAPRS*, 32(Part 1): 67-74.

Baltsavias, E.P.; Waegli, B. (1996). Quality Analysis and Calibration of DTP Scanners.

18 ISPRS Congress, Vienna, Austria. *IAPRS*, 31(B1): 13-19.

Belart, P. (1999). *Control de calidad en las clasificaciones ocupacionales del suelo mediante imágenes de sátelite en la provincia de Jaén.* Proyecto Fin de Carrera. Escuela Politécnica Superior. Universidad de Jaén.

Bie, S. (1971). Quality control in soil survey. Introduction: 1.- The choice of mapping unit. *Journal of Soil Science*, 22: 32-49.

Brassel, K.; Bucher, F.; Stephan, E.M., Vckovski, A. (1995). Completeness. En: Guptill, S. y Morrison, J. (Ed.) *Elements of Spatial data quality*. Pergamon-ICA.

Canavos, G. (1986). *Probabilidad y Estadística. Aplicaciones y Métodos*. McGraw-Hill.

Caridad, J.M. (1985). *Cálculo de probabilidades y análisis de datos I y II*. Universidad de Córdoba.

Carrasco, J.; Hernán, M. (1993). *Estadística Multivariante en las ciencias de la vida.* Editorial Ciencia.

Caspary, W. (1992). *Error bands as measures of geometrical accuracy. EGIS'92*. Munich.

Castillo, E. (1993). *Introducción a la Estadística Aplicada con Mathematica.* Paraninfo, S.A., Madrid.

Cid, A.; Delgado, C.A.; Leguey, S. (1999). *Introducción al Muestreo en Poblaciones Finitas.* Editorial Nuevas Estructuras. Madrid.

CMA, 1995. *Usos y coberturas de Andalucía a Escala 1/50.000*. Consejería de Medio Ambiente, Junta de Andalucía, Sevilla.

Coduras, A.; Costa, A. (1990). *Diseño muestral para un control de calidad de las revisiones catastrales realizadas para el Centro de Gestión catastral y Cooperación Tributaria de Barcelona.* Departamento de Econometría, Estadística y Economía Española. Universidad de Barcelona.

Congalton, R. (1988). A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54: 593-600.

Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37: 35-46.

Congalton, R. (1993). A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, 59(5): 641-644.

Congalton, R.; Green, K. (1998). *Assesing the accuracy of remotely sensed data: Principles and Practices*. Lewis Publishers.

CSG (1992). *Normas y condiciones técnicas para los levantamientos cartográficos a grandes escalas*. Consejo Superior Geográfico. Ministerio de Obras Públicas y Transportes. Instituto Geográfico Nacional. Madrid.

Cuatrecasas, L. (1999). *Gestión Integral de la calidad. Implantación, Control y Certificación*. Ed. Gestión 2000, Barcelona.

Cuening, R. (1971). *Cartographie Générale*. Eyrolles. Paris.

Curran, P. (1985). *Principles of remote sensing*. Longman Scientific and Technical.

Davis, J. C. (1986). *Statistics and Data Analysis in Geology*. Wiley & Sons. New York.

Delgado, J.; Cardenal, F.J. (1998). *Informe de Calibración del escáner Agfa Horizon Plus* (informe interno). Servicios Técnicos de Apoyo a la Investigación. Universidad de Jaén.

Deming, E. (1989). *Calidad, productividad y competitividad: la salida de la crisis*. Ed. Diaz de Santos. Madrid.

Deseilligny, M.P. (1994). *Lecture automatique de cartes scannees*. Institut Geographique National. Francia.

Domingo, J.; Arranz A. (2000). *Calidad y mejora continua*. Ed. Donostiarra, San Sebastián.

Drummond, J.; Ramlal, B. (1992). A prototype uncertainty sub-system implemented in ITC's ILWIS PC-based GIS, and tested in a Dutch land reallotment project. En *EGIS'92* (1): 234-243.

Earickson, R.J.; Harlin, J.M. (1994). *Geographic Measurement and Quantitative Analysis*. Macmillan College Publishing Company. New York.

Feigenbaum, A.(1990). *Control Total De Calidad*. Mcgraw-Hill.

FFII (1999). *Curso sobre mejoras de la Calidad. Herramientas avanzadas.* Fundación para el Fomento de la Innovación Industrial. Madrid

FGDC (1994). *Content Standards for Digital Geospatial Metadata.* Federal Geographic Data Committee, Washington, D.C.

FGDC (1998). *Geospatial Positioning Accuracy Standards*. Federal Geographic Data Committee. Washington, D.C.

Fienberg, S.E. (1970). An iterative procedure for estimation in contingency tables. *Annals of Mathematical Statistics*, 41: 907-917.

Fisher, N. (1995). *Statistical analysis of circular data.* Cambridge University Press.

Fogel, D.N.; Tinney, L.R. (1996). *Image registrattion using multiquadratic functions, the finite element method, bivariate mapping polynomials and the thin plate spline.* NCGIA. Santa Bárbara, California.

Foody, G. (1992). On the Compensation for Change Agreement in Image Classification Accuracy Assessment. *Photogrammetric Engineering and Remote Sensing*, 58(10): 1459-1460.

Galoway, D. (1998). *Mejora continua de procesos*. Ed. Gestión 2000, Barcelona.

García, M.; Martín, M. (1997). *La multiplicidad de p-valores: un riesgo a considerar.* JANO.

Goodchild, M. (1995). Attribute accuracy. En: *Elements of Spatial Data Quality.* International Cartographic Association.

GPO (1979). *GPO Contract Terms. Quality Assurance Through Attributes Program for Printing and Binding.* U. S. Goverment Printing Office. Publication 310.1.

Greenwalt, C.R.; Shultz, M.E. (1962). *Principles of Error Theory and Cartographic Applications.* ACIC Technical Report No. 96. Aeronautical Chart and Information Center, United States Air Force. St Louis.

Guptill, S.; Morrison, J. (ed.) (1995). *Elements of Spatial Data Quality.* Elsevier Science Ltd., Oxford.

Gutiérrez, R.; Martínez, A.; Rodríguez, C. (1993). *Curso Básico de Probabilidad.* Editorial Pirámide.

Gutiérrez, R.; Martínez, A.; Rodríguez, C. (1993). *Inferencia Estadística. Un enfoque clásico.* Editorial Pirámide.

Haering, S.; Kersten T.; Baltsavias, E.P. (1998). Quality Analysis of the LH Systems DSW300 Roll Film Scanner. ISPRS Commission I Symposium, Bangalore, India. *IAPRS*, 32(Part 1): 53-66.

Hair, J.F; Anderson, R.E.; Tatham, R.L.; Black, W.C. (2000). *Análisis Multivariante.* Prentice Hall.

Hansen, B.; Ghare, P. (1990). *Control de Calidad. Teoría y Aplicaciones.* Ediciones Díaz de Santos.

Hardin, P.; Shumway, M. (1997). Statistical significance and normalized confusion matrices. *Photogrammetric Engineering & Remote Sensing*, 63(6):735-740.

Hauser, J. R.; Clausing, D. (1989). *Otra novedad japonesa: la casa de la calidad.* Harvard-Deusto Business Review, 1er trimestre 1989.

Hay, A.M. (1979). Sampling Designs to Test Land-Use Map Accuracy. *Photogrammetric Engineering and Remote Sensing*, 45(4): 529-533.

Hermoso, J.; Hernández, A. (1994). *Curso de Estadística Económica y Empresarial.* Editorial Némesis.

Ishikawa, K (1989). *Introducción al Control de Calidad*, Editorial Díaz de Santos, Madrid, España, 1989.

Ishikawa, K. (1994). *¿Qué es control de Calidad?*. Parramón. Barcelona.

James, P. (1997). *La gestión de la calidad total: Un texto introductorio*. Ed. Prentice Hall, Madrid.

Juran, J.; Blanton, A.; Blanford, A. (1998). *The Juran's Quality HandBook.* McGraw-Hill Professional.

Juran, J.; Gryna, Jr (1980). *Quality Planning and Analysis.* Ed. McGraw-Hill, New York.

Kainz, W. (1995). Logical consistency. Guptill, S.C.; Morrison, J.L. (ed). *Elemments of Spatial Data Quality*. Pergamon Press, Oxford.

Keefer, B.J.; Smith, J.L.; Gregoire, T.G. (1991). Modeling and Evaluating the Effects of Stream Mode Digitizing Errors on Map Variables. *Photogrammetric Engineering & Remote Sensing*, 57(7): 957-963.

Kenneth, E. (1995). Error, accuracy and precision. *The Geographer's Craft Project.* University of Texas.

Langran, G. (1992). *Time in GIS.* Taylor & Francis. London.

Leung, Y.; Yan, J. (1998) A locational error model for spacial features. *International Journal of Geographical Information Science*, 12(6): 607-620.

MacQueen, J.B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*: 1: 281-297. Berkeley. University of California Press.

MacQueen, J.B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*: 1: 281-297. Berkeley. University of California Press.

Maling, D. (1989). *Measurements from Maps.* Pergamon Press. Oxford.

Marquardt, P.W. (1963). An Algorithm for least squares estimation of non linear parameters. *SIAM Journal on Applied Mathematics*, 2: 431-441.

Martín, Q.; Cabero, M.T.; Ardanuy, R. (1999). *Paquetes estadísticos SPSS 8.0. Bases teóricas. Prácticas propuestas, resueltas y comentadas.* Ed. Hespérides, Salamanca.

McDonald, B. (2002). A Teaching Note on Cook´s Distance - A Guideline. *Research Letters in the Information and Mathematical Sciences*, 3: 127-128.

McGwire, C., (1996). Cross Validated Assessment of geometric accuracy. *Photogrammetrin Engineering & Remote Sensing*, 62(10).

Medina, F. (1990). *Tamaño Óptimo de Muestra en Encuestas de Propósitos Múltiples.* Cepal.

Miller, D. (1994). Experience in Developing a Natural Resource Data Directory for New Soth Wales. *Proceedings AURISA '94.*

Minnesota Planning (1999). *Positional Accuracy Handbook.* Minnesota Planning - Land Management Information Center. Minnesota.

Montgomery, D. C.; Peck, E. A.; Vining, C. G. (2001). *Introduction to linear regression analysis.* Wiley.

Morant, T.; Lerma, J.L. (1999). *La Calidad en las bases de datos espaciales.* Universidad de Las Palmas de Gran Canaria.

Morrison, J. (1988). The proposed standard for digital cartographic data. *The American Cartographer*, 15(1): 9-140.

Pardo, L.; Valdés, T. (1987). *Simulación. Aplicaciones prácticas a la empresa.* Ediciones Díaz de Santos.

Peña, D. (1989). *Estadística: Modelos y Métodos, 2 (Modelos Lineales y Series Temporales).* Alianza Universidad Textos.

Pérez, C. (1999). *Técnicas de Muestreo Estadístico.* Ra-Ma. Madrid.

Pérez, C. (2001). *Técnicas Estadísticas con SPSS.* Prentice Hall.

Pindyck, R.; Rubinfeld, D. (1980). *Modelos econométricos.* Ed. Labor universitaria. Barcelona.

Podobnikar, T. (1999). Modelling and visualization of spatial data error. Monte Carlo Simulations in Slovenia. *Geomatics Info Magazine (GIM)*, 13(7): 47-49.

RAE (2001). *Diccionario de la lengua española.* 22.ª edición. Espasa Calpe. Madrid.

Ríos, D.; Ríos, S. (1997). *Simulación, métodos y aplicaciones.* Ra-Ma. Madrid

Ríos, D.; Ríos, S.; Martín, J. (1997). *Simulación. Métodos y aplicaciones.* Ra-Ma. Madrid.

Rodríguez, A.F. (1998). La Norma UNE 148001 EXP MIGRA: Mecanismo de Intercambio

de Información Geográfica Relacional formado por Agregación. *Topografía y Cartografía*, 91.

Ross, S. (1997). *Simulación*. Prentice Hall.

Rueda, M.; Arcos, A.; González, A. (1992). *Métodos de muestreo en poblaciones finitas*. Universidad de Granada.

SAI (1999). *Guideline for quality checking of ortho imagery*. Space Applications Institute, ISPRA.

Sebastián, M. A.; Bargueño, V.; Novo, V. (1999). *Gestión y Control de Calidad*. Universidad Nacional de Educación a Distancia. Madrid.

Shi, W.; Ehlers, M.; Tempfli, K. (1994). Modelling and visualizing uncertainties in multi-data-based spatial analysis. *EGIS'94 Conference proceedings of the 5th European conference and exhibition on GIS*: (1): 454-464.

Sokal, R. R.; Rohlf, F. J. (1979). *Biometría. Principios y Métodos Estadísticos en la Investigación Biológica*. H. Blume Ediciones. Madrid.

Spiegel, M. (1970). *Estadística*. McGraw-Hill. México.

Stehman, S. (1996). Estimating the Kappa Coefficient and its Variance under Stratified Random Sampling. *Photogrammetric Engineering and Remote Sensing*, 62(4): 401-407.

Stehman, S. (2003). A critical evaluation of the normalized error matrix in the map accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, 70(6): 743-751.

Tortora, R.D. (1978). A Note on Sample Size estimation for Multinomial Populations. *The American Statistican*, 32(3): 100-102.

U.S. Bureau of the Budget (1947). *United States National Map Accuracy Standards*. Washington, D.C.

Uriel, E. (1995). *Análisis de datos. Series temporales y análisis multivariante*. Alfa Centauro, Madrid.

Veregin, H. (1989). *A Taxonomy of Errors in Spatial Data Bases*. Technical Paper 89-12, National Center for Geographic Information & Analysis (NCGIA), Santa Barbara.

Veregin, H. (1992). *GIS Data Quality Assessment Tools*. Environmental Monitoring Systems Laboratory. US Environmental Agency.

Veregin, H. (1994). Integration of Simulation Modeling and Error Propagation for the Buffer Operation in GIS. *Photogrammetrin Engineering & Remote Sensing*, 60(4): 427-435.

Vilar, J.F. (1997). *Las 7 Nuevas Herramientas para la mejora de la Calidad*. Ed. Fundación Confemetal, Madrid.

Wischmeier, W.H. (1959). A Rainfall Erosion Index for a Universal Soil Loss Equation. *Soil Science Society of America. Proceedings*, 23: 246-249.

Zeithaml, V., Parasuraman, A.; Berry, L. (1993). *Calidad total en la gestión de servicios*. Ed. Díaz de Santos, Madrid.

Zhenkul, M. (1995). Tau Coefficients for Accuracy Assessment of Classification of Remote Sensing Data. *Photogrammetric Engineering and Remote Sensing*, 61(4): 435-439.

Zhuang, M. (1992). Accuracy of Spatial Data Used in Geographic Information Systems. *Photogrammetrig Engineering and Remote Sensing*, 58(6): 835-841.

Normas:

AENOR (1998). *UNE 148001:1998EX. Mecanismo de Intercambio de Información Geográfica Relacional formado por Agregación (MIGRA)*. Asociación Española de Normalización y Certificación. Madrid.

AENOR (2000). *UNE-EN ISO 9000:2000. Sistemas de Gestión de la calidad. Fundamentos y vocabulario*. Asociación Española de Normalización y Certificación. Madrid.

AENOR (2000). *UNE-EN ISO 9001:2000. Sistemas de Gestión de la calidad. Requisitos.* Asociación Española de Normalización y Certificación. Madrid.

AENOR (2001). *UNE 66020-1:2001 (equivalente a ISO 2859-1:1999). Procedimientos de muestreo para la inspección por atributos. Parte 1: Planes de muestreo para las inspecciones lote a lote, tabulados según el nivel de calidad aceptable (NCA).* Asociación Española de Normalización y Certificación. Madrid.

AENOR (2001). *UNE 66020-2:2001 (equivalente a ISO 2859-2:1985). Procedimientos de muestreo para la inspección por atributos. Parte 2: Planes de muestreo para las inspecciones de lostes independientes, tabulados según la calidad límite (CL).* Asociación Española de Normalización y Certificación. Madrid.

CEN (1996). *PrENV 14825. Geographic Information. Geographic data files.* CEN/TC 278. Comité Europeo de Normalización.

DOD (1989). *MIL-STD-105E. Military Standard. Sampling Procedures and tables for inspection by attributes.* Department of Defense. Washington.

ISO (2002). *ISO 19113. Geographic Information - Quality principles.* International Standardization Organization. ISO TC/211 Secretariat, Majorstua, Norway.

ISO (2003). *ISO 19114. Geographic Information - Quality evaluation procedures.* International Standardization Organization. Ginebra, Suiza.

ISO (2003). *ISO 19115. Geographic information - Metadata.* International Standardization Organization. Ginebra, Suiza.

Este libro se acabó de imprimir el 21 de octubre de 2004 en
los Talleres de Gráficas La Paz de Torredonjimeno (Jaén), coincidiendo con
el 199 aniversario de la muerte en Trafalgar, del insigne cordobés, marino,
explorador y cartógrafo D. Francisco Alcalá Galiano.

ISBN 84-8439-239-2
9 788484 392392

Manual de carácter práctico orientado al control de calidad en cartografía. Plantea 31 casos de gran interés y actualidad sobre: mejora de la calidad, tamaños muestrales, control de procesos, control de las componentes posicional y temática, simulación y control de bases de datos geográficas. Se incluye un anejo estadístico para facilitar el acceso a estos conocimientos. Los contenidos se complementan con información en un espacio WEB. Sus contenidos son de valor tanto para alumnos universitarios como para profesionales.

Universidad de Jaén
Servicio de Publicaciones