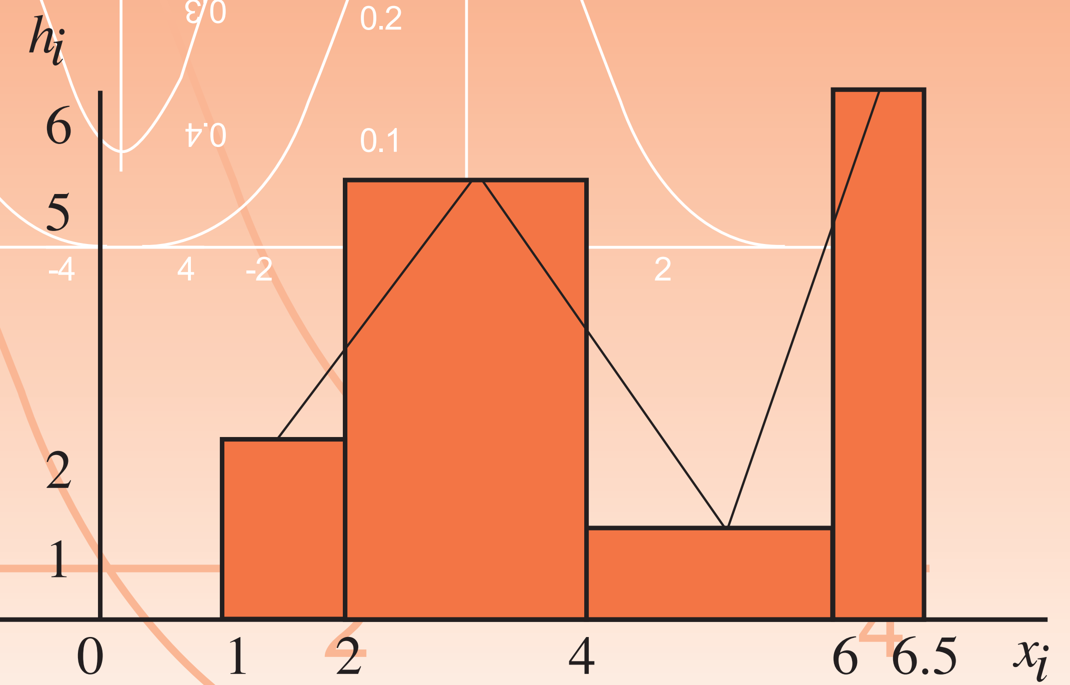
RAÚL AMOR PULIDO
CONCEPCIÓN AGUILAR PEÑA
ANTONIO MORALES LUQUE

ESTADÍSTICA DESCRIPTIVA Y CÁLCULO DE PROBABILIDADES



EDITORIAL GEU

RAÚL AMOR PULIDO
CONCEPCIÓN AGUILAR PEÑA
ANTONIO MORALES LUQUE
ESTADÍSTICA DESCRIPTIVA Y CÁLCULO DE PROBABILIDADES
GEU

ISBN 978-849915395-7
9 788499 153957

EDITORIAL GEU

RAÚL AMOR PULIDO
CONCEPCIÓN AGUILAR PEÑA
ANTONIO MORALES LUQUE

ESTADÍSTICA DESCRIPTIVA Y CÁLCULO DE PROBABILIDADES

Edita: Grupo Editorial Universitario

ISBN: 978-84-9915-395-7
Depósito Legal: GR-1.065-2011

Imprime: Lozano Impresores S.L.

Distribuye: Grupo Editorial Universitario
Telf.: (958) 80 05 80 Fax: (958) 29 16 15
http://www.editorialgeu.com
E-mail: info@editorialgeu.com

ÍNDICE:

SEGUNDA PARTE:
CÁLCULO DE PROBABILIDADES…….141

PRESENTACIÓN:

El libro cubre las partes de Estadística Descriptiva y Cálculo de Probabilidades del temario de las asignaturas de Estadística y Métodos Cuantitativos cursadas en distintas titulaciones de las Escuelas Politécnicas y Facultades de Ciencias Económicas y Empresariales.

Se encuentra dividido en dos partes, correspondiendo la primera a Estadística Descriptiva y la segunda a Probabilidad.

El desarrollo de cada parte viene estructurado en varios temas, en los cuales se realiza el desarrollo de la teoría acompañado de un ejemplo práctico de lo tratado. Cada tema, o conjunto de dos temas, va acompañado de una relación de problemas.

La finalidad del libro es que el alumno pueda seguir con más comodidad el desarrollo de las clases y tenga a mano los apuntes de la asignatura, para evitar los típicos errores en la toma de apuntes.

Esperamos que sirva para una mejor comprensión por parte de los alumnos de los contenidos.

Los autores.

PRIMERA PARTE:

ESTADÍSTICA DESCRIPTIVA

CAPÍTULO 1: INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA.

1.1. Introducción.

Comencemos haciendo una breve consideración histórica:
En sus orígenes, el concepto de *estadística* aparece ligado con la actividad gubernamental, y el término *estadístico* con el de estadista o político: los gobernantes deseaban conocer la extensión de sus dominios, la población residente en ellos y la cantidad de impuestos que podían esperar de dicha población; resultando así que la palabra *estadística* significaba recopilación de datos demográficos, económicos y políticos (los primeros datos estadísticos que se conocen datan del 2200 a. C. en el imperio chino). Pero la información crecía muy rápidamente en cantidad y extensión, teniendo que adoptar la forma de cuadros. Con el avance de las Matemáticas, se pasa de la descripción de los datos a intentar dar resultados sobre el comportamiento del fenómeno, a partir del siglo XVII; esta evolución se debe al Cálculo de Probabilidades, que tiene su origen en los juegos de azar.

Se llega así a que la palabra *estadística* significaba todo material numérico surgido de la observación del mundo exterior, aceptándose este uso a partir del siglo XIX.

Observando la realidad podemos distinguir dos tipos de fenómenos:

- Determinísticos: Son aquellos que presentan los mismos resultados si se realizan en idénticas condiciones (fenómenos físicos y químicos...).
- Aleatorios: Son aquellos que, al repetirlos en análogas condiciones, no podemos predecir los resultados (lotería primitiva, lanzamiento de un dado...).

La Estadística estudia los fenómenos aleatorios, pero la frontera entre ambos tipos de fenómenos no está perfectamente delimitada, aplicándose cada vez más las técnicas estadísticas en todo tipo de ciencias (Economía, Sociología, Psicología, Biología...).

Resumiendo, podemos 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, que es la que vamos a estudiar en esta primera parte, estudia y describe los fenómenos; mientras que la Inferencia Estadística se encarga de las predicciones.

1.2. Conceptos Fundamentales.

La Estadística Descriptiva se encarga del resumen de grandes cantidades de datos de forma que se da la máxima información posible de todo el conjunto. Veamos las definiciones de algunos conceptos importantes:

- Población: Es el conjunto de elementos objeto de estudio. Puede ser finita (alumnos de una clase) o infinita (resultado de lanzar un dado un número ilimitado de veces). Estudiaremos poblaciones finitas.
- Individuo o unidad estadística: Es cada uno de los elementos de la población.
- Tamaño: Es el número de elementos de la población. Lo denotaremos por N.
- Muestra: Es un subconjunto pequeño de la población, que debe ser representativa de toda ella. El estudio lo haremos sobre una muestra, ya que realizarlo sobre la población completa puede ser muy caro (encuestas electorales) o imposible (si es un proceso destructivo). Tiene n elementos.
- Caracteres y modalidades:
    - Caracteres son propiedades que se desean estudiar en los elementos de la población (color del pelo).
    - Cada carácter puede presentar distintas modalidades, que son incompatibles y exhaustivas, es decir, cada individuo presenta una y solo una de las modalidades del carácter (rubio, moreno...). Los caracteres pueden ser:
        1. Cualitativos: Con modalidades no medibles (color del pelo).

2. Cuantitativos: Con modalidades medibles (número de hijos).

- Variable estadística: Utilizamos este término para referirnos al carácter o caracteres en estudio. Las podemos clasificar en dos tipos:
    - Discretas: Son aquellas cuyos posibles valores son aislados (nº de hijos).
    - Continuas: Pueden tomar cualquier valor en un intervalo de la recta real (estatura de un grupo de personas).

Nota1.1.

Pero esta distinción es, muchas veces, arbitraria dado que toda variable es discreta, debido a la precisión limitada de los aparatos de medida, por lo que variables que pueden tomar un gran número de valores posibles y, aunque sean discretas, suelen tratarse como continuas.

Para describir una variable de tipo continuo se definen clases o grupos, que son las modalidades del carácter y se definen por intervalos. Llamaremos extremos de clase a los extremos de los intervalos y centro o marca de clase a los puntos centrales de esos intervalos. Los intervalos tienen que ser yuxtapuestos, con un extremo abierto y otro cerrado, para no dejar valores sin analizar. Cuando nos den los datos de una variable continua directamente, construiremos un número de intervalos aproximadamente igual a la raíz cuadrada del número de datos muestrales de los que disponemos, y la amplitud de cada intervalo será, aproximadamente, igual a la amplitud de los datos muestrales (diferencia entre los valores mayor y menor) dividido por el número de intervalos.

Ejemplo 1.1.

- Población: Todos los alumnos de la Universidad de Jaén.
- Muestra: Una clase.
- Individuo: Cada alumno.
- Características: Sexo, edad...
- Modalidades: Cada posibilidad de cada característica.

CAPÍTULO 2:

VARIABLES ESTADÍSTICAS UNIDIMENSIONALES.

En este capítulo vamos a estudiar variables estadísticas que estudian un solo carácter de los elementos de la población.

Una vez recogidos los datos correspondientes a la observación de un fenómeno estadístico, se procede a su descripción de la siguiente forma:

1. Clasificación.
2. Ordenación.
3. Representación.
4. Resumen de todos los datos en muy pocas cantidades.

2.1. Tabulación.

Tratamos de ordenar y clasificar los datos de forma clara en una tabla estadística.

Sea una muestra de la población de n individuos en la que se quiere estudiar un carácter C mediante una variable estadística X con k modalidades $(x_1,x_2,\ldots,x_k)$. La primera columna de la tabla va a estar formada por las distintas modalidades del carácter que estudia la variable, de manera que estén ordenadas, si es posible, de menor a mayor. Veamos las definiciones de las distintas frecuencias que vamos a utilizar en nuestro estudio y que, por tanto, van a ir incluidas en la tabla:

- La frecuencia absoluta de la modalidad *i*-ésima es el número de individuos de la muestra que presentan dicha modalidad. Lo notamos mediante n_i y, por consiguiente, se verifica que $n = \sum_{i=1}^{k} n_i$.
- La frecuencia relativa de la modalidad *i*-ésima es la proporción de individuos de la muestra que presentan esa modalidad expresada en tanto por uno. Lo notamos mediante f_i y, relacionado con la frecuencia absoluta, vale $f_i = {}^{n_i}\!/_{n}$, por lo que verifica que $\sum_{i=1}^{k} f_i = 1$. También lo podemos multiplicar por 100 para expresarlo en tanto por cien, recibiendo en este caso el nombre de proporción.
- La frecuencia absoluta acumulada de la modalidad *i*-ésima es el número de individuos que tienen una modalidad inferior o igual a la *i*-ésima, supuestas las modalidades ordenadas de menor a mayor. Lo notamos por N_i y, por tanto, es igual a $N_i = \sum_{j=1}^{i} n_j$.
- La frecuencia relativa acumulada de la modalidad *i*-ésima es la proporción de la frecuencia absoluta acumulada y lo notamos por F_i, es decir $F_i = \sum_{j=1}^{i} f_i = N_i / n$.

Estas dos últimas frecuencias no se consideran en variables cualitativas, puesto que no tienen sentido.

Veamos las tablas que recogen los datos estadísticos:

a) Discretas:

Modalidades	**Frec. Absoluta**	**Frec. Relativa**	**Frecuencias**	**acumuladas**
X_i	n_i	f_i	N_i	F_i
X_1	n_1	f_1	$n_1 = N_1$	$F_1 = f_1$
X_2	n_2	f_2	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
...	...	...	...	...
X_i	n_i	f_i	$N_i = \sum_{j=1}^{i} n_j$	$F_i = \sum_{j=1}^{i} f_j$
...	...	...	...	...
X_k	n_k	f_k	$N_k = \sum_{i=1}^{k} n_i$	$F_k = \sum_{i=1}^{k} f_i$

Ejemplo 2.1.

Sea X la variable estadística que mide el nº de hijos de una muestra de familias. Obtenemos una muestra, con los siguientes resultados:

Nº de hijos	Frec. Absoluta	Frec. Relativa	Frecuencias acumuladas	
X_i	n_i	f_i	N_i	F_i
0	2	0.2	2	0.2
1	3	0.3	5	0.5
2	3	0.3	8	0.8
3	1	0.1	9	0.9
4	1	0.1	10	1
Totales	10	1		

b) Continuas:

Introducimos en la tabla las columnas de las amplitudes y marcas de clase y definimos una nueva frecuencia.

- La densidad de frecuencias de la modalidad *i*-ésima se define como la frecuencia absoluta del intervalo por unidad de medida, es decir, el cociente entre la frecuencia absoluta y la amplitud del intervalo, $h_i = \dfrac{n_i}{a_i}$.

Intervalos	Marcas de clase	Amplitudes	n_i	f_i	N_i	F_i	h_i
$[e_0, e_1]$	$(e_0+e_1)/2$	$E_1 - e_0 = a_1$	n_1	f_1	N_1	F_1	h_1
$(e_1, e_2]$	$(e_1+e_2)/2$	$E_2 - e_1 = a_2$	n_2	f_2	N_2	F_2	h_2
...	...	...	...	...	...	...	...
$(e_{i-1}, e_i]$	$(e_{i-1}+e_i)/2$	$E_i - e_{i-1} = a_i$	n_i	f_i	N_i	F_i	h_i
...	...	...	...	...	...	...	...
$(e_{k-1}, e_k]$	$(e_{k-1}+e_k)/2$	$E_k - e_{k-1} = a_k$	n_k	f_k	N_k	F_k	h_k

Como ya hemos visto, los intervalos deben ser yuxtapuestos, ya que no se pueden dejar intervalos sin analizar; además, deben ser abiertos por un extremo y cerrados por otro para evitar que un punto esté en dos intervalos a la vez o en ninguno.

Ejemplo 2.2.

Sea X la variable estadística que mide la longitud en cm. Tomamos una muestra de 20 tornillos, obteniendo:

I_i	x_i	a_i	n_i	f_i	N_i	F_i	h_i
[1,2]	1.5	1	2	0.1	2	0.1	2
(2,4]	3	2	12	0.6	14	0.7	6
(4,6]	5	2	2	0.1	16	0.8	1
(6,6.5]	6.25	0.5	4	0.2	20	1	8
Totales			20	1			

Nota 2.1.

Podemos trabajar con variables discretas cuantitativas como si fueran continuas, simplemente tomando intervalos que cumplan que su marca de clase sea representativa de los valores posibles que pueda tomar cada modalidad de la variable.

Ejemplo 2.3.

Nº de hijos	0-1	2-3	4-5	6-7
	[-0.5, 1.5)	[1.5, 3.5)	[3.5, 5.5)	[5.5, 7.5)

Como se puede ver, la marca de clase de cada intervalo corresponde al valor medio de los valores que representan: 0.5, 2.5, 4.5 y 6.5.

También hay variables continuas que se nos pueden presentar con forma discreta.

Ejemplo 2.4.

Una altura de 1.75 se refiere a todo el intervalo [1.75, 1.76)

Veamos una definición más:

- Llamamos Distribución de Frecuencias al conjunto de los valores de la variable estadística junto a sus frecuencias correspondientes:
    1. En el caso discreto es: $\{(x_i,n_i)\}_{i=1,\ldots,k}$, o bien $\{(x_i,f_i)\}_{i=1,\ldots,k}$.
    2. En el caso continuo es: $\{(I_i,n_i)\}_{i=1,\ldots,k}$, o bien $\{(I_i,f_i)\}_{i=1,\ldots,k}$.

2.2. Representaciones Gráficas.

Pretenden visualizar el comportamiento global de la distribución, ya que a veces es conveniente resumir la información en un gráfico para realizar una síntesis o explicación visual. Veamos las distintas representaciones para cada tipo de caracteres:

2.2.1. Caracteres Cualitativos.

Los principales gráficos para las variables de tipo cualitativo son:

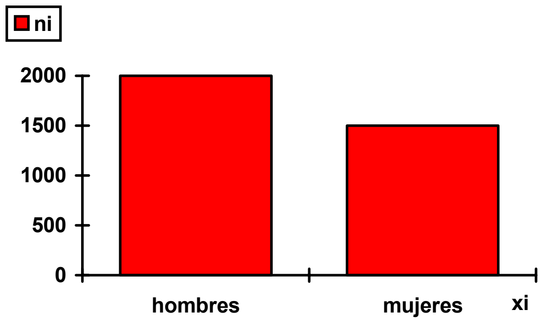
A) Diagrama de Rectángulos:

En el diagrama de rectángulos, a cada modalidad le corresponde una base constante y una altura proporcional a su frecuencia absoluta.

Ejemplo 2.5.

Sea X la variable que estudia el nº de personas con más de 20 años que viven con sus padres.

x_i	**hombres**	**mujeres**
n_i	2000	1500



B) Diagrama de Sectores.

El diagrama de sectores consiste en repartir el área de un círculo en sectores de tamaño proporcional a la frecuencia de cada modalidad. Si se representan las frecuencias absolutas, los grados de cada sector se obtienen resolviendo la siguiente proporción:

$$\left\{\begin{array}{l} 360^{\circ} \rightarrow n \\ x_i{}^{\circ} \rightarrow n_i \end{array}\right. \forall i$$

Ejemplo 2.6.

Estudiamos el nº de españoles fallecidos en el extranjero en un año determinado, clasificados según el estado civil y el sexo. Los resultados son:

Solteros		Casados		Viudos	
Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
80	18	75	40	57	30

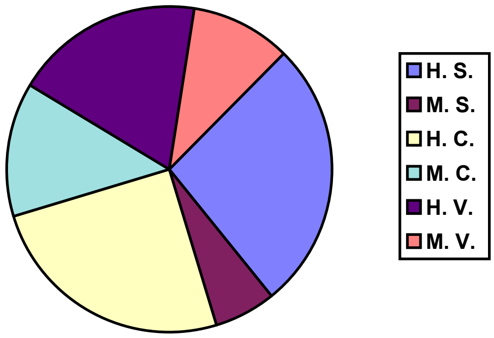
Tenemos $n = 300$ individuos.
Los grados en cada caso van a ser:

$$x_i^o = 360/300 \cdot n_i \ .$$

Por tanto:

$$96^0,\ 21.6^0,\ 90^0,\ 48^0,\ 68.4^0,\ 36^0$$

El diagrama de sectores es:

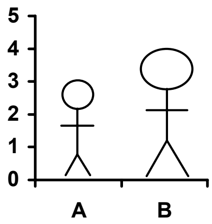


C) Pictograma.

En el pictograma podemos distinguir dos tipos. El primero dibuja figuras de tamaño proporcional a la frecuencia que representa y el segundo repite la figura tantas veces como indique la frecuencia.

Ejemplo 2.7.
Sea X la variable cualitativa que estudia el número de hombres y mujeres que trabajan en una empresa:

x_i	H	M
n_i	3	4



D) Cartograma.

En el cartograma, sobre un mapa se indica en cada región la intensidad de un carácter por medio de marcas.

Ejemplo 2.8.
Queremos estudiar el IPC en varias regiones:

0.2%	0.3%
0,4%	0.2%

donde color de la región señala el valor del IPC.

2.2.2. Caracteres Cuantitativos.

Dentro de los gráficos para variables de tipo cuantitativo, distinguimos entre los de variables discretas y continuas

A) Variables Discretas.

Los más importantes son:

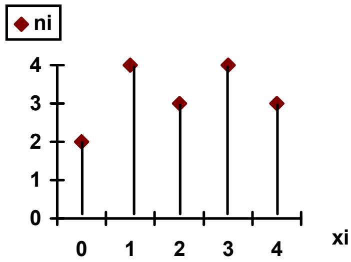
A.1) Diagrama de Barras.

En el diagrama de barras representamos para cada modalidad su frecuencia absoluta o relativa, es decir, representamos la Distribución de Frecuencias de la variable.

Ejemplo 2.9.
Sea X = nº de hijos por pareja.

x_i	0	1	2	3	4
n_i	2	4	3	4	3

Entonces:

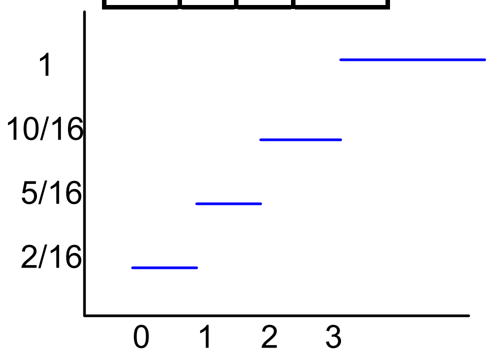


A.2) Curva Acumulativa o de Distribución.

Es la representación gráfica de las frecuencias acumuladas de la variable estudiada. Es constante en cada intervalo entre dos valores posibles consecutivos y su último valor es 1 (si la frecuencia representada es la relativa) o el tamaño de la muestra (si es la absoluta).

Ejemplo 2.10.
Sea X la variable que estudia el número de hermanos:

x_i	n_i	N_i	F_i
0	2	2	2/16
1	3	5	5/16
2	5	10	10/16
3	6	16	1



B) Variables Continuas.

Los principales gráficos para las variables continuas son:

B.1) Histograma.

El histograma está formado por rectángulos yuxtapuestos cuyas bases son las diferentes clases y sus alturas son:

- Si los intervalos son de amplitud constante, la altura será la frecuencia de cada intervalo.
- Si los intervalos tienen amplitudes desiguales, la altura correspondiente a cada intervalo será su densidad de frecuencias (h_i), es decir, el cociente de su frecuencia absoluta y su amplitud: $h_i = n_i / a_i$.

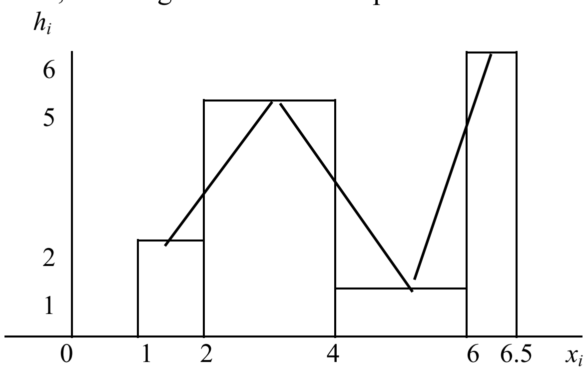
Uniendo los puntos medios de cada intervalo tenemos el Polígono de Frecuencias.

Ejemplo 2.11.

Sea X la variable que mide la longitud de una muestra de tornillos:

I_i	n_i	a_i	h_i	F_i
(1,2]	2	1	2	2/17
(2,4]	10	2	5	12/17
(4,6]	2	2	1	14/17
(6,6.5]	3	0.5	6	1

Entonces, el histograma viene dado por:

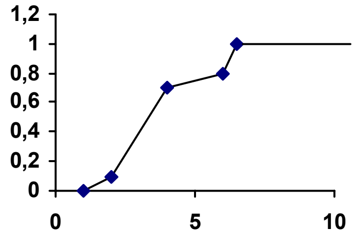


B.2) Curva de Distribución.

En la curva de distribución representamos las frecuencias acumuladas mediante los puntos (e_i, F_i), pero en este caso se desconoce la función en los puntos intermedios de los intervalos, pero aproximamos uniendo los puntos conocidos.

Ejemplo 2.11 (continuación).

Representamos la curva de distribución del ejemplo anterior.

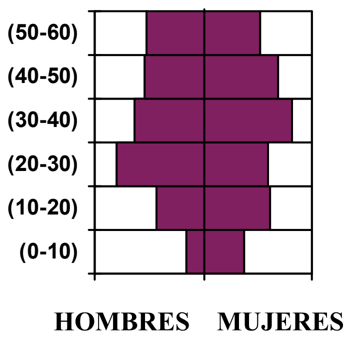


B.3) Pirámide de Población.

La pirámide de población consiste en la representación de una población humana según la edad y el sexo. Consta de dos histogramas, uno para cada sexo; la edad queda en la parte positiva del eje de ordenadas y los histogramas en el primer y segundo cuadrante.

Ejemplo 2.12.

Edad	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**
Hombre	10	20	40	30	20	10
Mujer	20	30	30	40	30	10



2.3. Medidas de Posición, Dispersión, Forma y Concentración.

En este apartado del tema tratamos de resumir todos los datos recogidos en una tabla estadística en pocos valores que nos informen sobre las principales características del fenómeno que estamos estudiando. Nos centramos en el estudio de características cuantitativas. Vamos a estudiar cuatro tipos de medidas:

- Medidas de posición: Informan sobre la situación de valores de la población. Hay dos tipos:
    - Medidas de tendencia central o promedios, que dan un valor central representativo.
    - Cuantiles, que suministran valores representativos de particiones de la distribución.
- Medidas de dispersión: Miden la desviación de los datos de una distribución con relación a una medida de tendencia central. Si se mide respecto de una medida de tendencia central, la medida de dispersión nos informará sobre la representatividad de dicha medida de tendencia central.
- Medidas de forma: Caracterizan la forma de la distribución sin necesidad de realizar su representación. Hay dos tipos:
    - Medidas de asimetría, que nos informan de la simetría de los datos de la distribución.
    - Medidas de apuntamiento, que indican la mayor o menor frecuencia de las observaciones más centrales.
- Medidas de concentración: Tienen origen y significado económico y miden la uniformidad en el reparto de una magnitud económica en un colectivo.

Todas las medidas deben cumplir una serie de características, que son:

1. Estar definidas de forma objetiva.
2. Usar todas las observaciones.
3. Tener un significado concreto y sencillo.
4. Ser fácil de calcular.
5. Prestarse al cálculo algebraico.
6. Ser poco sensible a fluctuaciones muestrales.

Es casi imposible encontrar medidas que satisfagan todas las condiciones anteriores, por lo que hay que procurar que cumplan la mayor parte de ellas.

2.3.1. Medidas de Posición.

Son valores en las mismas unidades que los datos estudiados, y miden la posición en torno a la que se distribuyen las observaciones, informándonos sobre el lugar central de la distribución, o bien nos indica la posición de una parte de ella.

A) Medidas de Tendencia Central o Promedios.

A.1) Media Aritmética (Media).

La media aritmética o, simplemente, media se define como el valor promedio de la variable estadística y no tiene por qué ser un valor de la variable. Es la más utilizada y se calcula como la suma de todos los valores que toma la variable dividida por el número total de datos:

$$\bar{x} = \sum_{i=1}^{k} \frac{n_i x_i}{n} = \sum_{i=1}^{k} f_i x_i \ ,$$

donde X es una variable estadística con k modalidades cuantitativas y, en el caso continuo, x_i es la marca de clase del intervalo i-ésimo.

Ejemplo 2.13.

Veamos un ejemplo en el caso discreto:

Sea X la variable que estudia el nº de hijos de determinadas parejas.

x_i	n_i	$n_i x_i$
0	3	0
1	2	2
2	4	8
	9	10

La media vale: $\bar{x} = \frac{10}{9}$.

Ejemplo 2.14.

Veamos un ejemplo en el caso continuo:

Sea X la variable que estudia la longitud en cm de un determinado producto.

I_i	n_i	x_i	$n_i x_i$
(0,2]	2	1	2
(2,3]	1	2.5	2.5
(3,5]	3	4	12
	6		16.5

La media de la longitud es: $\bar{x} = \frac{16.5}{6}$.

Las principales propiedades de la media aritmética son:

a) Las medias aritméticas de dos variables estadísticas en correspondencia lineal se obtienen por la misma correspondencia, es decir si $Y = aX + b$ entonces $\bar{y} = a \bullet \bar{x} + b$. Por tanto, dada una variable estadística como función lineal de otra de la que conocemos su media, podemos conocer fácilmente la media de esta nueva variable estadística.

Demostración:

$$\bar{y} = \sum_{i=1}^{k} f_i y_i = \sum_{i=1}^{k} f_i (ax_i + b) = \sum_{i=1}^{k} f_i a x_i + \sum_{i=1}^{k} f_i b = a \sum_{i=1}^{k} f_i x_i + b = a\bar{x} + b$$

Ejemplo 2.15.

Sea X una variable estadística con media 5 y la variable $Y = 3X+2$. Entonces $\bar{y} = 3 \bullet 5 + 2$.

b) Análogamente, si $Y = (X-a) / b$, entonces $\bar{y} = (\bar{x} - a) / b$.

c) Sean n observaciones agrupadas en R conjuntos con $n_1, \ldots, n_R$ observaciones cada uno y con medias $\bar{x}_1, \ldots, \bar{x}_R$. Entonces:

$$\bar{x} = \frac{\sum_{i=1}^{R} \bar{x}_i \bullet n_i}{n}$$

Ejemplo 2.16.

Estudiamos una variable en tres ciudades distintas, en las que conocemos la media y el tamaño de la muestra:

	n_i	$\bar{x}_i$
Jaén	54	18
Granada	34	22
Córdoba	27	45

$$\bar{x} = \frac{54 \bullet 18 + 34 \bullet 22 + 27 \bullet 45}{115} = 25.52 .$$

Las ventajas que tiene la media aritmética es que cumple las cinco primeras condiciones anteriores, pero tiene el inconveniente de ser muy sensible a cambios muestrales, ya que valores extremos muy dispares influyen notablemente en su valor, haciéndola menos representativa del conjunto de datos estudiado.

A.2) Media Geométrica.

La media geométrica se define como:

$$G = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$$

de donde se tiene que, tomando logaritmos para que sea más cómodo el cálculo práctico,

$$\log G = \frac{1}{n}\sum_{i=1}^{k} n_i \log x_i$$

Es menos sensible que la media aritmética a fluctuaciones muestrales, pero en ocasiones no queda determinada (si algún $x_i \leq 0$) y, además, tiene un significado poco intuitivo. Se usa para promediar tasas, porcentajes y en números índices. Para su cálculo necesitamos las siguientes columnas en la tabla:

x_i	n_i	$\log x_i$	$n_i \log x_i$
...	...	...	....

Ejemplo 2.17.

Calculamos la media geométrica de la siguiente variable:

x_i	n_i	$\log x_i$	$n_i \log x_i$
1	5	0	0
2	7	0.6931	4.8520
3	6	1.0986	6.5916
	18		11.443

Log G = 11.443 / 18 = 0.6357, por lo que $G = e^{Log\ G} = 1.8884$

A.3) Media Armónica.

La media armónica se define como:

$$H = \frac{n}{\sum_{i=1}^{k} n_i \frac{1}{x_i}}$$

por lo que para su cálculo necesitamos las siguientes columnas en la tabla estadística:

x_i	n_i	$1/x_i$	n_i/x_i
...	...	...	...

En ciertos casos es más representativa que la media aritmética, pero tiene el inconveniente de que en ella tienen mucha influencia los valores de la variable muy próximos a cero y no está definida en cero, por lo que no se debe utilizar en estos casos. Se usa para promedios de velocidades, tiempos, rendimientos...

Ejemplo 2.18.

x_i	n_i	$1/x_i$	n_i/x_i
1	5	1	5
2	7	0.5	3.5
3	6	0.3333	2
	18		10.5

Entonces: $H = \dfrac{18}{10.5} = 1.714$.

A.4) Media Cuadrática.

La media cuadrática se define como:

$$Q = \sqrt{\frac{\sum n_i x_i^2}{n}},$$

por lo que necesita las siguientes columnas en la tabla para su cálculo:

x_i	n_i	x_i^2	$n_i x_i^2$
...	...	...	...

Se usa muy poco y no da mucha información.

Ejemplo 2.19.

x_i	n_i	x_i^2	$n_i x_i^2$
1	5	1	5
2	7	4	28
3	6	9	54
	18		87

Entonces: $Q = \sqrt{\dfrac{87}{18}} = 2.198$.

La relación existente entre las medias es: $H < G < \bar{x} < Q$.

A.5) Mediana.

La mediana es el valor de la variable que deja por encima y debajo suyo el mismo número de individuos de la muestra, estando los datos ordenados de forma creciente, por lo que depende solo del orden de los datos y, a veces, es un valor más real que las medias, al no verse afectada por valores extremos.

$$F_{Me} = 1/2 \quad \text{, o bien, } \; N_{Me} = n/2.$$

Veamos cómo se calcula:

a) Dados directamente los datos, los ordenamos de menor a mayor y:

- Si n es impar, la mediana (Me), es el dato que ocupa el lugar $(n+1)/2$-ésimo.

Ejemplo 2.20.

6,7,8,10,12,15,16 $\rightarrow Me$ = lugar (7+1)/2 = lugar 4º $\Rightarrow Me = 10$.

- Si n es par, la mediana depende del tipo de la variable:
    - Discreta: la mediana son los valores de la variable estadística que ocupan la posición central, es decir, los están dentro del intervalo cerrado formado por los que ocupan los lugares $\frac{n}{2}$ y $\frac{n+2}{2}$ - ésimos.

Ejemplo 2.21.

6, 7, 8, 10 $\rightarrow Me = 7$ y 8.

- Continua: la mediana es la media aritmética de los datos que ocupan el lugar $\frac{n}{2}$ y $\frac{n+2}{2}$ - ésimos.

Ejemplo 2.22.

6, 7, 9, 10 $\rightarrow Me = 8$.

b) Dados los datos en una tabla de frecuencias.

- Si la variable es discreta y $N_i = N(x_i)$ es la frecuencia absoluta acumulada, la mediana es el primer valor de la variable que tiene frecuencia absoluta acumulada igual o superior a $n/2$, es decir:

$$N_i < n/2 < N_{i+1} \;\Rightarrow Me = x_{i+1}$$

Si tenemos que algún $N_i = n/2$, la mediana será el conjunto de valores de la variable situados en el intervalo $[x_i, x_{i+1}]$.

Ejemplo 2.23.

Sea una variable estadística discreta con la siguiente tabla de frecuencias:

x_i	n_i	N_i
0	2	2
1	1	3
2	4	7
3	3	10

La Mediana vale: $\frac{n}{2} = \frac{10}{2} = 5 \Rightarrow Me = 2$

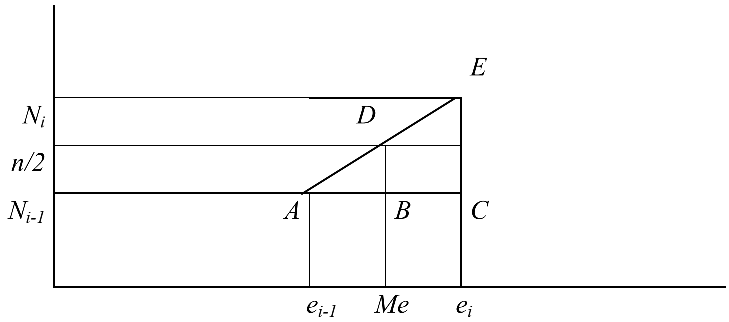
- Si la variable es continua buscamos el valor que ocupa el lugar $n/2$. Para ello, buscamos en primer lugar el intervalo mediano I_{Me}, que es aquel cuya frecuencia absoluta acumulada sobrepasa o iguala por primera vez el valor $n/2$. Entonces:

$$Me = \text{Extremo inferior de } I_{Me} + \frac{a_{Me}\left(\frac{n}{2} - N_{Me-1}\right)}{n_{Me}}, \text{ donde}$$

- a_{Me} = amplitud I_{Me},
- $N_{Me-1} = N_i$ del intervalo anterior al mediano y $n_{Me} = n_i$ de I_{Me} .

Demostración:

Dibujamos su curva de distribución:



Entonces, por semejanza de triángulos tenemos que $AC/AB = CE/BD$, que es equivalente a:

$$(e_i - e_{i-1}) / (Me - e_{i-1}) = (N_i - N_{i-1}) / (n/2 - N_{i-1}),$$

y, despejando, se obtiene el resultado buscado.

Ejemplo 2.24.

Sea una variable estadística continua con la siguiente tabla estadística:

I_i	n_i	N_i	a_i
(0,2]	2	2	2
(2,5]	3	5	3
(5,7]	4	9	2

Entonces la mediana vale: $\dfrac{n}{2} = 4.5 \;\Rightarrow I_{Me} = (2,5] \;\Rightarrow$

$$Me = 3\frac{(4.5-2)}{3} + 2 = 4.5.$$

Las ventajas de la mediana son:

1. Es sencilla de calcular.
2. Es de fácil interpretación al ser siempre un valor propio de la variable.
3. Solo influyen los datos centrales de la distribución, por lo que se puede calcular aunque no se conozcan los valores extremos de la distribución.

Sus inconvenientes son:

1. No se puede expresar con una fórmula sencilla.
2. No intervienen para su cálculo todos los valores de la variable.

A.6) Moda:

La moda es el valor más frecuente de la variable y no tiene por qué ser única. Se calcula de la siguiente forma:

- En una variable discreta, la moda será el valor de la variable que tenga mayor frecuencia absoluta.

Ejemplo 2.25.

Sea una variable discreta con distribución de frecuencias dada por:

x_i	n_i
0	2
1	3
2	7
3	2
4	5

Entonces la moda vale: $Mo = 2$.

Ejemplo 2.26.

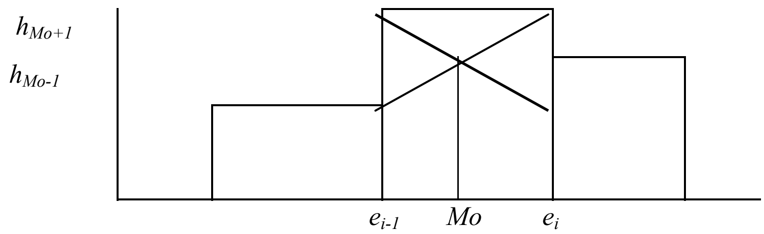
Sea una variable discreta con distribución de frecuencias dada por:

x_i	n_i
0	2
1	3
2	5
3	2
4	5

Entonces la moda vale: $Mo = 2$ y 4.

- En una variable continua, calculamos en primer lugar el intervalo modal I_{Mo}, que es aquel que tiene mayor densidad de frecuencias. A continuación, ya sí se puede calcular la moda:

Mediante un procedimiento gráfico, a partir del histograma, se sitúa la moda donde se supone que hay mayor frecuencia de observaciones:



Entonces, por proporcionalidad de bases y alturas de los triángulos creados, B/(B+b) = H/(H+h), que en nuestro caso es equivalente a:

$$\frac{h_{Mo} - h_{Mo-1}}{(h_{Mo} - h_{Mo-1}) + (h_{Mo} - h_{Mo+1})} = \frac{Mo - (Ext.\inf.I_{Mo})}{a_{mo}}.$$

Despejando, tenemos que:

$$Mo = \text{Extremo inferior de } I_{Mo} + \frac{(h_{Mo} - h_{Mo-1}).a_{Mo}}{(h_{Mo} - h_{Mo-1}) + (h_{Mo} - h_{Mo+1})}$$

Ejemplo 2.27.

Sea una variable continua con la siguiente tabla estadística:

I_i	n_i	a_i	h_i
(0,2]	2	2	1
(2,5]	4	3	4/3
(5,6]	2	1	2
(6,8]	3	2	1.5

$$I_{Mo} = (5,6] \qquad \Rightarrow \qquad Mo = 5 + \frac{(2-1.33)1}{(2-1.33)+(2-1.5)} = 5.57\ .$$

B) Cuantiles.

Los cuantiles suministran valores representativos de parte de la distribución.

Sea α un número entre 0 y 1. El cuantil de orden α, x_α es la raíz de la ecuación $F(x_\alpha) = \alpha$, es decir, una proporción igual a α de los individuos posee un valor de la variable inferior a x_α.

Los más utilizados son: cuartiles (Q_1 es el de orden 1/4, Q_3 es el de orden 3/4), deciles (D_i es el de orden $i/10$), percentiles (P_i es el de orden $i/100$)...

Se calculan de la misma forma que una mediana, tomando $n{\cdot}\alpha$, en lugar de $n/2$.

Ejemplo 2.28.

El P_{32} deja a la izquierda (por debajo) el 32% de los valores observados y el 68% a la derecha.

Calculamos el P_{32} de la variable continua en la que hemos calculado la mediana:

$n{\cdot}\alpha = 9 \cdot 0.32 = 2.88 \Rightarrow I_{32} = (2,5] \Rightarrow P_{32} = 2 + 3{\cdot}(2.88\text{-}2)/3 = 2.88.$

2.3.2. Medidas de Dispersión.

Miden el grado de esparcimiento de los datos de una distribución respecto al dato que se pretende que sea su síntesis, por lo que miden la representatividad de las medidas de posición, a las que deben de acompañar; ya que, aunque dos distribuciones tengan una medida de posición igual, pueden tener una representatividad muy distinta. Las dividimos en dos grupos:

A) Absolutas.

Las medidas de dispersión absolutas miden la variabilidad en la misma unidad de la variable que se está estudiando, por lo que no es posible realizar comparaciones respecto a la heterogeneidad de dos distribuciones. Las más importantes son:

A.1) Recorrido.

El recorrido es la diferencia entre el mayor y el menor valor de la variable. Está muy condicionada por las fluctuaciones, ya que solo utiliza dos valores y no nos proporciona mucha información acerca de la dispersión de los datos:

$$R = x_k - x_1.$$

Ejemplo 2.29.

Consideremos la siguiente distribución:

x_i	n_i
1	3
2	2
3	4
	9

Su recorrido es:

$$R = x_k - x_1 = 3 - 1 = 2$$

A.2) Recorrido Intercuartílico.

El recorrido intercuartílico indica la longitud del intervalo en el que está el 50% central de los valores de la distribución y evita la influencia de los datos extremos:

$$R_I = Q_3 - Q_1$$

Análogamente se pueden definir el recorrido interdecílico (D_9 - D_1) y el intercentílico (P_{99} - P_1).

Ejemplo 2.29 (continuación).

x_i	n_i	N_i
1	3	3
2	2	5
3	4	9
	9	

Los cuartiles son: Q_1=1 y Q_3=2

Por tanto, el recorrido intercuartílico vale:

$$R_I = Q_3 - Q_1 = 2 - 1 = 1$$

A.3) Desviación Absoluta Media.

Sea X una variable estadística y α una medida de posición central de dicha variable. Consideramos el valor $|x_i - \alpha|$, que es la distancia de los valores de la variable a la medida considerada (en valor absoluto para evitar que las dispersiones se compensen). La desviación absoluta media respecto a α viene dada por:

$$D_\alpha = \frac{\sum_{i=1}^{k} |x_i - \alpha| n_i}{n}.$$

Esto lo podemos hacer con todas las medidas de posición central y las más importantes son respecto a la media aritmética, mediana y moda: $D_{\bar{x}}$, D_{Me} y D_{Mo} .

En todos los casos cuanto menor sea el resultado, menor será la dispersión de los datos respecto de la medida considerada; por lo que se puede decir que mide la bondad de la medida considerada.

Tiene un grave inconveniente, que es el utilizar valores absolutos, por lo que la expresión es difícil de hacer su derivada. Por ello no se utiliza mucho.

Ejemplo 2.29 (continuación).

x_i	n_i	$x_i n_i$	$\lvert x_i - \bar{x}\rvert$	$n_i \lvert x_i - \bar{x}\rvert$
1	3	3	1.111	3.333
2	2	4	0.111	0.222
3	4	12	0.888	3.556
	9	19		7.111

$$\bar{x} = \frac{19}{9} = 2.111$$

$$D_{\bar{x}} = \frac{7.111}{9} = 0.7901.$$

A.4) Varianza y Desviación Típica.

Vamos a ver una medida que resuelve el problema de la desviación absoluta media, utilizando cuadrados en lugar de valores absolutos para que se pueda derivar la expresión, además de evitar que los errores se compensen.

Se define la desviación cuadrática media respecto a un valor α como:

$$Q(\alpha) = \sum_{i=1}^{k} (x_i - \alpha)^2 f_i \ ,$$

expresión que alcanza el mínimo cuando $\alpha = \bar{x}$.
Demostración:

$$Q'(\alpha) = \sum_{i=1}^{k} 2(x_i - \alpha)(-1) f_i = 0 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{k} (x_i - \alpha) f_i = \sum_{i=1}^{k} x_i f_i - \alpha \sum_{i=1}^{k} f_i = \bar{x} - \alpha = 0 \Rightarrow \ \bar{x} = \alpha \ .$$

$Q''(\alpha) = \sum_{i=1}^{k} (-2)(-1) f_i > 0$. Entonces $\bar{x}$ es un mínimo.

Definimos la varianza como:

$$Q(\bar{x}) = S^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 f_i = \frac{1}{n} \sum_{i=1}^{k} (x_i - \bar{x})^2 n_i$$

Definimos la desviación típica como la raíz cuadrada positiva de la varianza:

$$S = \sqrt{S^2} \ .$$

Ambas medidas miden la mayor o menor dispersión de los datos respecto a su media aritmética y cuanto más pequeñas sean, más representativa será $\bar{x}$.

Las principales propiedades de la varianza y desviación típica son:

a) La varianza se puede escribir como:

$$S^2 = \sum_{i=1}^{k} f_i x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 \ .$$

Demostración: $S^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 f_i = \sum_{i=1}^{k} (x_i^2 + \bar{x}^2 - 2 x_i \bar{x}) f_i =$

$$= \sum_{i=1}^{k} x_i^2 f_i + \bar{x}^2 \sum_{i=1}^{k} f_i - 2\bar{x} \sum_{i=1}^{k} x_i f_i = \sum_{i=1}^{k} x_i^2 f_i - \bar{x}^2 \ .$$

b) La varianza y la desviación típica son siempre mayores o iguales a cero. Además valen cero si y solo si todos los datos son iguales, es decir, no hay dispersión.

c) Sean $Y = aX + b$ y S_X^2 la varianza de la variable X. Entonces la varianza de la variable Y vale $S_Y^2 = a^2 S_X^2$ ($S_Y = |a| S_X$), es decir la varianza es invariable frente a cambios de origen.

Demostración: $S_Y{}^2 = \sum_{i=1}^{k} (y_i - \bar{y})^2 f_i = \sum_{i=1}^{k} (ax_i + b - a\bar{x} - b)^2 f_i =$

$$= \sum_{i=1}^{k} a^2 (x_i - \bar{x})^2 f_i = a^2 S_X{}^2$$

d) Análogamente, si $Y = (X-a)/b$, entonces $S_Y^2 = S_X^2/b^2$ ($S_Y = S_X/b$).

Para el cálculo de la varianza y la desviación típica, necesitamos en la tabla las siguientes columnas:

x_i	N_i	$n_i x_i$	$n_i x_i^2$
....	.....	......	......

Ejemplo 2.29 (continuación).

x_i	n_i	$n_i x_i$	$n_i x_i^2$
1	3	3	3
2	2	4	8
3	4	12	36
	9	19	47

Entonces, la varianza vale:

$$S^2 = \frac{1}{n} \sum_{i=1}^{k} n_i x_i{}^2 - \bar{x}^2 = \frac{47}{9} - 2.111^2 = 0.7654$$

B) Relativas.

Las medidas de dispersión relativas son valores sin dimensión, ya que son cocientes de magnitudes medidas en las mismas unidades y se utilizan para comparar la representatividad de una medida de posición central en varias distribuciones. La representatividad será mayor cuanto menor sea el valor de la medida de dispersión relativa. Los más importantes son:

B.1) Coeficiente de Apertura.

El coeficiente de apertura es el cociente del mayor y el menor valor de la variable y se utiliza en el estudio de la dispersión de salarios y variables de este tipo:

$$C_a = \frac{x_k}{x_1}.$$

Es muy sensible a las fluctuaciones al depender solo de dos valores.

Ejemplo 2.29 (continuación).

$$C_a = \frac{x_k}{x_1} = \frac{3}{1} = 3$$

B.2) Recorrido Relativo.

El recorrido relativo mide el número de veces que el recorrido contiene a la media aritmética:

$$R_R = \frac{R}{\overline{x}}$$

Ejemplo 2.29 (continuación).

$$R_R = \frac{R}{\overline{x}} = \frac{2}{2.111} = 0.9474$$

B.3) Recorrido Semi-intercuartílico.

Es el cociente de la diferencia de los cuartiles y la suma de ellos:

$$R_{SI} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Se utiliza cuando se usa la mediana como medida de posición central.

Ejemplo 2.29 (continuación).

$$R_{SI} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{2-1}{2+1} = 0.333$$

B.4) Coeficiente de Variación de Pearson.

El coeficiente de variación de Pearson representa el número de veces que la desviación típica contiene a la media aritmética:

$$C.V.(x) = \frac{S_x}{\overline{x}}$$

Es la medida de dispersión relativa más importante y mide el grado de homogeneidad de una distribución frente a otra (si nos piden ver qué distribución es más homogénea, se calcula en todas ellas y nos quedamos con la distribución que tenga el coeficiente más pequeño).

Ejemplo 2.29 (continuación).

$$C.V.(x) = \frac{S_x}{\bar{x}} = \frac{\sqrt{0.7654}}{2.111} = 0.4144$$

2.3.3. Momentos.

Sea X una variable estadística, r un número natural y a un número real. Definimos el momento de orden r respecto de a de la variable X como:

$$_{a}m_r = \sum_{i=1}^{k} (x_i - a)^r f_i \ .$$

Tenemos dos casos particulares:

- Si $a = 0$, tenemos los momentos no centrados, o centrados respecto al origen:

$$m_r = \sum_{i=1}^{k} x_i^{\,r} f_i \ .$$

Conocemos los resultados de algunos de ellos:

$$m_0 = 1 \ , m_1 = \bar{x} \ , \ m_2 = S^2 + \bar{x}^2 \ .$$

- Si $a = \bar{x}$, tenemos los momentos centrados respecto a la media:

$$M_r = \sum_{i=1}^{k} (x_i - \bar{x})^r f_i \ .$$

Conocemos los resultados de algunos de ellos:

$$M_0 = 1,\ M_1 = 0,\ M_2 = S^2.$$

También vamos a utilizar bastante a lo largo del curso la tipificación de variables, que consiste en que, dada una variable X y conocidas su media aritmética y su desviación típica, consideramos la variable $Z = \frac{X - \bar{x}}{S_X}$, que tiene media cero y desviación típica uno.

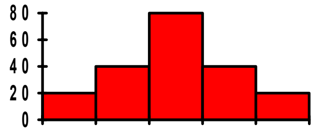
2.3.4. Medidas de Forma.

Dan valores que miden la forma de la representación gráfica de la distribución de frecuencias sin necesidad de dibujarla. Hay dos clases:

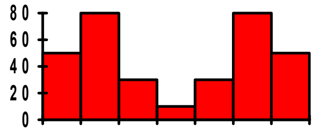
A) Medidas de Asimetría.

Una distribución de frecuencias es simétrica cuando lo es su representación gráfica respecto al eje vertical $x = \bar{x}$.

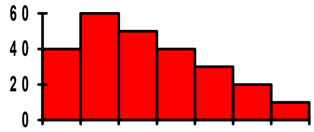
CASO 1



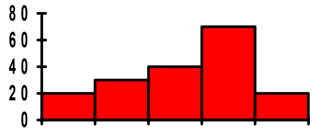
CASO 2



CASO 3



CASO 4



En el CASO 1: $\bar{x} = Me = Mo$, ya que tenemos una distribución simétrica unimodal.
En el CASO 2: $\bar{x} = Me$, ya que tenemos una distribución simétrica bimodal.
En el CASO 3: $\bar{x} - Mo > 0$, ya que tenemos una distribución asimétrica positiva.
En el CASO 4: $\bar{x} - Mo < 0$, ya que tenemos una distribución asimétrica negativa.

Tenemos dos coeficientes para estudiar la asimetría de las distribuciones:

- Coeficiente de Asimetría de Pearson:

$$G_p = \frac{\bar{x} - Mo}{S_X}$$

Si la distribución es asimétrica positiva entonces $G_P > 0$.
Si la distribución es simétrica entonces $G_P = 0$.
Si la distribución es asimétrica negativa entonces $G_P < 0$.

- Coeficiente de Asimetría de Fisher:

$$\gamma_1(x) = \frac{M_3}{S^3}.$$

Si la distribución es asimétrica a la derecha entonces $\gamma_1 > 0$.
Si la distribución es asimétrica a la izquierda entonces $\gamma_1 < 0$.

Si la distribución es simétrica entonces $\gamma_1 = 0$. El recíproco no tiene por qué ser cierto, por lo que si obtenemos que $\gamma_1 = 0$ podremos decir que la distribución puede ser simétrica.

La tabla que se plantea para su cálculo es la siguiente:

x_i	n_i	$n_i x_i$	$n_i x_i^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^3 n_i$
......	......	....	....	....	....

Ambas medidas deben llegar a la misma conclusión. En caso contrario, si los valores son de distinto signo pero próximos a cero, consideraremos que la distribución es simétrica; pero si los valores son de distinto signo y alguno de ellos no es próximo a cero, debemos dibujar la distribución para conocer su asimetría.

Ejemplo 2.29 (continuación).

x_i	n_i	N_i	$n_i x_i$	$n_i x_i^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^3 n_i$
1	3	3	3	3	-1.111	-4.114
2	2	5	4	8	-0.111	-0.0027
3	4	9	12	36	0.888	2.8104
	9		19	47		-1.3063

Los coeficientes de asimetría valen:

$$G_p = \frac{\bar{x} - Mo}{S_X} = \frac{2.111 - 3}{\sqrt{0.7654}} = -1.016$$

$$\gamma_1(x) = \frac{M_3}{S^3} = \frac{-1.3063/19}{\left(\sqrt{0.7654}\right)^3} = -1.026$$

Por consiguiente, la distribución es asimétrica negativa.

B) Medidas de Kurtosis o Aplastamiento.

Las medidas de kurtosis se aplican a distribuciones unimodales y casi simétricas para estudiar si los datos de una distribución están más o menos dispersos, para lo que comparamos con la distribución normal (que estudiaremos en la parte de Probabilidad).

Si los datos están agrupados en torno a las medidas de posición centrales, la gráfica será más alta y estrecha (apuntada).

Definimos el coeficiente de kurtosis como:

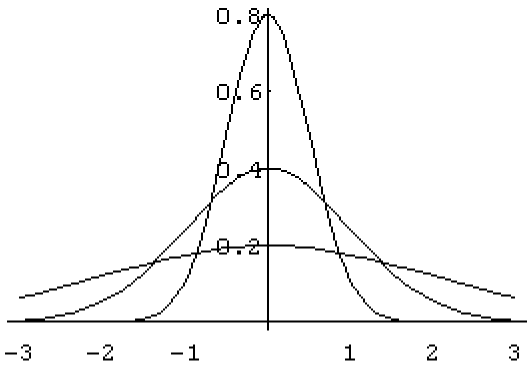
$$\gamma_2(x) = \frac{M_4}{S^4} - 3 \quad .$$

Restamos 3 porque es el valor del primer término en la distribución normal).

- Si $\gamma_2 > 0$ la distribución está más apuntada que la Normal, es decir, los datos están menos dispersos (leptocúrtica).
- Si $\gamma_2 < 0$ la distribución está menos apuntada que la Normal, es decir, los datos están más dispersos (platicúrtica).
- Si $\gamma_2 = 0$ la distribución está igual de dispersa que la Normal (mesocúrtica).

La tabla a representar es:

x_i	n_i	$n_i x_i$	$n_i x_i^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^4 n_i$
......	....	....	....	....	....



donde:

- La más apuntada tiene el coeficiente de kurtosis positivo.
- La intermedia es la distribución Normal(0,1).
- La menos apuntada tiene el coeficiente negativo.

Ejemplo 2.29 (continuación).

x_i	n_i	N_i	$n_i x_i$	$n_i x_i^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^4 n_i$
1	3	3	3	3	-1.111	4.5706
2	2	5	4	8	-0.111	0.0003
3	4	9	12	36	0.888	2.497
	9		19	47		7.0682

El coeficiente de kurtosis vale:

$$\gamma_2(x) = \frac{M_4}{S^4} - 3 = \frac{7.0682/19}{\left(\sqrt{0.7654}\right)^4} - 3 = -2.365$$

Por tanto, la distribución es platicúrtica.

2.3.5. Medidas de Concentración.

Son medidas fundamentalmente económicas y estudian el grado de igualdad en el reparto de la suma total de una variable, como por ejemplo el reparto del presupuesto entre las distintas áreas de una empresa o la distribución de los salarios entre los distintos empleados. Los casos extremos, entre los que encuentran la mayor parte de las situaciones reales, serían:

- Concentración máxima, en la que un elemento recibe el total de la variable y el resto, nada.
- Equidistribución, donde todos los elementos reciben la misma cantidad de la variable estudiada.

Las medidas de concentración más utilizadas son:

A) Curva de Lorenz

Se construye la tabla de frecuencias de la variable estudiada, añadiendo cuatro nuevas columnas:

L_{i-1}-L_i	x_i	n_i	$x_i n_i$	N_i	u_i	p_i	q_i
L_0-L_1	x_1	n_1	$x_1 n_1$	N_1	u_1	p_1	q_1
...	...	...	...	...	...	...	...
L_{i-1}-L_i	x_i	n_i	$x_i n_i$	N_i	u_i	p_i	q_i
...	...	...	...	...	...	...	...
L_{k-1}-L_i	x_k	n_k	$x_k n_k$	N_k	u_k	p_k	q_k
		N	u_k				

- Cantidad total percibida $x_i n_i$ por los n_i elementos con cantidades entre L_{i-1} y L_i.
- Cantidad total acumulada u_i, que es la cantidad total acumulada por los N_i elementos con menores cantidades:

$$u_i = \sum_{j=1}^{i} x_j n_j$$

- Frecuencias relativas acumuladas en tanto por ciento de los cantidades p_i:

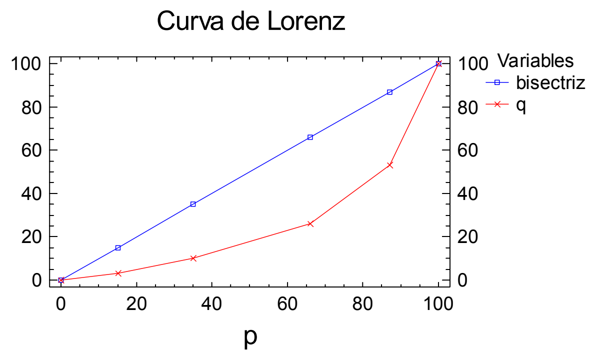
$$p_i = \frac{N_i}{N}100$$

- Cantidad acumulada en tanto por ciento sobre la cantidad total (u_k) q_i:

$$q_i = \frac{u_i}{u_k}100$$

La concentración en el reparto de la variable estudiada la obtenemos comparando estas dos últimas columnas.

La curva de concentración o curva de Lorenz es la curva que pasa por los puntos (p_i,q_i) para i=1,…,k, es decir, la representación gráfica de la distribución de elementos y cantidades.



Su interpretación es la siguiente:

- Si hay equidistribución, la curva coincide con la bisectriz del primer cuadrante.
- Si existe concentración máxima, la curva formaría prácticamente un triángulo rectángulo con la bisectriz del primer cuadrante hasta la cantidad total.

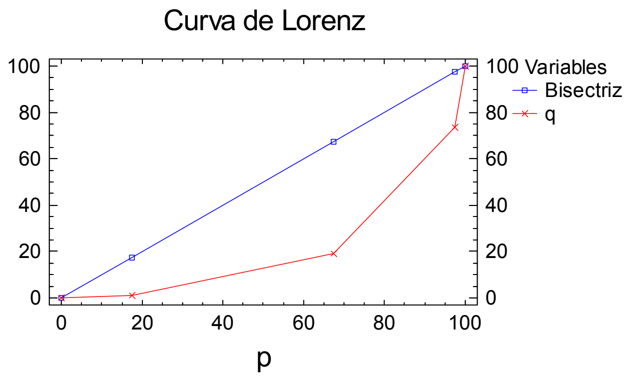
Ejemplo 2.30.

Los salarios mensuales de una gran compañía vienen dados en la siguiente tabla:

Salario (€)	Nº de empleados
0-200	175
200-1000	500
1000-5000	300
5000-30000	25
	1000

Construimos su curva de Lorenz:

L_{i-1}-L_i	x_i	n_i	$x_i n_i$	N_i	u_i	p_i	q_i
0-200	100	175	17500	175	17500	17.5	1.06
200-1000	600	500	300000	675	317500	67.5	19.18
1000-5000	3000	300	900000	975	1217500	97.5	73.56
5000-30000	17500	25	437500	1000	1655000	100	100
		1000	1655000			277.77	232.05



B) Índice de Gini.

El índice de Gini utiliza la curva de Lorenz para medir la concentración. Se define como el cociente entre el área formada por la bisectriz y la curva de Lorenz y el área del triángulo formado por la bisectriz y toma valores entre cero y uno. Si vale uno, existe concentración máxima y hay equidistribución si su valor es uno.

Tiene la dificultad de tener que calcular el área de una figura complicada, por lo que se define otro índice de Gini más sencillo de utilizar:

$$I_G = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i} = 1 - \frac{\sum_{i=1}^{k-1} q_i}{\sum_{i=1}^{k-1} p_i}$$

que también mide la proximidad de la curva de Lorenz a la bisectriz pero basándose en una propiedad distinta a la anterior. La interpretación de los resultados es similar a la comentada anteriormente:

- Si hay equidistribución, $p_i = q_i$, $i=1,\ldots,k$, por lo que I_G vale cero.

- Si existe concentración máxima, q_i= 0 para i=1,…,k-1, siendo el valor del índice uno.
- El índice de Gini vale entre cero y uno, e indica mayor concentración cuanto más próximo sea su valor a uno.

Ejemplo 2.30 (continuación).

Calculamos ahora su índice de Gini:

$$I_G = 1 - \frac{\sum_{i=1}^{k-1} q_i}{\sum_{i=1}^{k-1} p_i} = 1 - \frac{132.05}{177.77} = 0.2572$$

siendo, por tanto, la concentración de los salarios baja.

C) Mediala.

La mediala o valor medial Ml es la cantidad tal que los individuos que individualmente tienen menos de Ml tienen, en conjunto, tanto como los que tienen más de la mediala, es decir, es un concepto similar al a mediana, pues es el valor de la variable tal que la suma de las cantidades mayores que él es igual a la suma de las cantidades menores que él. Se calcula de manera similar a la mediana sobre los valores $n_i x_i$:

$$Ml = L_{i-1} + \frac{50 - q_{i-1}}{q_i - q_{i-1}} a_i$$

En cuanto a su interpretación, comparamos mediana y mediala, de manera que si hay equidistribución ambas son iguales. Cuanto más alejadas estén, menos concentración existirá. Para comparar de manera relativa, consideramos:

$$\frac{Ml - Me}{R}$$

la diferencia entre mediana y mediala dividida entre el recorrido.

Ejemplo 2.30 (continuación).

Calculamos ahora la mediala:

La Mediala se encuentra en el intervalo 1000-5000, pues es el primero con

$$u_i > \frac{1655000}{2} = 827500$$

Entonces:

$$Ml = L_{i-1} + \frac{50 - q_{i-1}}{q_i - q_{i-1}} a_i = 1000 + \frac{50 - 19.18}{73.56 - 19.18} 4000 = 3267.01$$

Para su interpretación, calculamos la mediana, que se encuentra en el intervalo 200-1000:

$$Ml = L_{i-1} + \frac{500 - N_{i-1}}{n_i} a_i = 200 + \frac{500 - 175}{500} 800 = 720$$

Entonces:

$$\frac{Ml - Me}{R} = \frac{3267.01 - 720}{30000} = 0.0849$$

por lo que la concentración es débil, como ya habíamos visto anteriormente.

D) Relación entre Concentración y Dispersión.

La concentración indica la forma de repartir una cantidad económica sobre un grupo de individuos, mientras que la dispersión se refiere a la homogeneidad de los valores que presenta una variable estadística. Aunque se da el caso de que si hay nula concentración también existe nula dispersión, no existe relación entre las medidas de ambas características, pues se puede ver que concentración máxima no corresponde a ningún tipo predeterminado de dispersión.

RELACIÓN DE ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL.

1.- Sea X una variable estadística con media 5 y varianza 4. Calcular la media y la varianza de la variable $6X + 4$.

2.- Completa la siguiente tabla estadística:

x_i	n_i	N_i	f_i
1	4		0.08
2	4		
3		16	
4	7		
5	5		
6		38	
7	7		
8	5		

3.- Se consideran las puntuaciones obtenidas por los alumnos de una clase:
1, 5, 4, 3, 8, 7, 5, 3, 2, 0, 9, 8, 1, 2, 0, 1, 3, 2, 8, 6, 4, 7, 2, 4, 5, 4, 7, 10, 3, 6, 6, 3, 9, 7, 3, 1, 4, 5, 7, 10, 5, 10, 8, 6, 2, 5, 0, 2, 9, 5.
a) Construir una tabla de frecuencias con estos datos.
b) Representarlos gráficamente mediante un diagrama de barras y su función de distribución.

4.- En una empresa se realizan unas pruebas de selección para un puesto de responsabilidad. Sumando los resultados de los distintos tests que se hacen a cada persona se han obtenido los siguientes resultados:
174, 185, 186, 176, 145, 166, 191, 175, 158, 156, 156, 187, 162, 172, 197, 181, 151, 161, 183, 172, 162, 147, 178, 176, 141, 170, 171, 158, 184, 173, 169, 162, 172, 181, 187, 177, 164, 171, 193, 183 173, 179, 188, 179,167, 178, 180, 168, 148, 173.
a) Agrupar los datos en intervalos de amplitud 5 (desde 140) y construir una tabla de frecuencias.
b) Representarlos gráficamente mediante un histograma y su curva de distribución.

5.- Cierto profesor dio la calificación de sus alumnos con el siguiente criterio :
suspensos: 40%, aprobados: 30%, notables: 15%, sobresalientes: 10% y matrículas de honor: 5%.

Si las notas fueron:

0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
34	74	56	81	94	70	41	28	16	4

a) Calcular entre qué notas estará cada una de las calificaciones.
b) ¿Qué porcentaje de alumnos presentan una puntuación inferior a 5.5? ¿Y superior a 5.5? ¿Y entre 3.5 y 6.75?

6.- En una Caja de reclutas se ha medido la altura de 110 jóvenes obteniéndose la tabla:

Altura	Nº jóvenes
1.50 - 1.60	18
1.60 - 1.70	31
1.70 - 1.80	24
1.80 - 1.90	20
1.90 - 2.00	17
	N=110

a) Se considera 'bajos' al 3% de jóvenes de menor altura. ¿Cuál es la altura máxima que pueden alcanzar?
b) Se consideran 'altos' al 18% de los más altos. ¿Cuál es su mínima altura?
c) ¿Qué valor divide a la población por la mitad?
d) ¿Cuál es la altura más frecuente? ¿Cuál es la altura media?

7.- La siguiente tabla muestra la distribución de coeficientes intelectuales de 120 alumnos:

I_i	n_i
60-70	2
70-80	3
80-90	25
90-100	46
100-110	35
110-120	5
120-130	3
130-140	1

a) Si se consideran bien dotados los alumnos cuya puntuación pertenece al 25% de las puntuaciones más altas, ¿qué puntuación mínima habrá de tener un alumno para ser considerado bien dotado?
b) ¿Qué porcentaje de alumnos presentan puntuación inferior a 109?
c) ¿Qué porcentaje de alumnos presentan puntuación superior a 124?

8.- La distribución de los salarios en € del año 2004 en la industria turística de un país es la de la siguiente tabla:

salarios	Número de trabajadores
<150	2145
150-200	1520
200-250	840
250-300	955
300-350	1110
350-400	2342
400-500	610
500-1000	328
>1000	150

Calcular:
a) El salario medio de un trabajador (marca del último intervalo 2000).
b) El salario más frecuente.
c) Un salario tal que la mitad de los restantes sea inferior a él.

9.- Dos vendedores del mismo producto tienen los registros expuestos en la tabla, durante un largo período de tiempo.
¿Cuál de los dos vendedores parece más constante en el volumen de ventas?

	Vendedor I	Vendedor II
Media de vol. Ventas/mes	3000 €	3000 €
Desviación típica	250 €	25 €

10.- A partir de la siguiente información sobre "Dependencia externa de la economía española", que figura en el Informe Económico-1986 del Banco de España:

DEPENDENCIA EXTERNA DE LA ECONOMÍA ESPAÑOLA
(Porcentaje sobre el PIB a precios de mercado)

Año	Importación de bienes y servicios	Exportación de bienes y servicios
1980	18.1	15.8
1981	20.2	18.1
1982	20.6	18.8
1983	21.9	21.3
1984	21.3	23.6
1985	21.0	23.3
1986	17.8	20.5

Determinar si es más estable en el tiempo, en relación con el PIB, la importación de bienes y servicios o su exportación.

11.- El presidente de Electrón S.A., compañía eléctrica, presenta en Junta General Ordinaria la siguiente tabla, sobre la que el Sr. Presidente afirma: "Como podéis observar, el recibo de consumo eléctrico importa, por término medio, 51.70 €, lo que aproximadamente corresponde al valor tal que la mitad de los hogares pagan menos. Sin embargo, la factura más frecuente está en torno a los 73 € al mes. Además, el 25% de los hogares pagan más de 70 €. y otro 25% paga menos de 32.75 €. El 18.25% de las facturas están entre 72.50 y 89.50 €/mes."

Importe por pagar	N_i
0-20	8
20-50	48
50-70	72
70-80	90
80-100	100

Discutir las afirmaciones hechas por el Sr. Presidente.

12.- En dos empresas se dan las siguientes distribuciones de salarios mensuales entre sus empleados:

Empresa A		Empresa B	
Salario	n_i	**Salario**	n_i
700	10	900	10
800	15	1000	15
900	40	1100	40
1000	25	1200	25
1100	10	1300	10

a) Comprobar:
- que el salario medio mensual en la empresa B es 200 € mayor que en la empresa A;
- que la varianza de ambas distribuciones es la misma.

b) ¿Para qué empresa resulta más representativo el salario medio?

13.- Un curso está dividido en cuatro grupos, de los cuales tenemos los siguientes datos:

Grupo	Nº alumnos	Nota media	Varianza
A	30	6	1
B	40	6.5	1.69
C	50	5	0.81
D	60	4	0.64

¿Qué grupo resulta más homogéneo?

14.- En un cierto barrio se ha constatado que las familias residentes se han distribuido, según su tamaño, de la forma siguiente:

Tamaño de familia	Nº de familias
1 - 2	110
3 - 4	200
5 - 6	90
7 - 8	75
9 - 10	25

Determinar:

a) ¿Cuál es el número medio de personas por familia?

b) ¿Cuál es el tipo de familia más frecuente?

c) Si sólo hubiera plazas de aparcamiento para el 50% de las familias de mayor a menor tamaño, ¿qué componentes tendría que tener una familia para entrar en el cupo? (Se supone que cada familia sólo tiene un coche).

d) Si el coeficiente de variación de Pearson de otro barrio es 1.8, ¿cuál de los dos barrios puede ajustar mejor sus previsiones en base al diferente tamaño de las familias que lo habitan?

15.- Dada la siguiente distribución de sueldos entre los empleados de una empresa. Calcular:

a) Los cuartiles.

b) La moda, media, varianza y el coeficiente de variación.

c) Los coeficientes de asimetría y kurtosis.

e_{i-1} - e_i	n_i
800 - 1000	10
1000 - 1200	30
1200 - 1500	40
1500 - 2000	15
2000 - 3000	5

16.- En una empresa, la distribución de sueldos mensuales es la siguiente:

Sueldo	**Porcentajes de empleados**
600 – 900	10
900– 1300	20
1300 - 2000	30
2000 - 3000	20
3000 - 5000	10
5000 - 10000	10

Calcular los salarios que definen el intervalo que agrupa el 50% central de la distribución.

CAPÍTULO 3:

VARIABLES ESTADÍSTICAS BIDIMENSIONALES.

En este capítulo vamos a estudiar variables estadísticas que estudian dos caracteres de la población.

3.1. Distribución de Frecuencias Bidimensional.

Sea una muestra con n individuos en la que vamos a estudiar simultáneamente dos caracteres X e Y. A la variable que representa dicho estudio se le llama variable estadística bidimensional (X,Y) , es decir, cada individuo viene determinado por el par $\{(x_i, y_j); i = 1,\dots,k ; j = 1,\dots,p\}$, donde:

- x_i indica la modalidad de la variable X que tiene el individuo.
- y_j indica la modalidad de la variable Y que tiene el individuo.
- k es el número de modalidades de la variable X.
- p es el número de modalidades de la variable Y.

Nota 3.1.

Nosotros vamos a trabajar con variables discretas, mientras que para variables continuas se trabaja de la misma forma con las marcas de clase de cada intervalo en lugar de los valores de la variable.

Definimos:

- n_{ij} como la frecuencia absoluta del par (x_i,y_j), es decir, indica el número de individuos que presentan el valor x_i de la variable estadística unidimensional X e y_j de la variable Y.
- f_{ij} como la frecuencia relativa del par (x_i,y_j): $f_{ij} = n_{ij}/n$.
- $n_{i.}$ es la frecuencia absoluta de los individuos con valor x_i de la variable X:

$$n_{i.} = \sum_{j=1}^{p} n_{ij}$$

- $f_{i.}$ es la frecuencia relativa de los individuos con valor x_i de la variable X:

$$f_{i.} = n_{i.} / n$$

- $n_{.j}$ es la frecuencia absoluta de los individuos con valor y_j de la variable Y:

$$n_{.j} = \sum_{i=1}^{k} n_{ij}$$

- $f_{.j}$ es la frecuencia relativa de los individuos con valor y_j de la variable Y:

$$f_{.j} = n_{.j} / n$$

Se verifica que:

$$n = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{p} n_{.j}$$

$$1 = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij} = \sum_{i=1}^{k} f_{i.} = \sum_{j=1}^{p} f_{.j}$$

Se denomina distribución bidimensional de frecuencias al conjunto de valores $\{(x_i,y_j),n_{ij}\}$ con $i=1,\dots,k$; $j=1,\dots,p$.

La representación numérica se realiza de la siguiente forma:
Sea una población estudiada según 2 caracteres X e Y y sea su distribución bidimensional de frecuencias $\{(x_i,y_j),n_{ij}\}$; entonces una forma de representar los resultados es la tabla de doble entrada, como la siguiente:

$X\backslash Y$	y_1	y_2	...	y_j	...	y_p	$n_{i.}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	$n_{2.}$
...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	$n_{i.}$
...	...	...	...	...	...	...	...
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{kp}	$n_{k.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.p}$	n

Ejemplo 3.1.

X\Y	15	25	35	$n_{i.}$
15	0	3	5	8
30	2	2	4	8
45	0	2	0	2
75	1	0	1	2
$n_{.j}$	3	7	10	20

Otra forma de presentar los resultados sería mediante una tabla simple, donde vemos la frecuencia de cada par, del tipo:

X	Y	n_{ij}
x_1	y_1	n_{11}
x_2	y_2	n_{12}
..	...	...
x_k	y_p	n_{kp}

Para distribuciones con pocas observaciones podemos escribir una tabla con pares observados:

	X	Y
1º	x_1	y_1
2º	x_2	y_2
...	...	...
i-ésimo	x_i	y_i
...	...	...
n-ésimo	x_n	y_n

Ejemplo 3.2.

Los tres valores muestrales siguientes se pueden presentar en una tabla simple:

X	Y
1	2
2	0
3	1

y también se puede escribir en una tabla de doble entrada:

$X\backslash Y$	**0**	**1**	**2**
1	0	0	1
2	1	0	0
3	0	1	0

3.2. Distribuciones Marginales y Condicionadas.

Veamos distribuciones unidimensionales que pueden obtenerse a partir de una distribución bidimensional.

3.2.1. Distribuciones Marginales.

Si de la variable bidimensional (X,Y) estudiamos solo una de las dos variables sobre el grupo de individuos que forman la muestra, los valores de dicha variable junto con sus frecuencias nos definen una distribución unidimensional, denominada distribución marginal de X o de Y (según la variable que se estudie).

Las distribuciones marginales de frecuencias son:

X	$n_{i.}$
x_1	$n_{1.}$
...	...
x_i	$n_{i.}$
...	...
x_k	$n_{k.}$
	n

Y	$n_{.j}$
y_1	$n_{.1}$
...	...
y_j	$n_{.j}$
...	...
y_p	$n_{.p}$
	N

Ejemplo 3.1 (continuación).

Las distribuciones marginales son:

X	$n_{i.}$
15	8
30	8
45	2
75	2
	20

Y	$n_{.j}$
15	3
25	7
35	10
	20

Como se trata de distribuciones unidimensionales se pueden calcular en ellas todas las medidas del tema anterior (medias, varianzas,...)

- Media marginal de X:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_{i.} = \sum_{i=1}^{k} f_{i.} x_i$$

- Media marginal de Y:

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{p} y_j n_{.j} = \sum_{j=1}^{p} f_{.j} y_j$$

- Varianza marginal de X:

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_{i.}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_{i.} x_i^2 - \bar{x}^2$$

- Varianza marginal de Y:

$$S_y^2 = \frac{1}{n}\sum_{j=1}^{p} n_{.j}(y_j - \bar{y})^2 = \frac{1}{n}\sum_{j=1}^{p} n_{.j} y_j^2 - \bar{y}^2$$

3.2.2. Distribuciones Condicionadas.

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.

Definimos:

- $n_{i/j}$ como la frecuencia absoluta de los individuos que toman el valor x_i de X cuando Y toma el valor y_j;
- $n_{j/i}$ como la frecuencia absoluta de los individuos que toman el valor y_j de Y cuando X toma el valor x_i;
- $f_{i/j}$ como la frecuencia relativa de los individuos que verifican que X toma el valor x_i dentro de los que cumplen la condición $Y = y_j$, y vale: $f_{i/j} = \dfrac{n_{ij}}{n_{.j}}$;
- $f_{j/i}$ como la frecuencia relativa de los individuos que verifican que Y toma el valor y_j dentro de los que cumplen la condición $X = x_i$, y vale: $f_{j/i} = \dfrac{n_{ij}}{n_{i.}}$.

$X \mid Y = y_j$

X	$n_{i/j}$
x_1	n_{1j}
...	...
x_i	n_{ij}
...	...
x_k	n_{kj}
	$n_{.j}$

$Y \mid X = x_i$

Y	$n_{j/i}$
y_1	n_{i1}
...	...
y_j	n_{ij}
...	...
y_p	n_{ip}
	$n_{i.}$

Ejemplo 3.1 (continuación).

Algunas distribuciones condicionadas son:

$X \mid Y = 25$

X	$n_{i\mid 25}$	$f_{i/25}$
15	3	3/7
30	2	2/7
45	2	2/7
75	0	0
	7	1

(j fijo)

$Y \mid X = 75$

Y	$n_{j/75}$	$f_{j/75}$
15	1	1/2
25	0	0
35	1	1/2
	2	1

(i fijo)

$X \mid Y = 25$ ó 35

X	$n_{i/25,35}$	$f_{i/25,35}$
15	8	8/17
30	6	6/17
45	2	2/17
75	1	1/17
	17	1

Como las distribuciones condicionadas son unidimensionales, se pueden definir en ellas los momentos vistos anteriormente, pero teniendo en cuenta que el número de individuos con los que se trabaja al condicionar no es el total, sino solo aquellos que cumplen la condición:

- Media condicionada de $X \mid Y = y_j$, que se nota:

$$\bar{x}_j = \frac{1}{n_{.j}} \sum_{i=1}^{k} n_{ij} x_i = \sum_{i=1}^{k} f_{i/j} x_i \quad .$$

- Media condicionada de $Y \mid X = x_i$, se nota:

$$\bar{y}_i = \frac{1}{n_{i.}} \sum_{j=1}^{p} n_{ij} y_j = \sum_{j=1}^{p} f_{j/i} y_j \, .$$

- Varianza condicionada de $X \mid Y = y_j$, que la vamos a escribir con la expresión:

$$S_j^2(X) = S_{x/y_j}^2 = \frac{1}{n_{.j}} \sum_{i=1}^{k} (x_i - \bar{x}_i)^2 n_{ij} = \frac{1}{n_{.j}} \sum_{i=1}^{k} n_{ij} x_i^2 - \bar{x}_j^2 \, .$$

- Varianza condicionada de $Y \mid X = x_i$,que la vamos a escribir mediante la expresión:

$$S_i^2(Y) = S_{y/x_i}^2 = \frac{1}{n_{i.}} \sum_{j=1}^{p} (y_j - \bar{y}_j) n_{ij} = \frac{1}{n_{i.}} \sum_{j=1}^{p} n_{j/i} y_j^2 - \bar{y}_i^2 \, .$$

3.2.3. Relación entre Distribuciones Marginales y Condicionadas.

Las frecuencias guardan la siguiente relación: $f_{ij} = f_{j/i} \cdot f_{i.} = f_{i/j} \cdot f_{.j}$

Demostración:

$$f_{ij} = \frac{n_{ij}}{n} = \frac{n_{ij}}{n_{i.}} \frac{n_{i.}}{n} = f_{j/i} f_{i.} \qquad\qquad f_{ij} = \frac{n_{ij}}{n_{.j}} = \frac{n_{ij}}{n_{.j}} \frac{n_{.j}}{n} = f_{i/j} f_{.j}$$

De este resultado obtenemos las siguientes relaciones entre sus principales momentos unidimensionales:

- Las medias marginales valen la media de las medias condicionadas:

$$\bar{x} = \sum_{j=1}^{p} \bar{x}_j f_{.j} \ , \qquad \bar{y} = \sum_{i=1}^{k} \bar{y}_i f_{i.}$$

- Las varianzas marginales valen la media de las varianzas condicionadas más la varianza de las medias condicionadas:

$$S_x^2 = \sum_{j=1}^{p} S_{x/y_j}^2 f_{.j} + \sum_{j=1}^{p} (\bar{x}_j - \bar{x})^2 f_{.j}$$

$$S_y^2 = \sum_{i=1}^{k} S_{y/x_i}^2 f_{i.} + \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 f_{i.}$$

3.3. Independencia Estadística y Dependencia Funcional.

Dada una variable estadística bidimensional vamos a estudiar las distintas posibilidades de dependencia e independencia entre las variables unidimensionales que la forman.

3.3.1. Independencia Estadística.

Decimos que X es estadísticamente independiente de Y si las distribuciones de $X \mid Y = y_j$ para $j = 1,...,p$ son todas iguales, es decir, no dependen de j , por lo que los $f_{i/j} = \frac{n_{ij}}{n_{.j}}$ son todos iguales para i fijo; en consecuencia, coinciden con la marginal (todo lo dicho anteriormente se puede resumir diciendo que todas las columnas de la tabla son proporcionales). La independencia es recíproca, es decir si X es independiente de Y, entonces Y es independiente de X.

Ejemplo 3.3.

X\Y	**1**	**2**
1	3	0
3	0	4

X	$n_{i/1}$	$f_{i/1}$
1	3	1
3	0	0
	3	

X	$n_{i/2}$	$f_{i/2}$
1	0	0
3	4	1
	4	

X e Y no son independientes pues los $f_{i/1}$ son distintos de los $f_{i/2.}$.

X\Y	**1**	**2**
1	3	5
3	6	10

X	$n_{i/1}$	$f_{i/1}$
1	3	1/3
2	6	2/3
	9	

X	$n_{i/2}$	$f_{i/2}$
1	5	1/3
2	10	2/3
	15	

X e Y son independientes pues los $f_{i/1} = f_{i/2} = f_{i.}$.

3.3.2. Dependencia Funcional.

Decimos que X depende funcionalmente de Y si a cada modalidad y_j de Y le corresponde una única modalidad posible de X (esto lo podemos observar en la tabla si en cada columna hay un único valor distinto de 0).

Decimos que Y depende funcionalmente de X si a cada modalidad x_i de X le corresponde una única modalidad posible de Y (esto lo podemos observar en la tabla si en cada fila hay un único valor distinto de 0).

La dependencia funcional no es recíproca.

Ejemplo 3.4.

Dada la siguiente distribución de frecuencias bidimensional:

X\Y	**1**	**2**	**3**
1	2	0	1
3	0	1	0

Y no depende funcionalmente de X porque para $X = 1$,Y=1,3
X depende funcionalmente de Y

X\Y	**1**	**5**
1	0	1
2	1	0
3	3	0

X no depende funcionalmente de Y porque para Y=1 , X= 2 y 3
Y depende funcionalmente de X

X\Y	**1**	**4**	**5**
2	0	3	0
6	4	0	0
7	0	0	2

X depende funcionalmente de Y
Y depende funcionalmente de X

X\Y	**2**	**3**	**5**
1	2	0	1
2	0	10	5
7	3	6	0

X no depende funcionalmente de Y
Y no depende funcionalmente de X

3.4. Momentos.

Sea (X,Y) una variable bidimensional.

- Se define el momento de órdenes r y s no centrado como:

$$m_{rs} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} x_i^r y_j^s n_{ij}$$

Como casos particulares tenemos:
1. $s = 0$ es el momento de orden r no centrado de X ;
2. $r = 0$ es el momento de orden s no centrado de Y;
3. $m_{11} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij}$.

- Se define el momento de órdenes r y s centrado como:

$$M_{rs} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} (x_i - \bar{x})^r (y_j - \bar{y})^s n_{ij}$$

Como casos particulares tenemos:
1. $M_{r0} = M_r(X)$
2. $M_{0s} = M_s(Y)$
3. $M_{11} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} (x_i - \bar{x})(y_i - \bar{y}) n_{ij}$ que se le da el nombre de covarianza y se nota S_{xy} , que también se puede escribir de la siguiente forma:

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij} - \overline{x}\overline{y} \ .$$

(que se ha obtenido simplemente desarrollando la expresión primitiva)

Ejemplo 3.5.

Cálculo de la covarianza:

$X \backslash Y$	5	10	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$x_i\left(\sum_{j=1}^{p} n_{ij}y_j\right)$
5	1	2	3	15	75	5(1·5+2·10)=125
15	3	1	4	60	900	15(3·5+1·10)=375
$n_{.j}$	4	3	7	75	975	500
$n_{.j}y_j$	20	30	50			
$n_{.j}y_j^2$	100	300	400			

$$\overline{x} = \frac{75}{7} = 10.714 \qquad S_x^2 = \frac{975}{7} - \overline{x}^2 = 24.49$$

$$\overline{y} = \frac{50}{7} = 7.1428 \qquad S_y^2 = \frac{400}{7} - \overline{y}^2 = 6.122 \qquad S_{xy} = \frac{500}{7} - \overline{x}\overline{y} = -5.10$$

Si X e Y son independientes, entonces $S_{xy} = 0$, aunque el contrario no es cierto.

La covarianza es una medida de la variabilidad conjunta y, por tanto, de la relación entre las variables X e Y:

- Si $S_{xy} > 0$, las dos variables varían en el mismo sentido (si una aumenta la otra también lo hace).
- Si $S_{xy} < 0$, las dos variables varían en sentido opuesto (si una aumenta la otra disminuye).

Pero un problema que presenta la covarianza es que puede ser cualquier número real y, al no estar acotado, no podemos asegurar solo con el conocimiento de ella si el grado de asociación es más o menos grande. Por esto, se define el coeficiente de correlación r como:

$$r = \frac{S_{xy}}{S_x S_y}$$

que es una medida adimensional y toma valores entre -1 y 1, como veremos en el próximo capítulo.

3.5. Representaciones Gráficas.

Veamos dos tipos de representaciones gráficas:

A) Nube de Puntos.

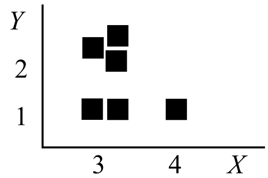
La nube de puntos consiste en un diagrama de ejes cartesianos, de forma que cada eje representa a una variable. Depende de cómo sean las variables:

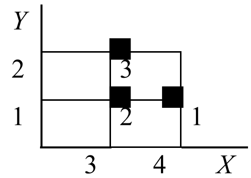
- Si las variables son de tipo discreto, dibujamos alrededor del punto (x_i, y_j) tantos puntos como indique su frecuencia absoluta n_{ij} o, al lado del punto, se coloca un número que indique la frecuencia .

Ejemplo 3.6.

Dada la siguiente distribución de frecuencias bidimensional:

*X**Y*	**1**	**2**
3	2	3
4	1	0

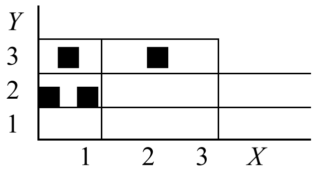




- Si las dos variables son continuas dibujamos en el rectángulo formado por los intervalos correspondientes tantos puntos como indique su frecuencia absoluta n_{ij}.

Ejemplo 3.7.

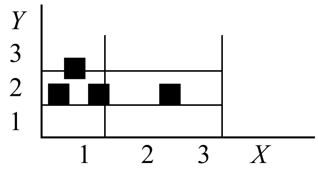
*X**Y*	**(1,2]**	**(2,3]**
(0,1]	2	1
(1,3]	0	1



- Si una variable es continua y otra discreta dibujamos en la intersección del punto e intervalo tantos puntos como indique su frecuencia absoluta n_{ij}.

Ejemplo 3.8.

$X \backslash Y$	**1**	**2**
(0,1]	2	1
(1,3]	0	1



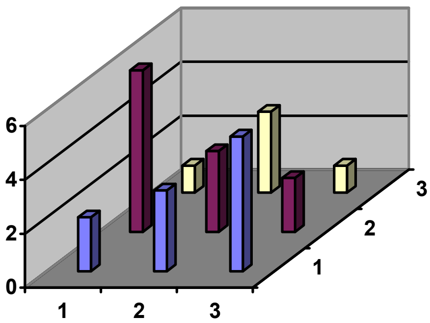
B) Estereograma.

El estereograma es la generalización del diagrama de barras, en el caso discreto, y del histograma, en el caso continuo. Consiste en una gráfica tridimensional en la que los ejes de la base van a ser las variables estadísticas que vamos a representar y la altura va a depender del tipo de las variables estadísticas:

- Dada una variable bidimensional (X,Y) cuyas variables unidimensionales son discretas, la altura del estereograma viene dada por la frecuencia absoluta de cada par.

Ejemplo 3.9.

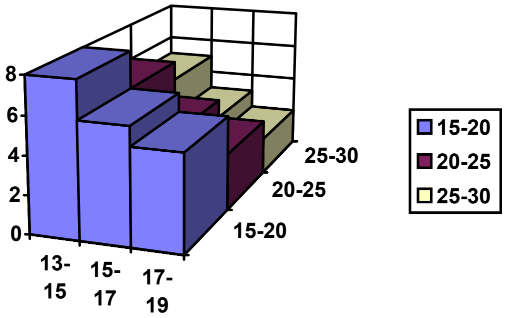
$X \backslash Y$	**1**	**2**	**3**
1	2	3	5
2	6	3	2
3	1	3	1



- Dada una variable bidimensional (X,Y) cuyas variables unidimensionales son continuas, la altura del estereograma viene dada por la densidad de frecuencias de cada par, es decir, la frecuencia dividido por el área del rectángulo de cruce correspondiente (puesto que lo que se pretende que sea igual a la frecuencia es el volumen del paralepípedo que construimos), por tanto, la frecuencia absoluta del par dividido entre el producto de las amplitudes de los intervalos respectivos.

Ejemplo 3.10.

*X**Y*	**(13,15]**	**(15,17]**	**(17,19]**
(15,20]	80 (8)	60 (6)	50 (5)
(20,25]	60 (6)	40 (4)	30 (3)
(25,30]	50 (5)	30 (3)	20 (2)



- El caso de una variable discreta y otra continua es una composición de los dos casos anteriores.

CAPÍTULO 4:

REGRESIÓN Y CORRELACIÓN. AJUSTE POR MÍNIMOS CUADRADOS.

4.1. Introducción.

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 vamos a tratar de realizar a lo largo de este capítulo.

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 lo que podemos decir, según lo visto en el tema anterior, que dichas variables son Funcionalmente Dependientes.

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,...; lo que quiere decir que estas variables son Estadísticamente Dependientes. Vamos a estudiar el grado de dependencia y la relación entre este tipo de variables:

1. La relación entre dos variables viene dada por la regresión.
2. El grado de dependencia entre las variables viene dado por la correlación.

3. Finalmente trataremos de dar predicciones de una variable.

La variable que queremos 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.

Nota 4.1.

Nosotros nos vamos a ocupar de estudiar la regresión y correlación simple, es decir, con una variable independiente. Todo esto se puede generalizar para explicar el comportamiento de una variable en función de otras, pero esto excede de los objetivos del curso.

4.2. Regresión.

La regresión consiste en ajustar lo mejor posible una función a una serie de datos observados; gráficamente queremos buscar una curva que se ajuste lo mejor posible a la nube de puntos. Por tanto, pretendemos buscar una función que exprese lo mejor posible la relación existente entre las variables.

Llamamos regresión de Y sobre X ($Y|X$) a la función que explica la variable Y a partir de X. En este caso, la variable explicativa será X e Y la variable explicada. La regresión de X sobre Y ($X|Y$) nos informa del comportamiento de la variable X a partir de Y, siendo, por tanto, Y la variable explicativa y X la explicada.

Tenemos dos criterios de regresión:

4.2.1. Regresión Tipo I.

Con este tipo de regresión obtenemos las curvas de regresión. Estudiamos los dos casos:

- Regresión de Y sobre X:

  El criterio es hacer corresponder a cada valor de la variable X (x_i), la media de la variable Y condicionada a que X toma el valor x_i, es decir, el valor $\bar{y}_i$.

  Por tanto, la gráfica

  $$\left\{\left(x_i, \bar{y}_i\right)\right\}_{i=1,\ldots,k}$$

  es la curva de regresión de Y sobre X.

  Esta curva tiene la propiedad de ser, entre todas las funciones, la que mejor se ajusta a los datos observados, pero tiene el grave inconveniente de no dar un funcional que de un valor de la variable Y para cada valor de la variable X, por lo que no es posible realizar predicción con ella.

- La regresión de X sobre Y da, análogamente, la curva de regresión

$$\left\{\left(\bar{x}_j, y_j\right)\right\}_{j=1,\dots,p}$$

Ejemplo 4.1.

Calculamos las curvas de regresión de las siguientes variables:

		[3,7)	**[7,13)**
	X \ Y	**5**	**10**
[0,10)	**5**	1	2
[10,20)	**15**	3	1

Las medias condicionadas son:

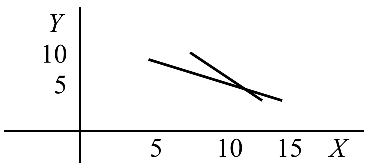
$$\bar{x}_5 = 12.5 \quad \bar{x}_{10} = 8.33 \qquad \bar{y}_5 = 8.33 \qquad \bar{y}_{15} = 6.25$$

Por tanto, las curvas de regresión son:

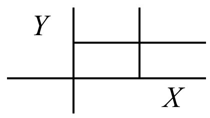
Y/X: {(5,8.33),(15,6.25)}

X/Y: {(12.5,5),(8.33,10)}

Gráficamente, las curvas son:



Si las variables son Estadísticamente Independientes las curvas de regresión son del tipo:



aunque el recíproco no es cierto.

4.2.2. Regresión Tipo II.

La regresión Tipo II da un funcional para poder predecir el valor de la variable explicada conocido un valor de la explicativa.

Sea (X,Y) una variable estadística bidimensional y consideremos la nube de puntos de la distribución de frecuencias asociada $\{(x_i,y_j);n_{ij}\}_{i,j}$. Pretendemos elegir una función matemática que se ajuste lo mejor posible a la nube de puntos, y el criterio a seguir es el de mínimos cuadrados:

Si queremos hallar la regresión de Y sobre X, consideramos $y' = h(x)$ la mejor predicción de Y en función de X si minimiza las diferencias al cuadrado entre los valores reales y los estimados por dicha función, es decir:

$$\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\left(y_j - h(x_i)\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} e_{ij}^{\ 2} \quad ,$$

donde $(y_j - h(x_i)) = e_{ij}$ es el residuo o error para el par (x_i, y_j) .

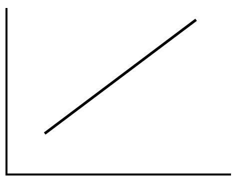
Nota 4.2.

También se podía haber considerado como criterio minimizar las siguientes cantidades:

1. Minimizar la suma de todos los residuos: $\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} e_{ij}$, pero no es un criterio adecuado puesto que los residuos positivos se podrían compensar con los negativos.
2. Minimizar la suma de los valores absolutos de los residuos: $\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} |e_{ij}|$, pero tiene el inconveniente de incluir el valor absoluto, que hace que la expresión no sea manejable de forma sencilla en expresiones algebraicas.

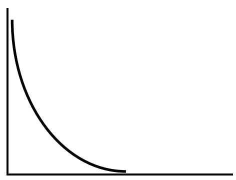
Por tanto el criterio de mínimos cuadrados se elige como el adecuado, ya que soluciona dichos inconvenientes.

Según sea la nube de puntos, podemos ajustar por distintas funciones. Las que nosotros vamos a utilizar son:



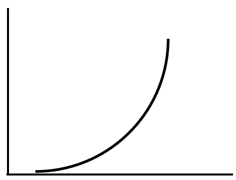
recta

$y' = ax + b$



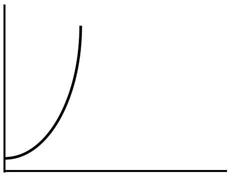
hipérbola equilátera

$$y' = \frac{a}{x} + b$$



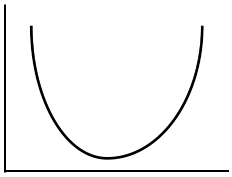
potencial

$y' = bx^a$



exponencial

$$y' = ba^X$$



parábola

$$y' = ax^2 + bx + c$$

Por tanto, para hacer la regresión se puede ver la nube de puntos y, a partir de ella, elegir el tipo de función que vamos a ajustar. A continuación tenemos que ver cómo calculamos los parámetros de la función que ajustamos.

A) Regresión Lineal (recta de regresión).

Pretendemos ajustar una recta a la nube de puntos, por lo que la regresión se va a denominar lineal. Debemos escoger de entre todas las rectas posibles, la mejor. Para ello utilizaremos el criterio de mínimos cuadrados como ya se ha indicado anteriormente. Vamos a trabajar con la regresión de *Y* sobre *X;* la regresión de *X* sobre *Y* será totalmente análoga.

La recta de mínimos cuadrados será la recta con expresión analítica:

y' = h(x) = ax+b,

que minimice la expresión:

$$S(a,b) = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\left(y_j - (ax_i + b)\right)^2 .$$

Vamos a buscar los valores de a y b que minimicen dicha expresión. Para ello derivamos la expresión respecto de los dos parámetros e igualamos dichas derivadas a cero:

$$\begin{cases} \dfrac{dS}{da} = 0 \\ \dfrac{dS}{db} = 0 \end{cases} \Rightarrow \begin{cases} \dfrac{dS}{da} = (-2)\sum_i\sum_j (y_j - b - ax_i)x_i n_{ij} = 0 \\ \dfrac{dS}{db} = (-2)\sum_i\sum_j (y_j - b - ax_i) n_{ij} = 0 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \sum_i\sum_j y_j x_i n_{ij} = a\sum_i\sum_j n_{ij}x_i^{\,2} + b\sum_i\sum_j n_{ij}x_i \\ \sum_i\sum_j y_j n_{ij} = a\sum_i\sum_j n_{ij}x_j + b\sum_i\sum_j n_{ij} \end{cases}$$

Dividiendo todo entre n, y sustituyendo los valores conocidos:

$$\begin{cases} \sum_i \sum_j y_j x_i f_{ij} = a \sum_i \sum_j f_{ij} x_i^2 + b\overline{x} \\ \overline{y} = a\overline{x} + b \Rightarrow b = \overline{y} - a\overline{x} \end{cases}.$$

Sustituyendo en la primera ecuación b por su valor:

$$\sum_i \sum_j y_j x_i f_{ij} = a \sum_i \sum_j f_{ij} x_i^2 + (\overline{y} - a\overline{x})\overline{x} = a \sum_i \sum_j f_{ij} x_i^2 + \overline{y}\overline{x} - a\overline{x}^2 =$$

$$= \overline{y}\overline{x} + a\left(\sum_i \sum_j f_{ij} x_i^2 - \overline{x}^2\right) = \overline{y}\overline{x} + a\, S_X^2 .$$

Entonces:

$$\sum_i \sum_j y_j x_i f_{ij} - \overline{x}\overline{y} = S_{XY} = a\, S_X^2 .$$

Despejando, tenemos: $\begin{cases} a = \dfrac{S_{XY}}{S_X^2} \\ b = \overline{y} - \dfrac{S_{XY}}{S_X^2}\overline{x} \end{cases}$,

que son los valores que minimizan la expresión $S(a,b)$.

Por consiguiente, la recta de regresión de Y sobre X es:

$$y' = h(x) = \frac{S_{XY}}{S_X^2} x + \overline{y} - \frac{S_{XY}}{S_X^2}\overline{x}$$

que se puede escribir de forma más cómoda como:

$$y' - \overline{y} = \frac{S_{XY}}{S_X^2}(x - \overline{x}) .$$

La recta de regresión de X sobre Y no se obtiene despejando x de la expresión anterior, sino que hay que seguir el mismo razonamiento, ya que el único punto común de ambas rectas es el par $(\overline{x}, \overline{y})$, obteniéndose la siguiente recta de regresión:

$$x' - \overline{x} = \frac{S_{XY}}{S_Y^2}(y - \overline{y}) .$$

Las dos rectas de regresión, podemos observar que se cortan en el punto $(\overline{x}, \overline{y})$, que es conocido como centro de gravedad de la nube de puntos.

Ejemplo 4.2.

Hacemos las rectas de regresión de esta distribución de frecuencias:

X\Y	**5**	**10**	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$x_i\left(\sum_{j=1}^{p} n_{ij}y_j\right)$
5	1	2	3	15	75	5(1·5+2·10)=125
15	3	1	4	60	900	15(3·5+1·10)=375
$n_{.j}$	4	3	7	75	975	500
$n_{.j}y_j$	20	30	50			
$n_{.j}y_j^2$	100	300	400			

$$\bar{x} = \frac{75}{7} = 10.714 \qquad S_x^2 = \frac{975}{7} - \bar{x}^2 = 24.49$$

$$\bar{y} = \frac{50}{7} = 7.1428 \qquad S_y^2 = \frac{400}{7} - \bar{y}^2 = 6.122 \qquad S_{xy} = \frac{500}{7} - \overline{xy} = -5.10$$

Las rectas de regresión son:

$$Y|X: \quad y' - 7.1428 = \frac{-5.10}{24.49}(x - 10.714) \Rightarrow y' = -0.2082x + 9.3788$$

$$X|Y: \quad x' - 10.714 = \frac{-5.10}{6.122}(y - 7.1428) \Rightarrow x' = -0.833y + 16.66$$

Ejemplo 4.3.

Calculamos las rectas de regresión de estas dos variables dados los datos como pares de frecuencia uno:

x_i	y_i	x_i^2	y_i^2	x_iy_i
0	1	0	1	0
1	3	1	9	3
2	4	4	16	8
3	8	5	26	11

Entonces, tenemos:

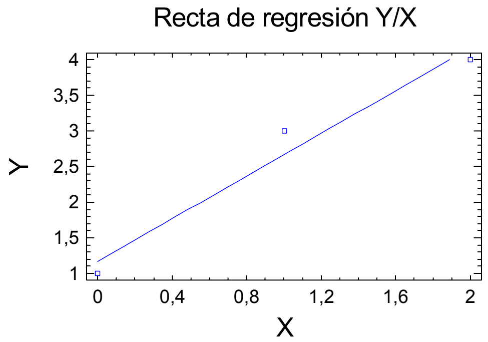
$$\bar{x} = 3/3 = 1 \qquad \bar{y} = 8/3 = 2.667 \qquad S_X{}^2 = 5/3 - 1 = 0.667$$

$$S_Y{}^2 = 26/3 - 2.667^2 = 1.555 \qquad S_{XY} = 11/3 - 1 \bullet 2.667 = 1.$$

Por tanto las rectas son:

$$Y|X: \quad y' - 2.667 = \frac{1}{0.667}(x - 1)$$

$$X|Y: \quad x' - 1 = \frac{1}{1.555}(y - 2.667)$$

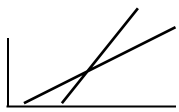


Los coeficientes de regresión

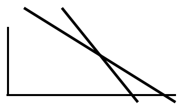
$$\begin{cases} Y|X: & \dfrac{S_{XY}}{S_X{}^2} \\ X|Y: & \dfrac{S_{XY}}{S_Y{}^2} \end{cases}$$

miden el grado de aumento de una variable al aumentar la otra. Ambos coeficientes tienen el mismo signo, que es el de la covarianza:

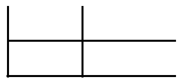
- Si la covarianza es positiva, ambos son positivos y las dos variables varían en el mismo sentido:



- 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: $\begin{cases} y = \overline{y} \\ x = \overline{x} \end{cases}$



Además, se verifica que si X e Y son independientes, entonces la covarianza vale cero, por lo que las rectas de regresión no explican nada; aunque el recíproco no es cierto.

B) Regresión no Lineal.

Aunque la regresión lineal tiene gran cantidad de aplicaciones, hay veces en las que la función puede no ser una recta. Nos centraremos en la función de regresión de Y sobre X, ya que el caso de X sobre Y es totalmente análogo. Consideramos las siguientes funciones de regresión no lineales:

B.1) Regresión Hiperbólica:

La función de regresión hiperbólica viene dada por:

$$y' = h(x) = a/x + b.$$

Consideramos el cambio de variable $Z = 1/X$, por lo que tenemos la recta:

$$y' = az + b.$$

Por tanto, estamos en el caso lineal, pero con las variables Y y Z. Entonces, tenemos que:

$$a = \frac{S_{YZ}}{S_Z{}^2} \qquad \text{y} \qquad b = \overline{y} - a\overline{z}\ ,$$

por lo que la función de regresión va a venir dada por:

$$y' - \overline{y} = \frac{S_{YZ}}{S_Z{}^2}\left(\frac{1}{x} - \overline{z}\right).$$

Por consiguiente, en la tabla bidimensional trabajaremos con las variables Y y Z.

Ejemplo 4.4.

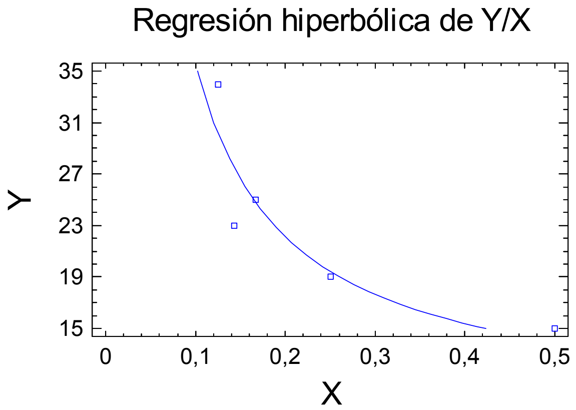
x_i	y_j	z_i	z_i^2	$z_i y_j$
1/2	15	2	4	30
1/4	19	4	16	76
1/6	25	6	36	150
1/7	23	7	49	161
1/8	34	8	64	272
	116	27	169	689

$\overline{y} = 116/5 = 23.2 \qquad \overline{z} = 27/5 = 5.4 \qquad S_Z^{\ 2} = 169/5 - 5.4^2 = 4.64$

$S_{YZ} = 689/5 - 23.2 \bullet 5.4 = 12.52$

Entonces la regresión hiperbólica de Y sobre X viene dada por la siguiente hipérbola equilátera:

$$y' - 23.2 = \frac{12.52}{4.64}\left(\frac{1}{x} - 5.4\right) \qquad \Rightarrow \qquad y' = \frac{2.70}{x} + 8.63$$



B.2) Regresión Potencial:

La función de regresión potencial viene dada por:

$$y' = h(x) = b \cdot x^a .$$

Para linealizarla, tomaremos logaritmos en ambos miembros, con lo que obtenemos lo siguiente:

$$ln(y) = ln(b \cdot x^a) = ln(b) + ln(x^a) = ln(b) + a \cdot ln(x)$$

Hacemos los cambios de variables $\widetilde{Y} = ln(Y)$, $\widetilde{X} = ln(X)$, $\widetilde{b} = ln(b)$, con lo que tenemos la recta:

$$\widetilde{y}' = a\,\widetilde{x} + \widetilde{b} .$$

Como tenemos una recta, le aplicamos la regresión lineal a las variables $\widetilde{X}$ e $\widetilde{Y}$, con lo que obtenemos los valores de los coeficientes:

$$\widetilde{y}' - \overline{\widetilde{y}} = \frac{S_{\widetilde{X}\widetilde{Y}}}{S_{\widetilde{X}}^{\ 2}}(\widetilde{x} - \overline{\widetilde{x}}),$$

que despejando $\widetilde{y}'$ nos da los coeficientes a y $\widetilde{b}$, por lo que el coeficiente b va a venir dado por $b = e^{\widetilde{b}}$.

Por tanto, en la tabla vamos a trabajar con las variables $\widetilde{X}$ e $\widetilde{Y}$.

Ejemplo 4.5.

Tras 7 años de funcionamiento, se ha observado en una piscifactoría el crecimiento de la población de truchas, donde *X* el año e *Y* es el número de truchas en miles. Ajustamos una función potencial:

x_i	y_j	$\widetilde{x}_i$	$\widetilde{y}_i$	$\widetilde{x}_i^{\,2}$	$\widetilde{x}_i\ \widetilde{y}_i$
1	3	ln(1)	ln(3)	0	0
2	7	ln(2)	ln(7)	0.4804	1.3487
3	15	ln(3)	ln(15)	1.2069	2.9753
5	100	ln(5)	ln(100)	2.5903	7.4116
6	250	ln(6)	ln(250)	3.2104	9.8929
7	650	ln(7)	ln(650)	3.7866	12.6036
24	1025	7.1388	22.3562	11.2746	34.2321

$$\overline{\widetilde{y}} = 22.3562 / 6 = 3.726 \ , \ \overline{\widetilde{x}} = 7.1388 / 6 = 1.1898$$

$$S_{\widetilde{X}}^{\ 2} = 11.2746 / 6 - \overline{\widetilde{x}}^{\,2} = 0.4634$$

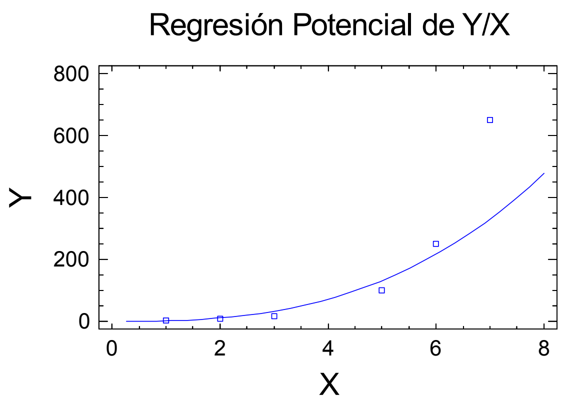
$$S_{\widetilde{X}\widetilde{Y}} = 34.2331 / 6 - \overline{\widetilde{x}}\,\overline{\widetilde{y}} = 1.2722 \ .$$

Entonces:

$$\widetilde{y}' - 3.726 = \frac{1.2722}{0.4634}(\widetilde{x} - 1.1898) \quad \Rightarrow \quad \widetilde{y}' = 2.745 \bullet \widetilde{x} + 0.46$$

Por tanto, $b = e^{\widetilde{b}} = 1.584$, por lo que:

$$y' = 1.584 \bullet x^{2.745}.$$



B.3) Regresión Exponencial.

La función de regresión exponencial viene dada por:

$$y' = h(x) = b \cdot a^x$$

Para linealizarla, tomaremos logaritmos en ambos miembros, con lo que obtenemos lo siguiente:

$$ln(y) = ln(b \cdot a^x) = ln(b) + ln(a^x) = ln(b) + x \cdot ln(a).$$

Hacemos los cambios de variables $\widetilde{Y} = ln(Y),\ \widetilde{a} = ln(a),\ \widetilde{b} = ln(b)$, con lo que tenemos la recta:

$$\widetilde{y}' = \widetilde{a}x + \widetilde{b}\ .$$

Como tenemos una recta, le aplicamos la regresión lineal a las variables X e $\widetilde{Y}$, con lo que obtenemos los valores de los coeficientes:

$$\widetilde{y}' - \overline{\widetilde{y}} = \frac{S_{X\widetilde{Y}}}{S_{X}^{\ 2}}(x - \overline{x})\ ,$$

que despejando $\widetilde{y}'$ nos da los coeficientes $\widetilde{a}$ y $\widetilde{b}$, por lo que los coeficientes a y b van a venir dados por $a = e^{\widetilde{a}}$ y $b = e^{\widetilde{b}}$.

Por tanto en la tabla vamos a trabajar con las variables X e $\widetilde{Y}$.

Ejemplo 4.5 (continuación).

A los datos del ejemplo 47, le ajustamos una regresión exponencial:

x_i	y_j	$\widetilde{y}_i$	x_i^2	$x_i\widetilde{y}_i$ '
1	3	ln(3)	1	1.0986
2	7	ln(7)	4	3.8918
3	15	ln(15)	9	8.124
5	100	ln(100)	25	23.025
6	250	ln(250)	36	33.128
7	650	ln(650)	49	45.338
24	1025	22.3562	124	114.608

$\overline{\widetilde{y}} = 22.3562 / 6 = 3.726 \qquad \overline{x} = 24 / 6 = 4$

$S_X^{\ 2} = 124 / 6 - \overline{x}^2 = 4.6667 \quad S_{X\widetilde{Y}} = 114.608 / 6 - \overline{x}\,\overline{\widetilde{y}} = 4.1973.$

Entonces:

$$\widetilde{y}' - 3.726 = \frac{4.1973}{4.6667}(x - 4) \quad \Rightarrow \quad \widetilde{y}' = 0.8994 \bullet x + 0.1283$$

Por tanto, $a = e^{\widetilde{a}} = 2.4581$ y $b = e^{\widetilde{b}} = 1.1370$, por lo que

$$y' = 1.1370 \bullet 2.4581^x\ .$$

B.4) Regresión Parabólica.

La función de regresión parabólica es:

$$y' = h(x) = ax^2 + bx + c$$

No la podemos reducir al caso lineal como en los casos anteriores, por lo que tenemos que tomar mínimos cuadrados directamente, es decir, tenemos que hacerlo mediante el caso general:

Minimizamos la expresión:

$$S(a,b,c) = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\, e_{ij}^{\,2} = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}(y_j - ax_i^{\,2} - bx_i - c)^2 \;:$$

Derivando respecto de los coeficientes a, b y c:

$$\begin{cases} \dfrac{dS}{da} = 0 \\ \dfrac{dS}{db} = 0 \\ \dfrac{dS}{dc} = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\, x_i^{\,2}\, e_{ij} = 0 \\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\, x_i\, e_{ij} = 0 \\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}\, e_{ij} = 0 \end{cases} \quad \Rightarrow$$

$$\Rightarrow \quad \begin{cases} \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}x_i^{\,2}y_j = c\sum_{i=1}^{k} n_{i.}x_i^{\,2} + b\sum_{i=1}^{k} n_{i.}x_i^{\,3} + a\sum_{i=1}^{k} n_{i.}x_i^{\,4} \\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}x_iy_j = c\sum_{i=1}^{k} n_{i.}x_i + b\sum_{i=1}^{k} n_{i.}x_i^{\,2} + a\sum_{i=1}^{k} n_{i.}x_i^{\,3} \\ \displaystyle\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}y_j = cn + b\sum_{i=1}^{k} n_{i.}x_i + a\sum_{i=1}^{k} n_{i.}x_i^{\,2} \end{cases}$$

Los coeficientes los obtendremos resolviendo el sistema de tres ecuaciones con tres incógnitas.

Ejemplo 4.6.

Ajustamos una parábola a los siguientes datos:

$X \backslash Y$	1	2	3	4	5	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$n_{i.}x_i^3$	$n_{i.}x_i^4$	$x_i\sum n_{ij}y_j$	$x_i^2\sum n_{ij}y_j$
1	3	2				5	5	5	5	5	7	7
2			4	4		8	16	32	64	128	56	112
4					3	3	12	48	192	768	60	240
$n_{.j}$	3	2	4	4	3	16	33	85	261	901	123	359
$n_{.j}y_j$	3	4	12	16	15	50						
ny_j^2	3	8	36	64	75	186						

El sistema a resolver es, por tanto:

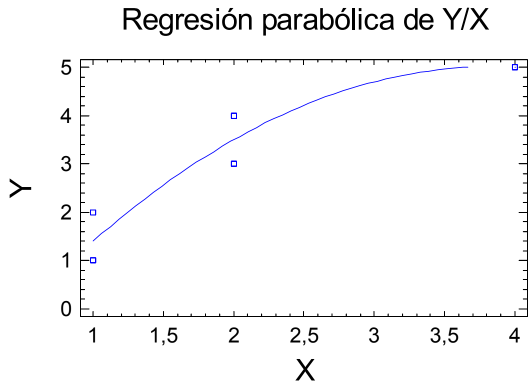
$$\begin{cases} 359 = c \bullet 85 + b \bullet 261 + a \bullet 901 \\ 123 = c \bullet 33 + b \bullet 85 + a \bullet 261 \\ 50 = c \bullet 16 + b \bullet 33 + c \bullet 85 \end{cases}$$

La solución del sistema es:

a = -0.45 b = 3.45 c = -1.6

Por tanto, la regresión viene dada por:

$$y' = -0.45x^2 + 3.45x - 1.6$$



En general, si queremos ajustar un polinomio de grado q superior a 2:

$$y' = a_0 + a_1x + a_2x^2 + ... + a_nx^n \quad ,$$

el método de mínimos cuadrados nos llevará a un sistema análogo al anterior de q+1 ecuaciones con q+1 incógnitas.

4.3. Correlación.

La correlación se encarga de estudiar el grado de asociación de las variables, por lo que nos va a indicar en qué medida la función obtenida mediante la regresión explica la variable dependiente en función de la independiente, es decir, cómo de bien o de mal se ajusta la función de regresión dada a la nube de puntos. 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.

Vamos 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 hemos utilizado para la regresión mide el error cometido por la función de regresión que hemos ajustado mediante la suma de los cuadrados de los residuos:

$$\sum_{i=1}^{k}\sum_{j=1}^{p}(y_j - h(x_i))^2 n_{ij} \ ,$$

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:

$$S_{rY}{}^2 = \sum_{i=1}^{k}\sum_{j=1}^{p}(y_j - h(x_i))^2 f_{ij} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p}(y_j - h(x_i))^2 n_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{p} e_{ij}{}^2 f_{ij}$$

(en funciones lineales en los parámetros como la recta, hipérbola equilátera y los polinomios la media de los residuos es cero, por lo que es la varianza de los residuos; pero en funciones no lineales en los parámetros, como la potencial o la exponencial la media de los residuos no es cero, pero se mantiene el mismo nombre). Indica lo que queda sin explicar de la variable Y 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

$$S_Y{}^2 = S_{eY}{}^2 + S_{rY}{}^2 \ ,$$

donde

$$S_{eY}{}^2 = \sum_i\sum_j (h(x_i) - \overline{y})^2 f_{ij}$$

es la varianza explicada por la regresión, pero que solo es válida dicha descomposición para funciones lineales en los parámetros (recta, hipérbola, polinomios) y no lo es para las demás curvas (potencial, exponencial,...), por lo que el coeficiente de determinación solo va a tener sentido para las funciones de la primera clase.

Demostración:

La vamos a hacer para la recta, pero es totalmente análoga para cualquier otra función lineal en los parámetros.

El valor de la variable será igual al valor estimado mediante la regresión más el error cometido en ésta:

$$y_j = h(x_i) + e_{ij} \Rightarrow$$

$$(y_j - \overline{y})^2 = \left((h(x_i) - \overline{y}) + e_{ij}\right)^2 = (h(x_i) - \overline{y})^2 + e_{ij}{}^2 + 2(h(x_i) - \overline{y})e_{ij}$$

Sumándolos todos, tenemos:

$$\sum_i\sum_j \left(y_j - \overline{y}\right)^2 f_{ij} =$$

$$= \sum_i\sum_j (h(x_i) - \overline{y})^2 f_{ij} + \sum_i\sum_j e_{ij}{}^2 f_{ij} + 2\sum_i\sum_j (h(x_i) - \overline{y})e_{ij} f_{ij}$$

por lo que tenemos que probar que el último término es cero:

$$\sum_i \sum_j (h(x_i) - \bar{y}) e_{ij} f_{ij} = \sum_i \sum_j (b + a x_i - \bar{y}) e_{ij} f_{ij} =$$

$$= b \sum_i \sum_j e_{ij} f_{ij} + a \sum_i \sum_j x_i e_{ij} f_{ij} - \bar{y} \sum_i \sum_j e_{ij} f_{ij} = 0$$

ya que los dos primeros sumandos son cero por las dos ecuaciones de mínimos cuadrados y el tercero es también cero porque la media de los residuos es cero.

Se define el Coeficiente de Determinación como la proporción de la varianza total de Y que aparece explicada por la regresión:

$$R^2 = \frac{S_{eY}{}^2}{S_Y{}^2} = 1 - \frac{S_{rY}{}^2}{S_Y{}^2} \ ,$$

que es un valor entre 0 y 1 que mide la fiabilidad del ajuste.
($1-R^2$ mide la proporción de varianza no explicada).

Por tanto, la varianza residual vale, despejando de la expresión anterior:

$$S_{rY}{}^2 = S_Y{}^2 (1 - R^2) \ .$$

Veamos la interpretación de los distintos valores del coeficiente de determinación:

- R^2 vale cero si y solo si $S_{rY}^2 = S_Y^2$, si y solo si la varianza explicada vale cero, es decir, la regresión no explica nada de Y a partir de X, por lo que el ajuste es el peor posible.
- R^2 vale uno si y solo si $S_{eY}^2 = S_Y^2$, 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.

Diremos que el ajuste que hemos realizado es aceptable para valores del coeficiente de determinación mayores de 0.49, es decir, para fiabilidad superior al 49%.

Veamos cómo se calcula la correlación en cada uno de los casos estudiados:

A) Regresión Tipo I.

En el caso de la regresión tipo I, sí podemos descomponer la varianza marginal de la variable dependiente como:

$$S_Y^{\,2} = \sum_{i=1}^{k} f_{i.} Var_i(Y) + \sum_{i=1}^{k} f_{i.} (\bar{y}_i - \bar{y})^2 = S_{rY}^{\,2} + S_{eY}^{\,2} \ .$$

Para medir la bondad del ajuste utilizamos la Razón de Correlación, que es el cociente de la varianza explicada entre la varianza marginal de la variable dependiente:

$$\eta_{Y|X}^{\,2} = \frac{\sum_{i=1}^{k} f_{i.} (\bar{y}_i - \bar{y})^2}{S_Y^{\,2}} = 1 - \frac{\sum_{i=1}^{k} f_{i.} Var_i(Y)}{S_Y^{\,2}},$$

que es un valor comprendido entre 0 y 1. La interpretación de su valor es la siguiente:

- $\eta_{Y|X}^{2} = 1$, quiere decir que todas las varianzas condicionadas valen cero, por lo que Y depende funcionalmente de X.
- $\eta_{Y|X}^{2} = 0$, quiere decir que las medias condicionadas coinciden con las marginales, por lo que las variables X e Y son independientes, es decir, son incorreladas y no existe dependencia entre ellas.
- $0<\eta_{Y|X}^{2}<1$, quiere decir que cuanto más próximo esté el valor a 1, más alta será la correlación entre las variables.

Decir además que, como es el mejor ajuste a los datos, tiene que ser la regresión que nos dé una mayor correlación entre las variables estudiadas.

Ejemplo 4.1 (continuación).

Calculamos la razón de correlación del ejemplo para el caso Y sobre X:

$$\eta_{Y|X}^{\,2} = \frac{\frac{3}{7} \bullet \left(8.33 - 7.1428\right)^2 + \frac{4}{7} \bullet \left(6.25 - 7.1428\right)^2}{6.122} = 0.2157 \, .$$

B) Regresión Lineal.

En el caso de la recta de regresión, vemos lo que vale el coeficiente de determinación:

La varianza residual vale:

$$S_{rY}^{\,2} = \sum_i \sum_j f_{ij} e_{ij}^{\,2} = \sum_i \sum_j f_{ij} (ax_i + b - y_j)^2 =$$

$$= a \sum_{i,j} f_{ij} (ax_i + b - y_j) x_i + b \sum_{i,j} f_{ij} (ax_i + b - y_j) - \sum_{i,j} f_{ij} (ax_i + b - y_j) y_j =$$

(por las condiciones de mínimos cuadrados)

$$= -\sum_{i,j} f_{ij}(ax_i + b - y_j)y_j = -a\sum_{i,j} f_{ij}x_i y_j - b\sum_{i,j} f_{ij}y_j + \sum_{i,j} f_{ij}y_j^{\,2} =$$

$$= -am_{11} - b\overline{y} + m_{02} = -\frac{S_{XY}}{S_X^{\,2}} m_{11} - \left(\overline{y} - \frac{S_{XY}}{S_X^{\,2}}\overline{x} \right)\overline{y} + m_{02} =$$

$$= -\frac{S_{XY}}{S_X^2}\left(m_{11} - \overline{x}\overline{y}\right) + m_{02} - \overline{y}^2 = -\frac{S_{XY}^{\,2}}{S_X^{\,2}} + S_Y^{\,2}$$

Entonces, el coeficiente de determinación vale:

$$R^2 = 1 - \frac{S_{rY}^{\,2}}{S_Y^{\,2}} = 1 - \frac{S_Y^{\,2} - \frac{S_{XY}^{\,2}}{S_X^{\,2}}}{S_Y^{\,2}} = \frac{S_Y^{\,2} - S_Y^{\,2} + \frac{S_{XY}^{\,2}}{S_X^{\,2}}}{S_Y^{\,2}} = \frac{S_{XY}^{\,2}}{S_Y^{\,2} S_X^{\,2}}$$

Por tanto, el coeficiente de determinación viene dado por:

$$R^2 = \frac{S_{XY}^{\,2}}{S_Y^{\,2} S_X^{\,2}}$$

Análogamente calculamos el coeficiente de determinación de la recta de regresión de X sobre Y, que, como caso excepcional, es exactamente el mismo que el de Y sobre X.

Definimos el Coeficiente de correlación lineal como:

$$r = \frac{S_{XY}}{S_Y S_X}$$

es decir, $\sqrt{R^2}$ con el signo de la covarianza, y mide la correlación lineal entre las variables. Toma valores entre -1 y 1. Veamos el significado de los distintos valores que puede tomar:

- 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.
Consideraremos 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.
Demostración:

$$R^2 = \frac{S_{XY}{}^2}{S_X{}^2 S_Y{}^2} = \frac{S_{XY}}{S_X{}^2}\frac{S_{XY}}{S_Y{}^2}$$

Ejemplo 4.2 (continuación).
El coeficiente de correlación lineal vale:

$$r = \frac{-5.10}{\sqrt{24.49 \bullet 6.122}} = -0.4165,$$

lo que nos dice que el ajuste no es bueno.

Ejemplo 4.3 (continuación).
El coeficiente de correlación vale:

$$r = \frac{1}{\sqrt{0.6667 \bullet 1.555}} = 0.9821,$$

por lo que podemos decir que el ajuste es bastante bueno, aunque la muestra puede no ser suficientemente representativa al ser de solo tres valores.

C) Regresión Hiperbólica.

Como la hipérbola es lineal en los parámetros, podemos definir el coeficiente de determinación de la hipérbola, que, tras su desarrollo, es igual a:

$$R_{Y/X}{}^2 = \frac{S_{YZ}{}^2}{S_Z{}^2 S_Y{}^2}$$

que toma valores entre 0 y 1.

D) Regresión Parabólica.

Como la parábola es lineal en los parámetros tiene sentido calcular su coeficiente de determinación. Para ello, calculamos en primer lugar la varianza residual:

$$S_{rY}{}^2 = \sum_{i,j} f_{ij} e_{ij}{}^2 = \sum_{i,j} f_{ij}(y_j - ax_i{}^2 - bx_i - c)e_{ij} =$$

$$= \sum_{i,j} f_{ij} e_{ij} y_j - a\sum_{i,j} f_{ij} e_{ij} x_i^{\,2} - b\sum_{i,j} f_{ij} e_{ij} x_i - c\sum_{i,j} f_{ij} e_{ij} =$$

(por las condiciones de mínimos cuadrados los tres últimos sumandos valen cero)

$$= \sum_{i,j} f_{ij} y_j \left(y_j - ax_i^2 - bx_i - c \right) =$$

$$= \sum_{i,j} f_{ij} y_j^{\,2} - a\sum_{i,j} f_{ij} y_j x_i^{\,2} - b\sum_{i,j} f_{ij} y_j x_i - c\sum_{i,j} f_{ij} y_j$$

$$= m_{02} - am_{21} - bm_{11} - cm_{01}$$

Entonces, el coeficiente de determinación parabólica viene dado por:

$$R^2 = 1 - \frac{S_{rY}^{\,2}}{S_Y^{\,2}} = 1 - \frac{m_{02} - am_{21} - bm_{11} - cm_{01}}{S_Y^{\,2}} .$$

Ejemplo 4.6 (continuación)

El coeficiente de determinación vale:

$$R^2 = 1 - \frac{\frac{186}{16} + 0.45 \bullet \frac{359}{16} - 3.45 \bullet \frac{123}{16} + 1.6\frac{50}{16}}{\frac{186}{16} - \left(\frac{50}{16}\right)^2} = 0.8924$$

por lo que el ajuste es muy bueno.

E) Regresión Potencial y Exponencial.

En los casos de las regresiones potencial y exponencial hemos obtenido los parámetros de la regresión tras realizar un ajuste mediante mínimos cuadrados de una recta de regresión después de realizar ciertas transformaciones.

Si calculamos el coeficiente de correlación de dicha recta, obtendremos la parte de regresión explicada por el ajuste entre las variables transformadas ($\widetilde{X}$ e $\widetilde{Y}$), pero no las originales (*X* e *Y*), por lo que con esto no podemos decir si el ajuste de nuestras 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, si dados varios ajustes por regresión de mínimos cuadrados de dos variables dadas (*Y/X*), el mejor ajuste será aquel cuya varianza residual sea menor; pero si no comparamos ni la regresión exponencial ni la potencial, podemos comparar con el coeficiente de determinación.

Ejemplo 4.5 (continuación).

Calculamos la varianza residual.

$$y' = 1.137 \cdot 2.4581^x.$$

Entonces:

x_i	y_i	$h(x_i)$	$(y_i - h(x_i))^2$
1	3	2.7946	0.04218
2	7	6.8694	0.01704
3	15	16.885	3.55612
5	100	102.02	4.1133
6	250	250.79	0.6326
7	650	616.48	1123.5806
			1131.9419

La varianza residual vale:

$$S_{rY}^{\ 2} = \frac{1}{6} 1131.9419 = 188.6570\,.$$

4.4. 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 y menos nos alejemos de ($\overline{x}, \overline{y}$).

Ejemplo 4.2 (continuación).

a) Queremos conocer el valor de Y cuando X=10.

Para ello sustituimos el valor x=10 en la recta de regresión de $Y|X$:

$$y'_{10} = -0.2082 \cdot 10 + 9.3388 = 7.2568$$

b) Queremos conocer el valor de X cuando Y=7.

Para ello sustituimos el valor y=7 en la recta de regresión $X|Y$ (no se puede sustituir en la otra):

$$x'_7 = -0.833 \cdot 7 + 16.66 = 10.829$$

RELACIÓN DE ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL.

1.- En una encuesta en la que se ha entrevistado a 480 familias, se han obtenido para las mismas los siguientes datos sobre ingresos mensuales (X) y depósitos a la vista en Bancos y Cajas de Ahorros (Y):

X\Y	**0-2000**	**2000-5000**	**5000-20000**	**20000-100000**
500 - 1000	40	12	8	-
1000 - 1500	16	48	12	4
1500 - 2500	8	80	92	20
2500 - 5000	4	40	72	24

a) Calcular $n_{1.}$, $n_{.3}$.
b) Expresar en porcentaje f_{13}, f_{44}, $f_{1.}$ y $f_{.4}$
c) Calcular la proporción de individuos que tienen $X= x_1$, condicionado a que $Y=y_2$.

2.- Se han estudiado los pesos (en Kg.) y las alturas (en cm.) de un grupo de individuos, obteniendo:

X\Y	**160**	**162**	**164**	**166**	**168**	**170**
48	3	2	2	1	0	0
51	2	3	4	2	2	1
54	1	3	6	8	5	1
57	0	0	1	2	8	3
60	0	0	0	2	4	4

a) Calcular peso y altura media.
b) ¿Qué variable es más homogénea?
c) Calcular el porcentaje de alumnos que pesan menos de 55 Kg. y miden más de 165 cm.
d) De aquellos que miden más de 165 cm., ¿cuál es el porcentaje de los que pesan más de 53 Kg.?
e) ¿Cuál es la altura más frecuente entre los individuos cuyo peso oscila entre 51 y 57 Kg.?
f) Decir qué peso medio es más representativo, el de los que miden 164 o el de los que miden 168 cm.

3.- En una población se ha procedido a realizar 100 observaciones sobre la pareja de variables X e Y, habiéndose obtenido los siguientes datos:

Y_i	X_i	**Frecuencias absolutas**
4	1	3
5	1	8
4	2	2
5	2	10
6	2	18
5	3	6
6	3	12
7	3	19
6	4	6
7	4	16

a) Estimar los parámetros de la recta de regresión de *Y/X* y los de la recta de *X/Y*.

b) Calcular el coeficiente de determinación entre *X* e *Y*.

c) Calcular la varianza residual de la regresión de *Y/X* y la de *X/Y*.

4.- Elegidos 50 matrimonios y preguntada la edad de los cónyuges al casarse, se obtuvo el siguiente resultado:

X* \ *Y	**(15,18]**	**(18,21]**	**(21,24]**	**(24,27]**
(15,20]	3			
(20,25]	2	4	7	
(25,30]	3	2	10	2
(30,35]		2	6	5
(35,40]			1	3

X = edad del hombre , *Y* = edad de la mujer.

a) Calcular la varianza de las distribuciones marginales.

b) Hallar la mediana de la edad de las mujeres que se casan con hombres con edades comprendidas entre 25 y 30 años.

c) Calcular las rectas de regresión de *Y* sobre *X* y de *X* sobre *Y*. ¿Es fiable el ajuste?

d) Si un hombre se casa a los 43 años, predecir la edad de su mujer.

5.- En un grupo de familias que habitan en pisos de alquiler en determinada zona se estudia el número de habitaciones de la vivienda y los ingresos mensuales de la familia. La información es:

X \ *Y*	**(0,200]**	**(200,500]**	**(500,1000]**	**(1000,1300]**
2	5	4	1	
3	2	6	5	2
4		2	10	8
5			10	15

a) ¿Cuál es el salario medio de esas familias?
b) Calcular el recorrido de la variable *X*.
c) Supuesta una relación lineal de las variables, predecir el salario de una familia que viva en un piso de 7 habitaciones; así como el número de habitaciones de una familia que cobre 2000 € ¿Es creíble la suposición de linealidad?.

6.- Sean dos variables estadísticas cuya distribución conjunta viene dada por la siguiente tabla de frecuencias absolutas:

X \ *Y*	**(4 - 6]**	**(6 - 8]**	**(8 - 10]**	**(10 - 12]**
< 100]	6	21	45	24
(100-200]	18	63	135	72
(200-300]	12	42	90	48

¿Son *X* e *Y* independientes?

7.- Se dispone de la siguiente información referente a 6 familias sobre el gasto en espectáculos (*G*) y la renta disponible mensual (*R*):

G_i	30	50	60	90	100	140
R_i	600	700	800	1000	1500	2000

Ambas variables medidas en miles de pesetas.

a) Estimar los parámetros de las siguientes funciones con las que se pretende hacer regresión:

- $G = b + aR$
- $Ln(G) = a\ Ln(R) + b$

b) ¿Qué función se ajusta mejor?

8.- Se pretende estudiar la distribución de la población de una nación en el territorio total. Para esto se recogen los siguientes datos:

*X**Y*	**0 - 10**	**10 - 25**	**25 - 30**
0 - 5	1	2	0
5 - 10	0	2	1
10 - 14	1	1	1

Y es la variable estadística que mide la extensión de las regiones (en miles de km^2) y *X* la población (en miles de personas).

a) ¿Qué cantidad de población es más frecuente que tenga una región?

b) Se ha propuesto dotar al país de mejoras en las comunicaciones. Se empezará por el 20% de las regiones de mayor extensión de entre las que tienen una mayor población, es decir, cuya población rebasa los 10 millones. ¿Qué extensión mínima han de tener esas regiones?

9.- Los habitantes de cierta localidad (en miles), cada cinco años, han sido:

Año	1985	1990	1995	2000
Habitantes	33	35	45	82

a) Ajustar una recta que exprese el número de habitantes en función del tiempo. Calcular el coeficiente de determinación. Estimar la población para 2007.

b) Suponer que las variables están relacionadas de la forma $H = b\ a^A$ (H = habitantes, A = años) y estimar la población para 2007.

10.- La siguiente tabla recoge la cantidad anual (*X* en miles de €) invertida en investigación y desarrollo y las utilidades anuales (*Y*, en miles de €) para 34 empresas:

X/Y	**0 - 2**	**2 – 4**	**4 - 5**
0 - 40	10	1	-
40 - 80	-	10	1
80 - 160	-	1	11

Discutir qué tipo de relación es la más adecuada para explicar la utilidad, a partir de la inversión, entre una relación lineal, una exponencial y una potencial.

11.- La presión *P* y el volumen *V*, de una masa gaseosa están ligadas por una ecuación de la forma: $PV^a = b$

donde a y b son constantes. Con los siguientes datos, obtenidos en un experimento, determinar los valores de a y b para explicar *P* a partir de *V*, por el método de los mínimos cuadrados.

P(kg/cm²)	0.5	1	1.5	2.0	2.5	3.0
V(litros)	1.65	1.03	0.74	0.61	0.53	0.45

12.- De los Tiempos de vuelo (*X*, expresado en horas) y consumo de combustible (*Y*, expresado en miles de libras) de una compañía aérea se han obtenido datos relativos a 24 trayectos distintos realizados por un avión. A partir de estos datos se ha obtenido la siguiente información:

$$\sum_j y_j = 219.719 \quad \sum_j y_j^2 = 2396.504 \quad \sum_i \sum_j x_i y_j = 349.486$$

$$\sum_i x_i = 31.470 \quad \sum_i x_i^2 = 51.075 \quad \sum_i \sum_j x_i^2 y_j = 633.993$$

$$\sum_i x_i^4 = 182.977 \quad \sum_i x_i^3 = 93.6$$

a) Estimar la función parabólica de *Y* sobre *X*.

b) ¿Qué consumo total se espera para un programa de vuelos compuesto por 100 vuelos de media hora, 200 de una hora y 100 de dos horas?

13.- Dada la siguiente distribución bidimensional:

*X**Y*	**1**	**2**	**3**	**4**
10	1	3	0	0
12	0	1	4	3
14	2	0	0	2
16	4	0	0	0

a) ¿Son X e Y estadísticamente independientes? Representar su nube de puntos.
b) Calcular y representar las curvas de regresión y estudiar su bondad.
c) ¿Están correladas linealmente? Dar sus rectas de regresión.

14.- Se hizo un estudio sobre la cantidad de azúcar transformada (Y) en cierto proceso, a varias temperaturas (X). Los datos obtenidos se recogen en la siguiente tabla:

X	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
Y	8.1	7.8	8.5	9.8	9.5	8.9	8.6	10.2	9.3	9.2	10.5

a) Obtener la recta de regresión que explique la cantidad de azúcar transformada en el proceso en función de la temperatura.
b) Estimar la cantidad de azúcar transformada en un proceso sometido a una temperatura de 11ºC. ¿Es fiable esta estimación?

15.- Los siguientes datos corresponden a una muestra de 200 hombres de un cierto pueblo a los cuales se les pesó (Y) y midió (X).

*X**Y*	**50-65**	**65-80**	**80-95**	**95-110**
159-161	2	1	0	0
161-163	2	4	1	0
163-165	3	12	4	0
165-167	5	20	15	4
167-169	8	25	15	2
169-171	1	9	20	12
171-173	1	5	13	6
173-175	0	1	3	4
175-177	0	0	0	2

a) Calcular media y moda de la altura.
b) ¿Qué distribución es más homogénea?

c) Si nos limitamos a los hombres que miden más de 171, ¿qué proporción de ellos pesa más de 60 Kg?
d) ¿Es razonable suponer la existencia de una relación lineal entre el peso y la altura de estos individuos?

16.- Se considera una colonia de 75 insectos, que se estudian atendiendo a su tiempo de vida (*X*) y al tiempo de permanencia en un hábitat extraño (*Y*) medidos en minutos:

Y\X	**0-5**	**5-10**	**10-15**	**15-20**
0-1	1	3	13	15
1-3	3	5	7	4
3-4	7	6	8	3

a) Calcular el tiempo medio de permanencia en el hábitat de los individuos de la muestra.
b) Calcular el tiempo medio de vida de los insectos cuyo tiempo de permanencia en el hábitat extraño estuvo entre 1 y 4 minutos.
c) ¿Qué intervalo referente al tiempo de permanencia en dicho hábitat es el más usual?
d) Estudiar la forma de la distribución del tiempo de vida.
e) Calcular el tiempo de vida a partir del cual se encuentra el 25% de la población, supuesto que el tiempo de permanencia en el hábitat extraño ha estado entre 1 y 3 minutos.
f) Obtener e interpretar el coeficiente de correlación lineal.

17.- En una expedición científica al Amazonas se encontró una variedad rara de algas, de la que se midió en cm. la longitud (*X*) y anchura (*Y*) de sus láminas.

X\Y	**2-4**	**4-6**	**6-10**
10-12	1	0	3
12-15	4	1	2
15-17	1	1	3

a) Calcular el porcentaje de algas con láminas de anchura inferior a 6 y longitud superior a 12.
b) Calcular la anchura media de las láminas con longitud entre 12 y 15 cm.
c) Longitud más frecuente de las algas con láminas de anchura entre 2 y 4 cm.
d) Calcular la longitud tal que la mitad de las restantes sea inferior a ella.
e) Estudiar la dependencia funcional e independencia estadística.

f) Dar la recta de regresión que estima la anchura de las láminas en función de su longitud. ¿Son fiables las estimaciones?

CAPÍTULO 5:

ANÁLISIS DESCRIPTIVO DE SERIES CRONOLÓGICAS.

A lo largo de este capítulo vamos a realizar un estudio descriptivo de una *serie temporal* o *cronológica*, es decir, de una variable que se observa a lo largo del tiempo y tiene como objetivo explicar la evolución de dicha variable a lo largo del tiempo y predecir sus valores futuros.

5.1. Comportamiento de una Serie Cronológica.

Se define una Serie Temporal o Cronológica como una sucesión de observaciones de una determinada variable tomadas en distintos instantes de tiempo.

Una serie cronológica no es nada más que un caso particular de una variable estadística bidimensional, donde la variable dependiente es la variable objeto de estudio y la variable independiente es el tiempo; y el objetivo fundamental del estudio de este tipo de variables es el poder hacer predicciones sobre valores futuros de la variable dependiente.

Al valor que toma la variable Y en un instante de tiempo t, lo representaremos mediante Y_t.

Los datos obtenidos de una serie cronológica se representan, habitualmente, mediante la siguiente tabla:

Tiempo	Observaciones
t_1	$Y(t_1)$
t_2	$Y(t_2)$
. . .	. . .

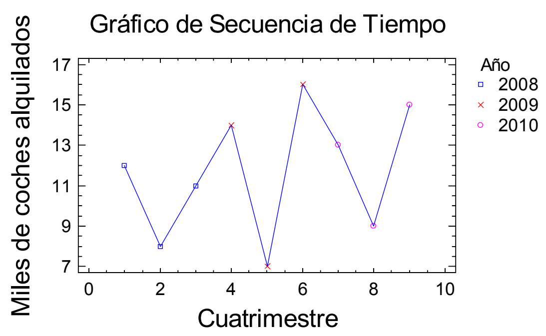
Si por cada intervalo de tiempo t se dispone de varias observaciones, los datos se recogen en una tabla del siguiente tipo:

Años\Estaciones	1	…	j	…	s
t_1	y_{11}	…	y_{1j}	…	y_{1s}
t_2	y_{21}	…	y_{2j}	…	y_{2s}
…	…	…	…	…	…
t_n	y_{n1}	…	y_{nj}	…	y_{ns}

Ejemplo 5.1.

Una empresa de alquiler de coches muestra el número de coches alquilados (en miles) por cuatrimestre durante 3 años:

Año\Cuatrimestre	**1**	**2**	**3**
2008	12	8	11
2009	14	7	16
2010	13	9	15



5.2. Componentes de una Serie Cronológica.

Los valores de una serie temporal son el resultado de la acción de cuatro factores que actúan sobre ella:

a) Tendencia secular $\tau(t)$: Es el movimiento de la serie a largo plazo e indica el comportamiento general del fenómeno, suponiéndolo aislado de cualquier tipo de oscilación.
b) Variaciones estacionales $E(t)$: Representa fluctuaciones de la serie que se repiten en intervalos cortos e iguales en el tiempo, cuyo periodo de repetición suele ser, en general, inferior al año; y recoge el crecimiento o decrecimiento de la serie por el hecho de estar en una determinada época del año.
c) Variaciones cíclicas $C(t)$: Son oscilaciones alrededor de la tendencia secular con periodicidad y amplitud variables, cuyo periodo de repetición suele ser superior al año, por lo que, a diferencia de la variación estacional, esta periodicidad no tiene por qué ser fija. Un ejemplo son los ciclos económicos con etapas de prosperidad, recesión y recuperación.
d) Variaciones Accidentales $\varepsilon(t)$: Es una fluctuación impredecible que ocurre de manera aleatoria en diferentes instantes de tiempo, como son las huelgas, catástrofes…

En este capítulo se describen diversos procedimientos de cálculo para la tendencia secular y la variación estacional, pero no se trabajará en el estudio de los ciclos.

5.2.1. Modelos.

Una vez estudiadas las distintas componentes de una serie, vamos a estudiar cómo se efectúa la composición de las mismas, para lo cual existen varios modelos, de los que los más comunes son el aditivo y el multiplicativo:

a) Modelo Aditivo:

Suponemos que las observaciones de la serie se generan como la suma de las cuatro componentes, es decir,

$$Y(t) = \tau(t) + E(t) + C(t) + \varepsilon(t)$$

Cada una de las componentes viene expresada en la misma unidad de medida que las observaciones.

La componente $\varepsilon(t)$ es independiente de las otras componentes, es decir, los valores que tome no tienen que ver con los valores de las otras componentes. Lo mismo ocurre con la variación estacional y la cíclica.

b) **Modelo Multiplicativo:**
Suponemos que las observaciones de la serie se generan como el producto de las cuatro componentes:

$$Y(t) = \tau(t) \cdot E(t) \cdot C(t) \cdot \varepsilon(t)$$

La tendencia secular viene expresada en las mismas unidades que las observaciones y las demás componentes en tantos por uno.
En este modelo no se cumple la hipótesis de independencia de la variación residual respecto de las demás componentes.
El **Modelo Multiplicativo Mixto** sí cumple la hipótesis de que la variación residual es independiente de las demás componentes**.** Suponemos que las observaciones de la serie se generan como el producto de las tres primeras componentes y la suma de la variación residual,:

$$Y(t) = \tau(t) \cdot E(t) \cdot C(t) + \varepsilon(t)$$

En este modelo, la tendencia y los residuos se expresan en la misma unidad que la serie, y el resto de componentes en tanto por uno.
Para resolver el cumplimiento de la hipótesis de independencia de la variación residual se toman logaritmos sobre el modelo multiplicativo, transformándolo en aditivo:

$$Ln\ Y(t) = Ln\ \tau(t) + Ln\ E(t) + Ln\ C(t) + Ln\ \varepsilon(t)$$

El modelo multiplicativo suele ser más adecuado que el aditivo cuando se describen series cronológicas de variables económicas.

5.2.2. Determinación del tipo de modelo.

Existen varios procedimientos para determinar el tipo de modelo al que corresponde una serie. Vamos a estudiar dos procedimientos basados en mostrar si las fluctuaciones de la serie son aproximadamente constantes o si se modifican en función del valor de la tendencia:

a) **Gráfico de la media-desviación típica:**
Se representa gráficamente el par (media, desviación típica) de cada año y si la nube de puntos resultante se distribuye en torno a una recta paralela al eje de abscisas, estamos ante un modelo aditivo; mientras que si las desviaciones típicas crecen claramente al aumentar las medias anuales, el modelo adecuado es el multiplicativo, pues es razonable pensar que la tendencia está multiplicada por las demás componentes.

b) **Análisis de la variabilidad de las diferencias y cocientes estacionales:**
Se observa una serie temporal en n años, de manera que las observaciones se realizan en m estaciones de dicho año (mensuales, trimestrales,…). Los pasos a seguir son los siguientes:

1. Se calculan las diferencias y cocientes estacionales, es decir, la diferencia y el cociente entre los datos de la misma estación *j* de dos años consecutivos:

$$d_{ij} = y_{ij} - y_{i-1,j}$$

$$k_{ij} = \frac{y_{ij}}{y_{i-1,j}}$$

para $i = 1,2,\ldots,n$ y $j = 1,\ldots,m$.

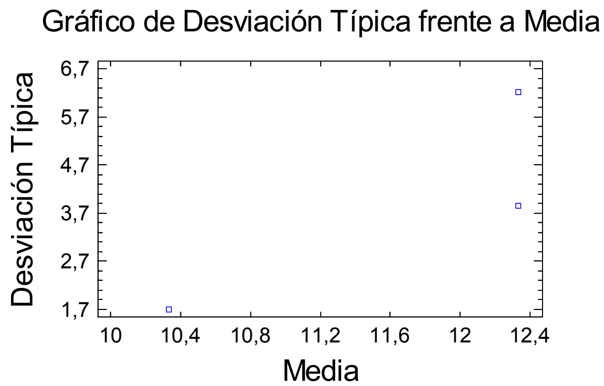
2. Se calculan los coeficientes de variación de Pearson para las diferencias y cocientes estacionales.
3. La selección del modelo se realiza comparando dichos coeficientes de variación:
   - Si $CV(d) < CV(k)$ se elige el modelo aditivo.
   - Si $CV(d) > CV(k)$ se elige el modelo multiplicativo.

Ejemplo 5.1. (continuación):

En el ejemplo anterior estudiamos qué modelo es más adecuado. Para ello, utilizaremos los dos procedimientos indicados:

a. Gráfico de la media-desviación típica:

Año\Cuatrimestre	1	2	3	Media anual	Des. Tip. anual
2008	12	8	11	10.33	1.70
2009	14	7	16	12.33	3.86
2010	13	9	15	12.33	6.22



Una vez observado el gráfico, podemos concluir que el modelo más adecuado es el multiplicativo, al aumentar las desviaciones típicas con el aumento de las medias.

b. Análisis de la variabilidad de las diferencias y cocientes estacionales:
Calculamos las diferencias y cocientes estacionales:

Diferencias estacionales				**Cocientes estacionales**			
Año\Cuatr.	**1**	**2**	**3**	**Año\Cuatr.**	**1**	**2**	**3**
2008				**2008**			
2009	2	-1	5	**2009**	1.17	0.875	1.45
2010	-1	2	-1	**2010**	0.93	1.29	0.94

$$\begin{cases} \overline{d} = 1 \\ S_d^{\ 2} = 5 \\ S_d = 2.236 \\ CV(d) = 2.236 \end{cases} \qquad \begin{cases} \overline{k} = 1.108 \\ S_k^{\ 2} = 0.0451 \\ S_k = 0.2124 \\ CV(k) = 0.192 \end{cases}$$

Como $CV(k) < CV(d)$, entonces es más adecuado el modelo multiplicativo.

5.3. Determinación de la Tendencia Secular.

Se consideran dos procedimientos para determinar la tendencia secular en una serie cronológica:

- Enfoque global: Se utiliza para las previsiones a largo plazo y consiste en el ajuste de una función matemática (mediante el método de mínimos cuadrados) para la obtención de la tendencia.
- Enfoque local: Se utiliza para las previsiones a corto plazo y utiliza parte de las observaciones para el estudio de la tendencia en cada momento (método de las medias móviles).

5.3.1. Método de Mínimos Cuadrados.

Queremos obtener una función que explique el comportamiento de la serie a largo plazo para captar aspectos más permanentes en la evolución de un fenómeno. Se presupone que la componente de mayor peso en las observaciones es la tendencia secular, suponiendo que tenemos un modelo solo con tendencia secular y variación aleatoria.

Tenemos que ajustar una función matemática que explique el funcionamiento de las observaciones Y_t en función del tiempo t. Para ello, utilizaremos el método de los mínimos cuadrados, ya estudiado en el capítulo 4. Vamos a utilizar el ajuste de una línea recta, pero debería utilizarse otro tipo de curva en la determinación de la tendencia si las características de la serie así lo requiriesen.

Representemos la tendencia mediante una función lineal:

$$Y_t = a + bt$$

Trabajaremos con datos anuales, ya que si fueran de otro tipo (mensuales, trimestrales,…), se calculan las medias anuales y se ajusta la recta o función a estas medias, eliminando así las oscilaciones debidas a la variación estacional.

Aplicando el método de los mínimos cuadrados sabemos que:

$$b = \frac{S_{yt}}{S_t^2} \text{ y } a = \bar{y} - b\bar{t}$$

Se pueden simplificar bastante los cálculos, haciendo un cambio de origen para la variable t:

- Si el número de años es impar, se trabaja con $t' = t - t_0$, siendo t_0 la media de los t_i.
- Si el número de años es par, se toma $t' = 2(t - t_0)$.

Se trabaja con $Y = a + bt'$ y, una vez calculados a y b, se tiene la recta:

- $Y = a + b\,(t - t_0)$, si el número de años es impar.
- $Y = a + 2b(t - t_0)$, si el número de años es par.

La pendiente de la recta (b) representa el incremento que tiene lugar en un año. Por lo tanto, cada mes la variable se incrementará $b/12$, cada trimestre $b/4$, cada cuatrimestre $b/3$,…

Este método permite predecir el comportamiento de la serie en el futuro. Para obtener dichas predicciones, basta con sustituir en la recta el valor del año para el que se quiere obtener la predicción. Si el periodo para el que se quiere obtener la predicción es inferior al año, se calcula la predicción para el año y a continuación para obtener la predicción para el primer mes, trimestre, cuatrimestre,… se le resta a ese valor la cantidad:

$$\frac{b}{k}\left(\frac{k}{2} - 1\right) + \frac{b}{2k}$$

donde b es la pendiente de la recta obtenida por el método de los mínimos cuadrados y k el número de meses, trimestres o cuatrimestres que tiene el año.

Para el resto de meses, trimestre, cuatrimestres,… se le suma al valor del primer mes predicho la cantidad $\frac{b}{k}$.

Ejemplo 5.1. (continuación):

En el ejemplo anterior, calculamos la tendencia secular utilizando el método de mínimos cuadrados:

Año\Cuatrimestre	**1**	**2**	**3**	**Media anual**
2008	12	8	11	10.33
2009	14	7	16	12.33
2010	13	9	15	12.33

Calculemos la ecuación de la recta $y = a+bt'$, siendo $t' = t - 2009$.

t'_i	y_i	$y_i \cdot t'_i$	t'^2_i
-1	10.33	-10.33	1
0	12.33	0	0
1	12.33	12.33	1

$$\bar{y} = 11.667\,,\ \bar{t'} = 0\,,\ S_{yt'} = 0.667\,,\ S_{t'}^{\,2} = 0.667$$

Por tanto, $b = 1$ y $a = 11.667$, y la recta queda:

$$y = 11.667 + t' = 11.667 + (t - 2009)$$

5.3.2. Método de las Medias Móviles.

Si estamos interesados en el estudio de la serie a corto plazo, el procedimiento idóneo para el estudio de la tendencia secular de la serie es el de las medias móviles, que es más flexible que el de mínimos cuadrados, pues no exige suponer una forma funcional para la tendencia.

Se trata de un método que consiste en sustituir la serie original por otra suavizada (de valores con menor fluctuación), que se toma como línea de tendencia, de manera que se reemplaza cada valor de la serie por otro que se obtiene como media aritmética de dicho valor y de un número de valores anteriores y posteriores a él. El número de valores adecuados (orden) para calcular dicha media aritmética viene dado por la forma de las observaciones, ya que pretendemos eliminar la variación estacional: si los datos se observan mensualmente se toma orden 12; si son trimestrales, orden 4; si son cuatrimestrales, orden 3,…

Este método tiene el inconveniente, de que no nos permite hacer predicciones, por lo que es menos utilizado que el método de los mínimos cuadrados.

A continuación detallaremos los pasos que hay que seguir para obtención de las medias móviles de orden o amplitud *h*:

1. Se forma un grupo con las primeras *h* observaciones y se calcula su media.
2. Se construyen los demás grupos excluyendo de cada grupo la primera observación del grupo anterior e incluyendo la primera del siguiente.
3. Se calcula la media de cada grupo.

Las medias obtenidas mediante este proceso se denominan medias móviles de orden o amplitud *h*.

Cada media móvil se asocia al punto medio del intervalo de tiempo sobre el que ha sido calculada.

Si el número de observaciones utilizado para calcular las medias es impar la serie obtenida estará centrada; si este número es par la serie obtenida estará descentrada y volveremos a hacer medias de orden dos, para obtener una serie centrada.

La línea que une estas medias recibe el nombre de línea de tendencia.

Ejemplo 5.1. (continuación):

En el ejemplo anterior, calculamos la tendencia secular utilizando el método de medias móviles:

Tomaremos orden 3, puesto que los datos vienen dados en cuatrimestres. La serie de valores de la tendencia es la siguiente:

Año\Cuatr.	1	2	3	Medias móviles	1	2	3
2008	12	8	11	**2008**		10.33	11
2009	14	7	16	**2009**	10.67	12.33	12
2010	13	9	15	**2010**	12.67	12.33	

5.4. Variación Estacional.

La variación estacional indica el aumento o disminución ocurrido en un periodo estacional concreto respecto del valor medio referido a todo el año. Nos interesa determinar las variaciones estacionales para eliminarlas del comportamiento general de la serie, de esta forma podemos conocer la evolución del fenómeno considerándolo aislado de posibles influencias estacionales.

- En el modelo multiplicativo, la componente estacional se mide mediante el índice adimensional denominado índice de variación estacional (*IVE*), que está expresado en tanto por ciento, y nos indica la fluctuación del valor de la serie respecto al valor de la tendencia media del año.
- En el modelo aditivo, la componente estacional indica la cantidad en que se ha superado o no ha alcanzado la tendencia media anual, y viene expresada en las mismas unidades que la variable observada.

Existen diversas técnicas para determinar la variación estacional.

5.4.1. Métodos para el cálculo de la Variación Estacional.

Vamos a estudiar dos métodos para calcular la variación estacional:

A) Método de la Razón (o diferencia) a la Tendencia:

El objetivo de este método consiste en eliminar de las observaciones el valor de la tendencia secular. En el modelo multiplicativo, esta eliminación la realizaremos mediante el cociente o razón:

$$\frac{Y(t)}{\tau(t)} = E(t)C(t)\varepsilon(t),$$

mientras que en el modelo aditivo, se realizará mediante la diferencia:

$$Y(t) - \tau(t) = E(t) + C(t) + \varepsilon(t)$$

A continuación, se promedian estos valores en cada una de las estaciones, de manera que los efectos de las variaciones cíclicas y aleatorias se anulan. La estimación para la tendencia se realiza mediante el ajuste de una recta por el método de mínimos cuadrados. Los pasos a seguir son los siguientes:

1. Si los datos no son anuales, se calculan las medias anuales.
2. Se estima la tendencia secular mediante el ajuste, por el método de mínimos cuadrados, de una recta a estos valores medios por año.
3. Se estiman los valores teóricos de la tendencia secular en cada una de las estaciones del año. Para ello, hay que tener en cuenta que:
   a) Tendencia: $\tau(t) = a + bt$
   b) Variación de la tendencia por año: b
   c) Variación de la tendencia por estación: b/s,
   donde s es el número de estaciones por año.

4. Se elimina de las observaciones el valor de la tendencia, realizando los cocientes o diferencias pertinentes, dependiendo del modelo considerado:
   a) Si el modelo es el multiplicativo, se calculan los cocientes
   $$\frac{Y(t)}{\tau(t)}$$
   b) Si el modelo es el aditivo, se calculan las diferencias
   $$Y(t)-\tau(t)$$
5. Se calcula la media por estación de los anteriores valores (E_j).
6. Se normalizan las medias anteriores, de manera que su media sea 1 en el modelo multiplicativo (dividiendo entre la media de todas las estaciones) ó 0 en el aditivo (restándole la media de todas las estaciones).

Ejemplo 5.1. (continuación):

En el ejemplo anterior, calculamos los índices de variación estacional:

- Modelo multiplicativo:

Año\Cuatrimestre	1	2	3
2008	12	8	11
2009	14	7	16
2010	13	9	15

Anteriormente, hemos calculado la ecuación de la recta.

$$y = 11.667 + t' = 11.667 + (t - 2009)$$

Calculamos el valor de la tendencia en cada año:

$$\begin{cases} y_{2008} = 11.667 + (2008 - 2009) = 10.667 \\ y_{2009} = 11.667 + (2009 - 2009) = 11.667 \\ y_{2010} = 11.667 + (2010 - 2009) = 12.667 \end{cases}$$

A partir de ellos, calculamos los valores teóricos de la tendencia, situando el valor estimado de la tendencia en la estación central y teniendo en cuenta que la variación estacional es $\frac{b}{s} = \frac{1}{3}$, obtenemos la siguiente tabla:

Año\Cuatrimestre	1	2	3
2008	10.333	10.667	11
2009	11.333	11.667	12
2010	12.333	12.667	13

Dividimos cada observación entre su tendencia teórica $\frac{Y(t)}{\tau(t)}$, calculamos la media en cada estación y normalizamos dividiendo entre la media de todos ellos:

Año\Cuatrimestre	1	2	3	Total
2008	1.161	0.750	1	
2009	1.235	0.600	1.333	
2010	1.054	0.710	1.154	
Media por estación (E_j)	1.15	0.687	1.162	0.99967
***IVE* normalizado**	115.04%	68.72%	116.24%	100%

- Modelo aditivo:
  Calculamos la diferencia de cada observación con su tendencia teórica $Y(t)-\tau(t)$, la media en cada estación y normalizamos, si es necesario, restándole la media total:

Año\Cuatrimestre	1	2	3	Total
2008	1.667	-2.667	0	
2009	2.667	-4.667	4	
2010	0.667	-3.667	2	
Media por estación (E_j)	1.667	-3.667	2	0
***IVE* normalizado**	1.667	-3.667	2	0

B) Método de la Razón (o diferencia) a las Medias Móviles:

Este método utiliza el suavizamiento de la serie proporcionado por las medias móviles.

Tenemos observaciones en n años y s estaciones en cada año y debemos seguir los siguientes pasos:

1. Se calcula la tendencia, aplicando el método de las medias móviles con amplitud s (si s es par, hay que centrar las medias móviles en cada estación), de manera que se eliminan las variaciones estacionales, por lo que la serie de las medias móviles solamente recoge el efecto de:
   - $\tau(t)C(t)$ en el modelo multiplicativo

- $\tau(t)+C(t)$ en el modelo aditivo.

2. Se aísla la componente estacional de la serie:
   - En el modelo multiplicativo:
$$\frac{Y(t)}{\tau(t)C(t)}=E(t)\varepsilon(t)$$
   - En el modelo aditivo:
$$Y(t)-\tau(t)+C(t)=E(t)+\varepsilon(t)\,.$$
3. Promediando, en ambos casos, se elimina la variación aleatoria $\varepsilon(t)$.
4. Finalmente, como en el modelo anterior, se normaliza la serie obtenida.

Se obtienen los correspondientes índices de variación estacional *IVE*, que nos indican el porcentaje de aumento o disminución , con relación a la tendencia, que corresponde a los valores de la serie en un determinado mes, trimestre…

Ejemplo 5.1. (continuación):

En el ejemplo anterior, calculamos los *IVE* utilizando el método de medias móviles. Tomamos orden 3, puesto que los datos vienen dados en cuatrimestres. La serie de valores de la tendencia es la siguiente:

Año\Cuatr.	1	2	3	Medias móviles	1	2	3
2008	12	8	11	**2008**		10.33	11
2009	14	7	16	**2009**	10.67	12.33	12
2010	13	9	15	**2010**	12.67	12.33	

Resolvemos el problema utilizando los dos modelos estudiados:

- Modelo multiplicativo:
  Dividimos cada valor de la serie original por su correspondiente media móvil, calculamos la media de cada cuatrimestre y la normalizamos dividiendo entre la media total, obteniendo la siguiente tabla:

Año\Cuatrimestre	1	2	3	Total
2008		0.774	1	
2009	1.312	0.568	1.333	
2010	1.026	0.730		
Media por estación (E_i)	1.169	0.690	1.167	1.0087
***IVE* normalizado**	115.7	68.4	115.7	300

- Modelo aditivo:
  Restamos a cada valor de la serie original su correspondiente media móvil, calculamos la media de cada cuatrimestre y la normalizamos restándole la media total, obteniendo la siguiente tabla:

Año\Cuatrimestre	1	2	3	Total
2008		-2.333	0	
2009	3.333	-5.333	4	
2010	0.333	-3.333		
Media por estación (E_j)	1.833	-3.667	2	0.055
***IVE* normalizado**	1.778	-3.722	1.945	0

C) Desestacionalización de una serie:

La determinación de la variación estacional es un aspecto importante en el análisis de una serie cronológica. En ocasiones nos interesa conocer las variaciones estacionales y eliminarlas del comportamiento global para observar mejor el movimiento de ésta ajeno a causas estacionales. A la eliminación de la componente estacional de la serie se le denomina desestacionalización y a partir de aquí se hacen comparables cantidades recogidas en distintas estaciones.

El método de desestacionalización depende del tipo de modelo seleccionado:

- Si el modelo es el multiplicativo, se obtiene la serie desestacionalizada dividiendo los valores originales entre el correspondiente índice de variación original en tanto por uno:
  $$\frac{Y(t)}{E(t)} = \tau(t)C(t)$$
  y a estos valores se les denomina componente extraestacional, y recogen el valor que presentaría la serie si no estuviera afectada por los periodos estacionales.
- Si el modelo es aditivo, la desestacionalización se realiza restando a la observación la correspondiente diferencia estacional, de manera que la componente extraestacional es del tipo:
  $$Y(t) - E(t) = \tau(t) + C(t)$$

Ejemplo 5.1. (continuación):

En el ejemplo anterior, desestacionalizamos la serie:

Año\Cuatrimestre	1	2	3
2008	12	8	11
2009	14	7	16
2010	13	9	15

- Modelo multiplicativo:
  Dividimos los valores observados entre el IVE normalizado, para obtener la serie desestacionalizada:

Año\Cuatrimestre	1	2	3
2008	10.431	11.813	9.463
2009	12.170	10.337	13.765
2010	11.300	13.290	12.904
***IVE* normalizado**	115.04%	68.72%	116.24%

- Modelo aditivo:
  Restamos a los valores observados el IVE normalizado, para obtener la serie descentralizada:

Año\Cuatrimestre	1	2	3
2008	10.333	11.667	9
2009	12.333	10.667	14
2010	11.333	2.667	13
***IVE* normalizado**	1.667	-3.667	2

5.5. Predicción.

El valor de la serie cronológica sería conocido en cualquier momento si se conoce el valor de las cuatro componentes de la serie y el modelo que las relaciona. Pero la componente aleatoria siempre es desconocida y, del resto de componentes, se puede obtener una estimación, por lo que es imposible obtener el valor exacto de la serie en un instante futuro y se busca una predicción lo más próxima posible a partir de la tendencia secular y la variación estacional, no teniendo en cuenta la variación cíclica:

- En el modelo aditivo, la predicción en un instante futuro t viene dada por:

$$\hat{Y}(t) = \tau(t) + E(t)$$

- En el modelo multiplicativo, la predicción en un instante futuro t viene dada por:

$$\hat{Y}(t) = \tau(t)E(t)$$

Ejemplo 5.1. (continuación):

En el ejemplo anterior, hacemos la predicción para el año 2012.

Utilizaremos la recta de mínimos cuadrados obtenida para estimar la tendencia secular:

$$y_{2012} = 11.667 + (t - 2009) = 14.667$$

Para estimar su valor en cada cuatrimestre, utilizamos los índices de variación estacional del método de la razón o diferencia a la tendencia en cada modelo:

- Modelo multiplicativo:

Año\Cuatrimestre	**1**	**2**	**3**
2012	14.667·1.1504= 16.873	14.667·0.6872= 10.079	14.667·1.1624= 17.049
***IVE* normalizado**	115.04%	68.72%	116.24%

- Modelo aditivo:

Año\Cuatrimestre	**1**	**2**	**3**
2012	14.667+1.667= 16.334	14.667-3.667= 11	14.667-2= 12.667
***IVE* normalizado**	1.667	-3.667	2

RELACIÓN DE SERIES CRONOLÓGICAS.

1.- La siguiente tabla representa el gasto de una empresa (en miles de euros) en los años 2006, 2007, 2008:

	Trimestre 1º	Trimestre 2º	Trimestre 3º	Trimestre 4º
2006	12	16	11	25
2007	14	19	13	26
2008	16	19	12	32

a) Estudiar la evolución del gasto de la empresa durante esos años.
b) Estimar la previsión del gasto para el año 2109. ¿Es fiable esa predicción?
c) Obtener la serie desestacionalizada utilizando el método de las medias móviles y el de los mínimos cuadrados.

2.- La siguiente tabla representa el precio de las acciones de determinada empresa, para cada cuatrimestre del periodo 2001-2004.

	Cuatrimestre 1º	Cuatrimestre 2º	Cuatrimestre 3º
2007	24,8	25	25.2
2008	25	25.5	25.3
2009	25,5	25.4	26
2010	25,8	26.2	26.5

a) ¿Qué modelo es el más adecuado?
b) ¿En qué época del año sería más beneficioso comprar acciones?, ¿y venderlas?
c) Estimar el precio de la acción para cada cuatrimestre del año 2012.

3.- La entrada de turistas en una localidad determinada, entre los años 2007 y 2010, y según las estaciones del año son:

	Invierno	Primavera	Verano	Otoño
2007	5520	8190	18247	6369
2008	5022	8606	19541	6961
2009	5616	9552	19670	7167
2010	5721	9412	18911	7220

a) Estudiar cómo se distribuyen los turistas a lo largo del año.

b) Predecir el número de turistas que entraran en cada época del año 2011.
c) ¿Es fiable esa predicción?.
d) Desestacionalizar la serie, utilizando el método de las medias móviles y el de los mínimos cuadrados.

4.- Las ventas de motocicletas (en miles) en un país han sido las siguientes:

	2005	**2006**	**2007**	**2008**	**2009**
Cuatrimestre1º	26	26	25	25	24
Cuatrimestre2º	52	53	53	52	51
Cuatrimestre3º	22	23	23	23	24

a) Calcular los índices de variación estacional de cada cuatrimestre por el método de la razón a la media móvil, bajo el supuesto del esquema multiplicativo.
b) Desestacionalizar la serie.

5.- Queremos estudiar la evolución de las ventas de una determinada empresa de prendas de abrigo durante los años comprendidos entre 2007 y 2010 y predecir lo que pasaría para cada cuatrimestre del año 2011. Para ello utilizamos la siguiente tabla (el número de prendas viene dado en miles):

	2007	**2008**	**2009**	**2010**
Cuatrimestre 1º	16	19	24	31
Cuatrimestre 2º	19	26	34	45
Cuatrimestre 3º	24	31	41	54

¿Qué modelo sería el más adecuado para este estudio?

6.- Una empresa de automóviles presenta las siguientes cifras de ventas (en miles de unidades)

	2007	**2008**	**2009**	**2010**
Primavera	2	2,2	2,2	2,4
Verano	3,1	3	3,5	3,6
Otoño	2,6	2,8	4,3	4,5
Invierno	1,8	2	2,1	2,2

a) ¿En qué época del año se venden mayor número de vehículos? Utilizar el método multiplicativo.

b) ¿Cuál hubiera sido el número de vehículos vendidos si la influencia de la estación no hubiera existido en esos años?
c) Estimar el número de automóviles vendidos para el año 2011.

CAPÍTULO 6:

NÚMEROS ÍNDICES.

En general, las magnitudes socioeconómicas varían en el espacio y/o en el tiempo. Muchas veces surge la necesidad de hacer comparaciones en el espacio y/o en el tiempo, tanto por separado como por grupos, de las mismas. Para ello, se elaboran unos indicadores socioeconómicos, denominados números índices, que dan respuesta, por ejemplo, a cuestiones como si la coyuntura económica es positiva o negativa. Se utiliza mucho en Economía, aunque su uso se extiende cada vez más al resto de las Ciencias Sociales.

6.1. Índice Elemental.

Los números índices simples o elementales surgen cuando se estudia la evolución a lo largo del tiempo de una magnitud que tiene una única componente.

Se considera la evolución temporal de una magnitud G con una sola componente (por ejemplo, el precio de la gasolina), y sean G_0, $G_1,\ldots,G_t,\ldots$ los valores de G en los instantes sucesivos 0, $1,\ldots,t,\ldots$ Para describir la evolución de la magnitud G, se calcula el crecimiento que experimenta en cada instante con relación a un instante temporal de referencia, denominado periodo base del índice y cuya elección es arbitraria, y que representa un momento temporal de referencia que se toma como origen de las comparaciones.

Se denomina índice elemental de la magnitud simple *G* en el periodo corriente o actual *t* respecto al periodo base (0) al siguiente cociente:

$$I_{t/0}(G) = \frac{G_t}{G_0}$$

El índice de una magnitud simple *G* es un número sin dimensión, que permite comparar la evolución de una misma magnitud en dos periodos y la evolución de dos o más magnitudes, de naturaleza diferente, medidas en unidades diferentes y en un mismo periodo.

Ejemplo 6.1.

El presupuesto de un club de fútbol durante los años 2010 y 2011 es:

Año	Presupuesto
2010	1.200.000
2011	1.150.000

Para calcular el aumento o descenso del presupuesto:

$$I_{2011/2010}(G) = \frac{G_{2011}}{G_{2010}} = \frac{1150000}{1200000} = 0.958333$$

El presupuesto en 2011 es el 95.8333% del presupuesto de 2010, es decir, ha disminuido un 4.1666%

6.1.1. Propiedades de un Índice Elemental.

De la definición del índice elemental, se obtienen las siguientes propiedades:

a) Identidad:

Si hacemos coincidir el periodo base con el de comparación, el número índice vale 1:

$$I_{0/0}(G) = 1$$

b) Circular

$$I_{t/0}(G) = I_{t/t'}(G) I_{t'/0}(G)$$

Esta propiedad permite comparar, no sólo los periodos 0 y *t* de una parte y 0 y t' de otra, sino también *t* y t', realizándose la comparación de manera independiente a la elección del periodo de referencia 0:

$$I_{t/t'}(G) = \frac{I_{t/0}(G)}{I_{t'/0}(G)}$$

c) Inversión:

$$I_{0/t}(G) = \frac{1}{I_{t/0}(G)}$$

Esta propiedad se utiliza cuando tenemos un criterio distinto al tiempo.

d) Encadenamiento

$$I_{t/0}(G) = I_{t/t-1}(G) I_{t-1/t-2}(G) \cdot \ldots \cdot I_{1/0}$$

El índice del periodo *t* respecto al periodo 0 se obtiene mediante el producto de los índices de un período respecto al período anterior.

e) Multiplicación

El índice elemental de un producto es el producto de los índices elementales.

$$I_{t/0}(AB) = I_{t/0}(A) I_{t/0}(B)$$

f) División

El índice elemental de un cociente es el cociente de los índices elementales.

$$I_{t/0}\left(\frac{A}{B}\right) = \frac{I_{t/0}(A)}{I_{t/0}(B)}$$

6.2. Índice Compuesto o Sintético.

Los índices sintéticos nos permiten estudiar la evolución conjunta de los precios de varios productos, como por ejemplo, el crecimiento de los precios de los productos alimenticios en España en los últimos años.

Se considera una magnitud compleja G constituida por n magnitudes simples G^1, $G^2, \ldots, G^n$ (por ejemplo, los precios de un conjunto de productos alimenticios). A partir de estas componentes, se pueden construir, para el periodo t, sus correspondientes índices simples:

$$I_{t/0}\left(G^i\right) = \frac{G_t^i}{G_0^i}, \quad i = 1,2,\ldots,n$$

El objetivo de los índices sintéticos es sintetizar en un índice único los índices elementales de las componentes de G. Para ello, hallamos un promedio de dichos índices.

Para simplificar la notación vamos a denominar al periodo t como periodo 1.

Existen dos grandes bloques de números índices sintéticos, los ponderados y los que no utilizan ponderación. Los ponderados surgen cuando a cada uno de los componentes de la magnitud compleja estudiada se le asigna un determinado coeficiente de ponderación u^i, asociado a su importancia, de manera que:

$$\sum_i u_0^i = \sum_i u_1^i = 1$$

Los principales índices son:

a) Índice de Bradstreet y Dudot:

Este índice es una media aritmética y viene dado por:

$$D_{1/0}(G) = \frac{\sum_i G_1^i}{\sum_i G_0^i}$$

b) Índice de Sauerbeck:

Es la media aritmética de índices simples con la misma ponderación 1/*n*, es decir:

$$S_{1/0}(G) = \frac{1}{n}\sum_i I_{1/0}(G^i) = \frac{1}{n}\sum_i \frac{G_1^i}{G_0^i}$$

c) Índice de Laspeyres:

Es la media aritmética de los índices elementales con factores de ponderación de período base, y viene dado por:

$$L_{1/0}(G) = \sum_i I_{1/0}(G^i)u_0^i = \sum_i \frac{G_1^i}{G_0^i}u_0^i$$

d) Índice de Paasche:

Es la media armónica de los índices elementales con factores de ponderación en el periodo actual.

$$\frac{1}{P_{1/0}(G)} = \sum_i \frac{u_1^i}{I_{1/0}(G^i)} = \sum_i \frac{G_0^i}{G_1^i}u_1^i$$

e) Índice de Fischer

Es la media geométrica de los índices de Laspeyres y Paasche.

$$F_{1/0}(G) = \sqrt{L_{1/0}(G)P_{1/0}(G)}$$

Este índice presenta algunos problemas en su interpretación:

- No es muy claro el sistema de ponderaciones de los índices elementales que utiliza.
- No está claro a qué responde este índice, excepto que es la media geométrica de los dos índices.

Ejemplo 6.2.

En la siguiente tabla tenemos el precio y la ponderación considerada de dos productos, patatas y tomates, en los años 2008 y 2011:

	Patatas		Tomates	
Año	**Precio**	**Ponderación**	**Precio**	**Ponderación**
2008	63	60	51	40
2011	70	55	48	45

Los valores de los distintos índices son:

$$D_{2011/2008}(G) = \frac{\sum_i G_1^i}{\sum_i G_0^i} = \frac{70+48}{63+51} = 1.0351$$

$$S_{1/0}(G) = \frac{1}{n}\sum_i \frac{G_1^i}{G_0^i} = \frac{1}{2}\left(\frac{70}{63}+\frac{48}{51}\right) = 1.0261$$

$$L_{1/0}(G) = \sum_i \frac{G_1^i}{G_0^i}u_0^i = 0.6\frac{70}{63}+0.4\frac{48}{51} = 1.0431$$

$$\frac{1}{P_{1/0}(G)} = \sum_i \frac{G_0^i}{G_1^i}u_1^i = 0.55\frac{63}{70}+0.45\frac{51}{48} = 0.973$$

$$P_{1/0}(G) = \frac{1}{0.973} = 1.0276$$

$$F_{1/0}(G) = \sqrt{L_{1/0}(G)P_{1/0}(G)} = \sqrt{1.0431 \cdot 1.0276} = 1.0353$$

6.2.1. Propiedades de un Índice Sintético.

Sus propiedades son las siguientes:

1) El índice de Fischer está entre el de Laspeyres y el de Paasche, ya que es su media geométrica.

2) Los índices de Sauerbeck, Bradstreet-Dudot, Laspeyres y Paasche están entre los índices elementales extremos, ya que son medias de los índices elementales.

3) Todos los índices sintéticos son iguales cuando lo sean los índices elementales.

4) El índice de Paasche suele ser inferior al índice de Laspeyres. Si la ponderación fuese igual, el índice de Paasche sería siempre menor que el de Laspeyres, ya que estaríamos ante las medias armónica y aritmética de unos mismos datos.

5) El índice la Laspeyres tiene la ventaja sobre el de Paasche de exigir para el cálculo de las ponderaciones una información menos actualizada.

6) Los índices de Sauerbeck, Laspeyres, Paasche y Fischer no cumplen la propiedad circular, mientras que el de Bradstreet-Dudot sí lo hace.

7) Los índices de Sauerbeck, Laspeyres y Paasche no cumplen la propiedad de inversión, mientras que los de Fischer y Bradstreet-Dudot sí lo hacen.

8) Los índices de Laspeyres y Paasche poseen una propiedad de agregación: el índice de Laspeyres de un conjunto es igual al índice de Laspeyres de los índices de Laspeyres de cada subconjunto de componentes:

$$L_{1/0}(G) = \sum_i u_0^i L_{1/0}(G^i)$$

Análogo con el índice de Paasche.

6.2.2. Índices de Precios y Cantidades.

Consideramos la información relativa a precios y cantidades de los periodos base y actual de un conjunto de n artículos, donde tanto los precios como las cantidades tienen un amplio significado (las cantidades pueden ser consumidas, producidas, compradas, vendidas,…) y llamaremos presupuesto a cada conjunto de cantidades del periodo base o periodo actual:

$$\text{Periodo Base}\begin{cases} \text{Pr}\,ecios: & p_{10} \quad \cdots \quad p_{i0} \quad \cdots \quad p_{n0} \\ Cantidades: & q_{10} \quad \cdots \quad q_{i0} \quad \cdots \quad q_{n0} \end{cases}$$

$$\text{Periodo Actual}\begin{cases} \text{Pr}\,ecios: & p_{11} \quad \cdots \quad p_{i1} \quad \cdots \quad p_{n1} \\ Cantidades: & q_{11} \quad \cdots \quad q_{i1} \quad \cdots \quad q_{n1} \end{cases}$$

El valor de cada una de las cantidades de un artículo i viene dado por $p_{i0}q_{i0}$ para el periodo base y por $p_{i1}q_{i1}$ para el periodo actual donde $i = 1,\ldots,n$. Por tanto, el valor global del presupuesto es:

$$\begin{cases} Periodo\ Base: & \sum_i p_{i0}q_{i0} \\ Periodo\ Actual: & \sum_i p_{i1}q_{i1} \end{cases}$$

Se utiliza como factor de ponderación el valor relativo de cada artículo respecto del valor global del presupuesto, es decir:

$$\left\{\begin{array}{l} u_0^i = \dfrac{p_{i0}q_{i0}}{\sum_i p_{i0}q_{i0}} \\ u_1^i = \dfrac{p_{i1}q_{i1}}{\sum_i p_{i1}q_{i1}} \end{array}\right. \text{ para } i = 1,\ldots,n.$$

A partir de aquí, obtenemos los índices de Laspeyres y de Paasche para precios y cantidades.

A) Precios.

Obtenemos los índices de precios de Laspeyres y Paasche:

1) Índice de Laspeyres:

Es el cociente entre el presupuesto del periodo base a precios del periodo actual y el presupuesto del periodo base a precios del periodo base.

$$L_{1/0}^p = \sum_i u_0^i \frac{p_{i1}}{p_{i0}} = \sum_i \frac{p_{i0}q_{i0}}{\sum_i p_{i0}q_{i0}} \frac{p_{i1}}{p_{i0}} = \frac{\sum_i p_{i1}q_{i0}}{\sum_i p_{i0}q_{i0}}$$

2) Índice de Paasche:

Es el cociente entre el presupuesto del periodo actual a precios del periodo actual y el presupuesto del periodo actual a precios del periodo base.

$$\frac{1}{P_{1/0}^p} = \sum_i u_1^i \frac{p_{i0}}{p_{i1}} = \sum_i \frac{p_{i1}q_{i1}}{\sum_i p_{i1}q_{i1}} \frac{p_{i0}}{p_{i1}} = \frac{\sum_i p_{i0}q_{i1}}{\sum_i p_{i1}q_{i1}}$$

$$P_{1/0}^p = \frac{\sum_i p_{i1}q_{i1}}{\sum_i p_{i0}q_{i1}}$$

Claramente el índice de Paasche exige una información más actualizada para su cálculo que el de Laspeyres.

Un caso especial es cuando los presupuestos del periodo base y actual coinciden, es decir:

$$q_{i0} = q_{i1} \text{ para } i = 1,\ldots,n$$

En este caso, si el conjunto de artículos y sus cantidades se considera representativo de un determinado nivel de vida, a este índice de precios se le llama índice de coste de vida o índice de precios al consumo (IPC) y al conjunto de artículos y cantidades se le denomina cesta de la compra.

B) Cantidades.

Obtenemos los índices de cantidades de Laspeyres y Paasche:

1) Índice de Laspeyres:

Es el cociente entre el presupuesto del periodo actual a precios del periodo base y el presupuesto del periodo base a precios del periodo base.

$$L_{1/0}^{q} = \sum_i u_0^i \frac{q_{i1}}{q_{i0}} = \sum_i \frac{p_{i0}q_{i0}}{\sum_i p_{i0}q_{i0}} \frac{q_{i1}}{q_{i0}} = \frac{\sum_i q_{i1}p_{i0}}{\sum_i q_{i0}p_{i0}}$$

2) Índice de Paasche:

Es el cociente entre el presupuesto del periodo actual a precios del periodo actual y el presupuesto del periodo base a precios del periodo actual.

$$\frac{1}{P_{1/0}^{q}} = \sum_i u_1^i \frac{q_{i0}}{q_{i1}} = \sum_i \frac{p_{i1}q_{i1}}{\sum_i p_{i1}q_{i1}} \frac{q_{i0}}{q_{i1}} = \frac{\sum_i q_{i0}p_{i1}}{\sum_i q_{i1}p_{i1}}$$

$$P_{1/0}^{q} = \frac{\sum_i q_{i1}p_{i1}}{\sum_i q_{i0}p_{i1}}$$

Tanto para cantidades como para precios, el índice de Fisher es la media geométrica de los índices de Laspeyres y Paasche.

En la siguiente tabla aparecen todos los cálculos necesarios para obtener los índices de precios y cantidades de Laspeyres y Paasche:

Año Base		Año Actual					
p_{i0}	q_{i0}	p_{i1}	q_{i1}	$p_{i0}\, q_{i0}$	$p_{i1}\, q_{i0}$	$p_{i0}\, q_{i1}$	$p_{i1}\, q_{i1}$
p_{10}	q_{10}	p_{11}	q_{11}	$p_{10}\, q_{10}$	$p_{11}\, q_{10}$	$p_{10}\, q_{11}$	$p_{11}\, q_{11}$
…	…	…	…	…	…	…	…
p_{n0}	q_{n0}	p_{n1}	q_{n1}	$p_{n0}\, q_{n0}$	$p_{n1}\, q_{n0}$	$p_{n0}\, q_{n1}$	$p_{n1}\, q_{n1}$
				Suma 1	**Suma 2**	**Suma 3**	**Suma 4**

$$L_{1/0}^{p} = \frac{Suma\,2}{Suma\,1} \quad L_{1/0}^{q} = \frac{Suma\,3}{Suma\,1} \quad P_{1/0}^{p} = \frac{Suma\,4}{Suma\,3} \quad P_{1/0}^{p} = \frac{Suma\,2}{Suma\,1}$$

- Una nueva propiedad de los índices sintéticos, considerando precios y cantidades, es:

$$L^{p} P^{q} = L^{q} P^{p} = F^{q} F^{p}$$

Ejemplo 6.3.

Consideramos precios y cantidades de tres productos durante los años 2009 y 2011:

2009		**2011**	
p_{i0}	q_{i0}	p_{i1}	q_{i1}
1000	10	1100	15
500	20	600	25
400	50	500	40

Completamos la tabla para poder realizar todos los cálculos:

	2009		**2011**					
	p_{i0}	q_{i0}	p_{i0}	q_{i0}	$p_{i0}\,q_{i0}$	$p_{i1}\,q_{i0}$	$p_{i0}\,q_{i1}$	$p_{i1}\,q_{i1}$
Jamón	1000	10	1000	10	10000	11000	15000	16500
Aceite	500	20	500	20	10000	12000	12500	15000
Pollo	400	50	400	50	20000	25000	16000	20000
					40000	**48000**	**43500**	**51500**

$$L_{11/09}^{p} = \frac{48000}{40000} = 1.2 \qquad L_{11/09}^{q} = \frac{43500}{40000} = 1.0875$$

$$P_{11/09}^{p} = \frac{51500}{43500} = 1.1839 \qquad P_{11/09}^{p} = \frac{51500}{48000} = 1.0729$$

$$F_{11/09}^{p} = \sqrt{1.2 \cdot 1.1839} = 1.1919 \qquad F_{11/09}^{q} = \sqrt{1.0875 \cdot 1.0729} = 1.08$$

6.3. Enlace de Series de Números Índices con distinta Base.

Con frecuencia tenemos dos series de números índices, relativos al mismo fenómeno, pero construidos con distinto periodo base. Para tener una única serie y facilitar comparaciones, necesitamos un procedimiento para unirlas, obteniendo una sola serie con todos los números índices construidos con el mismo periodo base.

El problema es que, como se vio en la propiedad 6 de los índices sintéticos, los principales índices no cumplen la propiedad circular. Pese a ello, trabajaremos como si se verificase dicha propiedad.

Consideramos la siguiente situación:

- Tenemos una serie de números índices con base en t_1.
- Tenemos otra serie de números índices con base en t_2, donde $t_1 < t_2$.
- Conocemos el valor del índice en t_2 respecto de t_1.

Si pretendemos obtener una única serie con base en t_2 , solo tenemos que modificar la primera serie, teniendo en cuenta la propiedad:

$$I_{t/t_1} = I_{t/t_2} I_{t_2/t_1} \text{ o, equivalentemente, } I_{t/t_2} = \frac{I_{t/t_1}}{I_{t_2/t_1}}$$

Ejemplo 6.4.

Tenemos la serie del IPC una parte con base en 2004 y otra con base 2007. Vamos a obtenerla con ambas bases:

Años	IPC (base 2004)	IPC (base 2007)
2004	**1.00**	0.5556
2005	**1.15**	0.6389
2006	**1.35**	0.75
2007	**1.80**	**1.00**
2008	1.89	**1.05**
2009	2.16	**1.20**
2010	2.70	**1.50**

$$I_{04/07} = \frac{1.00}{1.80} = 0.5556 \quad I_{05/07} = \frac{1.15}{1.80} = 0.6389$$

$$I_{06/07} = \frac{1.35}{1.80} = 0.75 \quad I_{08/07} = 1.05 \cdot 1.80 = 1.89$$

$$I_{09/07} = 1.20 \cdot 1.80 = 2.16 \quad I_{10/07} = 1.50 \cdot 1.80 = 2.7$$

6.4. Índices de Valor.

Vamos a obtener un número índice que recoja la variación en el valor agregado de fenómenos caracterizados por un conjunto de precios y cantidades que varían a lo largo del tiempo, como, por ejemplo, la producción.

Sean:

- $V_0 = \sum_i p_{i0} q_{i0}$ el valor en el año base, y
- $V_1 = \sum_i p_{i1} q_{i1}$ el valor en el año actual

Llamaremos índice de valor agregado al cociente:

$$\frac{V_1}{V_0} = \frac{\sum_i p_{i1} q_{i1}}{\sum_i p_{i0} q_{i0}}$$

Este número índice se puede obtener también como un producto de un índice de precios por un índice de cantidad:

$$\frac{V_1}{V_0} = \frac{\sum_i p_{i1} q_{i1}}{\sum_i p_{i0} q_{i1}} \frac{\sum_i p_{i0} q_{i1}}{\sum_i p_{i0} q_{i0}} = P_{1/0}^{p} L_{1/0}^{q}$$

$$\frac{V_1}{V_0} = \frac{\sum_i p_{i1} q_{i1}}{\sum_i p_{i1} q_{i0}} \frac{\sum_i p_{i1} q_{i0}}{\sum_i p_{i0} q_{i0}} = P_{1/0}^{q} L_{1/0}^{p}$$

Ejemplo 6.3. (continuación)

Calculamos los índices de valor y el índice de valor agregado:

$$V_0 = \sum_i p_{i0} q_{i0} = 40000 \quad V_1 = \sum_i p_{i1} q_{i1} = 51500$$

$$\frac{V_1}{V_0} = P_{1/0}^{p} L_{1/0}^{q} = P_{1/0}^{q} L_{1/0}^{p} = \frac{51500}{40000} = 1.1839 \cdot 1.0875 = 1.0729 \cdot 1.2 = 1.2875$$

6.5. Deflación de Series Económicas.

Una de las funciones del dinero es servir como unidad de cuentas, pues nos permite expresar la producción total valorada en dinero. Pero cuando se comparan magnitudes económicas en distintos periodos de tiempo, debido a las alteraciones de los precios, a las unidades monetarias de los distintos periodos les corresponde un poder adquisitivo distinto. Para poder comparar dichas cantidades hay que homogeneizarlas en el sentido de que los valores estén expresados en los mismos precios.

El procedimiento utilizado para obtener la homogeneización se denomina deflación y consiste en dividir los valores de la serie económica por un índice de precios adecuado, denominado deflactor. No existe un deflactor de validez universal, cada fenómeno exige un deflactor adecuado.

Examinemos las posibilidades como deflactores, de los índices de precios de Laspeyres y Paasche:

- Consideremos valores agregados a precios corrientes:

$$V_0 = \sum_i p_{i0} q_{i0} \quad V_1 = \sum_i p_{i1} q_{i1}$$

- Deflectamos V_1 con el índice de Laspeyres:

$$\frac{V_1}{L^p_{1/0}} = \frac{\sum_i p_{i1} q_{i1}}{\dfrac{\sum_i p_{i1} q_{i0}}{\sum_i p_{i0} q_{i0}}} = \left(\sum_i p_{i0} q_{i0}\right) \frac{\sum_i p_{i1} q_{i1}}{\sum_i p_{i1} q_{i0}} = V_0 P^q_{1/0}$$

- Deflectamos V_1 con el índice de Paasche:

$$\frac{V_1}{P^p_{1/0}} = \frac{\sum_i p_{i1} q_{i1}}{\dfrac{\sum_i p_{i1} q_{i1}}{\sum_i p_{i0} q_{i1}}} = \sum_i p_{i0} q_{i1}$$

En el segundo caso se obtiene el valor de la producción actual a precios del periodo base, por lo que el índice de Paasche es el deflactor adecuado; aunque frecuentemente se utiliza el índice de Lasperyres por ser el único disponible.

Ejemplo 6.5.

Consideramos el presupuesto de una compañía en millones de € corrientes y los Índices de Precios al Consumo en el periodo 2004-2010 y calculamos el porcentaje, en términos reales, en que ha variado el presupuesto en enseñanza en el periodo de tiempo considerado:

Años	Presupuesto	IPC (Base 2004)	IPC (Base 2007)
2004	150	1.00	
2005	230	1.15	
2006	240	1.35	
2007	290	1.80	1.00
2008	300		1.05
2009	330		1.20
2010	400		1.50

Para obtener el presupuesto en enseñanza en términos reales hay que deflactar el dinero corriente, para transformar los euros corrientes de cada año en euros del año base utilizado.
En primer lugar, como tenemos el IPC relativo a dos años bases distintos, enlazamos ambas series de índices, mediante las expresiones:

$$I_{t/2004} = I_{t/2007} I_{2007/2004}\text{, o bien, } I_{t/2007} = \frac{I_{t/2004}}{I_{2007/2004}}$$

Años	IPC (Base 2004)	IPC (Base 2007)
2004	**1.00**	1/1.80 = 0.5555
2005	**1.15**	1.15/1.80 = 0.6389
2006	**1.35**	1.35/1.80 = 0.75
2007	**1.80**	**1.00**
2008	1.05·1.80 = 1.89	**1.05**
2009	1.20·1.80 = 2.16	**1.20**
2010	1.5·1.8 = 2.70	**1.50**

Si expresamos el presupuesto en € de 2004 y de 2007, tenemos:

Años	Presupuesto en € de 2004	Presupuesto en € de 2007
2004	150/1 = 150	150/0.5555 = 270
2005	230/1.15 = 200	230/0.6389 = 359.94
2006	240/1.35 = 177.78	240/0.75 = 320
2007	290/1.80 = 161.11	290/1 = 290
2008	300/1.89 = 158.73	300/1.05 = 285.71
2009	330/2.16 = 152.77	330/1.20 = 275
2010	400/2.7 = 148.15	400/1.50 = 266.66

- En euros de 2004, ha habido una disminución absoluta en términos reales de:

  148.15-150 = 1.852 millones de €,

  lo que supone una disminución porcentual del:

  1.852/150 · 100 = 1.2345%
- En euros de 2007, ha habido una disminución absoluta en términos reales de:

  266.666-270 = 3.333 millones de €,

  lo que supone una disminución porcentual del:

  3.333/270 · 100 = 1.2345% ,

  que, por supuesto coincide con la obtenida respecto de 2004.

6.6. Dependencia de un Índice General de un grupo de productos.

Vamos a ver cómo afecta a un índice general la variación de uno o un grupo de los artículos considerados en su construcción. Consideramos la expresión del índice de Laspeyres como una suma ponderada de índices simples o de índices de grupos, en la que destacamos el periodo actual:

$$\tilde{I}_t = \sum_i I(i,t)u^i \ ,$$

donde el índice i se refiere al artículo o grupo concreto.

La variación del índice en el periodo de t-1 a t es

$$\Delta\tilde{I}_t(t-1,t) = \sum_i \left(I(i,t) - I(i,t-1)\right)u^i = \sum_i \Delta I(i;t-1,t)u^i$$

Diremos que la repercusión del artículo o grupo i en la variación del índice general es:

$$\frac{\Delta I(i;t-1,t)u^i}{\sum_i \Delta I(i;t-1,t)u^i}100 \quad \text{para } i = 1,2,\ldots$$

Claramente la suma de repercusiones extendida a todos los artículos o todos los grupos es 100.

6.7. Índice de Precios al Consumo.

El Índice de Precios al Consumo (IPC) se calcula mensualmente por el Instituto Nacional de Estadística (INE) y refleja la evolución de los precios para la familia media española. Es un índice de Laspeyres que se realiza atendiendo a los precios de un determinado número de artículos que están asociados en diversos grupos. Se determina el índice de precios para cada uno de estos grupos y, posteriormente utilizando la propiedad de agregación, se calcula el IPC general. Por tanto, compara el valor de cada mes de un conjunto de artículos con su valor en el periodo de referencia, el cual se modifica aproximadamente cada diez años, para tener en cuenta los cambios en la estructura de consumo de la población española. El IPC se suele utilizar como un indicador de la inflación ya que se construye mensualmente y se publica una o dos semanas después de haber finalizado el mes.

El problema que presenta el IPC es que mide la variación de los precios en bienes y servicios destinados al consumo final y la inflación se suele referir a todos los bienes y servicios que se intercambian en un país. Hay una variante del IPC que consiste en excluir de la cesta de la compra los artículos referentes a la energía (por la fuerte dependencia de fuentes no nacionales) y los artículos agrarios no elaborados (por la fuerte estacionalidad). Con este IPC modificado se mide la denominada inflación subyacente.

Además, para tener una medida común de inflación que permita realizar comparaciones en la Unión Europea, se determina en cada uno de los países miembros el Índice de Precios de Consumo Armonizado (IPCA), que se obtiene a partir del IPC nacional con adaptaciones y ajustes para que las cestas de la compra sean lo más similares posibles en todos los países.

Ponderaciones año 2009

Grupos	IPC	IPCA español
1. Alimentos y bebidas no alcohólicas	18,07	18,14
2. Bebidas alcohólicas y tabaco	2,54	2,57
3. Vestido y calzado	8,81	8,86
4. Vivienda	10,74	10,91
5. Menaje	7,20	7,12
6. Medicina	3,13	3,14
7. Transporte	15,29	14,66
8. Comunicaciones	3,72	3,65
9. Ocio y cultura	7,81	7,9
10. Enseñanza	1,32	1,33
11. Hoteles, cafés y restaurantes	12,33	14,85
12. Otros bienes y servicios	9,04	6,87
GENERAL	100	100

Ejemplo 6.6.

Las ponderaciones de los siguientes grupos en el IPC con base una año cualquiera son:

Grupo	Ponderación
Alimentos	35
Vestidos	9
Vivienda	17
Menaje y hogar	7
Servicios de salud	3
Transporte	12
Ocio	5
Enseñanza	8
Otros	4
	100

Calculamos el IPC para el año siguiente si:

a) El índice de alimentos se incrementa un 10%, permaneciendo iguales el resto de los índices:
Como el IPC está basado en el índice de Laspeyres, teniendo en cuenta sus propiedades, tenemos:

$$IPC = \sum_i u^i I_i \text{ y } \Delta IPC = \sum_i u^i \Delta I_i$$

y, por tanto:

$$\Delta IPC = 0 + 0.09 \cdot 0.10 + 0 = 0.009,$$

por lo que el nuevo IPC con base el año considerado será:
IPC (año siguiente)=IPC (año considerado)+ΔIPC=1+0.009=1.009

b) Los índices de vestidos y transporte se incrementan un 5 y 8%, respectivamente; los índices de ocio y salud disminuyen un 3 y 6%, respectivamente; el resto de índices no varía.
Tenemos:

$$\Delta IPC = 0.09 \cdot 0.05 + 0.12 \cdot 0.08 - 0.05 \cdot 0.03 - 0.03 \cdot 0.06 = 0.0108,$$

por lo que el nuevo IPC con base el año considerado será:
IPC (año siguiente) = 1 + 0.0108 = 1.0108

RELACIÓN DE NÚMEROS ÍNDICES.

1.- Una nave industrial ha sido alquilada para su explotación al precio de 7500€ mensuales en el año 2007. Si el IPC ha evolucionado de la siguiente forma:

Año	**IPC (base 2007)**
2007	1
2008	1.06
2009	1.11
2010	1.20

¿Cuál será el precio del alquiler para 2011, si ese año se revisa el precio de acuerdo con el incremento del IPC?

2.- El valor de un artículo en 2011 es un 25% superior al de 2005 y un 5% superior a su valor en 2008. ¿Cuál es su valor relativo en 2008 respecto de 2005?

3.- Dada la estadística sobre la contratación efectiva en las bolsas españolas, en millones de €:

Años	**Madrid**	**Barcelona**	**Bilbao**	**Valencia**
2005	67993	28878	26694	2817
2006	100049	43360	19782	3865
2007	113385	40658	21198	6892
2008	102500	31116	23582	6837
2009	131180	35426	14350	4775
2010	74279	17253	16724	7839

a) Calcular los índices simples para estos años con base 2005.
b) Calcular los índices de Sauerbeck para estos años con base 2005.

4.- Dada la serie del IPC 2000-2006 con base 1998 y 2006-2010 con base 2006, construir una base homogénea con base 2006:

IPC (base 1998)		IPC (base 2006)	
Año	**Índice**	**Año**	**Índice**
2000	108	2006	100
2001	116	2007	124
2002	126	2008	149
2003	141	2009	172
2004	163	2010	199
2005	190		
2006	224		

5.- Una fábrica de automóviles produce cuatro modelos, cuyos precios de venta, costes y número de unidades producidas en los últimos dos años fueron:

	2009			2010		
Modelos	**Precio**	**Coste**	**Unidades**	**Precio**	**Coste**	**Unidades**
A	9000	4000	3200	11000	5000	4100
B	13000	6000	4200	13000	7000	3000
C	19000	10000	2300	20000	11000	2400
D	38000	22000	1700	41000	25000	1500

Calcular:

a) Los índices de producción (precio de venta - coste) de Laspeyres, Paasche y Fischer para 2010 con base 2009.

b) Los índices de cantidad de de Laspeyres, Paasche y Fischer, con la misma base.

c) Los índices de precios con la misma base.

d) El valor a precios constantes de 2009 de la producción los dos años.

6.- Dadas las ventas de una empresa, estudiar el crecimiento real de dichas ventas en euros de 2008:

Año	Ventas	Índice
2008	150000	100
2009	180000	131.42
2010	210000	161.43

7.- Dados los ingresos de un individuo y el IPC entre 2000 y 2010, estudiar la evolución real de sus ingresos en el tiempo y la inflación en dicho periodo:

Año	Ingresos	IPC 1990	IPC 2004	IPC 2007
2000	15000	115		
2001	15400	123		
2002	16000	129		
2003	18000	132		
2004	18200	135	100	
2005	19000		103	
2006	21200		107	
2007	22500		113	100
2008	23000			110
2009	24000			120
2010	24500			123

8.- El contrato de alquiler de un apartamento es de 450€ mensuales desde 2007. Si para 2011 se quiere revisar en base a los incrementos del grupo vivienda del IPC en esos años, con índices:

Años	Índice grupo vivienda
2007	118.3
2008	130.5
2009	147.3
2010	167.8

Calcular el nuevo alquiler mensual para 2011.

9.- El IPC creció de 107 a 115 durante un periodo de tiempo.

a) Si una persona dedica 240€ al mes para servicios al comienzo de dicho periodo, ¿qué cantidad debe presupuestar para el mismo nivel de servicios al final del periodo?

b) Si el salario de una persona pasa de 600 a 662.50€ durante ese periodo, ¿Cuál es el cambio real de su poder adquisitivo?

10.- Tenemos el precio del menú de un determinado restaurante entre 2006 y 2010, junto con la evolución del IPPC en ese periodo:

Años	IPC (2002)	Precio
2006	199.3	30
2007	228.4	32.5
2008	261.3	37.5
2009	293.1	40
2010	326.1	45

Estudiar la evolución del precio en euros constantes de 2006.

11.- Las ponderaciones de los siguientes grupos en el IPC con base una año cualquiera son:

Grupo	Ponderación
Alimentos	20.61
Vestidos	8.81
Vivienda	10.74
Menaje y hogar	7.20
Servicios de salud	3.13
Transporte	19.01
Ocio	7.81
Enseñanza	1.32
Otros	21.37
	100

Calculamos el IPC para el año siguiente si:

a) El índice de vivienda se incrementa un 15%, permaneciendo iguales el resto de los índices:

b) Los índices de vestidos y transporte se incrementan un 4 y 9%, respectivamente; los índices de alimentos y enseñanza disminuyen un 1.5 y 5%, respectivamente; el resto de índices no varía.

SEGUNDA PARTE:

CÁLCULO DE PROBABILIDADES

CAPÍTULO 7:

TEORÍA DE LA PROBABILIDAD.

Para empezar con la Probabilidad recordamos un poco la combinatoria.

7.1. Combinatoria.

Repasemos algunas nociones vistas en cursos anteriores sobre combinatoria:
Cuando tengamos que resolver algún problema en el que intervenga la combinatoria podemos preguntarnos lo siguiente:

- ¿Interviene el orden?
    - No: Entonces tenemos Combinaciones.
    - Sí: Entonces nos preguntamos: ¿Intervienen todos los elementos?
        - Sí: Entonces tenemos Permutaciones.
        - No: Entonces tenemos Variaciones.

En caso de que se pueda repetir algún elemento se tomarían con repetición.

Las expresiones de cada una de ellas son:

- Variaciones de n elementos tomados de p en p:

$$V_n^p = \frac{n!}{(n-p)!}$$

Ejemplo 7.1.

¿Cuántos números de 3 cifras distintas se pueden escribir?

$$V_{10}^{3} = \frac{10!}{7!} = 10 \bullet 9 \bullet 8 = 720$$

- Variaciones con repetición de *n* elementos tomados de *p* en *p*:

$$VR_{n,p} = n^{p}$$

Ejemplo 7.2.

¿Cuántos números de 3 cifras se pueden escribir?

$$VR_{10,3} = 10^{3} = 1000$$

- Permutaciones de *n* elementos:

$$P_{n} = n!$$

Ejemplo 7.3

¿De cuántas formas se pueden colocar 5 personas en una fila?

$$P_{5} = 5! = 120$$

- Permutaciones con repetición de *n* elementos con α_i repeticiones del elemento *i*:

$$P_{n}^{\alpha_1,\ldots,\alpha_r} = \frac{n!}{\alpha_1!\ldots\alpha_r!}$$

Ejemplo 7.4.

¿Cuántas palabras distintas se pueden formar con las letras de la palabra *aritmética*?

$$P_{10}^{2,2,2,1,1,1,1} = \frac{10!}{2!2!2!} = 10 \bullet 9 \bullet 8 \bullet 7 \bullet 6 \bullet 5 \bullet 3 = 453600$$

- Permutaciones circulares:

$$PC_{n} = (n-1)!$$

Ejemplo 7.5.

¿De cuántas formas se pueden sentar 7 personas en una mesa redonda de 7 sillas?

$$PC_{7} = 6! = 6 \bullet 5 \bullet 4 \bullet 3 \bullet 2 \bullet 1 = 720$$

- Combinaciones de n elementos tomados de p en p:

$$C_n^p = \frac{n!}{(n-p)!\,p!}$$

Ejemplo 7.6.

De un grupo de 12 alumnos, ¿de cuántas maneras puede un profesor escoger un comité formado por 4 alumnos?

$$C_{12}^4 = \frac{12!}{8!4!} = 495$$

- Combinaciones con repetición de n elementos tomados de p en p.

$$CR_{n,p} = C_{n+p-1}^n$$

Ejemplo 7.7.

De un conjunto de 5 colores, ¿de cuántas formas se pueden combinar para hacer una bandera de 3 colores sin considerar la posición de estos y de forma que los colores se puedan repetir?

$$CR_{5,3} = C_{5+3-1}^5 = \binom{7}{5} = 21$$

7.2. Introducción al Cálculo de Probabilidades.

La *Estadística* utiliza los datos para obtener conclusiones sobre las características de una variable en una población a la que no podemos acceder de manera completa. Normalmente tenemos información parcial de dicha variable eligiendo al azar algunos elementos de la población. Para conocer la intención de voto, se selecciona aleatoriamente una muestra de la población y se trabaja con ella. Por tanto, tenemos un experimento aleatorio en el que interviene el azar y la *Probabilidad* es quien estudia estos fenómenos.

En los ejemplos estudiados de *Estadística*, las distribuciones de frecuencias repiten ciertos patrones o formas, por lo cual dichas observaciones vienen de variables cuya distribución tiene cierta estructura, es decir, responden a un modelo. La *Probabilidad* proporciona y estudia modelos para las distribuciones de algunas variables.

Por tanto, podemos resumir diciendo que la *Probabilidad* se ocupa del estudio de los fenómenos aleatorios y su objetivo es la modelización matemática de estos fenómenos.

Sus orígenes son muy remotos y están muy relacionados con los juegos de azar. Hoy día muchos investigadores se dedican al descubrimiento y puesta en marcha de nuevas aplicaciones de la Probabilidad en campos como Medicina, Meteorología, Mercadotecnia,...

7.3. Distintas concepciones de Probabilidad. Definición Axiomática.

Veamos los distintos conceptos de Probabilidad que han existido a lo largo de la historia hasta llegar a la definición axiomática:

7.3.1. Concepción Clásica.

Consideremos un experimento en el que tenemos un número finito de resultados n que son igualmente factibles y excluyentes (no se puede obtener más de un resultado de forma simultánea). Un ejemplo es el lanzamiento de un dado equilibrado.

Consideremos A un resultado posible del experimento que se puede presentar en m de los n resultados posibles. En estas condiciones definimos la probabilidad del resultado A como:

$$P(A) = \frac{n^{\circ}\, casos\ favorables}{n^{\circ}\, casos\ posibles} = \frac{m}{n} = \text{Regla de Laplace.}$$

Ejemplo 7.8.

Si consideramos el lanzamiento de un dado:

Sea A el resultado obtener un 2, entonces $P(A)$ =1/6

Sea B el resultado obtener un número par, entonces $P(B)$ =3/6 = 0.5

Esta definición tiene el problema de obligar al experimento a tener un número finito de resultados y a que todos los resultados sean igualmente factibles, cosa que no ocurre siempre.

7.3.2. Concepción Frecuentista.

En muchas ocasiones los posibles resultados de un experimento no son igualmente factibles; por ejemplo, en una fábrica la probabilidad de que un producto sea defectuoso debe de ser distinta a la de que el producto sea correcto.

Esta definición viene de la idea de repetir el experimento muchas veces bajo las mismas condiciones y observar los resultados obtenidos. Entonces, 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:

$$P(A) = \lim_{N\to\infty} f_N(A) = \lim_{N\to\infty} \frac{m}{N},$$

donde N es el número de veces que realizamos el experimento y m es el número de veces que hemos obtenido el resultado A en todas las realizaciones del experimento.

Ejemplo 7.9.

Lanzamos un dado trucado:

Lo lanzamos 1000 veces, con los siguientes resultados:

El 1 sale 250 veces; el 2, 120 veces; el 3, 148 veces; el 4, 150 veces; el 5, 162 veces; y el 6, 170 veces.

Entonces:

$$P(1) = \frac{250}{1000} = 0.25 \qquad P(2) = \frac{120}{1000} = 0.12 \qquad P(3) = \frac{148}{1000} = 0.148$$

$$P(4) = \frac{150}{1000} = 0.15 \qquad P(5) = \frac{162}{1000} = 0.162 \qquad P(6) = \frac{170}{1000} = 0.17$$

Este tratamiento de la Probabilidad tiene el problema de que hay que repetir muchas veces el experimento en las mismas condiciones, lo cual no siempre es posible.

7.3.3. Concepción Subjetiva.

Hay fenómenos que no se pueden repetir en las mismas condiciones; por ejemplo, cuando se asegura una obra de arte, el precio del seguro debe ir relacionado con la probabilidad de que la obra resulte dañada.

En estos casos la probabilidad se interpreta como el grado de creencia o convicción respecto a la ocurrencia de un resultado, pero el problema que presenta es la subjetividad en la asignación de la Probabilidad.

7.3.4. Desarrollo Axiomático de la Probabilidad.

Vamos a realizar el desarrollo junto con un ejemplo. Realizamos un experimento aleatorio, por ejemplo, el lanzamiento de un dado:

- Cada uno de los posibles resultados de dicho experimento recibe el nombre de **suceso elemental**.
  En el ejemplo, son sucesos elementales: $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$.
- El conjunto formado por todos los sucesos elementales se llama **espacio muestral** del experimento aleatorio (Ω).
  En el ejemplo, $\Omega = \{1,2,3,4,5,6\}$.
  El espacio muestral puede ser:
  - Finito, como por ejemplo, la variable X =“nº de asientos vacíos de un autobús”, donde $\Omega = \{x \in \aleph / x <$ nº asientos del autobús$\}$.
  - Infinito numerable, como por ejemplo la variable Y = “nº de llegadas a una cola”, donde $\Omega = \{y \in \aleph\}$.
  - Infinito no numerable, como la variable Z = “Tiempo de vida de una pieza”, donde $\Omega = \{z \in \Re^+\}$.

  Un espacio muestral es discreto si su conjunto de resultados es finito o infinito numerable, y es continuo si es infinito no numerable (los continuos son aquellos que pueden tomar cualquier valor dentro de un intervalo).
- Un **suceso** de un espacio muestral es cualquier subconjunto del espacio muestral. En el ejemplo del dado son sucesos: $\{1,2,5\},\{2,6\},\{1,2,4,5\}$...
  Destacamos dos sucesos:
  - El suceso seguro, que coincide con el espacio muestral (Ω) y es el suceso que ocurre siempre que se realice el experimento.
  - El suceso imposible ($\emptyset$), que es aquel suceso que no ocurre nunca.

Estudiamos, a continuación, distintas operaciones entre sucesos. Dados dos sucesos distintos E_1 y E_2 definimos:

1. **Unión:** La unión de dos sucesos E_1 y E_2 es un suceso que ocurre cuando lo hacen E_1, E_2 o ambos.
   Como ejemplo, si $E_1 = \{1,2\}$ y $E_2 = \{2,5\}$, entonces $E_1 \cup E_2 = \{1,2,5\}$.
2. **Intersección:** La intersección de dos sucesos E_1 y E_2 es un suceso que ocurre cuando lo hacen E_1 y E_2.
   En el ejemplo anterior, $E_1 \cap E_2 = \{2\}$.

3. **Contrario:** El contrario de un suceso E_1 es el suceso $\overline{E_1}$ ó E_1^C formado por los sucesos elementales que no están en E_1.
   En el ejemplo, $\overline{E_1} = \{3,4,5,6\}$
4. **Diferencia de sucesos:** Definimos el suceso diferencia de dos sucesos E_1-E_2 como el suceso que ocurre si lo hace E_1 y no E_2, es decir, $E_1\text{-}E_2 = E_1 \cap \overline{E}_2$.
   En el ejemplo, si $E_1=\{1,3,5\}$ y $E_2=\{3\}$, entonces E_1 -E_2 = $\{1,5\}$.
5. **Inclusión:** Decimos que el suceso E_2 está incluido en el suceso E_1 *($E_2 \subset E_1$)* si cualquier resultado de E_2 lo es de E_1.
   En nuestro ejemplo, si $E_2 = \{2,4\}$ y $E_1 = \{2,4,6\}$, entonces $E_2 \subset E_1$.

Decimos que dos sucesos E_1 y E_2 son mutuamente excluyentes, disjuntos o incompatibles si su intersección es nula, es decir, $E_1 \cap E_2 = \emptyset$. Por lo tanto, los sucesos contrarios son sucesos incompatibles.

Además, tenemos las Leyes de Morgan, por las cuales se verifica:

$$\overline{E}_1 \bigcap \overline{E}_2 = \left(E_1 \cup E_2\right)^C$$

$$\overline{E}_1 \bigcup \overline{E}_2 = (E_1 \cap E_2)^C$$

Representamos los resultados de un experimento aleatorio por un par $(\Omega,\mathscr{A})$, donde Ω representa el espacio muestral y $\mathscr{A}$ una clase de subconjuntos de Ω que verifican ciertas propiedades (es cerrada para uniones numerables, para formación de contrarios, y además contiene al conjunto $\emptyset$).

Vemos a continuación la definición axiomática de Probabilidad:
Dado el par $(\Omega,\mathscr{A})$ y sea E un suceso cualquiera, definimos la probabilidad sobre $(\Omega,\mathscr{A})$ como una aplicación:

$P: (\Omega,\mathscr{A}) \rightarrow \Re$ verificando:

1. $P(E) \geq 0$,$\forall E \in \mathscr{A}$
2. $P(\Omega) = 1$
3. Si $E_1,E_2,...$ son infinitos sucesos incompatibles, es decir, con intersección vacía entre ellos ($E_i \cap E_j = \emptyset \quad \forall i \neq j$), entonces se verifica que la probabilidad de la unión de los sucesos es la suma de las probabilidades de los distintos sucesos:

$$P(E_1 \cup E_2 \cup E_3 \cup ...) = P(E_1) + P(E_2) + P(E_3) + ...$$

Podemos 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 las vamos a asignar utilizando cualquiera de los métodos ya vistos con anterioridad.

A la terna $(\Omega,\mathscr{A},P)$ se le llama espacio probabilístico y $P(E)$ es la probabilidad de cualquier suceso E.

Como consecuencia de la definición, tenemos las siguientes propiedades:

1. La probabilidad de la unión de un número finito de sucesos disjuntos ($E_i \cap E_j = \emptyset \quad \forall\ i \neq j$), es igual a la suma de sus probabilidades:
   $$P(E_1 \cup E_2 \cup ... \cup E_n) = P(E_1) + P(E_2) + ... + P(E_n).$$
   Es consecuencia inmediata de la tercera condición para que P sea una función de probabilidad.
2. Sea E un suceso cualquiera. Entonces,
   $$P(\overline{E}) = 1 - P(E).$$
   Demostración:
   $$P(E \cup \overline{E}) = P(E) + P(\overline{E}) = P(\Omega) = 1 \Rightarrow P(\overline{E}) = 1 - P(E).$$
3. $P(\emptyset) = 0$.
   Demostración:
   Utilizando la propiedad anterior, $P(\emptyset) = 1 - P(\Omega) = 1\text{-}1 = 0$
4. P es no decreciente, es decir, si $E_1 \subset E_2$ entonces $P(E_1) \leq P(E_2)$
   Demostración: Como $E_1 \subset E_2$ podemos escribir
   $E_2 = E_1 \cup (E_2 - E_1) \ \Rightarrow\ P(E_2) = P(E_1) + P(E_2 - E_1) \geq P(E_1)$.
5. La probabilidad de cualquier suceso E está entre 0 y 1 $(0 \leq P(E) \leq 1)$.
   Demostración: Es consecuencia inmediata de las propiedades anteriores.
6. Dados dos sucesos cualesquiera, la probabilidad de su unión viene dada por:
   $$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$
   Demostración:

   Podemos expresar $E_1 \cup E_2$ como unión de sucesos incompatibles, es decir,
   $E_1 \cup E_2 = E_1 \cup (E_2 - E_1)$ y también E_2 se puede expresar como $E_2 = (E_1 \cap E_2) \cup (E_2 - E_1)$
   Entonces, teniendo en cuenta las dos expresiones anteriores
   $P(E_1 \cup E_2) = P(E_1) + P(E_2 - E_1)$ y como
   $P(E_2) = P(E_1 \cap E_2) + P(E_2 - E_1)$

$P(E_2\text{-}E_1) = P(E_2) - P(E_1 \cap E_2)$, entonces sustituyendo en una de las expresiones anteriores

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

7. $$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 \angle i_2}^{n} P(E_{i1} \cap E_{i2}) +$$
$$+ \sum_{i1<i2<i3}^{n} P(E_{i1} \cap E_{i2} \cap E_{i3}) \;\; + \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} E_i\right)$$

Demostración: Se realiza por inducción utilizando la propiedad anterior.

8. $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

Demostración:

Como $A = (A \cap B) \cup (A \cap \overline{B})$, entonces

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

que nos da el resultado buscado.

Veamos algunos ejemplos en los que aplicamos algunas de las propiedades mencionadas anteriormente:

Ejemplo 7.10.

Se extrae una carta de una baraja de 40 cartas.

a) La probabilidad de que sea figura o copa vale:

Ω = {todas las cartas de la baraja},

A = {figura}, B = {copa}

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{12}{40} + \frac{10}{40} - \frac{3}{40} = \frac{19}{40}$$

b) La probabilidad de que sea figura pero no copa:

$$P(A \cap \overline{B}) = \frac{9}{40}$$

c) La probabilidad de que la carta no sea de copas:

$$P(\overline{B}) = 1 - \frac{10}{40} = \frac{3}{4}$$

Ejemplo 7.11.

a) Calcular la probabilidad de la unión de tres sucesos:

$P(A \cup B \cup C) = P(A)+P(B)+P(C)-[P(A \cap B)+P(A \cap C)+P(B \cap C)]+P(A \cap B \cap C)$

b) Calcular la probabilidad de la unión de cuatro sucesos:

$P(A \cup B \cup C \cup D) = P(A)+P(B)+P(C)+P(D)$ -

$- [P(A\cap B)+P(A\cap C)+P(A\cap D)+P(B\cap C)+P(B\cap D)+P(C\cap D)] +$
$+ [P(A\cap B\cap C)+P(A\cap C\cap D)+P(A\cap B\cap D)+P(B\cap C\cap D)] -$
$-P(A\cap B\cap C\cap D)$

7.4. Probabilidad Condicionada.

Considerramos el espacio probabilístico $(\Omega,\mathscr{A},P)$. Vamos a estudiar la probabilidad de un suceso cuando tenemos información de otro suceso que ya ha ocurrido.

Sea un suceso E_1 con probabilidad no nula. Definimos la probabilidad del suceso E_2, condicionada a E_1, y la notaremos por $P(E_2/E_1)$, como la probabilidad de que ocurra el suceso E_2 supuesto que ha ocurrido el suceso E_1, que viene dada por:

$$P(E_2 / E_1) = \frac{P(E_1 \bigcap E_2)}{P(E_1)}, \quad \text{con } P(E_1)>0$$

Fijado E_1, $P(E_2/E_1)$ con $E_2\in\mathscr{A}$ verifican los axiomas y por lo tanto $(\Omega,\mathscr{A},P(./E_1))$ es un espacio probabilístico.

Análogamente, se define $P(E_1 / E_2) = \dfrac{P(E_1 \cap E_2)}{P(E_2)}$, con $P(E_2)>0$.

De ambas definiciones obtenemos el Teorema del Producto, que dice que dados dos sucesos A y B de $(\Omega,\mathscr{A},P)$ con $P(A)$ y $P(B)>0$, se tiene que:

$$P(A\cap B) = P(A)\cdot P(B/A) = P(B)\cdot P(A/B)$$

La definición de probabilidad condicionada a un suceso se puede extender a cualquier número de sucesos:

$$P(B / C \cap D) = \frac{P(B \cap C \cap D)}{P(C \cap D)} \quad \text{con } P(C\cap D) > 0,$$

por lo que podemos extender el Teorema del Producto:

- Dados los sucesos $E_1,...,E_n$ con $P(P\left(\bigcap_{i=1}^{n} E_i)\right) > 0$, entonces:

$$P\left(\bigcap_{i=1}^{n} E_i\right) = P(E_1) \bullet P(E_2 / E_1) \bullet \ldots \bullet P\left(E_n / \bigcap_{i=1}^{n-1} E_i\right)$$

Ejemplo 7.12.

Se hace una encuesta para ver el número de lectores de los diarios Marca y El País. Los resultados fueron: El 20% de los encuestados leen Marca, el 16% leen El País y el 1% lee ambos.

Si se selecciona al azar un lector de Marca, ¿cuál es la probabilidad de que también lea El País?

Sean los sucesos:

A= leer Marca, B= leer El País.

$P(A)=0.2$, $P(B)=0.16$, $P(A \cap B)=0.01$,

$P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{0.01}{0.2} = 0.05$ es la probabilidad de que un lector de Marca lea también El País.

$P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.01}{0.16} = 0.0625$ es la probabilidad de que un lector de El País lea también Marca.

7.5. Independencia de Sucesos.

Al hablar de probabilidad condicionada de un suceso A dada la ocurrencia de un suceso B, hablamos de una cierta dependencia entre A y B, es decir, la información sobre la ocurrencia de B afecta a la probabilidad de la ocurrencia de A. Si esta información no tiene ningún efecto, definimos la independencia de sucesos.

Dados A y B dos sucesos, decimos que A es independiente de B si y solo si la probabilidad condicionada coincide con la probabilidad sin condicionar, es decir:

$$P(A/B) = P(A).$$

Fácilmente se puede comprobar que si A es independiente de B, entonces B es independiente de A.

Demostración:

$$P(A \cap B) = P(B)P(A/B) = P(B)P(A) = P(A)P(B/A) \Rightarrow P(B) = P(B/A).$$

Como consecuencia inmediata de la definición, tenemos que A y B son independientes si y solo si:

$$P(A \cap B) = P(A)\,P(B)$$

Si dos sucesos A y B son independientes entonces tenemos que:

i) A y $\overline{B}$ son independientes

ii) $\overline{A}$ y B son independientes

iii) $\overline{A}$ y $\overline{B}$ son independientes

Sean $A_1,...,A_n$ sucesos de un experimento aleatorio. Estos sucesos decimos que son mutuamente independientes si y solo si:

i) $P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i \neq j$

ii) $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \quad \forall i \neq j \neq k \neq i$

....

n-i) $P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$.

Ejemplo 7.13.

En el experimento del lanzamiento de un dado comprobamos si los sucesos A y B son independientes, siendo $A = \{5,6\}$ y $B = \{3,6\}$.

Calculamos $P(A \cap B) = \dfrac{1}{6}$ y $P(A)P(B) = \dfrac{1}{9}$

y, puesto que son valores distintos, podemos decir que no son independientes.

Ejemplo 7.14.

Se lanzan dos dados.

Sean A= salir impar en el primer dado.

B = salir impar en el segundo dado.

C = la suma de los puntos obtenidos entre los dos datos sea impar

Entonces $P(A)$=1/2 , $P(B)$=1/2 , $P(C)$=1/2, $P(A \cap B)$= ¼ (impar,impar), $P(A \cap C)$=1/4 (impar,par), $P(B \cap C)$=1/4 (par,impar).

Entonces son independientes dos a dos.

Pero si calculamos $P(A \cap B \cap C)$= 0

(ya que no es posible "impar + impar = impar"), y como

$$0 \neq P(A)P(B)P(C) = 1/8 ,$$

entonces los tres sucesos no son mutuamente independientes.

7.6. Teoremas Fundamentales del Cálculo de Probabilidades.

Vemos dos teoremas importantes para la Probabilidad:

7.6.1. Teorema de la Probabilidad Total.

Sea $(\Omega,\mathscr{A},P)$ un espacio probabilístico y $A_1,A_2,...,A_n$ un sistema completo de sucesos (o también se puede decir que constituyen una partición de Ω), es decir:

- son mutuamente excluyentes ($A_i \cap A_j=\varnothing \ \forall i \neq j$) y
- son exhaustivos ($\bigcup_{i=1}^{n} A_i = \Omega$).

Supongamos que se conocen las probabilidades de todos los sucesos A_i, y sea B un suceso cualesquiera. Entonces:

$$P(B)=\sum_{i=1}^{n} P(B/A_i)P(A_i)\,.$$

Demostración:

$$P(B)=P(B\cap\Omega)=$$

$$=P(B\cap(\bigcup_{i=1}^{n} A_i))=P(\bigcup_{i=}^{n}(B\cap A_i))=\sum_{i=1}^{n} P(B\cap A_i)=\sum_{i=1}^{n} P(B/A_i)P(A_i).$$

7.6.2. Teorema de Bayes.

Con las mismas hipótesis del Teorema de la Probabilidad Total se verifica:

$$P(A_i/B)=\frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}\,,\ \text{donde}$$

$P(A_i)$ son probabilidades a priori, $P(B/A_i)$ son verosimilitudes y $P(A_i/B)$ son probabilidades a posteriori.
Demostración:

Como $P(A_i \cap B) = P(A_i/B)P(B)$ y $P(A_i \cap B) = P(B/A_i)P(A_i)$.

Entonces $$P(A_i/B)=\frac{P(A_i\cap B)}{P(B)}=\frac{P(B/A_i)P(A_i)}{\sum_{i=}^{n} P(B/A_i)P(A_i)}$$

Como se puede observar, estos teoremas se van a utilizar en fenómenos que tienen dos experimentos claramente diferenciados. Los sucesos del primer experimento son los A_i, que forman una partición del espacio muestral de todos sus posibles resultados. B es un suceso del segundo experimento.

Utilizaremos el Teorema de la Probabilidad Total cuando necesitemos la probabilidad de un suceso del segundo experimento, si conocemos las probabilidades condicionadas de dicho suceso a los posibles distintos resultados del primero.

Utilizaremos la Fórmula de Bayes cuando nos pidan la probabilidad de un resultado del primer experimento, sabiendo el resultado final del segundo experimento.

Ejemplo 7.15.

Se tienen dos urnas A_1 y A_2 .

La urna A_1 tiene 3 bolas blancas y dos negras, y la urna A_2 tiene 2 blancas y 3 negras.

a) Se elige una urna al azar y se saca una bola. La probabilidad de que la bola sea blanca vale:

$$P(B) = \sum_{i=1}^{2} P(A_i)P(B/A_i) = \frac{3}{5}\frac{1}{2} + \frac{2}{5}\frac{1}{2} = \frac{1}{2} \ .$$

b) Suponemos que realizada la extracción, la bola obtenida es blanca. La probabilidad de que la bola se haya extraído de A_1 vale:

$$P(A_1/B) = \frac{P(A_1)P(B/A_1)}{P(B)} = \frac{1/2 \bullet 3/5}{1/2} = \frac{3}{5} \ .$$

RELACIÓN DE INTRODUCCIÓN A LA PROBABILIDAD.

1.- Sean A, B y C tres sucesos cualesquiera. Expresar formalmente los siguientes sucesos:

a) Ocurren A y B, pero no C.
b) Ocurre al menos uno de los tres.
c) Ocurren al menos dos.
d) No ocurre ninguno de los tres.
e) Ocurre solamente uno de los tres.

2.- Un operario de una fábrica observa de tres en tres las piezas producidas, anotando si cada una de ellas es defectuosa o no. Sean:

A el suceso la primera pieza observada es defectuosa,
B el suceso la segunda pieza es defectuosa, y
C el suceso la tercera pieza es defectuosa.

a) Escribir el espacio muestral correspondiente a esta situación.
b) Describir los siguientes sucesos: $A\cup B$, $A\cup B\cup C$, $A\cap B^C\cap C^C$, $A^C\cap B^C\cap C^C$.

3.- Se tiene una baraja de 40 cartas. Se extraen 5 cartas una a una, devolviendo la carta una vez vista. Calcular:

a) La probabilidad de que las 5 cartas sean oros.
b) La probabilidad de que el primer oro sea la quinta carta.

4.- En el interior de un círculo se selecciona un punto al azar. Calcular la probabilidad de que el punto quede más cercano al centro que a la circunferencia.

5.- En el lanzamiento sucesivo de dos dados, calcular las probabilidades de:

a) La suma de los números obtenidos sea 7.
b) La suma de los números sea múltiplo de 3 o 4.
c) El número obtenido en el segundo dado sea superior al obtenido en el primero.

6.- Una urna contiene 20 bolas, de las cuales 8 son rojas, 3 verdes y 9 negras. Se extraen sucesivamente y sin reemplazamiento 3 bolas. Calcular las probabilidades de:

a) Las tres bolas son negras.
b) Por lo menos una es verde.
c) Las tres son de distinto color.
d) Se obtienen en el orden roja, verde y negra.

7.- Se extraen, sin reemplazo, dos cartas de una baraja de 40. ¿Cuál es la probabilidad de que las dos sean ases?

8.- Se lanza una moneda 10 veces y en todos los lanzamientos el resultado es cara. ¿Cuál es la probabilidad de que el 11º lanzamiento sea cruz?

9.- De entre 20 tanques de combustible fabricados para un transbordador espacial, tres están defectuosos. Si se seleccionan aleatoriamente 4:

a) ¿Cuál es la probabilidad de que ninguno de los tanques esté defectuoso?
b) ¿Cuál es la probabilidad de que uno de los tanques sea defectuoso?

10.- Se extrae una carta de una baraja española de 40. Comprobar qué pares de sucesos son independientes:

a) A= rey , B= espadas
b) A= figuras , B= espadas
c) A= rey , B= figuras

11.- Los sucesos A, B y C son independientes, verificando:

$$P(A \cap B) = 1/12 \ , \ P(A \cap C) = 1/15 \ , \ P(B \cap C) = 1/20.$$

Calcular $P(A)$, $P(B)$, $P(C)$ y $P(A^c \cap B^c \cap C^c)$.

12.- En una batalla naval 3 destructores disparan simultáneamente a un submarino. La probabilidad de que lo destruya cada destructor es: 0.6, 0.3 y 0.1 respectivamente. ¿Cuál es la probabilidad de que el submarino sea destruido?

13.- Los empleados de una compañía se encuentran separados en tres divisiones: administración, operación de planta y ventas. La siguiente tabla indica el nº de empleados en cada división separados por sexo:

	Mujer (*M*)	**Hombre (*H*)**	**Totales**
Admón. (*A*)	20	30	50
O. de planta(*O*)	60	140	200
Ventas (*V*)	100	50	150
Totales	180	220	400

a) ¿Son mutuamente excluyentes los sucesos O y M?
b) Si se elige aleatoriamente un empleado:

1. ¿Cuál es la probabilidad de que sea mujer?

2. ¿Cuál es la probabilidad de que trabaje en ventas?
3. ¿Cuál es la probabilidad de que sea hombre y trabaje en administración?
4. ¿Cuál es la probabilidad de que trabaje en operación de planta, si es mujer?
5. ¿Cuál es la probabilidad de que sea mujer, si trabaja en operación de planta?

c) ¿Son los sucesos V y H independientes?
d) ¿Son los sucesos A y M independientes?
e) Determinar las probabilidades de:
1. $P(A \cup M)$
2. $P(A \cup M^c)$
3. $P(M/A)$

14.- Durante un año, las personas de una ciudad utilizan 3 tipos de transportes: metro (M), autobús (A) y coche (C). Las probabilidades de usar unos u otros son:
M: 0.3, A: 0.2, C: 0.15, M y A: 0.1, M y C: 0.05, A y C: 0.06, M,A y C: 0.01
Calcular las probabilidades de:
a) Que una persona tome, al menos, dos medios de transporte.
b) Que una persona viaje en M y no en A.
c) Que viaje en M o C y no en A.
d) Que viaje en M o A y en C.
e) Que vaya andando.

15.- Se sacan 2 bolas sucesivamente sin devolución de una urna con 3 rojas y 2 blancas. Se pide:
a) El espacio muestral.
b) Asociar una probabilidad a cada suceso elemental.
c) Descomponer en sucesos elementales y asociarles una probabilidad a los siguientes sucesos:
A= La 1ª bola es blanca.
B= La 2ª bola es blanca.

16.- Una urna tiene 5 bolas blancas y 3 rojas. Extraemos 2 bolas simultáneamente. Se pide:
a) El espacio muestral.
b) Asociar una probabilidad a cada suceso elemental.

17.- En una ciudad el 40% de las personas son rubias, el 25% tienen ojos azules y el 5% tienen pelo rubio y ojos azules. Calcular las probabilidades de:

a) Tener pelo rubio si tiene ojos azules.
b) Tener ojos azules si tiene pelo rubio.
c) No tener pelo rubio ni ojos azules.
d) Tener solo una de las características

18.- En una Universidad en la que solo hay estudiantes de Arquitectura, Ciencias y Letras, terminan la carrera el 5% de Arquitectura, el 10% de Ciencias y el 20% de Letras; y se sabe que el 20% estudian Arquitectura, el 30% Ciencias y el 50% Letras. Eligiendo un estudiante al azar, calcular:

a) La probabilidad de que sea de Arquitectura y haya terminado la carrera.
b) Nos dice que ha terminado la carrera; probabilidad de que sea de Arquitectura.
c) La probabilidad de que no estudie Letras.
d) La probabilidad de no terminar si está estudiando Letras.
e) ¿Son dependientes estudiar Arquitectura y Ciencias?
f) La probabilidad de que el alumno haya terminado la carrera.

19.- En un sistema de alarma, la probabilidad de que se produzca un peligro es 0.1. Si éste se produce, la probabilidad de que la alarma funcione es de 0.95. La probabilidad de que funcione la alarma sin haber peligro es 0.03. Hallar:

a) La probabilidad de que, habiendo funcionado la alarma, no hubiera peligro.
b) La probabilidad de que haya peligro y no funcione la alarma.
c) La probabilidad de que, no habiendo funcionado la alarma, haya peligro.
d) La probabilidad de que funcione la alarma.

20.- La probabilidad de que cierto componente eléctrico funcione es de 0.9. Un aparato contiene 2 de estos componentes. El aparato funcionará mientras lo haga, por lo menos, uno de ellos. ¿Cuál es la probabilidad de que el aparato funcione?

21.- Una planta armadora recibe microcircuitos provenientes de tres fabricantes distintos *A*, *B* y *C*. El 50% del total se compra a *A* y un 25% a cada uno de los otros fabricantes. El porcentaje de circuitos defectuosos es 5, 10 y 12% respectivamente. Si los circuitos se almacenan sin tener en cuenta quién fue el proveedor:

a) Determinar la probabilidad de que una unidad armada en la planta contenga un circuito defectuoso.

b) Si el circuito no está defectuoso, ¿cuál es la probabilidad de que halla sido vendido por el proveedor *B*?

CAPÍTULO 8:

VARIABLES ALEATORIAS.

Nuestro problema consiste en estudiar el comportamiento de fenómenos aleatorios. Para modelizar estos fenómenos y poderlos tratar adecuadamente introducimos el concepto de variable aleatoria.

8.1. Definición.

Anteriormente hemos estudiado la variable estadística, que representa el conjunto de resultados observados al realizar un experimento aleatorio en una población. Los experimentos se conciben de manera que los resultados del espacio muestral son cualitativos (el resultado del lanzamiento de una moneda, que puede ser cara o cruz) o cuantitativos (el número de cruces obtenidas en el lanzamiento de tres monedas).

Para el estudio de los experimentos aleatorios se hace necesario introducir una función que haga corresponder un número a cada suceso elemental del fenómeno. El concepto de variable aleatoria nos proporciona un medio para relacionar cualquier resultado con una medida cuantitativa; es decir, pretendemos realizar un proceso de abstracción mediante el cual a conceptos experimentales le asociamos entes matemáticos con los cuales se puede trabajar analíticamente:

- Frecuencia relativa → Probabilidad.
- Variable estadística → Variable aleatoria.
- Distribución de frecuencias → Distribución de probabilidad.
- Media aritmética → Esperanza Matemática.

Definimos el concepto de variable aleatoria de la siguiente forma:
Dado un fenómeno aleatorio con espacio probabilístico asociado $(\Omega,\mathscr{A},P)$, entendemos por variable aleatoria a toda función X definida sobre el espacio muestral Ω en el campo de los números reales:

$$X: \quad \Omega \to \Re$$
$$\omega \mapsto X(\omega)$$

tal que transforme los resultados del espacio muestral (sucesos) en valores reales.

Por tanto, una variable aleatoria es una función con base en el resultado de un experimento aleatorio, y es un fenómeno aleatorio con resultados numéricos. Además, vamos a tener asociada a cada variable aleatoria una función de probabilidad.

Ejemplo 8.1.

Sea X la variable aleatoria que estudia el lanzamiento de una moneda.
El espacio muestral está formado por los dos posibles resultados del experimento $\Omega = \{c,+\}$.
Como la variable aleatoria tiene que tomar valores reales, tenemos que nombrar la variable aleatoria de manera que sus valores sean reales:
X: número de cruces en un lanzamiento de la moneda. Entonces:

$$\begin{cases} X = 0 \equiv cara. & P(cara) = P(X = 0) = 1/2. \\ X = 1 \equiv cruz. & P(cruz) = P(X = 1) = 1/2. \end{cases}$$

Ejemplo 8.2.

Sea X la variable aleatoria que estudia el número de cruces obtenidas en el lanzamiento de una moneda tres veces.
El espacio muestral está formado por los 8 posibles resultados del experimento, cada uno con una probabilidad de $(1/2)^3 = 1/8$.

$$\Omega = \{ccc,+cc,c+c,cc+,++c,+c+,c++,+++\}$$

Los valores que toma la variable aleatoria X son:

X: $\Omega \to \Re$
ccc $\to$ 0
+cc $\to$ 1
c+c $\to$ 1
cc+ $\to$ 1
++c $\to$ 2
+c+ $\to$ 2
c++ $\to$ 2
+++ $\to$ 3

Entonces, la probabilidad de los distintos valores que toma la variable aleatoria es:

$$P[X=0] = P[(ccc)] = 1/8$$
$$P[X=1] = P[(+cc)\cup(c+c)\cup(cc+)] = 3/8$$
$$P[X=2] = P[(++c)\cup(+c+)\cup(c++)] = 3/8$$
$$P[X=3] = P[(+++)] = 1/8$$
$$P[X=x] = 0 \quad \forall x\neq 0,1,2,3.$$

En estos ejemplos, el número de valores posibles es finito, pero se pueden definir variables aleatorias cuyos valores sean contables o no. Por ello, definimos dos tipos de variables aleatorias:

- Discretas, si la variable aleatoria solo puede tomar un conjunto numerable de valores aislados, es decir, si existe un conjunto numerable E tal que $P[X\in E]=1$. Por tanto, asociamos a cada suceso un único número real. Como ejemplo podemos tomar el de la variable aleatoria que estudia el número de cruces en el lanzamiento de tres monedas.
- Continuas, si la variable aleatoria puede tomar cualquier valor dentro de un intervalo. Como ejemplo se puede tomar la variable aleatoria que estudia el tiempo de llegada a un establecimiento entre dos llegadas consecutivas, que se da en intervalos de tiempo.

8.2 Función de Distribución.

La función de distribución de una variable aleatoria calcula probabilidades de que la variable aleatoria tome valores en determinados subconjuntos.

Definimos la función de distribución F_X de una variable aleatoria X como una función $F_X:\Re\rightarrow[0,1]$ tal que

$$F_X(x) = P[X \leq x].$$

Las propiedades que caracterizan a toda función de distribución son:

i) $F_X(-\infty) = 0$ y $F_X(\infty) = 1$.
ii) F_X es una función no decreciente.
iii) F_X es una función continua a la derecha.

Ejemplo 8.2 (continuación).

Hacemos la función de distribución:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 4/8 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}.$$

Además, toda función de distribución verifica estas otras propiedades, que vamos a ver con el ejemplo 65:

i) $P[x_1 < X \le x_2] = F(x_2) - F(x_1)$.

$P[1 < X \le 3] = F(3) - F(1) = 1 - 4/8 = 1/2$.

ii) $P[X = x] = F(x) - F(x\text{-}0)$.

$F(x\text{-}0)$ representa el límite por la izquierda de $F(x)$ en el punto x.

$P[X=2] = F(2) - F(2\text{-}0) = 7/8 - 4/8 = 3/8$.

iii) $P[x_1 < X < x_2] = F(x_2\text{-}0) - F(x_1)$.

$P[1 < X < 3] = F(3\text{-}0) - F(1) = 7/8 - 4/8 = 3/8$.

iv) $P[x_1 \le X < x_2] = F(x_2\text{-}0) - F(x_1\text{-}0)$.

$P[1 \le X < 3] = F(3\text{-}0) - F(1\text{-}0) = 7/8 - 1/8 = 3/4$.

v) $P[x_1 \le X \le x_2] = F(x_2) - F(x_1\text{-}0)$.

$P[1 \le X \le 3] = F(3) - F(1\text{-}0) = 1 - 1/8 = 7/8$.

8.3. Variables Aleatorias Discretas y Continuas.

Estudiamos por separado cada tipo de variable aleatoria.

8.3.1. Discretas.

Sea X una variable aleatoria discreta. 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:

$$p_i = P[X = x_i]$$

Ejemplo 8.3.

En el lanzamiento de un dado:

$p_1 = P[X=1] = 1/6$, $p_2 = P[X=2] = 1/6$, ..., $p_6 = P[X=6] = 1/6$.

El conjunto de valores que toma una variable aleatoria discreta X con sus probabilidades respectivas $\{(x_i,p_i)\}_{i=1,2,...}$ se conoce como la distribución de probabilidad de la variable aleatoria X.
En el ejemplo 66, la distribución de probabilidad va a ser $\{(i,1/6)\}_{i=1,...,6}$

Dada una variable aleatoria discreta X, diremos que $P(x) = P[X=x]$ es función de probabilidad de X si verifica:

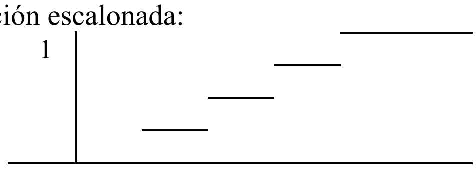
i) $P(x) \geq 0$ para todo valor real x.

ii) $\sum_{i=1}^{\infty} P(x_i) = 1$.

También se puede caracterizar la variable aleatoria discreta mediante su función de distribución:

$$F(x) = P[X \leq x] = \sum_{x_i \leq x} P(x_i),$$

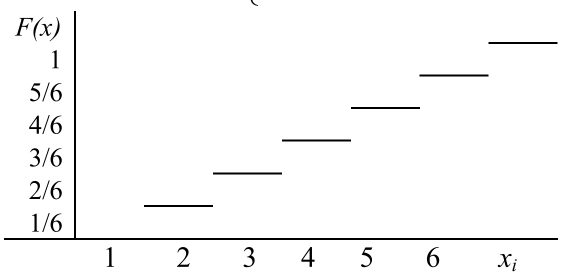
que es una distribución escalonada:



Ejemplo 8.3 (continuación).

La función de distribución viene dada por:

$$F(x) = \begin{cases} 0 & x < 1 \\ 1/6 & 1 \leq x < 2 \\ 2/6 & 2 \leq x < 3 \\ 3/6 & 3 \leq x < 4 \\ 4/6 & 4 \leq x < 5 \\ 5/6 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{cases}$$

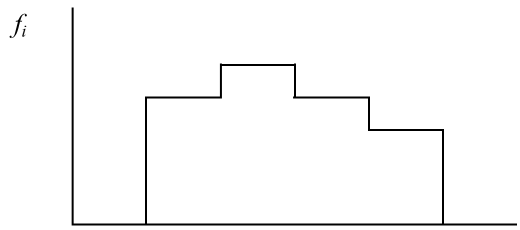


8.3.2. Continuas.

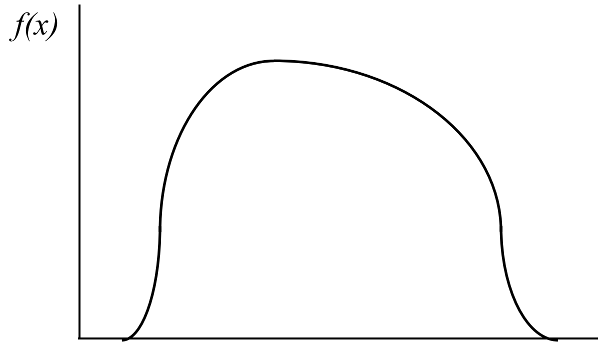
En el caso discreto se asignan probabilidades positivas a todos los valores puntuales de la variable aleatoria, siendo la suma de todos ellos igual a uno, aunque el conjunto de valores sea infinito numerable. Para el caso continuo, esto no es posible, ya que si aplicamos la Regla de Laplace la probabilidad de que la variable aleatoria tome un determinado valor va a ser $1/\infty = 0$. Entonces, podemos observar claramente que la probabilidad de que la variable tome un valor concreto va a ser nula, como veremos posteriormente.

Veámoslo con un ejemplo:
Estamos observando el peso de una determinada población, con un peso que únicamente nos da los Kg., y que, además, los da redondeados. Entonces, si el peso marca 76 Kg., lo que indica es que el peso verdadero de esa persona está entre 75.5 y 76.5 Kg., por lo que en el caso de variables aleatorias continuas es más lógico trabajar con las propiedades de los intervalos que con las propiedades de los puntos particulares. La distribución de probabilidad de una variable aleatoria continua va a venir caracterizada por la función de densidad, que notaremos mediante $f(x)$.

Veamos cómo se construye esta función de densidad:
Dados los datos de una variable aleatoria, tomados en intervalos todos con amplitud uno, consideramos la frecuencia relativa de todos los intervalos, la cual debe sumar uno. Consideramos el histograma de la variable para las frecuencias relativas (como todos los intervalos tienen amplitud uno, las alturas serán las frecuencias relativas de cada intervalo y la suma de todas las áreas será uno).



Si seguimos aumentando el número de observaciones y disminuyendo la amplitud de los intervalos, obtendremos una curva límite:



Esta curva límite será la función de densidad de la variable aleatoria, que definimos como:

Dada una variable aleatoria continua X, existe una función $f(x)$ llamada función de densidad, tal que

$$P[a \leq X \leq b] = \int_a^b f(x)dx \ ,$$

que además verifica:

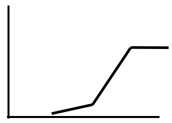
i) $f(x) \geq 0$ para cualquier valor real de x.

ii) $\int_{-\infty}^{\infty} f(x)dx = 1$.

La función de distribución asociada a la variable aleatoria viene dada por:

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(x)dx \ ,$$

que cumple las propiedades de función de distribución.
En este caso, la función de distribución es continua:



y además, se verifica que:

$$F'(x) = f(x)$$

Como ya se ha comentado anteriormente, la probabilidad de que la variable aleatoria continua tome un valor puntual vale cero, ya que:

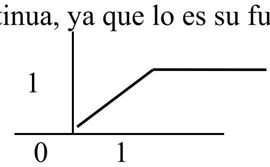
$$P[X = a] = \int_a^a f(x)dx = 0 .$$

Ejemplo 8.4.

Sea X una variable aleatoria con función de distribución:

$$F(x) = \begin{cases} 0 & x \le 0 \\ x & 0 < x \le 1 \\ 1 & x > 1 \end{cases}$$

Se puede ver que X es continua, ya que lo es su función de distribución:



Su función de densidad es:

$$f(x) = F'(x) = \begin{cases} 0 & x \le 0 \\ 1 & 0 < x \le 1 \\ 0 & x > 1 \end{cases} .$$

Ejemplo 8.5.

Sea X una variable aleatoria con función de densidad dada por:

$$f(x) = \begin{cases} x & 0 < x \le 1 \\ 2 - x & 1 < x \le 2 \\ 0 & resto \end{cases} .$$

Entonces, su función de distribución viene dada por:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} \int_{-\infty}^{x} 0 \bullet dx = 0 & x \le 0 \\ \int_{-\infty}^{0} 0 \bullet dx + \int_{0}^{x} x \bullet dx = 0 + \frac{x^2}{2} = \frac{x^2}{2} & 0 < x \le 1 \\ \int_{-\infty}^{0} 0 \bullet dx + \int_{0}^{1} x \bullet dx + \int_{1}^{x} (2 - x)dx = 2x - \frac{x^2}{2} - 1 & 1 < x \le 2 \\ \int_{-\infty}^{0} 0 \bullet dx + \int_{0}^{1} x \bullet dx + \int_{1}^{2} (2 - x)dx + \int_{2}^{x} 0 \bullet dx = 1 & 2 < x \end{cases}$$

Entonces:
$P[X<1.7] = F(1.7) = 2\cdot 1.7-1.7^2/2 -1 = 0.955$
$P[0.5<X<1.7] = F(1.7) - F(0.5) = 0.955 - 0.5^2/2 = 0.830$

8.4. Transformación de una Variable Aleatoria.

Veamos lo que sucede si tomamos una nueva variable aleatoria que sea función de otra ya conocida.

8.4.1. Transformación Lineal.

Sea X una variable aleatoria donde $F_X(x)$ es su función de distribución. Consideremos $Y=aX+b$. Entonces tenemos dos posibilidades:

- Si $a > 0$, tenemos que la función de distribución de Y vale:

$$F_Y(y) = P[Y \leq y] = P[aX + b \leq y] = P\left[X \leq \frac{y-b}{a} \right] = F_X\left(\frac{y-b}{a} \right).$$

Si la variable es continua, su función de densidad vale:

$$f_Y(y) = F_Y'(y) = \frac{1}{a} f_x\left(\frac{y-b}{a} \right).$$

- Si $a < 0$, tenemos que la función de distribución de Y vale:

$$F_Y(y) = P[Y \leq y] = P[aX + b \leq y] = P\left[X \geq \frac{y-b}{a} \right] = 1 - P\left[X < \frac{y-b}{a} \right] =$$

$$= 1 - P\left[X \leq \frac{y-b}{a} \right] + P\left[X = \frac{y-b}{a} \right] = 1 - F_X\left(\frac{y-b}{a} \right) + P\left[X = \frac{y-b}{a} \right]$$

Si la variable es continua, como la probabilidad puntual es nula, tenemos que:

$$\begin{cases} F_Y(y) = 1 - F_X\left(\dfrac{y-b}{a} \right) \\ f_Y(y) = \dfrac{-1}{a} f_X\left(\dfrac{y-b}{a} \right) \end{cases}$$

Resumiendo, en el caso continuo tenemos que

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Ejemplo 8.6.

Sea X una variable aleatoria con función de densidad:

$$f_X(x) = \begin{cases} \frac{1}{5} e^{-x/5} & x > 0 \\ 0 & resto \end{cases}.$$

Entonces, la función de densidad de $Y = 60 \cdot X$ vale:

$$f_Y(y) = \begin{cases} \frac{1}{60} f_X\left(\frac{y}{60}\right) = \frac{1}{300} e^{-y/300} & y > 0 \\ 0 & resto \end{cases}.$$

8.4.2. Transformación General.

Separamos los casos de variable aleatoria discreta y continua.

A) Variable Aleatoria Discreta.

Sea X una variable aleatoria discreta con función masa de probabilidad dada por los $\{p_i\}_{i=1,\ldots}$; y consideremos otra variable aleatoria función de la anterior $Y = h(X)$. Entonces

$$P\left[Y = y\right] = P\left[h(X) = y\right] = \sum_{\{x:h(x)=y\}} P\left[X = x\right].$$

Ejemplo 8.7.

Sea X una variable aleatoria discreta con función masa de probabilidad dada por:

$P[X=-2] = 1/5$, $P[X=-1] = 1/6$, $P[X=0] = 1/5$, $P[X=1] = 1/15$, $P[X=2] = 11/30$.

Entonces, la función masa de probabilidad de la variable aleatoria $Y = X^2$, que toma los valores $\{0,1,4\}$ es:

$P[Y=0] = P[X=0] = 1/5$

$P[Y=1] = P[X=-1] + P[X=1] = 1/15 + 1/6 = 7/30$

$P[Y=4] = P[X=-2] + P[X=2] = 1/5 + 11/30 = 17/30$

B) Variable Aleatoria Continua.

Vemos dos resultados importantes:

- Sea X una variable aleatoria con función de densidad $f(x)$ y sea $Y=h(X)$ otra variable aleatoria. Entonces, la función de distribución y la función de densidad de Y vienen dadas por:

$$F_Y(y) = P[Y \leq y] = P[h(X) \leq y] = \int_{\{h(x)\leq y\}} f(x)dx$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} \ .$$

- Sea X una variable aleatoria continua con función de densidad $f(x)$ y sea $Y=h(X)$ otra variable aleatoria continua con la función $h(X)$ continua y estrictamente monótona. Entonces, la función de densidad de Y es:

$$f_Y(y) = \begin{cases} f(h^{-1}(y))\left|\dfrac{dh^{-1}(y)}{dy}\right| & \alpha < y < \beta \\ 0 & resto \end{cases},$$

donde α y β son las imágenes de $h(x)$ en los extremos del intervalo de definición de X.

Ejemplo 8.8.

Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & resto \end{cases},$$

y sea $Y=1-X^2 = h(X)$.

Vamos a calcular la función de densidad de Y:

Claramente, $h(x)$ es continua y estrictamente decreciente en el intervalo donde X tiene densidad no nula.

$y = 1-x^2 \Rightarrow x = (1-y)^{1/2}$ con $0<x<1 \Rightarrow 0< (1-y)^{1/2} <1 \Rightarrow 0< 1-y <1 \Rightarrow 0<y<1$.

$h^{-1}(y) = x = (1-y)^{1/2}$

$$\left|\frac{d\,h^{-1}(y)}{dy}\right| = \frac{1}{2\sqrt{1-y}}$$

Entonces:

$$f_Y(y) = \begin{cases} f(h^{-1}(y))\left|\dfrac{dh^{-1}(y)}{dy}\right| = & \\ = 3\left(\sqrt{1-y}\right)^2 \bullet \left|\dfrac{1}{2\sqrt{1-y}}\right| = \dfrac{3\sqrt{1-y}}{2} & 0 < y < 1 \\ 0 & resto \end{cases}$$

Si existiera más de una inversa, la función de densidad de la variable aleatoria Y vale:

$$f_Y(y) = \sum_{i=1}^{n} f(h_i^{-1}(y))\left|\frac{d\,h_i^{-1}(y)}{dy}\right| .$$

Ejemplo 8.9.

Sea X una variable aleatoria con función de densidad:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ para } x \in \Re$$

y sea $Z = h(X) = 2X^2 + 1$.

Entonces, por construcción, Z toma valores entre 1 e infinito.

Además,

$$h^{-1}(z) = \begin{cases} \sqrt{\dfrac{z-1}{2}} \\ -\sqrt{\dfrac{z-1}{2}} \end{cases} ,$$

por lo que la función tiene dos inversas al tomar x valores positivos y negativos.

Entonces:

$$f_Z(z) = \sum_{k=1}^{2} f(h_k^{-1}(z))\left|\frac{dh_k^{-1}(z)}{dz}\right| =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{z-1}{4}} \frac{1/2}{2\sqrt{\dfrac{z-1}{2}}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{z-1}{4}} \frac{1/2}{2\sqrt{\dfrac{z-1}{2}}} =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{z-1}{4}} \frac{1/2}{\sqrt{\frac{z-1}{2}}} \quad \text{para} \quad z > 1$$

8.5. Características de una Distribución de Probabilidad.

Igual que hacíamos en Estadística Descriptiva, vamos a definir una serie de características que nos van a dar información resumida, en este caso de la distribución de probabilidad.

8.5.1. Esperanza Matemática.

La esperanza matemática de una variable aleatoria tiene su origen en los juegos de azar, ya que los apostadores deseaban conocer cuál era la esperanza de ganar repetidamente en el juego, puesto que en este caso representa la cantidad de dinero promedio que el jugador espera ganar o perder después de un número grande de apuestas.

Veamos la definición formal:

El valor esperado o Esperanza matemática de una variable aleatoria X es el valor promedio de X obtenido mediante:

- En el caso discreto, donde la variable aleatoria toma los valores $x_1, x_2, \ldots$ con probabilidades $p_1, p_2, \ldots$, la Esperanza matemática, si existe, vale:

$$E[X] = \sum_{i=!}^{\infty} x_i p_i \ .$$

- En el caso continuo, con $f(x)$ la función de densidad de X, la Esperanza matemática, si existe, vale:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \ .$$

Como se puede observar, la Esperanza matemática no es una función, sino que es un número real, que existirá si:

- En el caso discreto, la serie $\sum_{i=1}^{\infty} |x_i| p_i$ es convergente.

- En el caso continuo, la integral $\int_{-\infty}^{\infty} |x| f(x)dx$ es convergente.

Como caso particular, siempre que la variable aleatoria esté acotada ($P[a<X<b] = 1$ para a y b reales), existe su Esperanza matemática.

Ejemplo 8.10.

- Sea X una variable aleatoria discreta con función masa de probabilidad dada por:

  $P[X= -2] = 0.1 \quad P[X=0] = 0.4 \quad P[X=1] = 0.3 \quad P[X=4] = 0.2$

  Entonces, su Esperanza matemática vale:

  $$E[X] = -2\cdot 0.1 + 0\cdot 0.4 + 1\cdot 0.3 + 4\cdot 0.2 = 0.9$$

- Sea X una variable aleatoria con función de densidad:

  $$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & resto \end{cases}$$

  Entonces, su Esperanza matemática vale:

  $$EX = \int_{-\infty}^{0} x \bullet 0dx + \int_{0}^{1} x \bullet 1dx + \int_{1}^{\infty} x \bullet 0dx = \left.\frac{x^2}{2}\right]_0^1 = 1/2$$

El concepto de Esperanza matemática se puede también definir sobre una función de una variable aleatoria:

Sea $h(X)$ una función definida sobre todos los valores de la variable aleatoria X. Entonces, definimos la Esperanza matemática de $h(X)$ como:

- En el caso discreto,

  $$E(h(X)) = \sum_{i=1}^{\infty} h(x_i) p_i$$

  que existe si la serie converge de forma absoluta.

- En el caso continuo,

  $$E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x)dx$$

  que existe si es convergente la integral del valor absoluto.

Veamos las propiedades más importantes de la esperanza matemática:

1. El valor esperado de una constante es dicha constante:

   $$E[c]= c.$$

Demostración: $E[c]=\int_{-\infty}^{\infty} cf(x)dx = c\int_{-\infty}^{\infty} f(x)dx = c\bullet 1 = c.$

2. La esperanza de una combinación lineal es la combinación lineal de la esperanza:

$$E[aX+b] = a\,EX + b.$$

Demostración:

$$E[aX+b]=\int_{-\infty}^{\infty}(ax+b)f(x)dx=\int_{-\infty}^{\infty}axf(x)dx+b\int_{-\infty}^{\infty}f(x)dx=$$

$$=a\int_{-\infty}^{\infty}xf(x)dx+b\int_{-\infty}^{\infty}f(x)dx=aEX+b\,.$$

3. $E[h(X)+g(X)] = E[h(X)] + E[g(X)]$.

Demostración: $E[h(X)+g(X)]=\int_{-\infty}^{\infty}(h(x)+g(x))f(x)dx=$

$$=\int_{-\infty}^{\infty}h(x)f(x)dx+\int_{-\infty}^{\infty}g(x)f(x)dx=E[h(X)]+E[g(X)].$$

8.5.2. Momentos. Varianza.

Igual que en Estadística Descriptiva, los momentos son valores que recogen y resumen las propiedades de la variable aleatoria asociada:

- Definimos el momento no centrado de orden r como:

$$m_r = E[X^r]\,.$$

  Para $r = 1$, tenemos la esperanza matemática de la variable aleatoria.

- Definimos el momento centrado respecto de la esperanza de orden r como:

$$\mu_r = E[(X-EX)^r]\,.$$

De manera similar a la Estadística Descriptiva, definimos la varianza de la variable aleatoria X como:

$$Var(X) = \sigma_X^2 = E[(X-EX)^2] = E[X^2]-(EX)^2,$$

que mide la dispersión de los valores que toma la variable aleatoria respecto de su valor esperado; y la segunda igualdad se obtiene de forma análoga a como se obtenía en Estadística Descriptiva. Además, definimos la desviación típica de una variable aleatoria σ_X como la raíz cuadrada positiva de la varianza.

Las propiedades de la varianza son análogas a las de Estadística Descriptiva:

i) La varianza de una variable aleatoria es cero si y solo si la variable aleatoria solo puede tomar un valor.

ii) $Var(aX+b) = a^2\, Var(X)$.

8.5.3. Desigualdades relativas a los Momentos.

Vamos a obtener cotas para determinadas probabilidades en términos de los momentos de una variable aleatoria X.

- Teorema de Chebyshev:

  Sea X una variable aleatoria y $g(x)$ una función positiva de la variable aleatoria tal que existe su esperanza. Entonces:

  $$P\left[g(X) \geq k\right] \leq \frac{E[g(X)]}{k}.$$

  Demostración: La vemos en el caso discreto:

  Sea $A=\{x_i$ tales que $g(x_i)\geq k\}$.

  $E[g(X)] = \sum_{i=1}^{\infty} g(x_i)p_i = \sum_{A} g(x_i)p_i + \sum_{A^C} g(x_i)p_i \geq$ (porque el 2º sumando es siempre positivo al ser $g(x)$ siempre positiva) $\geq \sum_{A} g(x_i)p_i \geq k\sum_{A} p_i = kP(A) = kP\left[g(X) \geq k\right]$.

  Despejando, obtenemos el resultado.

Este teorema tiene dos casos particulares:

- Desigualdad de Markov:

  Sea X una variable aleatoria tal que existe su momento no centrado de orden $r>0$.

  Entonces, para todo valor t estrictamente positivo:

  $$P\left[|X| \geq t\right] \leq \frac{E|X|^r}{t^r}.$$

  Demostración:

  En el teorema anterior, tomamos:

  $g(x)=|x|^r$ y $k = t^r$, ya que $P[|X|\geq t] = P[|X|^r \geq t^r]$.

- Desigualdad de Chebishev:

Sea X una variable aleatoria tal que existe su varianza. Entonces, para todo valor t estrictamente positivo:

$$P\left[|X-EX|\geq t\right]\leq \frac{Var(X)}{t^2}.$$

En particular, $P\left[|X-EX|\geq k\sigma_X\right]\leq \frac{1}{k^2}$.

Demostración:

Tomamos en el teorema anterior:

$g(x)=(X\text{-}EX)^2$ y $k=t^2$, ya que $P[|X\text{-}EX|\geq t] = P[|X\text{-}EX|^2\geq t^2]$.

Ejemplo 8.11.

- Sea X una variable aleatoria que toma valores positivos con valor esperado igual a 6. Entonces:

$$P[X\geq 13] \leq 6/13.$$

- Sea X una variable aleatoria con esperanza 5 y varianza 1. Entonces:

$$P[|X\text{-}5|\geq 2] \leq 1/2^2 = 1/4$$

8.5.4. Otras Medidas de una Variable Aleatoria.

Veamos otras medidas análogas a las que veíamos en Estadística Descriptiva:

a) Mediana:

Es el valor de la variable aleatoria que deja a ambos lados el mismo número de valores, por lo que es cualquier valor de la variable aleatoria que:

- En el caso discreto: $1/2 \leq F(x) \leq 1/2 + P[X=x]$.
- En el caso continuo: $F(x) = 1/2$, que, en este caso, es único.

b) Moda:

Es el valor más frecuente de la variable aleatoria, es decir:

- En el caso discreto, es el valor o valores con mayor probabilidad.
- En el caso continuo es el máximo de la función de densidad.

c) Coeficiente de variación:

Mide en términos relativos la dispersión de la variable aleatoria respecto de su valor esperado, y vale:

$$C.V.(X)=\frac{\sqrt{Var(X)}}{EX}$$

d) Coeficiente de asimetría de Fisher:

Mide la simetría de la distribución, y viene dado por:

$$\gamma_1(X) = \frac{E\left[(X - EX)^3\right]}{{\sigma_X}^3}$$

e) Coeficiente de kurtosis:

Mide el apuntamiento de la distribución y viene dado por:

$$\gamma_2(X) = \frac{E\left[(X - EX)^4\right]}{{\sigma_X}^4} - 3$$

8.5.5. Función Generatriz de Momentos de una Variable Aleatoria.

Vamos a obtener a partir de ella los momentos no centrados de la variable aleatoria:

Definimos la función generatriz de momentos de la variable aleatoria X como:

$$g(t) = E[e^{tx}] = \begin{cases} \sum_{i=1}^{\infty} e^{tx_i} p_i & v.a.discretas \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & v.a.continuas \end{cases}$$

para $t \in \Re$ definido en un entorno de cero, y existirá siempre que exista la esperanza.

A partir de esta función podemos obtener todos los momentos no centrados que existan de la variable aleatoria estudiada, mediante:

$$EX = g'(0)$$
$$EX^2 = g''(0)$$
$$\ldots$$
$$EX^{i)} = g^{i)}(0)$$

Ejemplo 8.12.

Sea X una variable aleatoria con función de densidad:

$$f(\mathrm{x}) = \begin{cases} e^{-x} & x > 0 \\ 0 & resto \end{cases}.$$

Su función generatriz de momentos vale:

$$g(t) = E(e^{tx}) = \int_0^{\infty} e^{tx} e^{-x} dx = \int_0^{\infty} e^{-(1-t)x} dx = \left.\frac{-1}{1-t} e^{-(1-t)x}\right]_0^{\infty} = \frac{1}{1-t} \quad , \ t<1$$

Entonces:

$$g'(t) = \frac{1}{(1-t)^2}$$

$$g''(t) = \frac{2(1-t)}{(1-t)^4}$$

$$\ldots$$

Por tanto:

$$EX = g'(0) = 1$$

$$EX^2 = g''(0) = 2$$

$$\ldots$$

Veamos lo que vale la función generatriz de momentos de una transformación lineal de una variable aleatoria:

Sea X una variable aleatoria con $g_X(t)$ su función generatriz de momentos y sea $Y=aX+b$. Entonces, la función generatriz de momentos de Y vale:

$$g_Y(t) = e^{tb} g_X(at)$$

Demostración:

$$g_Y(t) = E[e^{ty}] = E[e^{t(ax+b)}] = E[e^{atx} \bullet e^{tb}] = E[e^{(at)x}]E[e^{tb}] = g_X(at)e^{tb}$$

RELACIÓN DE VARIABLES ALEATORIAS.

1.- Sea X una variable aleatoria que representa el número de llamadas telefónicas que recibe una oficina en un intervalo de 5 minutos y cuya función de probabilidad está dada por:

$$p(x)= e^{-3}\, 3^x / x! \quad \text{para} \quad x = 0,1,2,3,...$$

a) Determinar las probabilidades de que X sea igual a 0,1,2,3,4,5,6 y 7.
b) Determinar la función de distribución para estos valores de X.

2.- Sea X una variable aleatoria discreta. Determinar el valor de k para que la función $p(x)= k/x$, con $x = 1,2,3,4$ sea la función de probabilidad de X.
a) Calcular $P[1 \leq X \leq 3]$.
b) ¿Cuál es la probabilidad de que X tome el valor 2?
c) Calcular la probabilidad de que X sea mayor que 3.
d) Hallar la esperanza y varianza de X.

3.- Sea X una variable aleatoria que representa el número de clientes que llega a una tienda en un período de 1 hora. Dada la siguiente información:

x	0	1	2	3	4	5	6	7	8
$p(x)$	0.05	0.1	0.1	0.1	0.2	0.25	0.1	0.05	0.05

a) Calcular la esperanza y la varianza.
b) ¿Cuál es la probabilidad de lleguen menos de 5 clientes en 1 hora?
c) ¿Cuál es la probabilidad de que lleguen entre 4 y 7?
d) ¿Cuál es la probabilidad de que lleguen más de 6?
e) Si en la tienda no hay más de 2 dependientes, ¿cuál es la probabilidad de que al llegar un cliente, tenga que esperar?

4.- Sea X una variable aleatoria continua con función de densidad:

$$f(x) = \begin{cases} 1/6 & 1 < x \leq 2 \\ 1/3 & 3 < x \leq 4 \\ 1/2 & 5 < x \leq 6 \\ 0 & resto \end{cases}$$

Calcular:
a) La función de distribución.
b) $P[X \leq 1.75]$, $P[X \leq 3.5]$ *y* $P[1.8 \leq X \leq 3.5]$.
c) Esperanza y varianza.

5.- Una empresa estima que la demanda de sus clientes se comporta semanalmente de acuerdo a la ley de probabilidad dada por:

$$f(x)=\begin{cases} 3/8(4x-2x^2) & 0 \le x < 2 \\ 0 & resto \end{cases}$$

donde X representa miles de unidades.
¿Qué cantidad deberá tener puesta en venta al comienzo de cada semana para poder satisfacer la demanda con probabilidad 0.5?

6.- La dimensión de una serie de piezas es una variable aleatoria con:

$$f(x)= k/x^2 \qquad 1 \le x < 10$$

a) Calcular k.
b) Calcular la función de distribución.
c) Calcular la probabilidad de que la dimensión de la pieza esté entre 2 y 5.
d) Calcular a de forma que el 95% de las piezas tengan dimensión menor o igual que a.

7.- Sea X una variable aleatoria con función de densidad:

$$f(x)=\begin{cases} k\exp(-x/5) & x>0 \\ 0 & resto \end{cases}$$

Calcular:
a) k, de forma que $f(x)$ sea función de densidad.
b) La función de distribución.
c) $P[X \le 5], P[0 \le X \le 8]$.

8.- La variable aleatoria que mide la duración en horas de un componente eléctrico tiene la siguiente función de distribución:

$$F(x)= 1-\exp(-x/100),\ x>0.$$

Calcular:
a) La función de densidad.
b) La probabilidad de que la componente trabaje más de 200 horas.

9.- Sea X una variable aleatoria con función de densidad:

$$f(x)= 2(1-x),\ \ 0<x<1$$

Calcular su esperanza y la varianza.

10.- Sea X una variable aleatoria con función de densidad:

$$f(x)= 1/4 \exp(-x/4)\ ,\ x>0.$$

Calcular su esperanza y varianza.

11.- Sea X la variable aleatoria que mide el tiempo entre dos llegadas consecutivas a una tienda, que tiene

$$f(x) = \begin{cases} k\exp(-x/2) & x > 0 \\ 0 & resto \end{cases}.$$

Calcular:
a) k.
b) La función de distribución.
c) $P[2<X<6]$, $P[X<8]$.

12.- Sea X la variable aleatoria con función masa de probabilidad:

X	-2	-1	0	1	2
$p(x)$	1/5	1/10	1/5	2/5	1/10

Calcular la función masa de probabilidad de Y:
a) $Y= X+2$.
b) $Y= X^2$.

13.- Sea X una variable aleatoria con función de densidad:

$$f(x)=1/2 \,,\ \ 0<x<2.$$

Calcular la función de densidad de Y para:
a) $Y=2X+1$
b) $Y=X^2$.

14.-a) Estimar la probabilidad de que una variable aleatoria X se desvíe de su media no menos de 2 veces la desviación típica.

b) Sea X una variable aleatoria con $\sigma^2 = 0.01$. Estimar la probabilidad de que:

$$P[|X\text{-}EX|<0.15]$$

15.- Sea X una variable aleatoria continua con función de densidad:

$$f(x)= 1/2 \exp(-x/2) \,,\ x>0$$

y sea $Y=2X+2$. Calcular la función generatriz de momentos de Y.

CAPÍTULO 9:

VARIABLE ALEATORIA BIDIMENSIONAL.

En este capítulo vamos a tratar de estudiar simultáneamente más de un fenómeno aleatorio. La ampliación será a dos, considerando una variable aleatoria bidimensional.

9.1. Definición y Conceptos Asociados.

Una variable aleatoria bidimensional (*X,Y*) es una aplicación del espacio muestral Ω, asociado a los fenómenos aleatorios en estudio, en $\Re^2 = \Re x \Re$:

$$\Omega \xrightarrow{XxY} \Re x \Re$$
$$\omega_i \rightarrow (X(\omega_i), Y(\omega_i))$$

es decir, la variable aleatoria bidimensional es un mecanismo que nos permite cuantificar simultáneamente dos fenómenos aleatorios.

Por tanto, entendemos por variable aleatoria bidimensional a un par (*X,Y*) donde cada una de las componentes es una variable aleatoria unidimensional.

Ejemplo 9.1.

Sean *X* el número de cruces en 3 lanzamientos de una moneda.

Y el número de caras en los 2 primeros lanzamientos.

$$\Omega \xrightarrow{XxY} \Re x \Re$$
$$ccc \to (0,2)$$
$$cc+ \to (1,2)$$
$$c+c \to (1,1)$$
$$+cc \to (1,1)$$
$$++c \to (2,0)$$
$$+c+ \to (2,1)$$
$$c++ \to (2,1)$$
$$+++ \to (3,0)$$

El espacio muestral Ω de este experimento tiene los anteriores 8 sucesos elementales, cada uno de ellos con una probabilidad de 1/8 = 1/2·1/2·1/2 (los lanzamientos son independientes), lo que induce una probabilidad sobre los valores de (X,Y), que se recogen en la siguiente tabla :

$$p_{ij} = P[X = x_i, Y = y_j]$$

X\Y	**0**	**1**	**2**
0	0	0	1/8
1	0	2/8	1/8
2	1/8	2/8	0
3	1/8	0	0

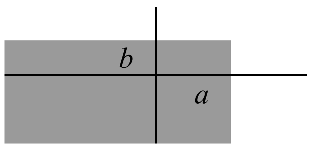
9.2. Función de Distribución y Función de Densidad.

Dada una variable aleatoria bidimensional (*X,Y*), se define la función de distribución conjunta de (*X,Y*) como:

$$\Re x \Re \xrightarrow{F} [0,1]$$
$$(a,b) \to F(a,b)$$

donde $$F(a,b) = P\left[w / X(w) \le a, Y(w) \le b\right],$$

es decir, *F*(*a,b*) es la probabilidad del conjunto de puntos que se aplican en un conjunto de la forma:



Las principales propiedades de la función de distribución son:

i) $F(+\infty, +\infty) = 1$.

ii) $F(-\infty, y) = F(x, -\infty) = 0$.

iii) El incremento de la función de distribución es siempre mayor o igual que cero:

$$\Delta F(x,y) = P[x<X\leq x',\ y<Y\leq y'] \geq 0 \quad \forall x,y.$$

iv) La función de distribución es continua a la derecha en cada variable.

Estas 4 propiedades caracterizan a toda función de distribución de una variable aleatoria bidimensional.

Aquí vamos a considerar solo las situaciones en las que ambas variables son discretas o la función de distribución es una función continua (si ambas variables son continuas):

- Si las dos variables son discretas:
  La variable aleatoria bidimensional (X,Y) toma un conjunto discreto de valores (x_i,y_j) en $\Re^2$ con unas probabilidades asociadas $p_{ij}= P[X =x_i , Y =y_j]$, tales que:
  1. $p_{ij} \geq 0$ para cualquier valor de i y de j.
  2. $\sum_i \sum_j p_{ij} = 1$.

  Al conjunto de puntos y probabilidades asociadas $((x_i,y_j),p_{ij})$ se le denomina distribución de probabilidad conjunta de X e Y.
  La función de distribución conjunta de X e Y para variables de tipo discreto está definida por:
  $$F(x,y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}$$

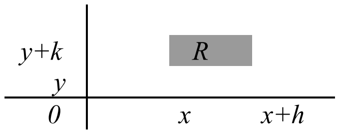
Ejemplo 9.2 (continuación).

$F(-0.5,2) = 0$, ya que X nunca toma valores menores que -0.5.

$F(2.2,1) = p_{00} + p_{01} + p_{10} + p_{11} + p_{20} + p_{21} = 0+0+0+2/8+1/8+2/8=5/8$.

$F(3.6,7) = 1$, ya que los valores de X y de Y son menores que 3.6 y 7 respectivamente.

- En el segundo caso se considera la función de densidad conjunta de la variable aleatoria bidimensional (X,Y) con el mismo significado al de la variable aleatoria unidimensional:
  $$f(x,y) = \lim_{(h,k)\to(0,0)} \frac{P[(X,Y)\in R]}{hk} = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$



Igual que en el caso univariante, la probabilidad de que la variable aleatoria tome valores en una región R viene dada por:

$$P[(X,Y) \in R] = \iint_R f(x,y)dxdy$$

y, en particular, la función de distribución es:

$$F(x,y) = P[X \leq x, Y \leq y] = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dydx \ .$$

Las propiedades de la función de densidad son las mismas que en el caso unidimensional, es decir:

1. $f(x,y) \geq 0$ para cualquier valor real de x e y.
2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dydx = 1.$

Ejemplo 9.3.

Sea (X,Y) una variable aleatoria bidimensional con la siguiente función de distribución:

$$F(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & 0 < x; 0 < y \\ 0 & resto \end{cases},$$

donde:

X = duración de llamadas comerciales,
Y = duración de llamadas personales.

Entonces, la función de densidad es:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y}\begin{Bmatrix} e^{-x} - e^{-x-y} & ; 0 < x, 0 < y \\ 0 & resto \end{Bmatrix} =$$

$$= \begin{cases} e^{-x-y} & 0 < x, 0 < y \\ 0 & resto \end{cases}$$

Entonces podemos calcular:

$$P[X \leq 1, Y \leq 2] = 1 - e^{-1} - e^{-2} + e^{-1-2} = 1 - e^{-1} - e^{-2} + e^{-3}$$

9.3. Distribuciones Marginales y Condicionadas.

Estudiamos ahora las distribuciones de las variables aleatorias unidimensionales tanto marginales como condicionadas:

9.3.1. Distribuciones Marginales.

En general, se definen las distribuciones marginales de X y de Y como las distribuciones por separado de cada una de las componentes de la variable aleatoria bidimensional, es decir:

$$F_X(x) = F(x,+\infty) = P[X \le x, Y \le \infty]$$
$$F_Y(y) = F(+\infty, y) = P[X \le \infty, Y \le y]$$

- En el caso discreto, se tiene:

$$p_{ij} = P[X = x_i, Y = y_j]$$

La distribución marginal de X se obtiene como:

$$p_X(x) = P[X = x_i] = P[X = x_i, Y\, arbitrario] = \sum_j p_{ij} = p_{i.} = \sum_y p(x,y)$$

y su función de distribución viene dada por:

$$F_X(x) = \sum_{x_i \le x} P[X = x_i] = \sum_{x_i \le x} \sum_j P[X = x_i, Y = y_j] = \sum_{x_i \le x} \sum_j p_{ij} = \sum_{x_i \le x} p_{i.}$$

Análogamente, tenemos resultados similares para la variable Y:

$$p_Y(y) = P[Y = y_j] = P[X\ arbitrario, Y = y_j] = \sum_i p_{ij} = p_{.j} = \sum_x p(x,y)$$

y su función de distribución marginal es:

$$F_Y(y) = \sum_{y_j \le y} P[Y = y] = \sum_i \sum_{y_j \le y} P[X = x_i, Y = y_j] = \sum_i \sum_{y_j \le y} p_{ij} = \sum_{y_j \le y} p_{.j}$$

Ejemplo 9.4.

X\Y	**0**	**1**	**2**	$p_{i.}$
0	0	0	1/8	1/8
1	0	2/8	1/8	3/8
2	1/8	2/8	0	3/8
3	1/8	0	0	1/8
$p_{.j}$	2/8	4/8	2/8	1

$$F_X(x) = \sum_{x_i \le x} p_{i.} = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 4/8 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases} \qquad F_Y(y) = \sum_{y_j \le y} p_{.j} = \begin{cases} 0 & x < 0 \\ 2/8 & 0 \le x < 1 \\ 6/8 & 1 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

- En el caso continuo:
  Sean *X* e *Y* dos variables aleatorias unidimensionales continuas con función de densidad conjunta *f(x,y)*. Las funciones de densidad marginales están dadas por:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{y} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

  y las correspondientes funciones de distribución marginales son:

$$P[X \le x] = F_X(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(x,y)dydx = \int_{-\infty}^{x} f_X(x)dx = F(x,\infty)$$

$$P[Y \le y] = F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{-\infty}^{y} f_Y(y)dy = F(\infty,y)$$

Ejemplo 9.5.

$$f(x,y) = \begin{cases} 3x(1-xy) & 0 \le x, y \le 1 \\ 0 & resto \end{cases}$$

$$f_X(x) = 3\int_0^1 x(1-xy)dy = 3\left(xy - \frac{x^2y^2}{2} \right)\Bigg]_0^1 = 3x\left(1 - \frac{x}{2}\right) \qquad 0 \le x \le 1$$

$$f_Y(y) = 3\int_0^1 x(1-xy)dx = 3\left(\frac{x^2}{2} - \frac{x^3y}{3} \right)\Bigg]_0^1 = (3-2y)/2 \quad 0 \le y \le 1$$

$$F(x,y) = 3\int_0^x\int_0^y x(1-xy)dydx = 3\int_0^x \left(xy - \frac{x^2y^2}{2} \right)dx = x^2y(3-xy)/2 \ ,$$

$$0 \le x, y \le 1$$

$$F_X(x) = F(x,1) = x^2(3-x)/2 \quad 0 \le x \le 1$$
$$F_Y(y) = F(1,y) = y(3-y)/2 \quad 0 \le y \le 1$$

9.3.2. Distribuciones Condicionadas.

Como ya sabemos, la probabilidad de un suceso condicionada a otro suceso viene dada por:

$$P(A/B)=\frac{P(A\cap B)}{P(B)}$$

En nuestra situación, lo que vamos a hacer es condicionar entre variables:

- En el caso discreto:

  Dada la variable aleatoria bidimensional (X,Y), se define la función de distribución de X condicionado a que Y toma un valor concreto y_j, $X/Y=y_j$, mediante:

$$F(x/y_j)=P[X\le x/Y=y_j]=\frac{P[X\le x,Y=y_j]}{P[Y=y_j]}=\frac{\sum_{x_i\le x}p_{ij}}{p_{.j}}=\sum_{x_i\le x}\frac{p_{ij}}{p_{.j}},$$

  siempre que $p_{.j}=P[Y=y_j]>0$,

  y su función de probabilidad es:

$$P[X=x/Y=y_j]=\frac{P[X=x,Y=y_j]}{P[Y=y_j]}$$

  Para $Y/X=x_i$ tenemos el resultado análogo:

$$F(y/x_i)=P[Y\le y/X=x_i]=\frac{P[X=x_i,Y\le y]}{P[X=x_i]}=\frac{\sum_{y_j\le y}p_{ij}}{p_{i.}}=\sum_{y_j\le y}\frac{p_{ij}}{p_{i.}}$$

$$P[Y=y/X=x_i]=\frac{P[X=x_i,Y=y]}{P[x=x_i]},$$

  siempre que $p_{i.}=P[X=x_i]>0$.

Ejemplo 9.4 (continuación)

Consideramos el caso $X/Y=1$.

La función de probabilidad es:

$$P[X=1/Y=1]=\frac{P[X=1,Y=1]}{P[Y=1]}=\frac{2/8}{4/8}=\frac{1}{2}$$

$$P[X=2/Y=1]=\frac{P[X=2,Y=1]}{P[Y=1]}=\frac{2/8}{4/8}=\frac{1}{2}$$

La función de distribución es:

$$F(x/Y=1)=\begin{cases} 0 & x<1 \\ 1/2 & 1\le x<2 \\ 1 & x\ge 2 \end{cases}$$

- En el caso continuo:
  Dada la variable aleatoria bidimensional (X,Y) con X e Y dos variables aleatorias unidimensionales continuas con $f(x,y)$ su función de densidad conjunta, definimos las funciones de densidad condicionadas como:
  1. $f(x/y)=\dfrac{f(x,y)}{f_Y(y)}$, siempre que $f_Y(y)>0$, es la función de densidad condicionada de X dado que Y vale y.
  2. $f(y/x)=\dfrac{f(x,y)}{f_X(x)}$, siempre que $f_X(x)>0$, es la función de densidad condicionada de Y dado que X vale x;

  y sus funciones de distribución condicionadas son:
  1. $F(x/y)=\displaystyle\int_{-\infty}^{x}\frac{f(x,y)dx}{f_Y(y)}$, siempre que $f_Y(y)>0$
  2. $F(y/x)=\displaystyle\int_{-\infty}^{y}\frac{f(x,y)dy}{f_X(x)}$, siempre que $f_X(x)>0$

Ejemplo 9.5. (continuación)

Consideramos X/Y:

$$f(x/y)=\frac{f(x,y)}{f_Y(y)}=\frac{3x(1-xy)}{(3-2y)/2}=\frac{6x(1-xy)}{3-2y} \qquad 0\le x,y\le 1$$

$$F(x/y)=\int_{-\infty}^{x}\frac{f(x,y)dx}{f_Y(y)}=\int_0^x\frac{6x(1-xy)}{3-2y}dx=\frac{6}{3-2y}\int_0^x\left(x-x^2y\right)dx=$$

$$=\frac{6}{3-2y}\left[\frac{x^2}{2}-y\frac{x^3}{3}\right]_{x=0}^{x}=\frac{6}{3-2y}\left(\frac{x^2}{2}-y\frac{x^3}{3}\right)\qquad 0\le x,y\le 1$$

9.4. Independencia de Variables Aleatorias.

Sabemos que dos sucesos A y B son independientes si se cumple que $P(A\cap B)=P(A)\cdot P(B)$.

Sean X e Y dos variables aleatorias unidimensionales con una distribución de probabilidad conjunta. Entonces, decimos que estas variables son estadísticamente independientes si y solo si se verifica que:

- $p(x,y)=p_X(x)\cdot p_Y(y)$ si X e Y son discretas
- $f(x,y)=f_X(x)\cdot f_Y(y)$ si son continuas,

es decir, si se cumple que:

$$F(x,y)=F_X(x)\cdot F_Y(y).$$

Por tanto, si las variables aleatorias unidimensionales X e Y son estadísticamente independientes, se cumple que:

$P[a<X<b,\ c<Y<d]=P[a<X<b]\cdot P[c<Y<d]$

Ejemplo 9.6.

Comprobar si son independientes las variables X e Y procedentes de la variable bidimensional (X,Y) con la siguiente función de cuantía:

X\Y	**0**	**1**	$p_{i.}$
0	¼	0	¼
1	¼	¼	½
2	0	¼	¼
$p_{.j}$	½	½	1

Para que las variables sean independientes debe ocurrir que

$P[X=x_i,Y=y_j]=P[X=x_i]\cdot P[Y=y_j]$ para todo x_i,y_j

Consideramos un par cualquiera:

$¼ = P[X=0,Y=0]\neq P[X=0]\cdot P[Y=0]=1/4\cdot 1/2=1/8$

Por tanto, no son independientes.

Ejemplo 9.7.

Ver si son independientes las variables aleatorias con función de densidad conjunta dada por:

$$f(x,y)=\begin{cases}\dfrac{3}{64}xy^2 & \begin{cases}0<x<4\\0<y<2\end{cases}\\ 0 & resto\end{cases}$$

$$f_X(x)=\frac{3}{64}\int_0^2 xy^2dy=\frac{3}{64}x\frac{y^3}{3}\Big]_{y=0}^{2}=\frac{1}{64}x\cdot 8=\frac{1}{8}x \quad 0<x<4$$

$$f_Y(y)=\frac{3}{64}\int_0^4 xy^2dx=\frac{3}{64}y^2\frac{x^2}{2}\Big]_{x=0}^{4}=\frac{3}{64}x^2\cdot 8=\frac{3}{8}y^2 \quad 0<y<2$$

Para que las variables sean independientes debe ocurrir que

$f(x,y)=f_X(x)\cdot f_Y(y)$ para todo x,y

$$f(x,y)=\frac{3}{64}xy^2=f_X(x)\cdot f_Y(y)=\frac{1}{8}x\cdot\frac{3}{8}y^2=\frac{3}{64}xy^2$$

Por consiguiente, ambas variables son independientes.

9.5. Esperanza Matemática.

Sea (X,Y) una variable aleatoria bidimensional. El valor esperado de una función $\varphi(x,y)$ se define como:

- $E[\varphi(x,y)]=\sum_i\sum_j \varphi(x_i,y_j)p_{ij}$ siempre que $\sum_i\sum_j\left|\varphi(x_i,y_j)\right|p_{ij}<\infty$, cuando X e Y son discretas.
- $E[\varphi(x,y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x,y)f(x,y)dxdy$ siempre que $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\varphi(x,y)\right|f(x,y)dxdy<\infty$, cuando X e Y son continuas.

En particular, podemos escribir algunas esperanzas concretas:

- $E[X] = \begin{cases} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i p(x_{i,} y_j) = \sum_{i=1}^{\infty} x_i p_x(x_i) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dydx = \int_{-\infty}^{\infty} xf_x(x)dx \end{cases}$, dependiendo de que X sea discreta o continua.

- $E[Y^2] = \begin{cases} \sum_i \sum_j y_j^2 p\ (x_{i,} y_j) = \sum_i y_j^2 p_x(x_i) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f(x,y)dydx = \int_{-\infty}^{\infty} y^2 f_y(y)dy \end{cases}$, dependiendo de que Y sea discreta o continua.

Algunas propiedades de la esperanza son:

i) $E[aX+bY+c]=aE[x]+bE[y]+c$.

ii) Si X e Y son independientes:

$$E[X\cdot Y]=E[X]\cdot E[Y],$$

por lo que la covarianza de ambas variables será cero y, en general,

$$E[\varphi(X)\phi(Y)] = E[\varphi(X)]E[\phi(Y)]$$

9.6. Momentos.

Sea (X,Y) una variable aleatoria bidimensional. Definimos:

- Momentos no centrados:
  El momento no centrado de órdenes (r,s) es $E[X^r Y^s]$
- Momentos centrados:
  El momento centrado de órdenes (r,s) es $E[(X\text{-}EX)^r(Y\text{-}EY)^s]$.

El momento bivariante centrado más utilizado es el de órdenes (1,1), también llamado covarianza, que es:

$$\sigma_{XY} = \sigma(x,y) = E[(X-EX)(Y-EY)] = E[XY] - E[X]E[Y]$$

Claramente, si X e Y son independientes, tenemos que la covarianza entre ambas variables es nula.

Además, sabemos que:

$VarX=EX^2\text{-}(EX)^2$

$VarY=EY^2\text{-}(EY)^2$

$Var[g(x,y)]= E[g(x,y)^2]\text{-}(E[g(x,y)])^2$

Finalmente, tenemos la siguiente propiedad:

$$\sigma^2(aX \pm bY) = a^2\sigma^2(X) + b^2\sigma^2(Y) \pm 2ab\sigma(X,Y)$$

9.7. Funciones Generatrices. Reproductividad.

Definimos la función generatriz de momentos de la variable aleatoria bidimensional (X,Y) como:

$$g(t_1,t_2) = E\left[e^{t_1X+t_2Y}\right],$$

si dicha esperanza existe en algún entorno de (0,0).

En caso de existencia, las funciones generatrices de momentos de las variables aleatorias marginales X e Y vienen dadas por:

$$g_X(t)=g(t,0) \qquad g_Y(t)=g(0,t).$$

Ejemplo 9.8.

Consideremos la variable aleatoria bidimensional con la siguiente función de probabilidad:

Y\X	**0**	**1**
2	1/4	1/4
4	3/8	1/8

Entonces:

$$g(t_1,t_2) = E[e^{t_1X+t_2Y}] = \sum_i \sum_j p_{ij} e^{t_1x_i+t_2y_j} =$$

$$= \frac{1}{4}e^{0t_1+2t_2} + \frac{3}{8}e^{0t_1+4t_2} + \frac{1}{4}e^{1t_1+2t_2} + \frac{1}{8}e^{1t_1+4t_2} =$$

$$= \frac{1}{4}e^{2t_2} + \frac{3}{8}e^{4t_2} + \frac{1}{4}e^{t_1+2t_2} + \frac{1}{8}e^{t_1+4t_2}$$

$$g_X(t) = g(t,0) = \frac{5}{8} + \frac{3}{8}e^t$$

$$g_Y(t) = g(0,t) = \frac{1}{2}e^{2t} + \frac{1}{2}e^{4t}$$

Ejemplo 9.9.

Dada la variable aleatoria bidimensional (X,Y) con función de probabilidad:

$$P[X=x,Y=y]=\frac{e^{-2\lambda}\lambda^{x+y}}{x!\,y!};x,y=0,1,2,\ldots$$

Entonces:

$$g(t_1,t_2)=E[e^{t_1X+t_2Y}]=\sum_{x=1}^{\infty}\sum_{y=1}^{\infty}P[X=x,Y=y]e^{t_1x+t_2y}=$$

$$=\sum_x\sum_y e^{t_1x+t_2y}\ \frac{e^{-2\lambda}\lambda^{x+y}}{x!\,y!}=e^{\lambda(e^{t_1}+e^{t_2}-2)}\qquad\forall t_1,t_2\in\Re$$

ya que $e^{\lambda}=\dfrac{\sum_x \lambda^x}{x!}$ y $e^{e^{t_1}\lambda}=\dfrac{\sum_x \left(e^{t_1}\lambda\right)^x}{x!}$

$$g_X(t)=g(t,0)=e^{\lambda(e^t-1)}\qquad\forall t\in\Re$$

$$g_Y(t)=g(0,t)=e^{\lambda(e^t-1)}\qquad\forall t\in\Re.$$

Ejemplo 9.10.

Sea la función de densidad de una variable aleatoria bidimensional:

$$f(x,y)=\begin{cases}e^{-x-y} & x>0,y>0\\ 0 & resto\end{cases}$$

Entonces:

$$g(t_1,t_2)=E[e^{t_1X+t_2Y}]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x+t_2y}f(x,y)dydx=$$

$$=\int_0^{\infty}\int_0^{\infty}e^{t_1x+t_2y}e^{-x-y}dydx=\frac{1}{1-t_1}\bullet\frac{1}{1-t_2}\qquad\text{para } t_1<1,\ t_2<1$$

Si existe la función generatriz de momentos de (X,Y), entonces los momentos bidimensionales no centrados de órdenes (r, s) vienen dados por:

$$\frac{\partial^{r+s}}{\partial t_1^r\partial t_2^s}g(t_1,t_2)\Big|_{t_1=t_2=0}=E[X^rY^S]\quad\text{para }\ r,s\in\aleph.$$

Ejemplo 9.11.

Si la función generatriz de momentos de una variable aleatoria bidimensional es:

$$g(t_1,t_2)=\frac{1}{4}e^{2t_2}+\frac{3}{8}e^{4t_2}+\frac{1}{4}e^{t_1+2t_2}+\frac{1}{8}e^{t_1+4t_2}$$

entonces:

$$EX=\frac{\partial}{\partial t_1}g(t_1,t_2)\ \Big|_{t_1=t_2=0}=\frac{1}{4}e^{t_1+2t_2}+\frac{1}{8}e^{t_1+4t_2}\ \Big|_{t_1+t_2=0}=\frac{3}{8}$$

y de igual forma se podría comprobar que $EY=3$.

$$EX^2=\frac{\partial^2}{\partial t_1^2}g(t_1,t_2)\ \Big|_{t_1=t_2=0}=\frac{1}{4}e^{t_1+2t_2}+\frac{1}{8}e^{t_1+4t_2}\ \Big|_{t_1=t_2=0}=\frac{3}{8}$$

$EY^2=10$

$$E[XY]=\frac{\partial^2}{\partial t_1\partial t_2}g(t_1,t_2)\ \Big|_{t_1=t_2=0}=\frac{2}{4}e^{t_1+2t_2}+\frac{4}{8}e^{t\ +4t_2}\ \Big|_{t_1=t_2=0}=1\,.$$

Veamos algunas propiedades más:

- Las variables aleatorias X e Y son independientes si y solo si su función generatriz de momentos conjunta puede expresarse como el producto de sus funciones generatrices de momentos marginales:
$$g(t_1,t_2)=g_X(t_1)g_Y(t_2)\quad \forall t_1,t_2\in\Re.$$
- Sean X e Y dos variables aleatorias independientes con funciones generatrices de momentos $g_X(t)$ y $g_Y(t)$ respectivamente. Entonces:
$$g_{X+Y}(t)=g_X(t)\cdot g_Y(t)\quad \forall\ t\in\Re.$$

Este último resultado tiene una especial aplicación para contrastar la reproductividad de distribuciones:

Una familia $\mathscr{F}$ de distribuciones se dice reproductiva si, dadas dos variables aleatorias independientes X e Y de la familia, su suma $X+Y$ también es de la familia, es decir, sigue una distribución de la misma familia que los sumandos.

9.8. Esperanza Condicionada.

Vemos por separado los casos discreto y continuo:

- Sean X e Y dos variables aleatorias discretas con función de probabilidad dada por $P[X=x,Y=y]=p(x,y)$.
Definimos:
$$E[X/Y]=\sum_x x\cdot P[X=x/Y=y]$$
$$E[Y/X]=\sum_y y\cdot P[Y=y/X=x]$$

- Sean X e Y dos variables aleatorias continuas con $f(x,y)$ su densidad conjunta. Entonces definimos:

$$E[X/Y] = \int_{-\infty}^{\infty} x \cdot f(x/y)dx$$

$$E[Y/X] = \int_{-\infty}^{\infty} y \cdot f(y/x)dy\,.$$

- Para cualquiera de los casos, definimos:

$$Var(X/Y)=E[X^2/Y]-E^2[X/Y]$$

9.9. Regresión Bidimensional.

Sea (X,Y) una variable aleatoria bidimensional. Nos planteamos determinar una relación entre las variables para poder realizar conjeturas sobre una de las variables teniendo información de la otra.

Consideremos la situación Y sobre X, indicando que pretendemos determinar una función de X que nos permite efectuar conjeturas sobre Y. La situación X sobre Y es totalmente simétrica.

Con las ideas que ya tenemos de Estadística Descriptiva, planteamos el objetivo con los dos siguientes puntos:

- Determinar la función de la variable aleatoria X, notada $h(x)$, que hace mínima la expresión $E[(y-h(x))^2]$ (Análogo a la regresión tipo I de Estadística Descriptiva).
- Determinar la función lineal de la variable aleatoria X, notada $aX+b$, que hace mínima la expresión $E[(y-aX-b)^2]$ (Análogo a la regresión lineal de Estadística Descriptiva).

9.9.1. Curva de Regresión.

Nos planteamos el primer problema:

Sea (X,Y) una variable aleatoria bidimensional. Minimizando $E[(Y-h(x))^2]$ respecto de $h(.)$, se llega a que $h(x)=E[Y/X]$, que es conocida como la curva de regresión de Y sobre X. Análogamente, la curva de regresión de X sobre Y es $g(y)=E[X/Y]$.

Ahora nos interesa cuantificar la proximidad de la distribución a la curva de regresión, que se hace mediante la razón de correlación de Y sobre X:

$$\eta^2_{Y/X} = 1 - \frac{E[(Y - E(Y/X))^2]}{\sigma^2_Y} = \frac{E[(E(Y/X) - E[Y])^2]}{\sigma^2_Y}$$

y, de forma análoga se definiría $\eta^2_{X/Y}$.

Ambas razones de correlación están entre 0 y 1 y, en general, son diferentes:

- Si $\eta^2_{Y/X} = 0$, todas las medias condicionadas son iguales entre sí y, por lo tanto, la curva de regresión es una recta paralela al eje de abscisas.
- Si $\eta^2_{Y/X} = 1$, entonces *Y*=*E*[*Y/X*], es decir, *Y* depende funcionalmente de *X*, lo que significa que dado un valor de *X*, *Y* presenta un valor constante.

Ejemplo 9.12.

Sea (*X,Y*) una variable aleatoria bidimensional con función de densidad conjunta:

$$f(x,y) = \frac{4}{3}(x + xy) \quad \text{con} \quad 0 \le x \le 1, \quad 0 \le y \le 1 \ .$$

Entonces:

$$f_X(x) = 2x \quad 0 \le x \le 1$$

$$f_Y(y) = \frac{2}{3}(1 + y) \quad 0 \le y \le 1$$

$$f(x/y) = 2x \qquad 0 \le x \le 1$$

$$f(y/x) = \frac{2}{3}(1 + y) \quad 0 \le y \le 1$$

Entonces, la curva de regresión de *Y* sobre *X* será:

$$E[Y/X = x] = \int_0^1 y \cdot f(y/x)dy = \int_0^1 y\frac{2}{3}(1 + y)dy = 5/9 \ .$$

En este caso, la curva de regresión es constante y, por tanto, $\eta^2_{y/x} = 0$

Ejemplo 9.13.

Sea (*X,Y*) una variable aleatoria bidimensional con función de densidad conjunta:

$$f(x,y) = \frac{6 - x - y}{8} \quad \text{con} \quad 0 \le x \le 2\,, \quad 2 \le y \le 4\,.$$

Entonces:

$$f_X(x) = \frac{1}{4}(3-x); 0 \le x \le 2$$

$$f_Y(y) = \frac{1}{4}(5-y); 2 \le y \le 4$$

$$f(x/y) = \frac{6-x-y}{10-2y}; 0 \le x \le 2$$

$$f(y/x) = \frac{6-x-y}{6-2x}; 2 \le y \le 4$$

Por tanto, la curva de regresión de *Y/X* será:

$$E[Y/X = x] = \int_2^4 y\frac{6-x-y}{6-2x}dy = \frac{-18x+52}{3(6-2x)}$$

Para determinar la función de correlación, necesitamos calcular:

$$E[Y^2] = \int_2^4 y^2 \frac{1}{4}(5-y)dy = 8.3325$$

$$E[Y] = \int_2^4 y\frac{1}{4}(5-y)dy = 2.835$$

$$\sigma_Y^2 = E[Y^2] - (E[Y])^2 = 0.2955$$

$$E[(E[Y/X] - E[Y])^2] = \int_0^2 \left[\frac{-18x+52}{3(6-2x)} - 2.835\right]^2 \frac{1}{4}(3-x)dx =$$

$$= \frac{1}{144}\int_0^2 \frac{(-0.99x+0.97)^2}{3-x}dx = 0.003$$

Por tanto:

$$\eta_{Y/X}^2 = \frac{0.003}{0.2955} = 0.01$$

por lo que el ajuste es malo.

9.9.2. Rectas de Regresión.

Nos planteamos ahora el segundo problema:

Tenemos que encontrar *a* y *b* tales que minimicen $E[(Y-(aX+b))^2]$. La solución del sistema es:

$$b = EY - EX \bullet Cov(X,Y) / Var(X)$$
$$a = Cov(X,Y) / Var(X)$$,

por lo que la recta de regresión de Y sobre X será:

$$Y - EY = \left[Cov(X,Y) / Var(X)\right]\left(X - EX\right)$$

y, análogamente, la recta de regresión de X sobre Y será:

$$X - EX = \left[Cov(X,Y) / Var(X)\right]\left(Y - EY\right)$$

Para cuantificar la proximidad de la distribución a la recta de regresión definimos el coeficiente de correlación lineal:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Ejemplo 9.13 (continuación).

$E[Y] = 2.835$

$Var[Y] = 0.2955$

$$E\left[X\right] = \int_0^2 x\frac{1}{4}(3-x)dx = 0.833$$

$$E\left[X^2\right] = \int_0^2 x^2\frac{1}{4}(3-x)dx = 1$$

$Var\ (X)\ =\ 1\text{-}\ 0.8333^2 = 0.3055$

$$E\left[XY\right] = \int_0^2\int_2^4 xy\frac{1}{8}(6-x-y)dydx = 2.33$$

$$Cov(X,Y) = 2.33 - 0.833 \bullet 2.035 = 0.032$$

$$y - 2.835 = \frac{0.032}{0.3055}\left(x - 0.833\right) \Rightarrow y = 0.105x + 2.748$$

$$\rho = \frac{0.032}{\sqrt{0.3055}\sqrt{0.2955}} = 0.10$$

por lo que el ajuste no es muy bueno.

RELACIÓN DE VARIABLE ALEATORIA BIDIMENSIONAL.

1.- Sea (X,Y) la variable aleatoria bidimensional discreta con función de probabilidad dada por:

X/Y	1	2	3
0	0	0	1/5
1	1/5	1/5	0
2	1/5	0	1/5

Calcular:
a) El coeficiente de correlación entre X e Y.
b) Las dos rectas de regresión.
c) El valor esperado de X cuando Y vale 2.

2.- Sea (X,Y) la variable aleatoria bidimensional con función de densidad conjunta:

$$f(x,y) = kxy^2 ; 0 < x < y < 1$$

Calcular:
a) El valor de la constante k para que $f(x,y)$ defina una función de densidad de una variable aleatoria bidimensional.
b) Las distribuciones marginales.
c) Las distribuciones condicionadas.
d) La recta de regresión de X sobre Y.
e) Bondad de ajuste.

3.- Al estudiar la relación entre dos variables se obtiene la función de densidad:

$$f(x,y) = \begin{cases} k(1-xy) & -1 \le x \le 1, -1 \le y \le 1 \\ 0 & resto \end{cases}$$

Hallar:
a) El valor de k.
b) Funciones de densidad y de distribución marginales.
c) Coeficiente de correlación entre X e Y.
d) Rectas de regresión.
e) Funciones de densidad condicionadas.
f) Curva de regresión de Y sobre X.

4.- Sea (X,Y) una variable aleatoria bidimensional con función de densidad conjunta:

$$f(x,y)=\begin{cases}8xy & 0<y<x<1\\ 0 & resto\end{cases}$$

Calcular:

a) Rectas de regresión y el coeficiente de correlación lineal.

b) ¿Se puede afirmar que dichas rectas son las funciones que mejor explican la relación existente entre las variables *X* e *Y*? En caso de no ser cierto, encontrar dichas funciones.

5.- Sea la función de densidad conjunta:

$$f(x,y)=k\frac{y^x}{x!}e^{-4y};\quad x=0,1,2,...;y>0$$

a) Calcular *k* para que sea función de densidad, así como las distribuciones marginales de *X* e *Y*.

b) Calcular las distribuciones condicionadas.

c) Calcular las curvas de regresión.

CAPÍTULO 10:

ALGUNAS DISTRIBUCIONES DE PROBABILIDAD IMPORTANTES.

En este capítulo vamos a tratar de ajustar una función de probabilidad a los distintos experimentos.

10.1. Distribuciones Discretas Unidimensionales.

Iniciamos el capítulo con el estudio de las distribuciones de variables aleatorias discretas más importantes:

10.1.1. La Distribución Uniforme Discreta sobre *N* puntos.

Sea X una variable aleatoria discreta que solo toma valores en N puntos y, además, toma todos esos valores con la misma probabilidad. En estas condiciones, decimos que la variable aleatoria discreta X sigue una distribución Uniforme discreta sobre N puntos y su función masa de probabilidad viene dada por:

$$p_k = P[X = x_k] = \frac{1}{N}$$

para los N distintos valores que toma la variable ($k = 1,...,N$).

Ejemplo 10.1.

- Consideramos el lanzamiento de una moneda equilibrada ($x_1 \equiv$ cara $= 0$, $x_2 \equiv$ cruz$=1$). Entonces tenemos una distribución Uniforme en dos puntos, con probabilidad ½ en cada uno.

$$P[X=0] = ½$$
$$P[X=1] = ½$$

- Consideramos la variable aleatoria que representa el resultado del lanzamiento de un dado equilibrado, que se distribuye según una Uniforme discreta con probabilidad 1/6 para cada uno de los 6 valores posibles.

$$P[X=i] = 1/6 \quad \text{para } i = 1,...,6.$$

10.1.2. La Distribución de Bernouilli.

Consideremos la experiencia de Bernouilli, que es un experimento en las siguientes condiciones:

- El resultado es la ocurrencia o no de un suceso A; y llamamos éxito a la ocurrencia de este suceso.
- Sea p la probabilidad de éxito cuando realizamos el experimento, $p = P(A)$, y sea q la probabilidad de fracaso, $q = P(\overline{A})$, por lo que se verifica que $p + q = 1$.
- Sea X una variable aleatoria discreta que toma el valor 0 cuando no ocurre el suceso A y toma el valor 1 cuando ocurre A.

En estas condiciones decimos que la variable aleatoria X sigue una distribución de Bernouilli de parámetro p, donde su función masa de probabilidad es:

$$P[X=0] = q \quad \text{y} \quad P[X=1] = p$$

10.1.3. La Distribución Binomial.

Consideremos un experimento en las siguientes condiciones:

- El resultado es la ocurrencia o no de un suceso A; y llamamos éxito a la ocurrencia de este suceso.

- Sea p la probabilidad de éxito cada vez que realizamos el experimento, $p = P(A)$, y sea q la probabilidad de fracaso, $q = P(\overline{A})$, por lo que se verifica que $p + q = 1$.
- Realizamos n veces el experimento de forma independiente, es decir, p es la probabilidad de éxito en cada una de las repeticiones.
- Sea X la variable aleatoria que mide el número de éxitos en las n pruebas, que correspondería a la realización n veces de la experiencia de Bernouilli.

Entonces decimos que la variable aleatoria X sigue una distribución Binomial de parámetros el número de veces que se realiza el experimento (n) y la probabilidad de éxito (p), es decir, $X \sim B(n,p)$.

Un ejemplo puede ser considerar la variable aleatoria X que estudia el número de unos obtenidos en 7 lanzamientos de un dado. En estas condiciones, la variable aleatoria X sigue una distribución Binomial de parámetros 7 y 1/6 ($X \sim B(7,1/6)$).

La función de probabilidad de la variable aleatoria discreta que sigue una distribución Binomial viene dada por:

$$P[X=x] = \binom{n}{x} p^x q^{n-x}.$$

Veamos cómo se obtiene dicha función de probabilidad:
Consideramos el caso de $n = 3$. Si realizamos 3 pruebas sucesivas e independientes y en las mismas condiciones, los resultados posibles son $AAA, AA\overline{A}, A\overline{A}A, \overline{A}AA, A\overline{A}\overline{A}, \overline{A}A\overline{A}, \overline{A}\overline{A}A, \overline{A}\overline{A}\overline{A}$ y, por lo tanto, su función masa de probabilidad está dada por:

$$P[X=3] = p^3, \quad P[X=2] = 3p^2q, \quad P[X=1] = 3pq^2, \quad P[X=0] = q^3$$

Si consideramos ahora n repeticiones del experimento un suceso con x éxitos y n-x fracasos, tiene de probabilidad $p^x q^{n-x}$, pero las ordenaciones posibles de las n letras $AA..^{x)}..A$, $\overline{A}\overline{A}..^{n-x)}..\overline{A}$ son permutaciones con repetición de n elementos con x y n-x repeticiones en cada una de las dos opciones posibles:

$$P_n^{x,n-x} = \frac{n!}{x!(n-x)!} = \binom{n}{x}$$

La distribución se llama distribución Binomial, ya que su nombre proviene de que estas probabilidades son los términos del desarrollo de $(p+q)^n$, por lo que se verifica que la suma de todas las probabilidades vale 1 ($p + q = 1$).

La función de distribución de una Binomial es:

$$F(x) = P[X \le x] = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$$

que está tabulada (si p no estuviera en la tabla se puede interpolar o sustituir en la función de probabilidad).

Una posibilidad cuando disponemos de una variable aleatoria discreta X que sigue una distribución Binomial(n,p) que estudia el número de éxitos obtenidos en n realizaciones de un experimento es estudiar la variable aleatoria Y 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 ($Y= n-X$) y seguirá una distribución también Binomial, pero con parámetros (n,q).

Aprendemos a utilizar la tabla de la función de distribución de la distribución Binomial con unos ejemplos:

- Sea $X\sim B(6,0.2)$. Entonces:
  - $P[X\leq 2] = 0.9011$
  - $P[X=3] = P[X\leq 3] - P[X\leq 2] = 0.9830 - 0.9011 = 0.0819$
  - $P[X\geq 3] = 1-P[X<3] = 1-P[X\leq 2] = 1-0.9011 = 0.0989$
- Sea $X\sim$B(8, 0.7). En este caso, $p = 0.7$ no viene en las tablas, pero sí viene $q = 1-p = 0.3$ y, utilizando la propiedad anterior, consideramos la variable aleatoria Y que estudia el número de fracasos y que sigue la distribución $Y = n-X\sim B(8,1-0.7) = B(8,0.3)$, que ya si aparece en las tablas. Entonces:
  - $P[X\leq 2] = P[Y\geq 6] = 1-P[Y<6] = 1-0.9887 = 0.0113$
  - $P[X=5] = P[Y=3] = P[Y\leq 3]-P[Y\leq 2] = 0.8059-0.5518 = 0.2541$
  - $P[X\geq 3] = P[Y\leq 5] = 0.9887$

Ejemplo 10.2.

Un club empieza una campaña telefónica para aumentar su número de socios. Se sabe, por experiencia, que 5 de cada 20 personas llamadas se unen al club. Si un día se llama a 15 personas, ¿cuál es la probabilidad de que al menos 2 se apunten?

Sea X= "nº de personas que se apuntan al club, de las 15 llamadas" $\sim B(15,0.25)$.

Entonces $P[X\geq 2] = 1-P[X\leq 1] = 1-0.0802 = 0.9198$

Las principales características de la distribución Binomial de parámetros n y p son:

- Comenzaremos determinando la función generatriz de momentos:

$$g(t) = E[e^{tx}] = \sum_{x=0}^{n} e^{tx} p(x) =$$

$$= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n$$

- Calculemos ahora esperanza y varianza a partir de la función generatriz de momentos:

$$g'(t) = n\,(pe^t+q)^{n-1}\, p\, e^t$$

Entonces:

- $E[X] = g'(0) = np$
- $g''(t) = n\,(n-1)\,(pe^t+q)^{n-2}\,(pe^t)^2 + n\,(pe^t+q)^{n-1} pe^t$

$$g''(0) = np + np^2\,(n-1) = E[X^2].$$

$$Var(X) = E[X^2]-(E[X])^2 = np + n^2p^2 - np^2 - n^2p^2 = np(1-p) = npq$$

Basándonos en la función generatriz de momentos, podríamos demostrar la llamada propiedad de reproductividad o aditividad que dice:
Sean dos variables aleatorias con distribución Binomial con la misma probabilidad de éxito $X \sim B(n,p)$ e $Y \sim B(m,p)$ independientes. Entonces la variable aleatoria suma de las dos sigue también una distribución Binomial de parámetros $X+Y \sim B(n+m,p)$.

Ejemplo 10.3.

Sean X= "nº de caras en 10 lanzamientos" $\sim B(10,1/2)$,
Y= "nº de veces que sale par en 3 lanzamientos de un dado" $\sim B(3,1/2)$.
Entonces, la variable aleatoria suma $X+Y \sim B(13,1/2)$.

Un resultado importante para la distribución Binomial es el Teorema de Bernouilli, que dice:

La probabilidad de que la variable aleatoria $X \sim B(n, p)$ 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 $\frac{pq}{n\varepsilon^2}$, es decir,

$$P\left[\left|\frac{X}{n} - p\right| \geq \varepsilon\right] \leq \frac{pq}{n\varepsilon^2}.$$

Además, si no conocemos p, será necesario calcular un extremo superior de $\frac{pq}{n\varepsilon^2}$ y, por lo tanto, podemos expresar el teorema como:

$$P\left[\left|\frac{X}{n} - p\right| \geq \varepsilon\right] \leq \frac{1}{4n\varepsilon^2}$$

ya que $p \cdot q$ está acotado por el valor máximo, que lo toma en 1/4.

Ejemplo 10.4.

¿Cuántas veces tendremos que realizar un experimento para tener una probabilidad de 0.95 ó mayor de que la frecuencia relativa difiera de p en menos de 0.02?

Nos pide n tal que $P\left[\left|\frac{X}{n} - p\right| < 0.02\right] \geq 0.95$ (no conocemos p).

Entonces:

$0.95 = 1 - 1/4\ n\varepsilon^2$, por lo que, despejando, obtenemos que $n = 12500$ repeticiones.

10.1.4. La Distribución de Poisson.

Sea X una variable aleatoria que representa el número de sucesos aleatorios independientes que ocurren a velocidad constante sobre el tiempo o el espacio, y sea λ el número medio de ocurrencias por unidad de tiempo o espacio.

En estas condiciones, la variable aleatoria X, que puede tomar todos los valores enteros positivos (0,1,2,...), se dice que sigue una distribución de Poisson de parámetro λ ($X \sim \mathrm{P}(\lambda)$), y su función masa de probabilidad es:

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}$$

Como $\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$, esta distribución es una verdadera función de probabilidad.

Esta distribución 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”.

La función de distribución está tabulada y es:

$$F(x) = P[X \leq x] = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^i}{i!}$$

Aprendemos a usar las tablas con un ejemplo:

- Sea $X \sim \mathrm{P}(4)$;
    - $P[X \leq 2] = 0.2381$
    - $P[X \geq 5] = 1 - P[X \leq 4] = 1 - 0.6288 = 0.3712$

- $P[X=6] = P[X\leq 6]-P[X\leq 5] = 0.8893-0.7851 = 0.1042$

Ejemplo 10.5.

Una centralita recibe una media de 6 llamadas por minuto. Si la centralita puede atender 8 llamadas por minuto como máximo, ¿cuál es la probabilidad de que en un minuto no se puedan atender todas las llamadas?

X =”nº de llamadas por minuto”~P(6).

$$P[X>8] = 1-P[X\leq 8] = 1-0.8472 = 0.1528$$

Las principales características de la distribución de Poisson son:

- Función generatriz de momentos:

  $g(t) = E[\, e^{tx}] =$

$$= \sum_{x=0}^{\infty} e^{tx} P[X = x] = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

- Esperanza Matemática:

  $g'(t) = e^{\lambda(e^t-1)}\lambda e^t$. Entonces:

$$E[X] = g'(0) = \lambda.$$

- Varianza:

$$g''(t) = e^{\lambda(e^t-1)}(\lambda e^t)^2 + e^{\lambda(e^t-1)}\lambda e^t \;;$$

$$g''(0) = \lambda^2 + \lambda\,;$$

$$Var\,(X) = E[X^2] -(E[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

En este caso también tenemos la propiedad de reproductividad o aditividad, que dice:

Dadas dos variables aleatorias con distribución de Poisson (X~P(λ_1) e Y~P(λ_2)) independientes, entonces la variable aleatoria suma de las dos es otra distribución de Poisson con parámetro:

$$X + Y \sim \mathrm{P}\,(\lambda_1+\lambda_2).$$

Se demuestra que la distribución de Poisson es limite de la distribución Binomial, en determinadas condiciones, es decir, si la variable aleatoria X~$B(n,p)$ con $n \geq 25$, $p\leq 0.1$ y $np < 5$, entonces podemos aproximar X por una distribución de Poisson de parámetro $\lambda = np$ para calcular probabilidades de los valores de X, .

Por ejemplo si X~B(50,0.02) , entonces podemos aproximar por $X \cong$ P(1).

10.1.5. La Distribución Hipergeométrica.

Consideremos un experimento en el que:

- El número de elementos de la población es *N*.
- Llamamos éxito a la ocurrencia de un suceso.
- Sabemos que en la población hay *k* éxitos, por lo que la probabilidad de éxito vale $p = \dfrac{k}{N}$.
- Tomamos una muestra de *n* individuos, pero de forma que los extraemos de uno en uno, y el individuo extraído se deja fuera de la población, con lo cual las extracciones no son independientes y, por lo tanto, no estamos en el caso de la distribución Binomial. Tenemos muestreo sin reemplazamiento, mientras que en la Binomial se tiene muestreo con reemplazamiento.
- Sea *X* la variable aleatoria que estudia el número de éxitos en las *n* extracciones.

Entonces *X* sigue una distribución Hipergeométrica de parámetros *N, n* y *k*, que notaremos *X*~H(*N, n, k*) y cuya función masa de probabilidad viene dada por:

$$P[X = x] = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x = 0,1,...Min(n,k)$$

Utilizando propiedades de números combinatorios se comprueba que es una verdadera función masa de probabilidad.

La función de distribución viene dada por:

$$F(x) = P[X \leq x] = \sum_{i=0}^{x} \frac{\binom{k}{i}\binom{N-k}{n-i}}{\binom{N}{n}}$$

En este caso están tabuladas tanto la función de probabilidad como la función de distribución, aunque también podríamos calcular las probabilidades utilizando la fórmula.

Veamos mediante un ejemplo como se utiliza la tabla:

- Sea *X*~ H(8,4,3).
    - $P[X \leq 1] = 0.5 = P[X=0] + P[X=1] = 0.071428+0.4285$

Ejemplo 10.6.

Supongamos que tenemos una urna con 10 bolas y consideremos éxito sacar bola blanca.

Sabemos que hay 3 bolas blancas. Sacamos 6 bolas sin reemplazamiento.

Sea X = nº de bolas blancas obtenidas de entre las 6 que hemos extraído.

Entonces $X \sim$ H(10,6,3).

$$P[X=1] = 0.3$$

Las principales características de la distribución Hipergeométrica son:

$$EX = \frac{nk}{N} = np$$

$$Var(X) = \frac{nk(N-n)(N-k)}{N^2(N-1)} = \frac{N-n}{N-1}npq \ .$$

Veamos la diferencia existente entre una distribución Binomial y una Hipergeométrica:

En la Binomial, la probabilidad de que un individuo verifique el suceso éxito (A) es constante; mientras que en la Hipergeométrica, dicha probabilidad varía de extracción en extracción. De todas formas, la Hipergeométrica se puede aproximar mediante una Binomial (H(N,n,k)$\rightarrow$$B$($n,p$)) cuando el tamaño de la población sea ∞ ó muy grande (suele bastar con que $\frac{n}{N} < 0.1$),

Sus principales áreas de aplicación son:

El control estadístico de calidad y el muestreo de aceptación (N es el número de unidades de un lote, n el tamaño de la muestra tomada, k el número de defectuosas tomadas, y vamos a aceptar el lote o no, dependiendo del número de unidades defectuosas obtenidas en el muestreo); también se utiliza en las encuestas sobre un determinado producto,...

Ejemplo 10.7.

Un fabricante de coches compra los motores a una empresa. El fabricante recibe un lote de 40 motores. Su plan para aceptar el lote es seleccionar 8 motores de manera aleatoria y someterlos a una prueba. Si ninguno de los seleccionados falla, acepta el lote.

Si el lote tiene 2 motores defectuosos, la probabilidad de que el lote sea aceptado es:

X ="nº de motores defectuosos"~ H(40,8,2).

Entonces:

$$P[X=0]=\frac{\binom{2}{0}\binom{40-2}{8-0}}{\binom{40}{8}}=0.6359$$

10.1.6. La Distribución Geométrica.

Consideramos un experimento tal que:

- Llamamos éxito a la ocurrencia de un suceso *A*.
- Sea p la probabilidad de éxito ($p = P(A)$ y $q = P(\overline{A}) = 1\text{-}p$).
- Sea una sucesión de pruebas independientes del experimento, que hacemos hasta obtener el primer éxito.
- Consideramos la variable aleatoria *X* que mide el número de fracasos antes del primer éxito (no el número de veces que realizamos el experimento) y el último resultado obtenido es, por lo tanto, éxito.

Entonces, decimos que la variable aleatoria *X* sigue una distribución Geométrica de parámetro *p* ($X \sim \mathrm{Ge}(p)$), cuya función de masa de probabilidad está dada por:

$$P[X=x] = p(1\text{-}p)^x \qquad x = 0,1,2,....$$

Veamos que se trata de una verdadera función de probabilidad. Para ello comprobamos que la suma de las probabilidades vale 1:

$P[X=0]+P[X=1]+P[X=2]+...+P[X=k]+... = p(1+q+q^2+...+q^k+...) =$

(puesto que es la suma de una serie geométrica de razón q<1 $\Rightarrow$S= $\frac{a_1}{1-r}$)

$$= p\frac{1}{1-q} = \frac{p}{p} = 1.$$

La variable aleatoria *Y* que estudia el número de realizaciones del experimento hasta obtener el primer éxito es igual a $X+1$.

Las principales características de la distribución Geométrica son:

- Función generatriz de momentos:

$$g(t) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx} P[X=x] = \sum_{x=0}^{\infty} e^{tx} pq^x = p\sum_{x=0}^{\infty} (e^t q)^x = \frac{p}{1-qe^t}$$

con $|qe^t| < 1$

- Esperanza:

$$g'(t) = \frac{pqe^t}{(1-qe^t)^2}$$

$$g'(0) = \frac{pq}{(1-q)^2} = \frac{q}{p} = E[X]\,.$$

- Varianza:

$$g''(t) = \frac{pqe^t(1-qe^t)^2 - pqe^t\,2(1-qe^t)(-qe^t)}{(1-qe^t)^4}$$

$$g''(0) = \frac{q^2}{p^2} + \frac{q}{p^2} = E[X^2]$$

$$Var(X) = E[X^2]\text{-}E[X]^2 = \frac{q}{p^2}$$

Ejemplo 10.8.

Consideremos el experimento que consiste en el lanzamiento de un dado hasta obtener un 6.
Y = " nº de lanzamientos de un dado hasta obtener el primer 6".
X ="nº de lanzamientos de un dado antes de obtener un 6" ~ Ge(1/6).
Entonces:

$$P[Y=3] = P[X=2] = 1/6\ (\ 1\text{-}1/6)^2 = 0.1157$$

10.1.7. La Distribución Binomial Negativa o de Pascal.

Consideremos un experimento aleatorio tal que:

- Llamamos éxito a la ocurrencia de un suceso A.
- Sea p la probabilidad de éxito (p = P(A) y q = P($\overline{A}$) = 1-p).
- Sea una sucesión de pruebas independientes del experimento, que hacemos hasta obtener el éxito número k.
- Sea X la variable aleatoria que mide el número de fracasos antes del k-ésimo éxito, con k un número entero positivo (no el número de pruebas) y el último resultado obtenido es, por tanto, el k-ésimo éxito.

Entonces decimos que la variable aleatoria X sigue una distribución Binomial Negativa de parámetros k y p (X~BN(k,p)), donde p es la probabilidad de éxito y k el número de éxitos que tiene que haber.

Su función de masa de probabilidad está dada por:

$$P[X=x] = \binom{x+k-1}{x} p^k (1-p)^x, \qquad x = 0,1,2...$$

Por propiedades de números combinatorios se comprueba que es una verdadera función masa de probabilidad.

Su función de distribución viene dada por:

$$F(x) = P[X \le x] = \sum_{i=0}^{x} \binom{k+i-1}{i} p^k (1-p)^i \qquad x=0,1,2...$$

Para $k=1$ tenemos la distribución Geométrica.

Tenemos la propiedad de reproductividad de la distribución Binomial Negativa, que nos dice que:

- Si tenemos k variables aleatorias que siguen la misma distribución Geométrica (X_i ~Ge(p)) independientes, entonces la suma de dichas variables aleatorias sigue una distribución Binomial Negativa: $\sum_{i=1}^{k} X_i \sim$ BN(k,p); y toda variable aleatoria con distribución Binomial Negativa se puede escribir como suma de k variables aleatorias Geométricas con el mismo parámetro de probabilidad.
- La suma de dos variables aleatorias Binomial Negativa con la misma probabilidad de éxito (Y~BN(l,p) y X~BN(k,p)), es otra variable aleatoria Binomial Negativa, con parámetros la suma de los éxitos y la probabilidad de éxito, es decir, $X+Y \sim$ BN($k+l,p$).

  Por tanto, las características de la distribución Binomial Negativa están dadas por:

$$g(t) = \left(\frac{p}{1-qe^t}\right)^k \qquad \text{con } |qe^t| < 1$$

$$EX = \frac{kq}{p}.$$

$$Var(X) = \frac{kq}{p^2}.$$

Si X sigue una distribución Binomial Negativa (X~BN(k,p)), la variable aleatoria $Y = k+X$ indica el número de realizaciones del experimento hasta el k-ésimo éxito, que es una distribución de Pascal en la que $P[Y=m] = P[X=m-k]$, y cuyas características se obtienen mediante un cambio de variable a partir de las de X.

Ejemplo 10.9.

La probabilidad de realizar 12 lanzamientos de un dado para obtener el tercer 6 es:

Y = " número de lanzamientos hasta obtener el tercer 6".

X = "nº de fracasos hasta el tercer 6"~ BN(3,1/6).

$$P[Y=12] = P[X=9] = \binom{9+3-1}{9} (1/6)^3 (5/6)^9 = 0.0493$$

Podemos realizar una aproximación a la distribución Binomial de la siguiente forma:
Si X es una variable aleatoria que estudia el número de fracasos antes del k-ésimo éxito y sigue una distribución BN(k,p), entonces la $P[X=x]$ se puede aproximar mediante $P[Y=k]$, donde Y es una variable aleatoria que mide el número de éxitos cuando el experimento se realiza $k+x$ veces, y que sigue una distribución Binomial de parámetros ($k+x,p$).

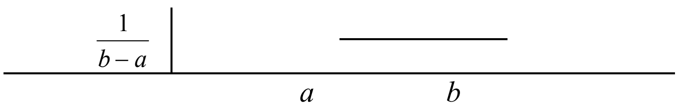
10.2. Distribuciones Continuas Unidimensionales.

Continuamos el tema con el estudio de las distribuciones de variables aleatorias continuas más importantes:

10.2.1. La Distribución Uniforme.

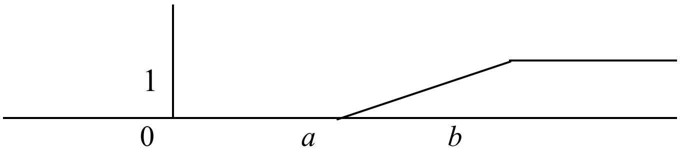
Sea una variable aleatoria continua que toma valores en un intervalo finito, de forma que estos se encuentran igualmente distribuidos sobre el intervalo, es decir, la probabilidad de que la variable aleatoria tome un valor en cada subintervalo de igual longitud es la misma. Entonces decimos que X está uniformemente distribuida en su intervalo de valores (a,b) y se dice que $X \sim U(a,b)$, y su función de densidad está dada por:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & resto \end{cases}$$



Se puede comprobar que es una verdadera función de densidad, es decir que se verifica que:
a) $f(x) \geq 0 \quad \forall x \in \Re$.
b) $\int_{-\infty}^{\infty} f(x)dx = 1$
y, por lo tanto, su función de distribución es:

$$F(x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$



Sus principales características son:

- $$EX = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a}\frac{(b^2-a^2)}{2} =$$
$$= \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2}$$
- $$\text{E}[\text{X}^2] = E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_a^b \frac{x^2}{b-a}dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)} =$$
$$= \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$
- $$Var(X) = E[X^2] - (EX)^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Su principal área de aplicación es proporcionar una representación adecuada para redondear diferencias que surgen al medir entre los valores observados y los reales (si el peso lo redondeamos al Kg. más cercano, la diferencia con el peso verdadero estará entre -0.5 y 0.5 Kg).

Ejemplo 10.10.

Una persona tarda entre 20 y30 minutos en llegar al trabajo. ¿A qué hora debe salir para tener una probabilidad 0.9 de no llegar tarde si entra a las 8 h.?

X = tiempo que tarda en ir al trabajo ~ $U(20,30)$

$P[X \leq x] = 0.9 \Rightarrow \dfrac{x-20}{30-20} = 0.9 \Rightarrow x = 29$ (debe salir a las 7:31 h.).

10.2.2. La Distribución Normal.

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...

Comenzaremos estudiando la distribución Normal de parámetros 0 y 1, para ver, a continuación, el caso general, ya que veremos que el caso general siempre se va a poder trasladar a este caso particular.

A) Normal (0,1).

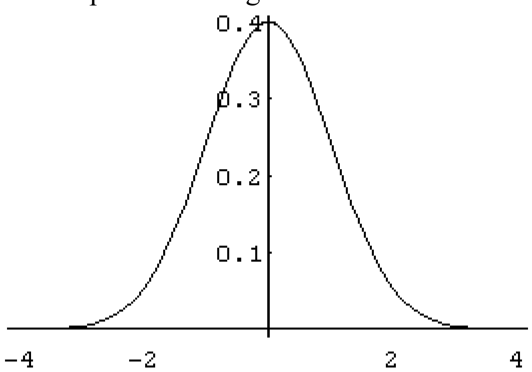
Una variable aleatoria continua Z se dice que sigue una distribución $N(0,1)$ si su función de densidad viene dada por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{para todo valor } z \in \Re.$$

Podemos observar que verifica las siguientes propiedades:

- La curva es simétrica respecto al eje de ordenadas: $f(z) = f(-z)$.
- La función tiene una asíntota horizontal en $y = 0$ cuando $z \to \pm\infty$.
- Para $z > 0$ la curva es decreciente, y para $z < 0$ es creciente.
- Para $z = \pm 1$ tenemos dos puntos de inflexión.

Por lo tanto su representación gráfica sería:



Las principales características son:

- La función generatriz de momentos es:

$$g_z(t) = E[e^{tz}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tz} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tz-z^2/2} dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{(2tz-z^2)/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-(z^2-2tz+t^2)}{2}+\frac{t^2}{2}} dz =$$

$$= \frac{1}{\sqrt{2\pi}} \bullet e^{t^2/2} \int_{-\infty}^{+\infty} e^{-(z-t)^2/2} dz = \begin{vmatrix} y = z - t \\ dy = dz \end{vmatrix} =$$

$$= \frac{1}{\sqrt{2\pi}} \bullet e^{t^2/2} \int_{-\infty}^{+\infty} e^{-y^2/2} dy = e^{t^2/2}$$

- La esperanza vale, puesto que $EZ = g'(0)$:

  $g'(t) = e^{t^2/2} t$.

  Entonces:

$$EZ = 0$$

- Varianza: Como $Var(Z) = EZ^2 - (EZ)^2$:

  $EZ^2 = g''(0)$

  $g''(t) = e^{t^2/2} t^2 + e^{t^2/2}$.

  Entonces $EZ^2=1$, y, por lo tanto,

$$Var(Z) = 1$$

A esta distribución se le llama de Gauss.

B) Normal (μ,σ).

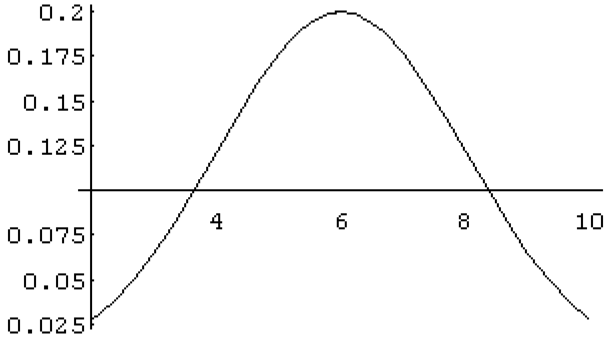
Una variable aleatoria continua X se dice que tiene una distribución $N(\mu,\sigma)$ si su función de densidad viene dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{para todo valor } x \in \Re.$$

Podemos observar que verifica las siguientes propiedades:

- La curva es simétrica respecto al eje vertical $x = \mu$.
- La función tiene una asíntota horizontal en $y = 0$ cuando $x \to +\infty$ ó $-\infty$.
- Para $x > \mu$, la curva es decreciente y para $x < \mu$, es creciente.
- En $x = \mu-\sigma$ y $x = \mu+\sigma$ la función tiene dos puntos de inflexión.

Por lo tanto su representación gráfica es:



Normal(6,2)

Consideramos una variable aleatoria $X \sim N(\mu,\sigma)$. Entonces la variable aleatoria Z definida como $Z = \dfrac{X-\mu}{\sigma}$, es decir, la variable aleatoria X tipificada, sigue una distribución $N(0,1)$, y, puesto que ya conocemos la función generatriz de momentos, la esperanza y la varianza de la $N(0,1)$, a partir de ellos vamos a calcular dichas características de la distribución $N(\mu,\sigma)$:

- La función generatriz de momentos:

  Sabiendo que $g_Z(t) = e^{t^2/2}$, calculamos

  $$g_X(t) = E[e^{x}] = E[e^{t(z\sigma+\mu)}] = e^{t\mu}E[e^{\sigma t z}] = e^{t\mu}g_Z(t\sigma) = e^{t\mu}e^{t^2\sigma^2/2} = e^{t\mu+\frac{t^2\sigma^2}{2}} \quad \forall t$$

- La esperanza:

  Puesto que $EZ = 0$, entonces

  $$EX = E[\sigma z+\mu] = \sigma E[Z] + \mu = \mu$$

- La varianza:

  Puesto que $Var[Z] = 1$, entonces

  $$Var[X] = Var[\sigma z+\mu] = \sigma^2 Var(Z) = \sigma^2$$

La función de distribución de una variable aleatoria con distribución $N(\mu,\sigma)$ viene dada por:

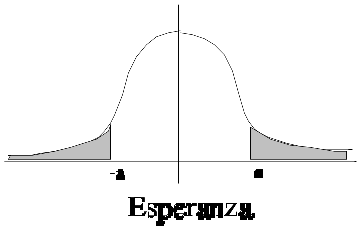
$$F(x) = P[X \le x] = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\Pi}} e^{-(x-\mu)^2/2\sigma^2}\,dx \quad \text{para todo } x \in \Re$$

y coincide con el área que deja la función de densidad entre $-\infty$ y x.

La función de distribución está tabulada, pero como no es posible disponer de una tabla para cada $N(\mu,\sigma)$, tenemos sólo la tabla para una distribución $N(0,1)$ y, tipificando, podemos pasar cualquier otra distribución a una $N(0,1)$.

Como la distribución $N(0,1)$ es simétrica respecto de cero, podemos calcular las probabilidades de que la variable sea mayor que un determinado valor como:

$$P[Z>a] = P[Z<-a]$$



Veamos mediante algunos ejemplos cómo se busca en la tabla:

- Si $Z \sim N(0,1)$, veamos cuánto vale:
  a) $P[Z< 0.3] = 0.6179$
  b) $P[Z< -0.76] = 0.2236$
  c) $P[Z>1.32] = P[Z<-1.32] = 0.0934$ ó
  $= 1-P[Z<1.32] = 1-0.9066 = 0.0934$
  d) $P[Z>-2.01] = P[Z<2.01] = 0.9778$ ó
  $=1-P[Z<-2.01] = 1-0.0222 = 0.9778$
- Si $X \sim N(2,3)$, veamos lo que vale:

$$P[X \le 6] = P\left[\frac{X-2}{3} \le \frac{6-2}{3}\right] = P[Z \le 4/3] = 0.9082$$

- Si $X \sim N(1,4)$, tenemos que:

$$P[X < 2] = P\left[Z < \frac{2-1}{4}\right] = P[Z < 0.25] = 0.5987$$

Vemos ahora la propiedad de reproductividad de la distribución Normal:

Sean n variables aleatorias normales con distintos parámetros $X_i \sim N(\mu_i, \sigma_i)$ e independientes. Entonces la variable aleatoria formada por la suma de dichas variables multiplicada cada una de ellas por un escalar es:

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i ; \sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2}\right)$$

Ejemplo 10.11.

La altura de los individuos en edad militar sigue una $N(170,10)$.

a) ¿Qué proporción de individuos mide más de 200 cm?

$X \sim N(170,10)$;

$$P[X > 200] = P\left[Z > \frac{200-170}{10}\right] = P[Z > 3] = P[Z < -3] = 0.0013$$

b) Si no se admite al 20% de los más altos, ¿qué limites hay que poner?

$P[X>B] = P[Z>b] = 0.2 \Rightarrow P[Z<-b] = 0.2 \Rightarrow -b = -0.84 \Rightarrow b = 0.84$

$$b = \frac{B-\mu}{\sigma} = \frac{B-170}{10} \Rightarrow 0.84 \cdot 10 = B - 170 \Rightarrow B = 178.4$$

10.2.3. La Distribución Gamma.

Una variable aleatoria continua X se dice que sigue una distribución Gamma de parámetros α y β ($G(\alpha,\beta)$) si su función de densidad viene dada por:

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta} \quad \text{para todo valor } x \in \Re^{+},$$

donde:

$\alpha > 0$ es un parámetro de forma,

$\beta > 0$ es la frecuencia constante de ocurrencia, y

$\theta = 1/\beta$ es el tiempo promedio entre dos ocurrencias consecutivas.

Como podemos observar en la definición de función de densidad aparece la función analítica del mismo nombre:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \ \alpha > 0$$

que verifica las siguientes propiedades:

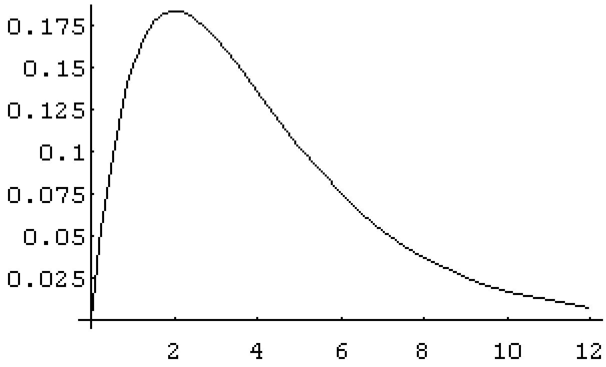
a) Si $n \in \aleph$, entonces $\Gamma(n) = (n-1)!$.

b) $\Gamma(½) = \sqrt{\pi}$.

c) $\Gamma(t+1) = t \cdot \Gamma(t)$.

La distribución Gamma tiene muchas aplicaciones en ingeniería, como por ejemplo:

Consideramos una pieza metálica que está sometida a cierta fuerza de manera que se romperá después de aplicar un número específico de ciclos de fuerza. Si estos ciclos ocurren de manera independiente y con una frecuencia promedio, el tiempo que debe transcurrir antes de que el material se rompa es una variable tipo Gamma.



Función de densidad de la Gamma(2,2)

Ejemplo 10.12.

Un muro de contención de una presa sabemos que se romperá después de sufrir una 2ª embestida de una riada, y sabemos que el número de riadas por siglo sigue una variable aleatoria de Poisson de promedio 4. ¿Cuál es la probabilidad de que dicho muro dure más de 20 años?

Sea X = tiempo que tarda en producirse la 2ª embestida ~ G(2,1/25).

Nos pide que calculemos la $P[X \geq 20]$.

Las principales características de la distribución Gamma son:

- La función generatriz de momentos vale:

$$g(t) = \frac{\beta^\alpha}{(\beta - t)^\alpha} = \left(\frac{1}{1-\theta t}\right)^\alpha \qquad |t| < \beta.$$

- La esperanza vale:

$$EX = \alpha/\beta = \alpha\theta$$

- La varianza vale:

$$Var(X) = \alpha/\beta^2 = \alpha\theta^2$$

Veamos la propiedad de reproductividad de la distribución Gamma:

Sean un conjunto de variables aleatorias con distribución Gamma $X_i \sim G(\alpha_i, \beta)$ independientes. Entonces la variable aleatoria suma sigue una distribución:

$$\sum_{i=1}^{n} X_i \sim G\left(\sum_{i=1}^{n} \alpha_i, \beta\right)$$

Vemos ahora un par de casos particulares de la distribución Gamma:

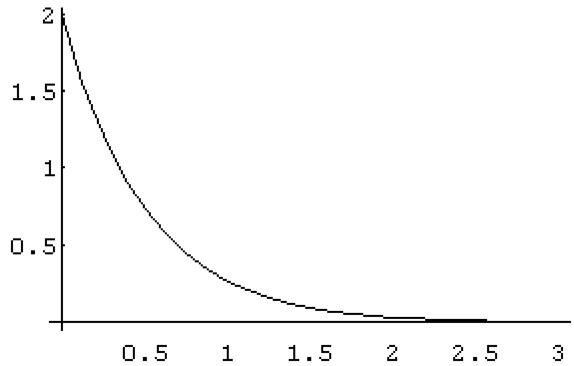
A) Erlang.

Si α es un número natural n y β vale $n\cdot\mu$, esta distribución Gamma recibe el nombre de distribución de Erlang de parámetros n y μ, y describe la distribución de la duración del intervalo de tiempo hasta la aparición de n sucesos de un proceso de Poisson de parámetro μ.

B) Exponencial Negativa.

Cuando $\alpha = 1$, tenemos la distribución Exponencial Negativa de parámetro β, que estudia fundamentalmente tiempos de vida, y cuyas propiedades principales son:

- $f(x) = \beta e^{-\beta x}$ para todo valor $x \in \Re^+$.
- $F(x) = \begin{cases} 0; x < 0 \\ 1 - e^{-\beta x}; x \in \Re^+ \end{cases}$
- $EX = 1/\beta$.
- $Var(X) = 1/\beta^2$.
- La propiedad de reproductividad para variables aleatorias Exponenciales Negativas es:
  Si tenemos n variables aleatorias Exponenciales Negativas con el mismo parámetro β e independientes, entonces la variable aleatoria $\sum_{i=1}^{n} X_i$ sigue una distribución G(n,β).



Función de densidad de la Exponencial Negativa(2).

10.2.4. La Distribución Beta (β).

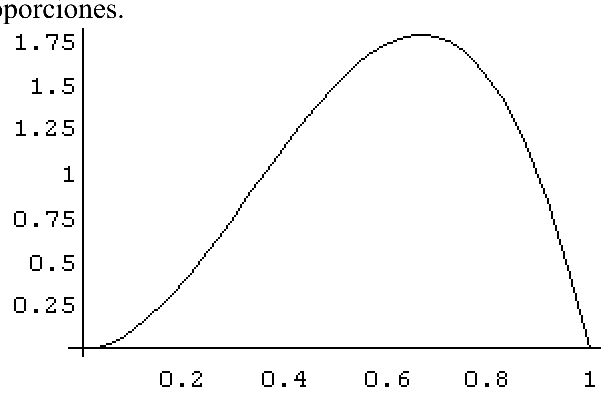
Una variable aleatoria X sigue una distribución $\beta(a,b)$ con $a>0$ y $b>0$ si su función de densidad viene dada por:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \text{ , para } 0 < x < 1.$$

Sus principales características son:

- $EX = \dfrac{a}{a+b}$.
- $Var(X) = \dfrac{ab}{(a+b+1)(a+b)^2}$.

Como su rango varía entre 0 y 1, es útil como modelo probabilístico que representa proporciones.



Función de densidad de la Beta(2,3)

10.2.5. Distribuciones asociadas a la Distribución Normal.

Estudiamos ahora algunas distribuciones continuas que están asociadas a la distribución Normal.

A) La Distribución χ^2.

Decimos que una variable aleatoria continua X sigue una distribución χ^2 con n grados de libertad si, dadas n variables aleatorias Normales(0,1) e independientes ($\left\{X_i\right\}_{i=1,\dots,n}$ ~$N(0,1)$), la variable aleatoria X es la suma de los cuadrados de las n variables aleatorias anteriores:

$$X = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n),$$

por lo que toma cualquier valor real positivo.

También se podía haber definido como una distribución Gamma en la cual $\alpha = \frac{n}{2}$ y $\beta = \frac{1}{2}$, es decir, X~G(n/2;1/2).

Por lo tanto, su función de densidad es:

$$f(x) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2} \;\; ; \;\; \text{para todo } x>0 \text{ con } n \in Z^+.$$

Sus principales características son:

- La función generatriz de momentos es:

$$g(t) = (1-2t)^{-n/2}$$

- La esperanza vale:

$$EX = n$$

- La varianza vale:

$$Var(X) = 2n$$

(Se pueden comprobar sin nada más que sustituir α y β por su valor en las características de la distribución Gamma)

La propiedad de reproductividad es:

Sean un conjunto de variables aleatorias $X_i \sim \chi^2(k_i)$ independientes. Entonces la variable aleatoria suma sigue una distribución:

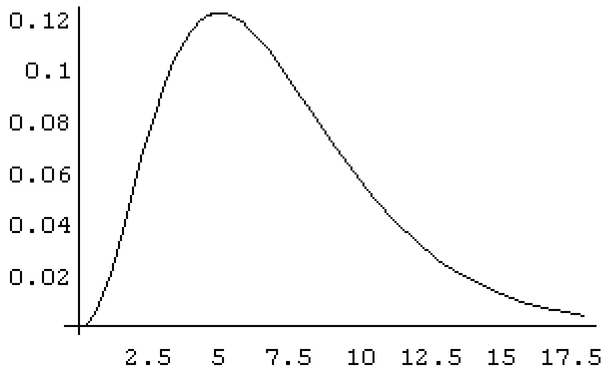
$$\sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} k_i\right).$$

Los valores de la función de distribución están tabulados $P[X \leq \chi^2_{n,p}] = p$. Veamos cómo se trabaja con las tablas mediante un ejemplo:

- Sea una variable aleatoria $X \sim \chi^2(5)$. Entonces:

  $P[X \leq k] = 0.025$. Entonces $k = 0.83 = \chi^2_{5,0.025}$

  $P[X \leq 9.24] = p$. Entonces $p = 0.9$ ($9.24 = \chi^2_{5,0.9}$)



Función de distribución de la $\chi^2(7)$

B) La Distribución *t* de Student.

Sean una variable aleatoria X~Normal (0,1) y una variable aleatoria $Y \sim \chi^2(n)$ independientes. Entonces la variable aleatoria T definida como

$$T = \frac{X}{\sqrt{Y/n}}$$

que puede tomar cualquier valor real, sigue una distribución t de Student con los mismos grados de libertad que la variable Y ($T \sim t(n)$).

Sus principales características son:

$$ET = 0$$

$$Var(T) = \frac{n}{n-2}$$

La función de densidad es simétrica respecto al eje de ordenadas ($t_{n,1-p} = -t_{n,p}$).

Los valores de la función de distribución están tabulados ($P[T < t_{n,p}] = p$). Veamos cómo se utiliza la tabla a partir de un ejemplo:

- Sea la variable aleatoria $X \sim t(20)$. Entonces:

  $P[X \leq k] = 0.8 \Rightarrow k = 0.86 = t_{20,0.8}$.

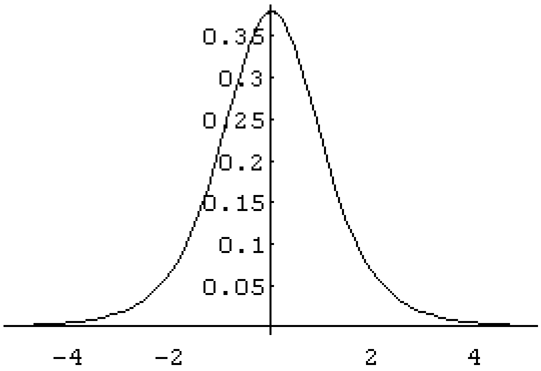
  $P[X \leq 2.086] = p \Rightarrow p = 0.975$.

  $P[X \leq k] = 0.025 \Rightarrow k = -2.086 = t_{20,0.025}$

  $P[X \leq -2.845] = p \Rightarrow p = 0.005$.

  $t_{1,0.8} = 1.376$.

  $t_{1,0.2} = -1.376$.



Función de densidad de $t(5)$

La propiedad de reproductividad es:
Sean un conjunto de variables aleatorias $X_i \sim t(k_i)$ independientes. Entonces la variable aleatoria suma sigue una distribución:

$$\sum_{i=1}^{n} X_i \sim t(\sum_{i=1}^{n} k_i)\,.$$

C) La Distribución F de Snedecor.

Sean las variables aleatorias $X\sim \chi^2(m)$ e $Y\sim \chi^2(n)$ independientes. Entonces la variable aleatoria F definida como

$$F = \frac{X/m}{Y/n}$$

que toma cualquier valor real positivo, sigue una distribución F de Snedecor de parámetros m y n (F~F(m,n))

Sus principales características son:

- $EF = \dfrac{n}{n-2}$, para $n > 2$.
- $Var(F) = \dfrac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$, para $n > 4$.

Una propiedad importante es:
Si F~F(m,n), entonces la variable $1/F \sim$ F(n,m).

Los valores de su función de distribución están tabulados ($P[F \leq F_{n1,n2,p}] = p$) para p = 0.9, 0.95, 0.975 y 0.99 .Vemos cómo se utilizan las tablas con un ejemplo:

- Sea la variable aleatoria X~F(7,3). Entonces:

  $P[X\leq k] = 0.95 \Rightarrow k = 8.89 = F_{7,3,0.95}$

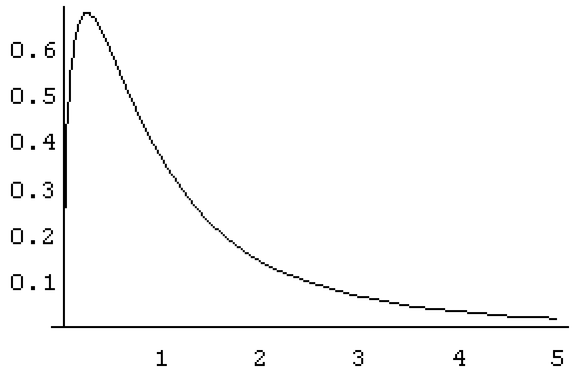
  $P[X \leq 5.27] = p \Rightarrow p = 0.9 = F_{7,3,0.9}$

Si p = 0.1, 0.025, 0.05 ó 0.01, entonces tendremos que aplicar la propiedad anterior:

$$F_{m,n,p} = \frac{1}{F_{n,m,1-p}}.$$

Continuamos con el ejemplo:

$$P[X\leq k] = 0.05 \Rightarrow k = F_{7,3,0.05} = \frac{1}{F_{3,7,0.95}} = \frac{1}{4.35} = 0.2298 \quad .$$



Función de densidad de F(3,5)

10.3. 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.

10.3.2. La 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 X_i que mide el número de veces que ocurre el suceso B_i en n realizaciones del experimento. X es pues una variable aleatoria k-dimensional $(X_1,...,X_k)$ que sigue una distribución Multinomial de parámetros n y $(p_1,...,p_k)$, donde:

$$\sum_{i=1}^{k} p_i = 1$$

es decir:

$$(X_1,...,X_k) \sim M\,(n;\,(p_1,...,p_k))$$

Su función de probabilidad está dada por la expresión:

$$P[X_1 = x_1,...,X_k = x_k] = \frac{n!}{x_1!\cdot...\cdot x_k!}\, p_1^{x_1}\cdot...\cdot p_k^{x_k}$$

donde:

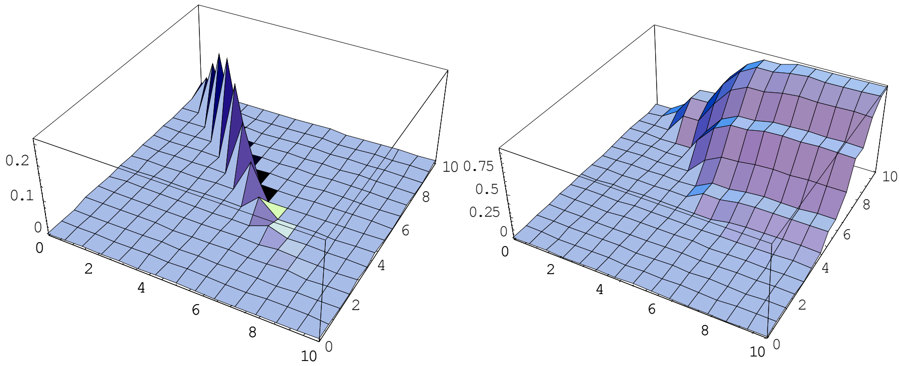
$$\sum_{i=1}^{k} x_i = n$$

Sus principales características son:

$$E[X] = [EX_1,...,EX_k]=[np_1,...,np_k]$$
$$Var[X_i] = np_i\,(1\text{-}p_i)$$
$$Cov(X_i,X_j) = -np_i\,p_j$$

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.



Función de densidad y de distribución de la Multinomial (10;0.3,0.7)

10.3.2. La Distribución Normal Multivariante.

Dada una variable aleatoria continua de dimensión k, $X=(X_1,...,X_k)$ se dice que sigue una distribución Normal k-dimensional de parámetros:

$$\mu = (\mu_1, \ldots, \mu_k)$$

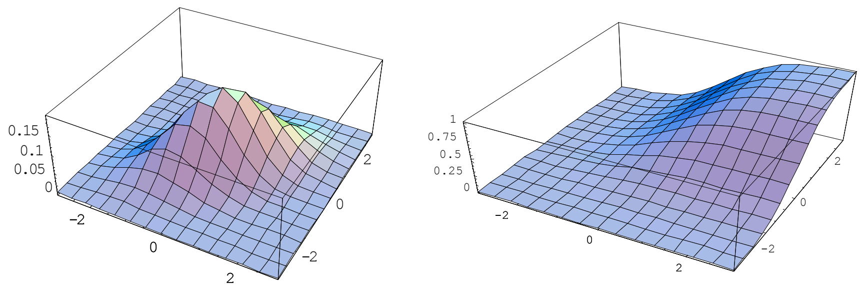
$$S = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \ldots & \sigma_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{1k} & \sigma_{2k} & \ldots & \sigma_k^2 \end{pmatrix}$$

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:

$$f(x=(x_1,\ldots,x_k)) = \frac{1}{(2\pi)^{k/2}(\det S)^{1/2}} e^{\left(\frac{-1}{2}(x-\mu)S^{-1}(x-\mu)^T\right)}$$

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.



Función de densidad y de distribución de la distribución Normal bivariante de parámetros μ = (0,0) y $S=\begin{pmatrix} 1 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & 1 \end{pmatrix}$

CAPÍTULO 11:

TEOREMA CENTRAL DEL LÍMITE.

El Teorema Central del Límite estudia el comportamiento asintótico de la suma de variables aleatorias, y nos va a servir para aproximar distintas distribuciones. Dice lo siguiente:

- Dada una sucesión $\{X_i\}_{i=1,\ldots,n}$ de variables aleatorias independientes e idénticamente distribuidas (es decir, variables aleatorias con resultados no relacionados y todas ellas con la misma distribución) con media μ y varianza σ^2 finitas, entonces la variable aleatoria suma de todas ellas $\sum_{i=1}^{n} X_i$ tipificada tiende a una distribución Normal de parámetros 0 y 1:

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{} Z \sim N(0,1)$$

ya que

$$\begin{cases} E\left[\sum_{i=1}^{n} X_i\right] = n\mu \\ Var\left[\sum_{i=1}^{n} X_i\right] = \sigma^2 n \end{cases}$$

o, equivalentemente,

$$\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow[n\to\infty]{} Z \sim N(0,1)$$

donde $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

Por tanto, 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 Normal(0,1).

No hacemos la demostración puesto que es un proceso muy extenso.

Como casos particulares tenemos:

1. Si X sigue una distribución Binomial $B(n,p)$, y no es posible realizar una aproximación mediante la distribución de Poisson, entonces:
$$\frac{X - np}{\sqrt{npq}} \xrightarrow[n\to\infty]{} Z \sim N(0,1),$$
o lo que es lo mismo,
$$X \cong X' \sim \mathrm{N}(\mu,\sigma) \text{ donde } \mu = np \;\; y \;\; \sigma = \sqrt{npq} \;\; ,$$
ya que podemos escribir $X = \sum_{i=1}^{n} X_i$ donde cada $X_i \sim \mathrm{B}(1,p)$.
2. Si X sigue una distribución de Poisson $\mathrm{P}(\lambda)$, entonces:
$$\frac{X - \lambda}{\sqrt{\lambda}} \xrightarrow[\lambda\to\infty]{} Z \sim N(0,1)$$
o lo que es lo mismo,
$$X \cong X' \sim N(\mu,\sigma) \text{ donde } \mu = \lambda \;\; y \;\; \sigma = \sqrt{\lambda} \;\; ,$$
ya que podemos escribir $X = \sum_{i=1}^{\lambda} X_i$ donde cada $X_i \sim \mathrm{P}(1)$.
3. Si X sigue una distribución $\chi^2(n)$, entonces:
$$\frac{X - n}{\sqrt{2n}} \xrightarrow[\lambda\to\infty]{} Z \sim N(0,1)$$
o lo que es lo mismo,
$$X \cong X' \sim N(\mu,\sigma) \text{ donde } \mu = n \;\; y \;\; \sigma = \sqrt{2n} \;\; ,$$

ya que podemos escribir $X = \sum_{i=1}^{\lambda} X_i$ donde cada $X_i \sim \chi^2(1)$.

4. Si X sigue una distribución t de Student $t(n)$, con $n \to \infty$, entonces

$$X \cong X' \sim N(0,1)$$

puesto que $EX = 0$ y $\sigma_X = \sqrt{\frac{n}{n-2}} \xrightarrow[n\to\infty]{} 1$.

Hay que tener un especial cuidado al aproximar una distribución discreta mediante una continua, ya que pasamos de probabilidades puntuales a probabilidades en intervalos:

Sea X una variable aleatoria discreta y X' su variable aleatoria normal aproximada. Queremos conocer la probabilidad de que X tome un valor cualquiera a natural. Entonces:

$P[X = a] = P[X' = a] = 0$, ya que X' 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, asignamos a cada valor de X natural el intervalo correspondiente a $[a\text{-}0.5, a\text{+}0.5)$ en la X' aproximada, de manera que todo valor real entre dos valores naturales consecutivos esté asignado a un valor. Por tanto:

$$P[X=a] \cong P[X' \in (a-0.5, a+0.5)]$$
$$P[X \le a] \cong P[X' \le a+0.5]$$
$$P[X \ge a] \cong P[X' \ge a-0.5]$$
$$P[X < a] \cong P[X' \le a-0.5]$$
$$P[X > a] \cong P[X' \ge a+0.5]$$

Ejemplo 11.1.

Las ventas diarias de gasolina se comportan según un modelo uniforme en el intervalo (10000,20000) expresado en litros.

1. La probabilidad de que las ventas en 30 días superen los 490000 litros viene dada por:

   Sea X_i la venta de gasolina el día i-ésimo.

   Sea $n = 30$.

   Nos piden $P\left[\sum_{i=1}^{30} X_i > 490000\right]$.

Sabemos, por el TCL que $\dfrac{\sum_{i=1}^{30} X_i - n\mu}{\sigma\sqrt{n}} \cong Z \sim N(0,1)$.

Como estamos en una distribución uniforme continua en [10000,20000]:

$$\begin{cases} \mu = \dfrac{a+b}{2} = 15000 \\ \sigma^2 = \dfrac{(b-a)^2}{12} = 8333333.33 \end{cases}.$$

Entonces $\dfrac{\sum_{i=1}^{30} X_i - 30 \bullet 15000}{\sqrt{8333333.33} \bullet \sqrt{30}} = \dfrac{\sum_{i=1}^{30} X_i - 450000}{15811.38} \cong Z \sim \mathrm{N}(0,1)$.

Por tanto,

$$P\left[\sum_{i=1}^{30} X_i > 490000\right] = P\left[Z > \frac{490000 - 450000}{15811.38}\right] =$$

$$= P[Z > 2.52] = 0.0059$$

2. La probabilidad de que las ventas medias por día en 30 días superen los 16000 litros vienen dadas por:
   Sea $n = 30$.
   Sea $\overline{X}_{30}$ las ventas medias por día en un mes de 30 días.
   Nos piden calcular $P\left[\overline{X}_{30} > 16000\right]$.
   Sabemos, por el TCL, que $\dfrac{\overline{X}_{30} - \mu}{\sigma/\sqrt{n}} = \dfrac{\overline{X}_{30} - 15000}{527.046} \cong Z \sim N(0,1)$.
   Entonces

$$P\left[\overline{X}_{30} > 16000\right] \cong P\left[Z > \frac{16000 - 15000}{527.046}\right] = P[Z > 1.897] = 0.0287.$$

RELACIÓN DE DISTRIBUCIONES DE PROBABILIDAD Y TEOREMA CENTRAL DEL LÍMITE.

1.- Se estudian las plantas de una zona en la que ha atacado un virus. La probabilidad de que una planta esté contaminada es de 0.35.

a) Definir la variable aleatoria que modeliza el experimento aleatorio de elegir una planta y ver si está contaminada. Dar su ley de probabilidad.

b) ¿Cuál es el número medio de plantas contaminadas que se pueden esperar en 5 análisis?

c) Calcular la probabilidad de encontrar 8 plantas contaminadas en 10 análisis.

d) Calcular la probabilidad de encontrar entre 2 y 5 contaminadas en 9 análisis.

e) Hallar la probabilidad de que en 6 análisis haya 4 no contaminadas.

2.- Sabiendo que el número medio de enfermos recibidos en un hospital cada 10 minutos es 1.8, calcular la probabilidad de que entre las 12:40 y 12:50 se reciba:

a) Ningún enfermo.

b) 1 enfermo.

c) 2 enfermos.

d) Al menos 2 enfermos.

e) Más de 2 enfermos.

3.- La distribución de un contaminante en una sustancia se distribuye de manera uniforme entre 4 y 20 partes por millón. Se considera tóxica una concentración de 15 o más.

a) ¿Cuál es la probabilidad de que, al tomarse una muestra, la concentración sea tóxica?

b) ¿Cuál es la probabilidad de que, al tomar una muestra, la concentración de contaminante oscile entre 5 y 10?

4.- La media de las calificaciones de un test es 400 y la desviación típica es 100. Si las calificaciones siguen una distribución normal, calcular:

a) El porcentaje de alumnos con calificaciones entre 300 y 500.

b) La probabilidad de que elegido un alumno al azar su calificación difiera de la media 150 puntos como máximo.

c) La puntuación máxima del 15% con menos puntuación.

d) Los límites del intervalo que contiene al 70% de la población más cercano a la media.

e) El porcentaje de la población que ha obtenido menos de 350 puntos.

5.- Los números 1,2,...,10 se escriben en 10 tarjetas y se colocan en una urna. Las tarjetas se extraen de una en una. Calcular:
a) Sin devolución, la probabilidad de que haya 3 números pares en 6 extracciones.
b) Con devolución:
1) La probabilidad de necesitar 5 extracciones para obtener 3 pares.
2) La probabilidad de obtener el primer 7 en la 4ª extracción.
3) La probabilidad de obtener un 7 en 4 extracciones.

6.- Se sabe que el 1% de los artículos incluidos en un embarque son defectuosos. Si se inspeccionan 30 artículos, obtener la probabilidad de que 2 o más sean defectuosos.

7.- Se sabe que un 2% de las ampollas que contienen una vacuna pierde su efectividad cuando se aplica en el 6º mes desde su elaboración y que un 50% lo hacen a partir del primer año.
a) ¿Cuál es la probabilidad de que una caja de 100 ampollas recibidas en un hospital y que van a ser usadas a los 6 meses de su elaboración tenga más de 3 ampollas ineficaces?
b) ¿Cuál es la probabilidad de más de 5 ineficaces si se utilizan a partir del primer año?

8.- Por experiencia se sabe que el promedio de errores tipográficos producidos por un operador manejando cierto tipo de máquina de impresión es de 2 errores cada 50 páginas. Calcular:
a) La probabilidad de que en un folleto de 100 páginas exista a lo sumo 1 errata.
b) La probabilidad de que en un cuadernillo de 50 páginas el número de erratas sea de 1 a 3.

9.- Sea X una variable aleatoria distribuida binomialmente con n=10 y p=0.5. Determinar las probabilidades de que X difiera de la media una cantidad menor que la desviación típica y una cantidad menor que el doble de la desviación típica.

10.- La probabilidad de que un satélite, después de colocado en órbita, funcione correctamente es de 0.9. Supongamos que 5 se colocan en órbita y operan de manera independiente:

a) ¿Cuál es la probabilidad de que, por lo menos, el 80% funcione adecuadamente?
b) ¿Y si n=10?
c) ¿Y si n=20?

11.- Para un volumen fijo, el número de células sanguíneas rojas es una variable aleatoria que se presenta con una frecuencia cte. Si el número promedio para un volumen dado es de 9 células para personas normales, determinar la probabilidad de que el número de células rojas para una persona se encuentre entre la media más la desviación típica y la media menos la desviación típica.

12.- En un cruce ocurren de manera aleatoria e independiente una media de 2 accidentes semanales. Determinar la probabilidad de que ocurra un accidente una semana y tres la semana siguiente.

13.- Sea X una variable aleatoria que sigue una distribución $N(10,5)$. Encontrar los valores que corresponden a las siguientes probabilidades:

$P[X<x] = 0.05 \quad P[X<x] = 0.95 \quad P[X<x] = 0.99 \quad P[X<x] = 0.01$

14.- Un departamento de una empresa pretende buscar financiación para un proyecto. Se sabe por situaciones parecidas que al tener una entrevista personal con un representante de las muchas empresas a las que han resuelto problemas, la probabilidad de encontrar apoyo económico es de 0.3.
¿Cuántas entrevistas habrá que realizar para conseguir el apoyo de al menos 2 empresas con una probabilidad de 0.9?

15.- Las horas productivas por mes de un departamento de administración se distribuyen mediante una Normal. Se sabe que el 2.56% de los meses las horas productivas son menos de 1305 y que el 15.87% de los meses las horas productivas son más de 1600.
a) ¿Cuál es el número de horas productivas por término medio?
b) ¿Cuál es la probabilidad de que en un mes se trabajen entre 1450 y 1550 horas?
c) Si se seleccionan 5 meses al azar, ¿cuál es la probabilidad de que en 3 de ellos se hayan trabajado entre 1450 y 1550 horas?

16.- La probabilidad de éxito de un experimento es 0.7. ¿Cuántas veces hay que realizar dicho experimento para tener al menos una probabilidad de 0.8 de que la frecuencia relativa difiera de la probabilidad de éxito en menos de 0.1?

17.- Una central telefónica de una gran ciudad recibe una media de 480 llamadas por hora. Si la central tiene capacidad para atender a lo sumo 12 llamadas por minuto, ¿cuál es la probabilidad de que en un minuto concreto no sea posible dar línea a todos los clientes?

18.- Un pescador desea capturar un ejemplar de sardina que se encuentra en su zona de pesca en un porcentaje del 15%. Calcular la probabilidad de que tenga que pescar 10 peces de otras especies antes de:

a) Pescar la sardina buscada.
b) Pescar 3 ejemplares de la sardina buscada.

19.- Se capturan 100 peces de un estanque que contiene 10000. Se les marca con una anilla y se les devuelve al agua. Unos días después se capturan 100 peces y se cuentan los anillados. Calcular la probabilidad de encontrar al menos un pez anillado.

20.- Un grupo de investigadores sabe que el 0.03% de los animales afectados por un virus fallecen. Calcular:

a) La probabilidad de que en una población de 10000 animales enfermos fallezcan más de tres.
b) El número esperado de fallecidos.
c) Si por un tratamiento equivocado el porcentaje de fallecidos se eleva al 0.18%, calcular la probabilidad de que fallezcan entre 25 y 40 animales.

21.- Si el número de piezas defectuosas en un proceso de fabricación sigue una ley de Poisson con media 26:

a) ¿Qué ley sigue de forma aproximada (aplicando el TCL) la variable aleatoria número de defectuosas?
b) Calcular la probabilidad aproximada de que haya 8 defectuosas.
c) Calcular la probabilidad aproximada de que haya entre 24 y 28 defectuosas.
d) Calcular el número de defectuosas que, como máximo, se puede encontrar con probabilidad aproximada 0.9772.
e) Calcular el número de defectuosas que, como mínimo, se puede encontrar con probabilidad aproximada 0.1587.

22.- Un dado se lanza 720 veces. Sea X la variable aleatoria que mide el nº de veces que sale 6.

a) Calcular la ley de probabilidad de X.

b) Calcular:

$$P[X = 30], P[X = 130], P[X = 300], P[100 \leq X \leq 125], P[X \geq 150]$$

23.- a) Calcular la probabilidad de que al lanzar un dado 100 veces, la media de los valores obtenidos sea mayor que 3.7.

b) Calcular la probabilidad de que la suma de los 100 lanzamientos sea mayor que 375.

APÉNDICE I:

SIMULACIÓN. GENERACIÓN DE NÚMEROS ALEATORIOS.

Se entiende por Simulación al proceso de diseñar un modelo de un sistema real y llevar a cabo experiencias con él, con la finalidad de aprender el comportamiento del sistema o evaluar ciertas estrategias para el funcionamiento del mismo. Por tanto, su objetivo es el analizar el comportamiento de un sistema y su estudio no da una solución matemática del problema.

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. Mediante procedimientos físicos, se creó una tabla de números aleatorios para evitar tenerlos que obtener mediante procedimientos de azar; pero esta tabla es eficiente para cálculos manuales, pero para su trabajo con ordenadores es bastante problemática, por lo que se crean 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, artiméticas,...

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 Normal (0,1), de la que, deshaciendo la tipificación, se puede pasar a cualquier otra distribución Normal.

Por tanto, los pasos a seguir son los siguientes:

AI.1. Generación de Números Pseudo-aleatorios.

Se propone la utilización del método congruencial mixto, al ser uno de los más avanzados. Para su utilización es necesaria la utilización de módulos:

Se define el módulo algebraico de orden m de un valor entero positivo d como el valor entero c entre 0 y m-1 tal que $d=k\cdot m+c$ para algún valor entero positivo de k.

Ejemplo AI.1

13 (mod 8) = 5, pues 13 = 8·1+5.

Con este método, se obtienen una serie de números enteros positivos entre 0 y m 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.
- $u_n = a\cdot u_{n-1} + b$ (mod m) , que es equivalente a

$$u_n = a^n \cdot u_0 + \left(\frac{a^n - 1}{a - 1}\right) b \ (\operatorname{mod} m)$$

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 se tarde lo máximo posible en repetir los valores se deben cumplir las siguientes condiciones:

- a debe de ser impar 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 Uniforme en el intervalo [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.

AI.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 distribución de 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}.

AI.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:

AI.3.1. 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 Uniforme en el intervalo [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.

AI.3.2. 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 Uniforme en el intervalo [0,1] y de una variable aleatoria con función de densidad $h(a)$.
- Si el valor generado $u_i \leq g(a_i)$, se acepta el valor de V como valor de la variable A. En caso contrario, se rechaza el par (u_i,a_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 Normal (0,1), para, a continuación pasarlos a la distribución Normal utilizada.

Sea Z una variable aleatoria con distribución Normal (0,1). Entonces, su función de densidad es:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \text{, para cualquier valor real de } z.$$

Se aplica el método de aceptación y rechazo:

- La función de densidad de Z se descompone como $f_Z(z) = c\cdot h(z)\ g(z)$, con $c\geq 1$, se toma $h(z) = \dfrac{e^{-z}}{\left(1+e^{-z}\right)^2}$, 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, $\dfrac{1}{\sqrt{2\pi}} e^{-z^2/2} \leq c\dfrac{e^{-z}}{\left(1+e^{-z}\right)^2}$, que, desarrollada, equivale a $\dfrac{-z^2}{2} + z + 2\ln\left(1+e^{-z}\right) \leq \ln\left(c\sqrt{2\pi}\right)$. 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 $H(z) = \dfrac{1}{1+e^{-z}}$:
  - Se genera u_1^* de una distribución Uniforme [0,1].
  - Entonces, $u_1^* = \dfrac{1}{1+e^{-z_1}}$ implica que $z_1 = -\ln\left(\dfrac{1-u_1^*}{u_1^*}\right)$.
- Se genera u_1 de una distribución Uniforme [0,1].
- Como $f_Z(z) = c \cdot h(z)\, g(z)$, entonces $g(z) = \dfrac{1}{\sqrt{2\pi}} e^{-z^2/2} \dfrac{\left(1+e^{-z}\right)^2}{1.5958\ e^{-z}}$.
- 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 Normal (μ,σ). Para generar números aleatorios de dicha distribución, se generan de la Normal (0,1) y se pasan a la Normal (μ,σ) mediante la transformación $a_i = \mu + z_i\sigma$.

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 $N(0,1)$:

  $z_1 = (-2 \ln u_1)^{1/2} \cos 2\pi u_2$

  $z_2 = (-2 \ln u_1)^{1/2} \operatorname{sen} 2\pi u_2$
- Se repite el proceso hasta obtener el número de valores normales deseados.

APÉNDICE II:

TABLAS ESTADÍSTICAS.

Tabla 1

Función de Distribución Binomial

		P										
n	***X***	**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

		P										
n	*X*	**0.01**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
9	**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
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

		P										
n	*X*	**0.01**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
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	03.9995	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
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

						P						
n	X	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
17	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

						P						
n	*X*	**0.01**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
20	**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
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

X	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

X	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

X	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.3796	0.3546	0.3309	0.3084	0.2873	0.2674	0.2487	0.2311	0.2146	0.1991
2	0.6496	0.6227	0.5960	0.5697	0.5438	0.5184	0.4936	0.4695	0.4460	0.4232
3	0.8386	0.8194	0.7993	0.7787	0.7576	0.7360	0.7141	0.6919	0.6696	0.6472
4	0.9379	0.9275	0.9163	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8153
5	0.9796	0.9751	0.9700	0.9643	0.9580	0.9510	0.9433	0.9349	0.9258	0.9161
6	0.9941	0.9925	0.9906	0.9884	0.9858	0.9828	0.9794	0.9756	0.9713	0.9665
7	0.9985	0.9980	0.9974	0.9967	0.9958	0.9947	0.9934	0.9919	0.9901	0.988!
8	0.9997	0.9995	0.9994	0.9991	0.9989	0.9985	0.9981	0.9976	0.9969	0.9962
9	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9993	0.9991	0.9989
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

X	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1847	0.1712	0.1586	0.1468	0.1359	0.1257	0.1162	0.1074	0.0992	0.0916
2	0.4012	0.3799	0.3594	0.3397	0.3208	0.3027	0.2854	0.2689	0.2531	0.2381
3	0.6248	0.6025	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4533	0.4335
4	0.7982	0.7806	0.7626	0.7442	0.7254	0.7064	0.6872	0.6678	0.6484	0.6288
5	0.9057	0.8946	0.8829	0.8705	0.8576	0.8441	0.8301	0.8156	0.8006	0.7851
6	0.9612	0.9554	0.9490	0.9421	0.9347	0.9267	0.9182	0.9091	0.8995	0.8893
7	0.9858	0.9832	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546	0.9489
8	0.9953	0.9943	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815	0.9786
9	0.9986	0.9982	0.9978	0.9973	0.9967	0.9960	0.9952	0.9942	0.9931	0.9919
10	0.9996	0.9995	0.9994	0.9992	0.9990	0.9987	0.9984	0.9981	0.9977	0.9972
11	0.9999	0.9999	0.9998	0.9998	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991
12	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999

	λ									
X	**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

X	**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

X	**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									
X	**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
X	**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									
X	**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

	λ									
X	**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	0.9S94	0.9317	0.8933	0.&432
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	0.9?05	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
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*	*x*	*F(x)*	*P(x)*
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	I	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

N	*n*	*k*	*x*	*F(x)*	*P(x)*
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	1	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
8	4	3	3	1.000000	0.071429
8	4	4	0	0.014286	0.014286
8	4	4	1	0.242857	0.228571
8	4	4	2	0.757143	0.5142S6
8	4	4	3	0.985714	0.228571
8	4	4	4	1.000000	0.014286
8	5	1	0	0.375000	0.375000
8	5	1	1	1.000000	0.625000
8	5	2	0	0.107143	0.107143
8	5	2	1	0.642857	0.5357Í4
8	5	2	2	1.000000	0.357143
8	5	3	0	0.017857	0.017857
8	5	3	1	0.285714	0.267857
8	5	3	2	0.821429	0.535714
8	5	3	3	1.000000	0.178571
8	5	4	1	0.071429	0.071429
8	5	4	2	0.500000	0.428571
8	5	4	3	0.928571	0.428571
8	5	4	4	1.000000	0.071429
8	5	5	2	0.178571	0.178571
8	5	5	3	0.714286	0.535714
8	5	5	4	0.982143	0:267857
8	5	5	5	1.000000	0.017857
8	6	1	0	0.250000	0.250000
8	6	1	1	1.000000	0.750000
8	6	2	0	0.035714	0.035714
8	6	2	1	0.464286	0.428571
8	6	2	2	1.000000	0.535714
8	6	3	1	0.107143	0.107143
8	6	3	2	0.642857	0.535714
8	6	3	3	1.000000	0.357143
8	6	4	2	0.214286	0.214286
8	6	4	3	0.785714	0.571429
8	6	4	4	1.000000	0.214286
8	6	5	3	0.357143	0.357143
8	6	5	4	0.892857	0.535714
8	6	5	5	1.000000	0.107143
8	6	6	4	0.535714	0.535714
8	6	6	5	0.964286	0.428571
8	6	6	6	1.000000	0.035714
8	7	1	0	0.125000	0.125000
8	7	1	1	1.000000	0.875000
8	7	2	1	0.250000	0.250000
8	7	2	2	1.000000	0.750000
8	7	3	2	0.375000	0.375000
8	7	3	3	1.000000	0.625000
8	7	4	3	0.500000	0.500000
8	7	4	4	1.000000	0.500000
8	7	5	4	0.625000	0.625000
8	7	5	5	1.000000	0.375000
8	7	6	5	0.750000	0.750000
8	7	6	6	1.000000	0.250000
8	7	7	6	0.875000	0.875000
8	7	7	7	1.000000	0.125000
9	1	1	0	0.888889	0.888889
9	1	1	1	1.000000	0.111111
9	2	1	0	0.777778	0.777778
9	2	1	1	1.000000	0.222222
9	2	2	0	0.583333	0.583333
9	2	2	1	0.972222	0.388889
9	2	2	2	1.000000	0.027778
9	3	1	0	0.666667	0.666667
9	3	1	1	1.000000	0.333333
9	3	2	0	0.416667	0.416667
9	3	2	1	0.916667	0.500000
9	3	2	2	1.000000	0.083333

N	*n*	*k*	*x*	*F(x)*	*P(x)*
9	3	3	0	0.238095	0.238095
9	3	3	1	0.773809	0.535714
9	3	3	2	0.988095	0.214286
9	3	3	3	1.000000	0.011905
9	4	1	0	0.555556	0.555556
9	4	1	1	1.000000	0.444444
9	4	2	0	0.277778	0.277778
9	4	2	1	0.833333	0.555556
9	4	2	2	1.000000	0.166667
9	4	3	0	0.119048	0.119048
9	4	3	1	0.595238	0.476190
9	4	3	2	0.952381	0.357143
9	4	3	3	1.000000	0.047619
9	4	4	0	0.039683	0.039683
9	4	4	1	0.357143	0.317460
9	4	4	2	0.833333	0.476190
9	4	4	3	0.992063	0.158730
9	4	4	4	1.000000	0.007936
9	5	1	0	0.444444	0.444444
9	5	1	1	1.000000	0.555556
9	5	2	0	0.166667	0.166667
9	5	2	1	0.722222	0.555556
9	5	2	2	1.000000	0.277778
9	5	3	0	0.047619	0.047619
9	5	3	1	0.404762	0.357143
9	5	3	2	0.880952	0.476190
9	5	3	3	1.000000	0.119048
9	5	4	0	0.007936	0.007936
9	5	4	1	0.166667	0.158730
9	5	4	2	0.642857	0.476190
9	5	4	3	0.960317	0.317460
9	5	4	4	1.000000	0.039683
9	5	5	1	0.039683	0.039683
9	5	5	2	0.357143	0.317460
9	5	5	3	0.833333	0.476190
9	5	5	4	0.992063	0.158730
9	5	5	5	1.000000	0.007936
9	6	1	0	0.333333	0.333333
9	6	1	1	1.000000	0.666667
9	6	2	0	0.083333	0.083333
9	6	2	1	0.583333	0.500000
9	6	2	2	1.000000	0.416667
9	6	3	0	0.011905	0.011905
9	6	3	1	0.226190	0.214286
9	6	3	2	0.761905	0.535714
9	6	3	3	1.000000	0.238095
9	6	4	1	0.047619	0.047619
9	6	4	2	0.404762	0.357143
9	6	4	3	0.880952	0.476190
9	6	4	4	1.000000	0.119048
9	6	5	2	0.119048	0.119048
9	6	5	3	0.595238	0.476190
9	6	5	4	0.952381	0.357143
9	6	5	5	1.000000	0.047619
9	6	6	3	0.238095	0.238095
9	6	6	4	0.773809	0.535714
9	6	6	5	0.988095	0.214286
9	6	6	6	1.000000	0.011905
9	7	1	0	0.222222	0.222222
9	7	1	1	1.000000	0.777778
9	7	2	0	0.027778	0.027778

N	*n*	*k*	*x*	*F(x)*	*P(x)*
9	7	2	1	0.416667	0.388889
9	7	2	2	1.000000	0.583333
9	7	3	1	0.083333	0.083333
9	7	3	2	0.583333	0.500000
9	7	3	3	1.000000	0.416667
9	7	4	2	0.166667	0.166667
9	7	4	3	0.722222	0.555556
9	7	4	4	1.000000	0.277778
9	7	5	3	0.277778	0.277778
9	7	5	4	0.833333	0.555556
9	7	5	5	1.000000	0.166667
9	7	6	4	0.416667	0.416667
9	7	6	5	0.916667	0.500000
9	7	6	6	1.000000	0.833333
9	7	7	5	0.583333	0.583333
9	7	7	6	0.972222	0.388889
9	7	7	7	1.000000	0.027778
9	8	1	0	0.111111	0.111111
9	8	1	1	1.000000	0.888889
9	8	2	1	0.222222	0.222222
9	8	2	2	1.000000	0.777778
9	8	3	2	0.333333	0.333333
9	8	3	3	1.000000	0.666667
9	8	4	3	0.444444	0.444444
9	8	4	4	1.000000	0.555556
9	8	5	4	0.555556	0.555556
9	8	5	5	1.000000	0.444444
9	8	6	5	0.666667	0.666667
9	8	6	6	1.000000	0.333333
9	8	7	6	0.777778	0.777778
9	8	7	7	1.000000	0.222222
9	8	8	7	0.888889	0.888889
9	8	8	8	1.000000	0.111111
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

N	*n*	*k*	*x*	*F(x)*	*P(x)*
10	4	4	3	0.995238	0.114286
10	4	4	4	1.000000	0.004762
10	5	1	0	0.500000	0.500000
10	5	1	1	1.000000	0.500000
10	5	2	0	0.222222	0.222222
10	5	2	1	0.777778	0.555556
10	5	2	2	1.000000	0.222222
10	5	3	0	0.083333	0.083333
10	5	3	1	0.500000	0.416667
10	5	3	2	0.916667	0.416667
10	5	3	3	1.000000	0.083333
10	5	4	0	0.023810	0.023810
10	5	4	1	0.261905	0.238095
10	5	4	2	0.738095	0.476190
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	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
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

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
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 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 6

Función de Distribución t-Student

	p						
g.l.	**0.80**	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**	**0.999**
1	1.376	3.078	6.31	12.70	31.82	63.65	318.39
2	1.061	1.886	2.920	4.303	6.965	9.925	22.32
3	0.978	1.638	2.353	3.182	4.541	5.841	10.21
4	0.941	1.533	2.132	2.776	3.747	4.604	7.173
5	0.920	1.476	2.015	2.571	3.365	4.032	5.893
6	0.906	1.440	1.943	2.447	3.143	3.707	5.208
7	0.896	1.415	1.895	2.365	2.998	3.499	4.785
8	0.889	1.397	1.860	2.306	2.896	3.355	4.501
9	0.883	1.383	1.833	2.262	2.821	3.250	4.297
10	0.879	1.372	1.812	2.228	2.764	3.169	4.144
11	0.876	1.363	1.796	2.201	2.718	3.106	4.025
12	0.873	1.356	1.782	2.179	2.681	3.055	3.930
13	0.870	1.350	1.771	2.160	2.650	3.012	3.852
14	0.868	1.345	1.761	2.145	2.624	2.977	3.787
15	0.866	1.341	1.753	2.131	2.602	2.947	3.733
16	0.865	1.337	1.746	2.120	2.583	2.921	3.686
17	0.863	1.333	1.740	2.110	2.567	2.898	3.646
18	0.862	1.330	1.734	2.101	2.552	2.878	3.610
19	0.861	1.328	1.729	2.093	2.539	2.861	3.579
20	0.860	1.325	1.725	2.086	2.528	2.845	3.552
21	0.859	1.323	1.721	2.080	2.518	2.831	3.527
22	0.858	1.321	1.717	2.074	2.508	2.819	3.505
23	0.858	1.319	1.714	2.069	2.500	2.807	3.485
24	0.857	1.318	1.711	2.064	2.492	2.797	3.467
25	0.856	1.316	1.708	2.060	2.485	2.787	3.450
26	0.856	1.315	1.706	2.056	2.479	2.779	3.435
27	0.855	1.314	1.703	2.052	2.473	2.771	3.421
28	0.855	1.313	1.701	2.048	2.467	2.763	3.408
29	0.854	1.311	1.699	2.045	2.462	2.756	3.396
30	0.854	1.310	1.697	2.042	2.457	2.750	3.385
35	0.852	1.306	1.690	2.030	2.438	2.724	3.340
40	0.851	1.303	1.684	2.021	2.423	2.704	3.307
45	0.850	1.301	1.679	2.014	2.412	2.690	3.281
50	0.849	1.299	1.676	2.009	2.403	2.678	3.261
60	0.848	1.296	1.671	2.000	2.390	2.660	3.232
70	0.847	1.294	1.667	1.994	2.381	2.648	3.211
80	0.846	1.292	1.664	1.990	2.374	2.639	3.195
90	0.846	1.291	1.662	1.987	2.369	2.632	3.183
100	0.845	1.290	1.660	1.984	2.364	2.626	3.174
200	0.843	1.286	1.652	1.972	2.345	2.601	3.131
500	0.842	1.283	1.648	1.965	2.334	2.586	3.107
1000	0.842	1.282	1.646	1.962	2.330	2.581	3.098

Esta distribución es simétrica: $t_{n,p}=t_{n,1-p}$

Tabla 7

Función de Distribución F de Snedecor

P = 0.9

	Grados de libertad 1 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
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
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

	Grados de libertad 1 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	2.72
7	2.68	2.67	2.63	2.59	2.57	2.56	2.54	2.52	2.50 2	2.47
8	2.52	2.50	2.46	2.42	2.40	2.38	2.36	2.35	2.32 2	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
21	1.90	1.87	1.83	1.78	1.74	1.72	1.69	1.67	1.63	1.59
22	1.88	1.86	1.81	1.76	1.73	1.70	1.67	1.65	1.61	1.57
23	1.87	1.84	1.80	1.74	1.71	1.69	1.66	1.64	1.59	1.55
24	1.85	1.83	1.78	1.73	1.70	1.67	1.64	1.62	1.58	1.54
25	1.84	1.82	1.77	1.72	1.68	1.66	1.63	1.61	1.56	1.52
26	1.83	1.81	1.76	1.71	1.67	1.65	1.61	1.59	1.55	1.51
27	1.82	1.80	1.75	1.70	1.66	1.64	1.60	1.58	1.54	1.50
28	1.81	1.79	1.74	1.69	1.65	1.63	1.59	1.57	1.53	1.48
29	1.80	1.78	1.73	1.68	1.64	1.62	1.58	1.56	1.52	1.47
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

P=0.95

	Grados de libertad 1 gl_1									
gl_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	161.4	199.50	215.7	224.5	230.1	233.9	236.7	238.8	240.5	241.8
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
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
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=0.95

	Grados de libertad 1 gl_1									
gl_2	**11**	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**100**	**1000**
1	242.9	243.9	245.9	248.0	249.2	250.0	251.1	251.7	253.0	254.1
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.'5)	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
21	2.28	2.25	2.18	2.10	2.05	2.01	1.96	1.94	1.88	1.82
22	2.26	2.23	2.15	2.07	2.02	1.98	1.94	1.91	1.85	1.79
23	2.24	2.20	2.13	2.05	2.00	1.96	1.91	1.88	1.82	1.76
24	2.22	2.18	2.11	2.03	1.97	1.94	1.89	1.86	1.80	1.74
25	2.20	2.16	2.09	2.01	1.96	1.92	1.87	1.84	1.78	1.72
26	2.18	2.15	2.07	1.99	1.94	1.90	1.85	1.82	1.76	1.70
27	2.17	2.13	2.06	1.97	1.92	1.88	1.84	1.81	1.74	1.68
28	2.15	2.12	2.04	1.96	1.91	1.87	1.82	1.79	1.73	1.66
29	2.14	2.10	2.03	1.94	1.89	1.85	1.81	1.77	1.71	1.65
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

P=0.975

	Grados de libertad 1 gl_1									
gl_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6
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
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53
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

	Grados de libertad 1 gl_1								
gl_2	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**1000**
1	976.7	978.9	993.1	997.2	1001	1006	1010	1014	1018
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

P=0.99

	Grados de libertad 1 gl_1									
gl_2	**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
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
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

	Grados de libertad 1 gl_1									
gl_2	**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	7.7.2	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.I6	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
21	3.24	3.17	3.03	2.88	2.78	2.72	2.64	2.58	2.48	2.37
22	3.18	3.12	2.98	2.83	2.73	2.67	2.58	2.53	2.42	2.32
23	3.14	3.07	2.93	2.78	2.69	2.62	2.54	2.48	2.37	2.27
24	3.09	3.03	2.89	2.74	2.64	2.58	2.49	2.44	2.33	2.22
25	3.06	2.99	2.85	2.70	2.60	2.54	2.45	2.40	2.29	2.18
26	3.02	2.96	2.81	2.66	2.57	2.50	2.42	2.36	2.25	2.14
27	2.99	2.93	2.78	2.63	2.54	2.47	2.38	2.33	2.22	2.11
28	2.96	2.90	2.75	2.60	2.51	2.44	2.35	2.30	2.19	2.08
29	2.93	2.87	2.73	2.57	2.48	2.41	2.33	2.27	2.16	2.05
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

BIBLIOGRAFÍA:

- ABAD MONTES F., VARGAS JIMÉNEZ M. (1991):
  Estadística. Gráficas JUFER.

- ÁLVAREZ GONZÁLEZ F. (1994):
  Estadística Aplicada. Mignon Libreros.

- AMOR PULIDO R., AGUILAR PEÑA C., MORALES LUQUE A. (2005):
  Estadística Aplicada. Grupo Editorial Universitario.

- CANAVOS G.C.(1986):
  Probabilidad y Estadística. Aplicaciones y Métodos. Editorial McGraw-Hill.

- GUTIÉRREZ JAIMEZ R., MARTÍNEZ ALMECIJA A., RODRÍGUEZ TORREBLANCA C. (1993):
  Curso Básico de Probabilidad. Editorial Pirámide.

- HERMOSO GUTIÉRREZ J. A., HERNÁNDEZ BASTIDA A.:
  Curso de Estadística Económica y Empresarial. Editorial Némesis (1994).
  Introducción a la Estadística. Gráficas JUFER (1990).

- HERNÁNDEZ BASTIDA A. (2007):
  Curso elemental de Estadística Descriptiva. Ediciones Pirámide.

- MUÑOZ VÁZQUEZ A., RODRÍGUEZ AVI J., LOZANO AGUILERA E.D., RUIZ MOLINA J.C. (1993):
  Problemas de Estadística. Estadística descriptiva. Imprenta Gutiérrez. Jaén.

- RODRÍGUEZ AVI, ALBA FDEZ (1995):
  Problemas de Cálculo de Probabilidades. Colección apuntes. Univ. de Jaén.