

PRO.T-15/143

UNIVERSIDAD DE GRANADA  
Facultad de Ciencias  
Fecha ..28-11-95  
SALIDA NUM. ..2644

T  
15  
76

**UNIVERSIDAD DE GRANADA**

**DEPARTAMENTO DE ESTADISTICA E  
INVESTIGACION OPERATIVA**

UNIVERSIDAD DE GRANADA  
Facultad de Ciencias  
Fecha ..28-9-95  
ENTRADA NUM. ..2051

**APORTACIONES A LAS TECNICAS  
GRAFICAS PARA EL ESTUDIO DE  
NORMALIDAD Y LAS CAUSAS DE SU  
PERDIDA.**

TESIS DOCTORAL

BIBLIOTECA UNIVERSITARIA  
GRANADA  
Nº Documento 619677042  
Nº Copia ..21223804

UNIVERSIDAD DE GRANADA  
22 SET. 1995  
COMISION DE DOCTORADO

Director: D. Andrés González Carmona

Doctorando: Emilio D. Lozano Aguilera


Granada, Septiembre de 1995.

**APORTACIONES A LAS TECNICAS GRAFICAS  
PARA EL ESTUDIO DE NORMALIDAD Y LAS  
CAUSAS DE SU PERDIDA**

Memoria que para optar al grado  
de Doctor en Ciencias, Sección de  
Matemáticas, presenta el Licenciado  
Emilio D. Lozano Aguilera



Vº Bº  
Director de la Tesis



Prof. Dr. D. Andrés González Carmona.

DEPARTAMENTO DE ESTADISTICA E INVESTIGACION OPERATIVA

FACULTAD DE CIENCIAS

UNIVERSIDAD DE GRANADA

1995





UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS

Núm. 2644

Tengo el gusto de remitirle 1 ejemplar de la Tesis Doctoral de D. Guillermo D. Losauro Aguilera, para su archivo en la Biblioteca de esta Facultad.

Granada, 28 -11- 95

EL SECRETARIO,



Gabriel Cardenete Hernández

\_\_\_\_ Sr. Director de la Biblioteca de esta Facultad.

*A mis padres, a Carmen y a mi hija Gloria*



*No hay una herramienta estadística simple que sea tan potente como un gráfico bien elegido.*

Chambers, Cleveland,  
Kleiner y Tukey

## INDICE

INTRODUCCION .....	1
CAPITULO 1.	
CONSIDERACIONES HISTORICAS SOBRE LA DISTRIBUCION NORMAL .....	9
CAPITULO 2.	
LOS CONTRASTES DE NORMALIDAD .....	17
Clasificación de los contrastes de normalidad ..	21
CAPITULO 3.	
CONTRASTES DE NORMALIDAD: BREVE RECORRIDO .....	27
El ambiente de un contraste de normalidad .....	27
Breve exposición de algunos contrastes de normalidad .....	29
Contrastes tipo chi-cuadrado .....	29
Tests basados en la función de distri- bución empírica .....	34
Tests basados en medidas descriptivas .....	42
A.Asimetría y Curtosis .....	42
B.Otros contrastes basados en me- didias descriptivas .....	50
El test de Shapiro y Wilk y sus modifi- caciones .....	53
A.El test W de Shapiro y Wilk .....	53



B.El test D de D'Agostino .....	55
C.El test W' de Shapiro y Francia ....	56
D.El test $\tilde{W}'$ de Weisberg-Bingham .....	57
E.El contraste del coeficiente de correlación en gráficos de proba- bilidad normal de Filliben .....	58
F.Extensión de La Brecque .....	59
G.Una generalización de W .....	63
H.Transformación de la distribución de W y W' a normalidad .....	64
Contraste de razón de verosimilitudes asociado con transformaciones para in- crementar la normalidad .....	68
Técnicas gráficas informales .....	70
<b>Consideraciones finales sobre los contrastes de normalidad univariantes .....</b>	<b>71</b>

**CAPITULO 4.**

<b>EL GRAFICO PROBABILISTICO NORMAL .....</b>	<b>77</b>
El gráfico de probabilidad empírico .....	79
El gráfico de probabilidad teórico .....	83
Algunos comentarios sobre los puntos de po- sición gráfica ( $p_i$ ) .....	86
Nuestra elección de los $p_i$ .....	93
Determinación de los $p_i$ a partir de la mediana de los estadísticos de orden de una distribución normal .....	95

**CAPITULO 5.**

<b>EL GRAFICO PROBABILISTICO NORMAL COMO TECNICA GRAFICA PARA EL ESTUDIO DE NORMALIDAD .....</b>	<b>117</b>
Causas de no normalidad apreciables en el gráfico probabilístico normal .....	123
A.Existencia de outliers .....	123
B.Existencia de pequeñas o grandes colas ..	125



C.Asimetría .....	127
D.Existencia de segmentos horizontales, grupos de puntos y discontinuidades .....	129
Ajuste de una recta al gráfico de probabilidad .	131
CAPITULO 6.	
ANALISIS DE LA VARIABILIDAD EN EL GRAFICO	
PROBABILISTICO NORMAL .....	135
Análisis de la variabilidad en el gráfico probabilístico normal a partir de la va- rianza de los estadísticos de orden .....	137
Análisis de la variabilidad en el gráfico probabilístico normal a partir de la dis- tribución exacta de los estadísticos de orden .....	146
Análisis de la variabilidad en el gráfico probabilístico normal a partir del compor- tamiento asintótico de los estadísticos de orden .....	153
APENDICE I.	
ESTADISTICOS DE ORDEN .....	165
Definición de estadístico de orden .....	166
Función de distribución empírica y función de distribución de $X_{(k)}$ .....	167
Distribución conjunta de los estadísticos de orden .....	173
Momentos de los estadísticos de orden .....	178
Algunos usos de los estadísticos de orden .....	181
Comportamiento asintótico de los estadís- ticos de orden .....	190
Máximo y Mínimo muestral .....	191
Otros estadísticos de orden extremos .....	204
Estadísticos de orden centrales e in- termedios .....	207



**APENDICE II.**

<b>APLICACIONES INFORMATICAS</b> .....	213
<b>Programas en S-PLUS</b> .....	214
Cálculo aproximado de la varianza de los estadísticos de orden de una distribución normal .....	214
Determinación de la variabilidad de los puntos sobre la recta del gráfico probabilístico normal basado en la varianza de los estadísticos de orden .....	215
Determinación de la banda de confianza en el gráfico probabilístico normal basada en la distribución exacta de los estadísticos de orden de una distribución normal .....	216
<b>Programas en UBASIC</b> .....	217
Determinación de los polinomios que definen a las medianas de los estadísticos de orden de una uniforme estándar, y de las raíces de éstos .....	217
<b>BIBLIOGRAFIA</b> .....	221

## INTRODUCCION

La cita con la que comenzamos la presentación de esta memoria es lo suficientemente breve y rotunda como para describir con claridad la importancia que las técnicas gráficas tienen en el contexto de un estudio estadístico.

La conjunción del *sentido de la vista* y la *capacidad de pensar* que posee el ser humano, puede ser considerada como el sistema de proceso de información más sofisticado que jamás pueda ser desarrollado. Los gráficos poseen la cualidad de poner en funcionamiento tal sistema y no sólo para describir y resumir el comportamiento de un conjunto de datos, sino incluso para obtener una idea precisa de su estructura.

Un gráfico puede recoger una gran cantidad de información. Nuestro sistema de proceso de información la resume rápidamente y extrae los aspectos más sobresalientes, con la ventaja añadida de ser también capaz de fijarse en los detalles más insignificantes.

En su gran mayoría, los métodos clásicos de la estadística están basados, implícita o explícitamente, en suposiciones hechas sobre los datos, y la validez de los análisis depende de la validez de tales suposiciones. Los métodos gráficos proporcionan potentes herramientas para analizar tales suposiciones y, cuando



éstas no sean confirmadas, sugerir ciertas correcciones.

Es precisamente éste el ambiente en el que desarrollamos este trabajo. En concreto, nos dedicaremos al estudio de una técnica gráfica diseñada para el análisis de la hipótesis de normalidad enunciada sobre un conjunto de observaciones.

La normalidad es un estado que podemos considerar deseable al enfrentarnos al estudio de un conjunto de observaciones. Si tenemos garantía de que tales observaciones provienen de una población distribuida según una normal, existen una gran variedad de técnicas que se pueden aplicar en el análisis de los datos, con la seguridad de obtener resultados del todo fiables. Hasta tal punto llega su importancia que, en la práctica, no tenemos inconveniente en perder algo de precisión y aproximar la verdadera distribución de los datos (ya sea continua o discreta) por una distribución normal.

Sin embargo, el desconocimiento habitual que se tiene sobre el comportamiento real de los datos nos hace establecer la hipótesis de normalidad sobre ellos, vistas las ventajas que esto reporta. Para validar esta hipótesis, necesitamos hacer uso de técnicas desarrolladas a tal efecto y que son denominadas contrastes de normalidad.

Desde principios de siglo, cuando a la distribución normal se le comienza a otorgar la importancia que tiene, se han venido introduciendo diversas técnicas encaminadas al contraste de normalidad. El interés por este tipo de estudios resurge allá por los años sesenta, coincidiendo con la introducción del test W de Shapiro y Wilk, de manera que actualmente el conjunto de técnicas estadísticas dedicadas al estudio de la posible normalidad de un conjunto de observaciones es amplísimo. Se observa que, aparte de los contrastes formales de normalidad, aparecen una serie de téc-



nicas calificadas como informales que, basadas en métodos gráficos, se dirigen también hacia el estudio de la normalidad de un conjunto de datos.

Dentro del conjunto de técnicas gráficas (informales) para el estudio de la normalidad, encontramos los que en la literatura estadística se han venido denominando gráficos de probabilidad normal o gráfico probabilístico normal (Q-Q plot).

El gráfico de probabilidad normal consiste, básicamente, en dibujar los cuantiles de la distribución empírica de los datos contra los correspondientes cuantiles de la distribución normal, de forma que si los puntos así obtenidos se presentan con una configuración aproximadamente rectilínea, podemos concluir que la distribución de los datos es compatible con la distribución normal.

El gráfico de probabilidad reúne una serie de características que hacen de él una técnica gráfica sencilla y de fácil interpretación. Esto es debido, en primer lugar, a que cada dato original es dibujado en el eje de ordenadas con la misma escala en la que la observación ha sido medida y, en segundo lugar, a que la conclusión obtenida en la observación de tal gráfico se fundamenta en una configuración que el ojo humano tiene gran facilidad de detectar: la linealidad.

Sin embargo, la escasa objetividad con la que se maneja el concepto de linealidad en el que se basan las afirmaciones con las que se puede concluir un estudio de normalidad haciendo uso del gráfico probabilístico normal, nos incita a analizarlo en profundidad, con el objetivo fundamental de formalizar los criterios que nos lleven a aceptar o rechazar la hipótesis de normalidad.

En particular, precisaremos la idea de que los puntos en el gráfico probabilístico normal tengan una configuración rectilí-



nea. Para ello introduciremos lo que hemos denominado *bandas de confianza* sobre el gráfico de probabilidad normal que, permitiendo ciertas fluctuaciones de los puntos sobre su configuración lineal, nos determinen una región en el gráfico que nos lleven a aceptar o rechazar la hipótesis de normalidad, con un nivel de confianza fijado.

En el desarrollo de esta nueva técnica gráfica, y con el estudio del gráfico probabilístico normal, proponemos también un método para determinar los puntos de posición gráfica (plotting position), basado en la determinación de la mediana de los estadísticos de orden de una distribución uniforme estándar.

Esta memoria ha quedado estructurada en seis capítulos (del Capítulo 1 al Capítulo 6) y dos apéndices (Apéndice I y Apéndice II).

En el Capítulo 1, se ha pretendido, haciendo una breve descripción sobre la historia de la distribución normal, justificar la importancia que tal distribución tiene, atendiendo sólo al tiempo que se ha dedicado a su estudio.

En el Capítulo 2 se establece el significado que tiene un contraste de normalidad, dando una serie de pinceladas sobre las ventajas que se obtienen al admitir una hipótesis distribucional enunciada sobre un conjunto de observaciones. Así mismo, partiendo de que tal hipótesis distribucional se refiera a la normalidad de los datos, proponemos una clasificación de los distintos contrastes de normalidad que se manejan en la literatura al uso.

En el Capítulo 3, y ajustándonos a la clasificación que de los contrastes de normalidad hemos propuesto en el capítulo anterior, hacemos un breve recorrido sobre aquellos contrastes de normalidad que, por su interés histórico o por las propiedades que



poseen, pueden ser considerados como los tests de normalidad más importantes que se conocen en la actualidad.

En el Capítulo 4 se introduce el gráfico probabilístico normal. La presentación de esta técnica gráfica se hace desde el gráfico de probabilidad empírico, lo que nos permite cargar de intuición al gráfico de probabilidad (teórico) normal, para, posteriormente, en el Capítulo 5 poder contemplar al gráfico probabilístico normal como una técnica gráfica encaminada al estudio de la posible normalidad sobre un conjunto de observaciones.

En el transcurso del Capítulo 4, y visto que los puntos de posición gráfica son elementos básicos para la obtención del gráfico de probabilidad, realizamos un estudio de éstos y concluimos haciendo una propuesta sobre su definición y determinación.

Una vez que el gráfico probabilístico normal ha sido introducido como una técnica gráfica para el estudio de la hipótesis de normalidad enunciada sobre un conjunto de datos (Capítulo 5), analizamos las distintas causas que podemos apreciar en dicho gráfico para dudar de tal hipótesis.

Por otra parte, debido a la propiedad de invarianza frente a cambios en los parámetros de localización y escala que tiene el gráfico probabilístico normal, vemos cómo éste puede ser usado para producir una estimación de tales parámetros de la distribución normal (media y desviación típica).

Por fin, dado que la conclusión sobre la posible normalidad de los datos, haciendo uso del gráfico probabilístico normal, se fundamenta en la configuración más o menos rectilínea de los puntos en dicho gráfico, y que ésta nunca será exacta debido a la aleatoriedad de las observaciones, establecemos, en el Capítulo 6, cómo serán las fluctuaciones que admitiremos sobre la recta para aceptar la normalidad. Esto lo hacemos introduciendo el con-



cepto de *banda de confianza* en un gráfico probabilístico normal. La determinación de tal elemento la sustentamos en tres ideas distintas que darán lugar a tres tipos de bandas de confianza. Los procedimientos seguidos se basan en: la varianza de los estadísticos de orden de una distribución normal, la distribución exacta de los estadísticos de orden de una distribución normal y el comportamiento asintótico de los estadísticos de orden de una distribución normal.

En el último epígrafe del Capítulo 1, donde analizamos la variabilidad en el gráfico probabilístico normal basándonos en el comportamiento asintótico de los estadísticos de orden, proponemos, usando el concepto de intervalo de tolerancia de una distribución, un método para clasificar los estadísticos de orden en *estadísticos de orden extremos*, *estadísticos de orden intermedios* y *estadísticos de orden centrales*.

Esta memoria, como ya hemos indicado, ha sido complementada introduciendo dos apéndices. El Apéndice I está dedicado al estudio de los estadísticos de orden. Su justificación se encuentra, por una parte, en el gran número de contrastes de normalidad que hacen uso de dichos estadísticos y, por otra, en que nuestra propuesta de *bandas de confianza* sobre el gráfico probabilístico normal se fundamenta en el comportamiento de los estadísticos de orden.

El apéndice II recoge una serie de aplicaciones informáticas que hemos desarrollado, tanto para la obtención de los puntos de posición gráfica ( $p_i$ ), como para los tres métodos seguidos en la determinación de las bandas de confianza en el gráfico probabilístico normal.

Esta memoria concluye con un listado de la bibliografía que hemos usado para su elaboración. En ella hemos recogido aquellos libros y monografías dedicados al estudio de la distribución



normal y de los contrastes de normalidad, y los artículos que, sobre el mismo tema, han sido consultados. Así mismo, se citan algunos artículos, que si bien no han sido consultados directamente por el doctorando, sí están, en su parte fundamental, recogidos en otros que han sido estudiados. La justificación de haberlos incorporado es simple desde el momento que aceptemos que las referencias bibliográficas en toda investigación deben considerarse como punto de partida para futuras investigaciones sobre el mismo tema y es conveniente, por tanto, hacer mención a la fuente original.

Antes de concluir esta introducción, creo que es de justicia referirme a una serie de personas que de forma directa o indirecta han colaborado en la elaboración de esta memoria.

Quiero agradecer a mis compañeros del Departamento, y en general, a los compañeros de las Universidades de Granada y Jaén las continuas muestras de apoyo y de ánimo que de ellos he recibido. En particular quiero referirme a mis "*maestros*" los profesores D. Andrés González Carmona y D. Ramón Gutiérrez Jáimez a los que debo, además de la dirección y el apoyo científico que he necesitado para elaborar este trabajo, las continuas muestras de aprecio que en todo momento me han brindado.

Por último, y no por ello menos importante, quiero agradecer las continuas muestras de comprensión y apoyo que de mi familia he recibido: de Carmen, mi mujer, que ha sabido aceptar y justificar mi ausencia durante tantos días del año y que junto con mi hija Gloria han sido el punto de apoyo necesario para no abandonar incluso en los momentos de más debilidad; de mis suegros que han sabido sustituirme durante bastante tiempo en mi papel de padre; de mis hermanos, de los que he recibido un aliento continuo; y, por supuesto, de mis padres a los que tanto debo, y que en gran medida, son los responsables de que este trabajo haya podido ser concluido.





## CAPITULO 1

# CONSIDERACIONES HISTORICAS SOBRE LA DISTRIBUCION NORMAL

Intentando situar el origen de la distribución normal, encontramos que ya, en el siglo XVII, Galileo expresó sus conclusiones sobre las distancias entre las estrellas medidas por los astrónomos. Encontró que en ellas aparecían dos tipos de errores: uno primero, de tipo aleatorio, producido por observaciones instrumentales, y otro segundo, de tipo sistemático, consecuencia de los cálculos realizados. Con respecto a los primeros, declaró que los errores pequeños eran más probables de ocurrir que los errores grandes, que las medidas estaban igualmente sujetas a errores en una dirección o en otra (exceso o defecto), y que la mayoría de las observaciones tendían a agruparse alrededor de un *valor cierto*. Fue este el modo mediante el que Galileo reveló muchas de las que hoy conocemos como características básicas de la Ley Normal de probabilidad.

Aunque el estudio de la probabilidad comenzó mucho antes, la estadística moderna encuentra su primer gran adelanto con la publicación en 1713 del trabajo de Jacob Bernoulli titulado "Ars Conjectandi". En esta obra aparece probada la conocida con el nombre de ley débil de los grandes números.

La distribución normal, como tal, aparece por primera



vez en 1733 como aproximación a la distribución binomial, en los trabajos recopilados por Abraham de Moivre en su obra "Miscelaneas Analytica". Este teorema fue de nuevo recogido en su libro "The Doctrine of Chances" publicado en 1738.

El desarrollo natural y el gran avance de la teoría de la probabilidad dentro de la estadística matemática, tiene su pilar en el matemático Pierre Simon de Laplace. Laplace introdujo la función característica como una herramienta para teoría de grandes muestras, y probó el primer teorema central del límite. Demostró que una clase de estimadores insesgados de coeficientes de regresión lineal están aproximadamente distribuidos según una normal. En 1812 probó que la distribución de probabilidad de la esperanza de vida, en cualquier año especificado, tiende a la normal. Derivó, también, la distribución asintótica de los estadísticos de orden en problemas de regresión lineal, cuando la distribución de partida es simétrica sobre cero y bien comportada, obteniendo en este caso un patrón normal. Así mismo, en 1818, probó que cuando la distribución de partida es normal, el estimador de mínimos cuadrados (LSE) tiene menos varianza que cualquier combinación lineal de las observaciones.

En 1805, problemas llegados de las colecciones de observaciones en astronomía, condujeron a Legendre a establecer el principio de mínimos cuadrados, minimizando la suma de los cuadrados de los errores de las observaciones. Por entonces, también obtuvo la ley normal.

En 1809, y a raíz de la publicación de la obra "Theoria Motus Corporum Coelestium" de Carl Friedrich Gauss, se produjo cierta polémica entre Gauss, Laplace y Legendre, debido a que el primero hizo uso del principio de mínimos cuadrados tal y como se utilizaba en 1795. (Este hecho está justificado por la falta de comunicación existente entre los investigadores en el siglo XIX).



En su "Theoria Motus Corporun Coelestium", Gauss probó que la distribución de errores, supuesto continuos, debería ser normal si el parámetro de localización tiene una distribución a priori uniforme, así que la media es la moda de la distribución a posteriori. El modelo lineal de Gauss de mínimos cuadrados era, por tanto, apropiado cuando los errores provenían de una distribución normal.

Un matemático americano, Rober Adrain (1808), que no conocía nada del trabajo de Gauss, pero que se sospecha pudo haber leído la obra de Legendre, obtuvo la distribución normal univariante y bivalente como una distribución de errores y, de aquí, el método de mínimos cuadrados, pero su trabajo no tuvo especial relevancia en el desarrollo del tema.

La publicación de una obra de Hagen (1837), que contenía una derivación de la ley normal como una aproximación a la distribución del error total cuando éste provenía del resultado de un gran número de igualdades de errores elementales, igualmente positivos o negativos, pudo haber conducido a Bessel, en 1838, a desarrollar las hipótesis de errores elementales. Así, Bessel obtuvo la ley normal como una aproximación para el error total, supuesto que aparecía como la suma de una gran número de errores elementales, mutuamente independientes, pero igualmente distribuidos, con propiedades de buen comportamiento, incluyendo la simetría en su distribución respecto al cero.

Las hipótesis de los errores elementales llegaron a estar firmemente establecidas, particularmente entre astrónomos como G. B. Airy (1861), que interpretó el teorema central del límite justo desde ese punto de vista.

En 1860, el físico-matemático escocés James Clerk Maxwell publicó el primero de sus dos importantes escritos sobre la teoría cinética de los gases. Haciendo uso de consideraciones de



carácter geométrico, obtuvo la distribución normal como la distribución de las componentes ortogonales de la velocidad de partículas moviéndose en el vacío. Sus resultados condujeron directamente a Boltzmann a la teoría moderna de la mecánica estadística, destacable como el primer intento de describir el movimiento de los gases por una función estadística, sin intervención determinística alguna: "Las velocidades se distribuyen entre las partículas de acuerdo a la misma ley según se distribuyen los errores entre las observaciones en la teoría del método de mínimos cuadrados".

Uno de los primeros en ajustar una curva normal a datos, fuera del campo de la astronomía, fue el científico belga Adolphe Quetelet (1846). Aparece también con él un nuevo concepto o método en el tratamiento de los datos. Antes se tomaban muchas medidas sobre el mismo individuo; este autor introduce la metodología de tomar una medida sobre mucho individuos.

Francis Galton, científico inglés, continuó en su trabajo "Natural Inheritance" (1889) la idea de Quetelet, aplicando la curva normal a medidas de naturaleza humana, desarrollando un modelo para describir la dependencia de tales medidas en componentes de una misma familia (padres e hijos). Dibujó sus datos en dos dimensiones, y notó que aquellos puntos de igual densidad aparecían sobre curvas elípticas. A partir de esto, desarrolló el modelo de regresión lineal, el concepto de correlación (1888) y la expresión de la distribución normal bivariante (1886), con la ayuda del matemático Hamilton Dickson. (Las distribuciones normales bivariantes y trivariantes habían sido ya desarrolladas por Bravais (1846), pero aún conociendo la noción de correlación, éste no le dio el grado de importancia que le otorgó Galton).

Galton indicó que ciertos conjuntos de datos podrían seguir bien algunas otras leyes de frecuencias. La media geométrica, escribió, podría representar mejor el valor más probable de una distribución y, así, cabría suponer que los logaritmos de las



observaciones siguieran una distribución normal. Esto le condujo a la distribución logarítmico-normal, que sirvió como estímulo a Pearson para desarrollar un sistema de curvas de frecuencias, dependiendo de un conjunto de parámetros, que podrían recoger todas las distribuciones que ocurrieran en la naturaleza, o al menos, aquellas que fuesen continuas.

Como Karl Pearson indicó, "En 1885 Galton había completado la teoría de la correlación normal bivalente", pero debido a que "él era muy modesto y hasta el final de su vida subestimó su propio potencial matemático, no escribió ni una vez la expresión" de la función de densidad normal bivalente. Este recato permitió que fuera Pearson (1896) quien diera una formulación matemática definitiva de la distribución normal bivalente.

Karl Pearson generalizó y dio precisión a la discusión sobre la correlación, dando origen a los estudios de análisis de regresión y correlación parcial y múltiple, que puede considerarse como el principio de la teoría de la distribución normal multivalente, tratada de forma amplia por primera vez por Edgeworth (1892).

En un intento de ajustar datos a sus curvas de frecuencias, Pearson vio la necesidad de contrastar la bondad de ajuste. Ernst Abbe había obtenido la distribución de  $\sum X_i^2$  donde  $X_1, X_2, \dots, X_n$  es una muestra aleatoria de una distribución normal de media cero, y Helber (1876) encontró la distribución de  $\sum (X_i - \bar{X})^2$ , donde  $\bar{X}$  es la media aritmética de  $X_1, X_2, \dots, X_n$ . Estas distribuciones correspondían, respectivamente, a una chi-cuadrado con  $n$  y  $(n-1)$  grados de libertad. (El trabajo de Abbe no fue conocido hasta mucho más tarde, por lo que, en un principio, este resultado fue atribuido a Helmer). La matriz de transformación de Helmer aún era usada como herramienta para obtener la distribución de  $\sum (X_i - \bar{X})^2$  y para establecer la independencia de  $\bar{X}$  y la varianza muestral  $\sum (X_i - \bar{X})^2 / (n-1)$ , en una muestra procedente de



una distribución normal.

Weldon había encontrado una notable excepción de ajuste de la curva normal a un conjunto de datos que relataban la anchura relativa frontal del cangrejo de Nápoles. Creyendo que podía estar compuesto de dos subespecies, ajustó una distribución normal compuesta o *doble curva ponderada*, como él le denominó, pero el método de ajuste que usó (1892) se basaba solamente en técnicas de ensayo y error. El primer escrito estadístico de Pearson en 1894, abordó este problema. Introdujo el método de los momentos como una técnica de ajuste a una distribución de frecuencias. La cuestión de si la distribución ajustada era razonable condujo a Pearson a un resultado (1900) donde estableció firmemente el contraste de bondad de ajuste chi-cuadrado (aunque anteriormente Bienaymé había trabajado sobre el tema), como un pilar de la estadística moderna. (Irónicamente, el primer "principio inamovible" que cayó con esta nueva herramienta fue la ley de los errores).

Pearson, y la escuela inglesa de biometría, prefirieron trabajar con grandes conjuntos de datos. Sin embargo, W. S. Gosset ("Student"), obligado por las circunstancias, llegó a resolver problemas en la fábrica de cervezas Guinness de Dublín, para los que sólo era posible tomar muestras pequeñas. Un año de trabajo, bajo la tutela de Pearson, condujo a Gosset a su famoso escrito titulado "The Probable Error of Mean" (Student, 1908), en el que dedujo la distribución t, la razón de la media muestral con la desviación típica en una muestra normal. (Es curioso que a este trabajo, Pearson prestara poco interés).

En contraposición a la escuela inglesa de biometría, la escuela rusa de Pafnuti Lvovich Tchebyshev y sus pupilos Markov y Lyapunov, emprendieron una serie de estudios en teoría de la probabilidad. Desde comienzos de la segunda mitad del siglo XIX, esta escuela aplicó rigor matemático a leyes de grandes números, sucesos dependientes y propiedades límite. Con la introducción del



concepto de variable aleatoria, el camino estaba preparado para establecer condiciones suficientes que arrojaran que sumas estandarizadas de variables aleatorias, tanto dependientes como independientes, convergieran a la ley normal. El primer enunciado referente a esto, junto con su demostración (que posteriormente requirió ser revisada y se dio una demostración adicional), fue dado por Tchebyshev en 1877, usando el método de los momentos. La importancia de los teoremas centrales del límite de Tchebyshev radicó en el carácter matemático claramente definido que atribuyó a la variable aleatoria. Estos teoremas fueron enunciados estableciendo restricciones en la aplicabilidad de los resultados en la teoría de la probabilidad, así que en cualquier conjunto de circunstancias, uno podría determinar si se mantenían o no los teoremas. En esta línea se dirigían los trabajos de Andrei Andreevich Markov (1898) y Alexander Mikhailovich Lyapunov (1901). El primero corrigiendo los teoremas de Tchebyshev, y el segundo, encargado de producir un teorema central límite de gran generalidad, rigurosamente probado con las herramientas del análisis clásico e incluyendo funciones características.

Estos trabajos tuvieron su continuidad en los realizados por Bernstein, Khinchine y Kolmogorov.

La temprana historia de la distribución normal es la historia de los comienzos de la Estadística como una ciencia. Podemos considerar que la *estadística histórica* acaba en este punto, cuando los trabajos de R. A. Fisher dan pie al nacimiento de la *estadística moderna*.

Desde 1915 en adelante, Fisher encontró la distribución del coeficiente de correlación, de la desviación absoluta  $\sum |X_i - \bar{X}|/n$ , en muestras normales, del coeficiente de regresión, de coeficientes de correlación múltiple y correlación parcial y del cociente F de varianzas muestrales de dos poblaciones normales. A su vez, desarrolló sus ideas de estimación, suficiencia, verosimilitud, inferencia, análisis de la varianza y diseños expe-



rimentales.

Las hipótesis de normalidad han jugado un papel fundamental, a través de los años, en el análisis estadístico, pero es a partir de los años sesenta cuando se ha prestado una considerable atención al problema que surge al cuestionarse tales hipótesis, requiriéndose, por una parte, una serie de pruebas que nos informen sobre la validez de estas hipótesis, y por otra, en caso de ser violadas, la presencia de estimadores robustos.

La distribución normal ha recibido varias denominaciones desde el comienzo de su historia, coincidiendo éstas con el nombre de muchos de aquellos que la trataron (aunque se ha de señalar que curiosamente nunca se la ha llamado con referencia al que en la actualidad se ha venido en considerar como su promotor, Abraham de Moivre). Hoy en día es conocida mundialmente como distribución normal, nombre que Karl Pearson declaró haber acuñado, pero que mucho antes (1877) Galton ya había usado.



## CAPITULO 2

### LOS CONTRASTES DE NORMALIDAD

Constituye una costumbre habitual al comenzar un estudio sobre un conjunto de datos el establecer una afirmación o suposición distribucional sobre la procedencia de éstos, indicando que los mismos surgen como una muestra de una distribución perteneciente a alguna familia conocida de distribuciones de probabilidad. Son distintas las razones que fundamentan la conveniencia de establecer tales suposiciones: en primer lugar, si podemos describir un conjunto de datos como una muestra proveniente de cierta distribución teórica, lograremos dar una valiosa descripción compacta de éstos, aprovechando las características generales que tal distribución teórica tenga; por ejemplo, en el caso normal, los datos pueden ser sucintamente descritos a partir de la media y la desviación típica, e indicando que la distribución empírica de los datos está bien aproximada por la distribución normal.

Una segunda razón es que tales afirmaciones distribucionales nos permiten hacer uso de útiles procedimientos estadísticos cuyos resultados no serían ni fácil ni claramente entendibles en el supuesto de no poder establecerlas; por ejemplo, al admitir que los datos están generados por una distribución normal, se nos permite aplicar, con garantías de éxito, técnicas como la del análisis de la varianza, o la del análisis factorial.

En tercer lugar, señalemos que las suposiciones distri-



bucionales nos permiten determinar la distribución en el muestreo de estadísticos cuyo uso será necesario durante el análisis de los datos y así, hacer inferencias y establecer sentencias probabilísticas sobre aspectos desconocidos del comportamiento de los datos bajo estudio; por ejemplo, el suponer que los datos provienen de una población distribuida normalmente, nos permite usar la distribución t de Student y la distribución  $\chi^2$  de Pearson para la construcción de intervalos de confianza para la media y la varianza, respectivamente, de la distribución teórica.

Por último, en cuarto lugar, y no por ello menos importante, citaremos una razón que podríamos denominar de retroalimentación (feedback). Son numerosas las ocasiones en las que el conocimiento de la distribución de un conjunto de datos puede explicar los mecanismos físicos involucrados en la generación de éstos, y que posibles anomalías detectadas en ellos, sean corregidas o analizadas y explicadas en su justa medida.

Si bien es cierto que, por lo anteriormente indicado, podemos obtener bastante ventaja en el análisis de los datos al establecer sobre ellos suposiciones distribucionales, no es menos seguro que aquellos análisis que estén basados en tales suposiciones, carecerán de validez si éstas no tienen un grado razonable de confianza. Es precisamente en este momento en el que nos planteamos la necesidad de manejar una serie de técnicas que nos permitan aceptar o rechazar las suposiciones establecidas respecto el comportamiento distribucional del conjunto de datos sobre el que estamos trabajando. Más concretamente, nos referimos al estudio de técnicas encaminadas a determinar la razonabilidad de la hipótesis de normalidad para un conjunto de observaciones. Tales técnicas nos permitirán, en unos casos, reafirmarnos en la validez de los procedimientos seguidos y, en otros, nos servirán como guía para el posterior análisis de los datos, quizá para sugerir la necesidad y naturaleza de las transformaciones que nos aproximen a la normalidad o, tal vez, para indicar las variaciones apropiadas de



los métodos para el posterior estudio y análisis de las observaciones.

Cuando la hipótesis de normalidad se establece sobre un conjunto de datos univariantes, y llegamos a la conclusión de que ésta no es válida, es interesante establecer cuál es la naturaleza de tal negativa. Las causas más frecuentes por las que un conjunto de datos no puedan aceptarse como generados por una distribución normal, están recogidas en las siguientes tres categorías:

- Por falta de simetría, presentándose en la distribución de los datos una cola, bien a la derecha, bien a la izquierda, más larga que en la distribución normal. Hablaremos, en estos casos, de asimetría positiva o asimetría negativa, respectivamente.
- Por aparecer en la distribución de los datos colas más o menos pesadas que en la distribución normal, presentándose así distribuciones muy "empinadas" o muy "aplastadas" (distribuciones leptocúrticas o platicúrticas, respectivamente).
- Por la existencia de más de una moda (distribuciones multimodales)

Cuando se trata de un estudio de datos multirespuesta (multivariantes), la hipótesis de normalidad puede no satisfacerse por otras muchas causas: unas por falta de normalidad en las distribuciones marginales de las variables que conforman el vector aleatorio (caso univariante anterior), que como sabemos constituye una condición necesaria de multinormalidad; y otras, propias del comportamiento multivariante de los datos.

En resumen, lo expuesto hasta el momento justifica la necesidad de una gran variedad de técnicas que sean sensibles a los distintos tipos de causas de no normalidad. Además, podemos



adelantarnos al calificar de "inútil" cualquier intento de búsqueda de la "mejor técnica" en términos absolutos, ya que cada una de ellas se conformará en "la mejor", en función de las distintas condiciones ambientales que aparezcan en el transcurso del análisis de los datos (tamaño de la muestra, restricciones temporales, disponibilidad de software y hardware para llevar a cabo complicados y laboriosos cálculos, técnicas estadísticas previstas a aplicar posteriormente a los datos, etc.).

Sin embargo, antes de aplicar alguna de las distintas técnicas existentes para el estudio de normalidad, en ocasiones resulta conveniente realizar una serie de pruebas elementales que, sin guardar total rigurosidad y vistas como meros tanteos, nos pueden indicar la posible tendencia de las observaciones en lo que se refiere a su probable normalidad y que, posteriormente, nos ayuden a decidir la técnica específica a aplicar, que generalmente será aquella que incida más directamente en las causas que nos llevan a sospechar la falta de normalidad en los datos. En este sentido, recomendamos, con cautela, el uso de diagramas de puntos (en el caso de disponer de pocos datos) o de histogramas (cuando el tamaño muestral es suficientemente elevado) que nos pueden resultar útiles para revelar posibles faltas de simetría que nos conducen a no aceptar la hipótesis de normalidad en los datos.

En la misma línea, teniendo en cuenta que la distribución normal univariante asigna al intervalo  $(\mu - \sigma, \mu + \sigma)$  una probabilidad de 0'683 y al intervalo  $(\mu - 2\sigma, \mu + 2\sigma)$  de 0'954, y supuesto que disponemos de una muestra suficientemente grande, en caso de normalidad, esperaremos que las frecuencias de pertenencia de los datos a los intervalos  $(\bar{x} - s, \bar{x} + s)$  y  $(\bar{x} - 2s, \bar{x} + 2s)$  se sitúen en torno a 0'683 y 0'954 respectivamente, lo que, de no ocurrir, nos indicaría un posible comportamiento leptocúrtico o platicúrtico en los datos que evidencien su falta de normalidad.



## CLASIFICACION DE LOS CONTRASTES DE NORMALIDAD

Sin pretender ser exhaustivos, si llevamos a cabo una revisión de los distintos contrastes de normalidad existentes, observamos que, en su gran mayoría, éstos se generan principalmente desde tres puntos de vista: en primer lugar, poniendo de relieve algunas de las causas principales de no normalidad ya mencionadas; en segundo lugar, usando propiedades que son características de la distribución normal; y, en tercer lugar, aprovechando las múltiples aproximaciones y caracterizaciones de la distribución normal.

En la bibliografía al uso, encontramos distintas clasificaciones de los variados contrastes de normalidad. Así, Daniel Peña Sánchez de Rivera y Juan Ignacio Peña Sánchez de Rivera (1986) proponen clasificar los contrastes de normalidad en cuatro grandes grupos:

-En el primero se mide si la distancia entre el vector de frecuencias observadas y el vector de frecuencias teóricas es lo suficientemente pequeña. Dentro de este grupo, el más popular es el de Pearson, que utiliza como estadístico la distancia de Mahalanobis

$$D_1 = (x-\mu)' \Sigma (x-\mu)$$

donde  $x$  es el vector de frecuencias observadas,  $\mu$  es el vector de frecuencias teóricas si las observaciones provienen de una distribución normal, y  $\Sigma$  es la matriz de varianzas y covarianzas de la variable multinomial definida por las clases. Un problema central de este contraste es cómo construir el vector  $x$ , esto es, cómo agrupar las frecuencias en clases. (En un planteamiento de este problema, tuvimos ocasión de elaborar un estudio y una propuesta en lo que se refiere a la división del recorrido de la variable en clases, que constituyó el contenido de la memoria que



presentamos para obtener el grado de Licenciado, recogido bajo el título "Aportaciones a las caracterizaciones y a los contrastes de normalidad").

-El segundo grupo utiliza la distancia entre la distribución empírica,  $F_n(x)$ , y la distribución teórica,  $F(x)$ . El estadístico más representativo de este tipo es el de Kolmogorov-Smirnov, que está dado por:

$$D_2 = \text{máx} |F_n(x) - F(x)|$$

Crámer y von Mises sugirieron definir la distancia entre estas dos funciones por:

$$D_3 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)$$

-El tercer grupo se basa en que una muestra de una población normal da lugar a una recta al dibujarla en papel probabilístico normal, siendo el coeficiente de correlación entre las observaciones ordenadas en forma creciente y sus valores esperados, una medida de la bondad de este ajuste. Los test de Shapiro y Wilk (1965) y Shapiro y Francia (1972), así como otras modificaciones de éstos, responden a este principio.

-El cuarto grupo se centra en el análisis de los coeficientes de asimetría y curtosis, estudiados ya por separado, ya conjuntamente.

El primer grupo y ciertas variaciones del tercero, sobre las que trataremos extensamente en este trabajo, pueden aplicarse al ajuste de cualquier tipo de distribución; el segundo, sólo para distribuciones continuas, y el cuarto grupo es específico para las distribuciones gaussianas. Por otro lado, los tres primeros dan lugar a contrastes de carácter general, mientras que el último va

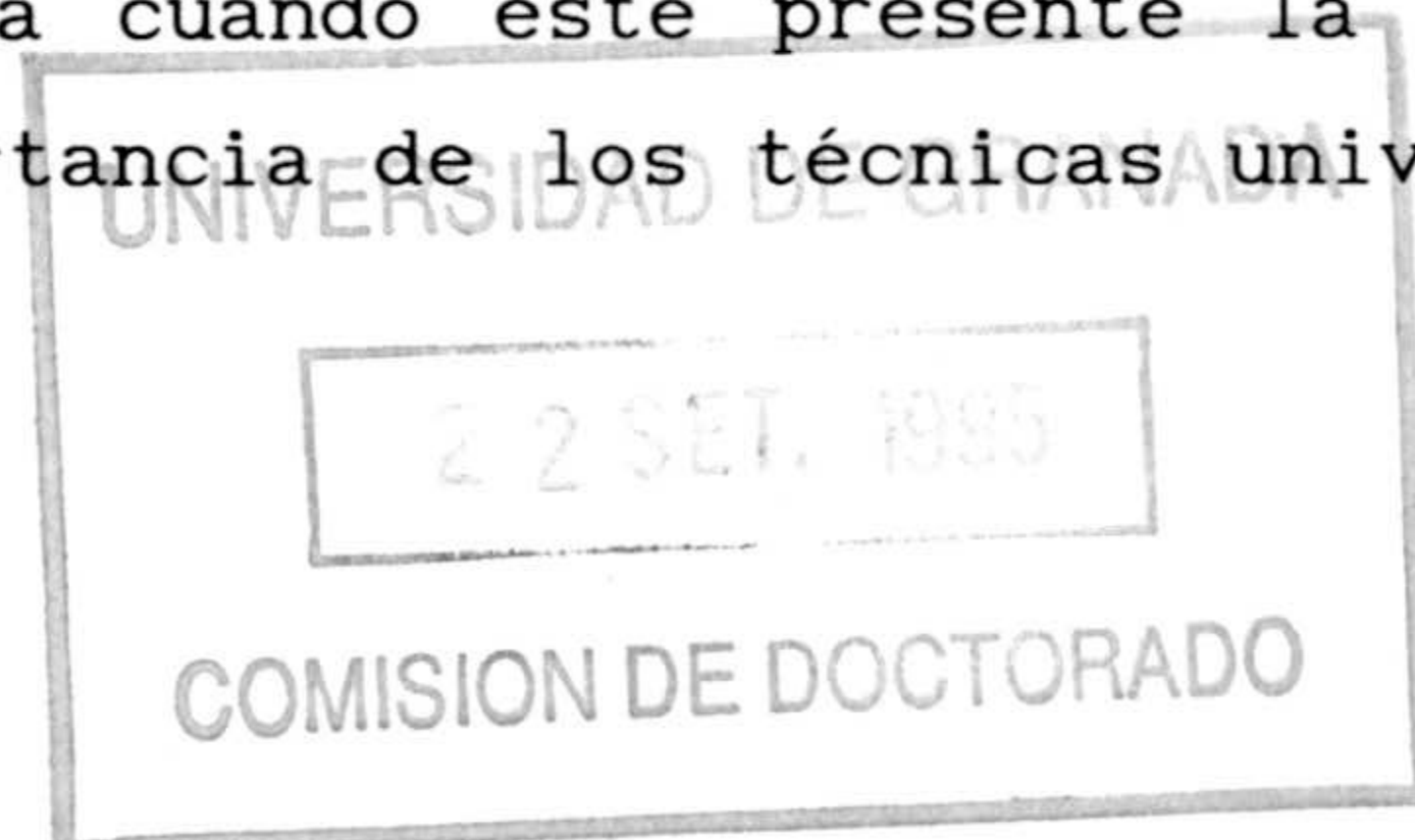


dirigido específicamente hacia la asimetría y la curtosis, estudiadas, como se ha indicado, conjunta o separadamente.

Por otra parte, Koziol (1986) advierte que la mayoría de las técnicas para contrastar la hipótesis de normalidad multivariante aparecen como generalizaciones de procedimientos o técnicas univariantes. Por ello, presenta una clasificación de los métodos univariantes destacando los más comunes y reuniéndolos en cuatro bloques: (i) procedimientos generales de bondad de ajuste, tales como  $\chi^2$ , Kolmogorov-Smirnov, Crámer-von Mises y Shapiro-Wilk; (ii) test de razón de verosimilitud con alternativas especificadas; (iii) contrastes basados en medidas de asimetría y curtosis; y (iv) procedimientos gráficos informales.

Andrews, Gnanadesikan y Warner (1973), al hablar de las distintas técnicas para asegurar normalidad, distinguen entre el caso univariante y el multivariante. El caso univariante es abordado tanto por ser el propio para el análisis de variables aleatorias, como por poder considerarlo como complemento en el estudio de vectores aleatorios, dado el carácter univariante de las variables que conforman el vector aleatorio, y teniendo en cuenta que aunque la normalidad marginal no implica normalidad conjunta, la presencia de muchos tipos de no normalidad es, a menudo, reflejada en las distribuciones marginales como tal. Así, engloban a los distintos métodos para asegurar normalidad univariante en los siguientes cuatro grupos: (i) Test de razón de verosimilitud asociado con transformaciones para incrementar la normalidad univariante; (ii) tests de asimetría y curtosis; (iii) contrastes generales de normalidad; y (iv) técnicas gráficas basadas en gráficos de probabilidad normal.

Al tratar el caso multivariante, se vuelve a poner de relieve que, en la práctica, y salvo casos raros o patológicos, la no normalidad marginal se presentará cuando esté presente la no normalidad conjunta. De ahí la importancia de los técnicas univa-





riantes. No obstante, éstas nunca llegan a explorar la naturaleza multivariante de los datos, y es por esto por lo que es necesario introducir técnicas apropiadas que sean sensibles al comportamiento multivariante de las observaciones.

Los diversos métodos encaminados a esta necesidad son recogidos en los siguientes apartados: (i) contrastes de razón de verosimilitud asociados con transformaciones encaminadas a incrementar la normalidad conjunta; (ii) contrastes de bondad de ajuste tales como el test  $\chi^2$  o el de Kolmogorov-Smirnov, y tests basados en densidades locales; y (iii) métodos gráficos informales combinados con una representación en radios y ángulos de los datos.

Por último, Mardia (1980) al tratar sobre contrastes de normalidad univariante y multivariante, establece una clasificación más amplia que las anteriores, distinguiendo los siguientes grupos: (i) Tests basados en medidas descriptivas (asimetría y curtosis, cumulantes, desviación media, etc.); (ii) el test de Shapiro-Wilk y sus modificaciones (W de Shapiro-Wilk, D de D'Agostino, W' de Shapiro-Francia,  $\tilde{W}$ ' de Weisberg-Bingham); (iii) contrastes de razón de verosimilitud asociado con transformaciones; (iv) tests de bondad de ajuste ( $\chi^2$ , Kolmogorov-Smirnov, Crámer-von Mises, Kuiper, etc.); y (v) otros tipos de contrastes (test basados en el principio de entropía, el estadístico U de Locke y Spurrier, gráficos de probabilidad, etc.).

Aunque ni las anteriores clasificaciones, ni otras muchas que podamos encontrar en la literatura al uso, coinciden plenamente en los grupos donde encuadrar las múltiples técnicas diseñadas para comprobar la validez de la hipótesis de normalidad de un conjunto de datos, si parecen estar de acuerdo en permitirnos establecer una clasificación resumen donde recoger las principales tests de normalidad. Esta puede ser:

- 1.- Contastes tipo chi-cuadrado.



- 2.- Contrastes basados en la función de distribución empírica.
- 3.- Contrastes basados en medidas descriptivas.
- 4.- El test de Shapiro y Wilk y sus modificaciones.
- 5.- Contraste de razón de verosimilitudes asociado con transformaciones para incrementar la normalidad.
- 6.- Técnicas gráficas informales.

Tanto si nos decidimos por alguna de las clasificaciones propuestas por los autores ya mencionados, como si aceptamos por válida la que proponemos en este trabajo, observamos que aparece un apartado relativo a técnicas gráficas ("informales") para el estudio de la normalidad.

Es precisamente en ese grupo donde vamos a desarrollar el estudio principal de esta memoria. Concretamente, nos dedicaremos al análisis de algunas técnicas gráficas, en particular, incidiremos en aquella que denominaremos gráfico probabilístico normal, que usaremos para analizar la hipótesis de normalidad enunciada sobre un conjunto de datos.







## CAPITULO 3

### CONTRASTES DE NORMALIDAD: BREVE RECORRIDO

En este capítulo expondremos aquellos contrastes de normalidad univariantes que hoy en día son considerados como los más relevantes, tanto por méritos propios, como por el largo uso que de ellos se ha dado en el transcurso de la historia.

Conscientes del elevado número de técnicas existentes encaminadas a la prueba de la normalidad, advertimos que está muy lejos de nuestra intención presentar un catálogo exhaustivo de todas ellas; antes, por el contrario, realizaremos una presentación de los más importantes, y de aquellos aspectos que le otorgan tal calificación.

#### EL AMBIENTE DE UN CONTRASTE DE NORMALIDAD

El ambiente en el que vamos a presentar los contrastes de normalidad univariantes es el siguiente:

Supongamos que  $X_1, X_2, \dots, X_n$  es una muestra aleatoria de tamaño  $n$ , tomada de una población con función de densidad  $f(x)$  y función de distribución  $F(x)$ . Sean  $\phi(x)$  y  $\Phi(x)$  la función de densidad y de distribución, respectivamente, de una distribución normal de parámetros  $\mu$  y  $\sigma$ . Esto es,

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \quad \begin{array}{l} -\infty < x < \infty \\ -\infty < \mu < \infty \\ 0 < \sigma < \infty \end{array}$$



y

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2} dt$$

Nuestra hipótesis nula,  $H_0$ , es:

$$H_0: f(x) = \phi(x) \quad \text{o} \quad H_0: F(x) = \Phi(x)$$

Al contrastar la hipótesis nula  $H_0$ , pretendemos comprobar si la variable  $X$ , bajo observación, está distribuida como una variable normal, o expresado con otras palabras, si la variable  $X$  tiene como función de densidad a  $\phi(x)$  (o como función de distribución a  $\Phi(x)$ ).

A estas alturas hemos de distinguir dos situaciones diferentes: si los valores de  $\mu$  y  $\sigma$  están especificados, diremos que la hipótesis nula es "simple", y ésta se referirá a una distribución normal en particular. Si  $\mu$ , o  $\sigma$ , o ambas, no están especificadas en la hipótesis nula, diremos que se trata de una hipótesis "compuesta". En general, no suele existir información disponible sobre  $\mu$  y  $\sigma$ , por lo que la hipótesis nula de normalidad suele ser una hipótesis compuesta. Por otra parte, desde el punto de vista de la potencia del contraste, no resulta ventajoso tener conocimiento sobre el valor de  $\mu$  y  $\sigma$  [ver, por ejemplo, Viollaz y Bru de Labanda (1992)].

En lo que se refiere a la hipótesis alternativa,  $H_1$ , usualmente es una hipótesis compuesta, enunciada en la forma "X no está distribuida normalmente". Existen casos en los que puede haber cierta información disponible concerniente a desviaciones de normalidad, generalmente enunciada en términos de asimetría y/o curtosis. En estos casos aconsejaremos el uso de aquellos contrastes que tengan en cuenta tal aspecto.



**BREVE EXPOSICION DE ALGUNOS CONTRASTES DE NORMALIDAD**

Enunciado el ambiente en el que se desarrollan los contrastes de normalidad, realizamos a continuación una breve exposición de algunos de ellos, destacando en cada uno sus aspectos más relevantes. Para llevar a cabo esta presentación de forma ordenada, nos ajustaremos a la clasificación que sobre contrastes de normalidad proponemos en el capítulo anterior.

Contrastes tipo chi-cuadrado.

Estos contrastes, bien conocidos, deben su origen a K. Pearson. Son adecuados tanto para hipótesis simples como compuestas. Para la distribución normal, su mecanismo consiste en discretizar la hipótesis distribucional (con parámetros conocidos o estimados) en una distribución multinomial con  $k$  celdas, contando el número de observaciones en cada celda y contrastando éste, vía el estadístico chi-cuadrado o un estadístico cociente de verosimilitud, con el número esperado de observaciones en cada celdilla. Este número esperado es determinado suponiendo que los datos provienen de la distribución normal. Más precisamente:

Supongamos que el espacio de la variable aleatoria  $X$  se descompone en un número finito de  $r$  partes,  $S_1, S_2, \dots, S_r$ , disjuntas. Sean  $p_1, p_2, \dots, p_r$  tales que

$$p_i = P\{S_i/H_0\} \quad i = 1, 2, \dots, r$$

verificando que:

$$\sum_{i=1}^r p_i = 1 \quad \text{y} \quad p_i > 0 \quad i = 1, 2, \dots, r$$

Considerada esta partición del espacio, sean  $\nu_1, \nu_2, \dots, \nu_r$  las fre-



cuencias de pertenencia de los datos de la muestra a cada uno de los conjuntos  $S_i$ . Tendremos que

$$\sum_{i=1}^r v_i = n$$

Dado que el objetivo es dar una medida de la desviación de la distribución hipotética con respecto a la real, y atendiendo a la teoría de mínimos cuadrados, obtenemos que ésta adoptará la forma:

$$\sum_{i=1}^r c_i \left[ \frac{v_i}{n} - p_i \right]^2$$

donde los coeficientes  $c_i$  pueden ser elegidos de distintas formas.

Pearson demostró que tomando  $c_i = (n/p_i)$ , se obtiene una medida de desviación con propiedades especialmente buenas. Aparece de esta manera el estadístico  $\chi^2$  de Pearson, que adopta la expresión

$$\chi^2 = \sum_{i=1}^r \frac{n}{p_i} \left[ \frac{v_i}{n} - p_i \right]^2 = \sum_{i=1}^r \frac{(v_i - np_i)^2}{np_i}$$

De esta forma, la medida de desviación obtenida con  $\chi^2$  queda expresada en función de las frecuencias observadas,  $v_i$ , y las frecuencias esperadas,  $np_i$ , en cada uno de los conjuntos  $S_i$ .

En el caso en que la hipótesis nula sea simple, es decir,  $\mu$  y  $\sigma$  conocidas, Pearson probó que el estadístico  $\chi^2$  sigue, asintóticamente una distribución chi-cuadrado con  $(r-1)$  grados de libertad.



En el supuesto de que los parámetros de la distribución bajo  $H_0$  fueran desconocidos, Pearson propuso estimarlos mediante cualquier estimador (como función de  $X_1, X_2, \dots, X_n$ ) y utilizar tales estimadores para determinar los  $p_i$ , apostando de nuevo por el estadístico  $\chi^2$ , creyendo que tendría el mismo comportamiento asintótico que antes, al menos, cuando los estimadores fueran consistentes.

No fue hasta 1924 cuando Fisher demostró que la distribución límite del estadístico  $\chi^2$ , con parámetros estimados, no era una chi-cuadrado con  $(r-1)$  grados de libertad, sino que ésta dependía del método usado para estimar los parámetros. (Este resultado está basado en el hecho de que los  $p_i$  pasan de ser constantes, a ser funciones de las observaciones muestrales, dado que dependen de las estimaciones de los parámetros desconocidos).

Si los parámetros desconocidos de la distribución son  $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ , Fisher argumentó que un método apropiado de estimación de los parámetros sería el de máxima verosimilitud, considerando la función de verosimilitud una vez que los datos están agrupados, esto es,

$$L = [p_1(\theta)]^{v_1} [p_2(\theta)]^{v_2} \dots [p_r(\theta)]^{v_r}$$

De esta forma, los estimadores serán solución del sistema de ecuaciones

$$\sum_{i=1}^r \frac{v_i}{p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_k} = 0 \quad k = 1, 2, \dots, s$$

Fisher, además, apuntó que el estadístico logaritmo del cociente de verosimilitud

$$G^2 = 2 \sum_{i=1}^r v_i \log \frac{v_i}{np_i}$$



era asintóticamente equivalente a  $\chi^2$ .

Así mismo, Fisher observó que un estimador asintóticamente equivalente al de máxima verosimilitud con los datos agrupados, era aquel que se obtenía eligiendo  $\theta$  de forma que se minimizase la expresión de  $\chi^2$ . Este método de estimación se conoce con el nombre de "método de estimación de la mínima  $\chi^2$ ", y los estimadores se obtienen como solución al sistema

$$\sum_{k=1}^r \frac{\partial p_k(\theta)}{\partial \theta_j} \left[ \frac{v_k - np_k(\theta)}{p_k(\theta)} + \frac{[v_k - np_k(\theta)]^2}{2np_k^2(\theta)} \right] = 0 \quad j = 1, 2, \dots, s$$

Aún en casos sencillos, el sistema anterior resulta difícil de resolver. Sin embargo, se puede demostrar que para valores grandes de  $n$ , la influencia del segundo término incluido en el corchete se hace despreciable, por lo que cuando  $n$  es suficientemente grande, bastará resolver el sistema

$$\sum_{k=1}^r \frac{\partial p_k(\theta)}{\partial \theta_j} \frac{v_k - np_k(\theta)}{p_k(\theta)} = 0 \quad j = 1, 2, \dots, s$$

que, por lo general, resulta más sencillo de manejar. Este método de estimación de los parámetros  $\theta_j$  recibe el nombre de "método modificado de la  $\chi^2$ ".

En el caso en que la estimación de los parámetros desconocidos se realice por alguno de los métodos anteriormente citados, se obtiene que el estadístico  $\chi^2$  de Pearson se comporta asintóticamente como una variable chi-cuadrado con  $(r-s-1)$  grados de libertad y, de aquí se desprende el famoso resultado de "pérdida de un grado de libertad por cada parámetro estimado".

Chernoff y Lehmann (1954) estudiaron las consecuencias



del uso de los estimadores de máxima verosimilitud obtenidos sin tener en cuenta la agrupación de los datos. Determinaron que la distribución límite del estadístico de Pearson, bajo la hipótesis nula, es

$$\chi^2_{(r-s-1)} + \sum_{k=1}^s \lambda_k(\theta) \chi^2_{k(1)}$$

donde  $\chi^2_{(r-s-1)}$  y  $\chi^2_{k(1)}$  son variables aleatorias chi-cuadrado independientes, con el número de grados de libertad indicados. Los números  $\lambda_k(\theta)$  satisfacen  $0 \leq \lambda_k(\theta) < 1$ . Así, el comportamiento asintótico del estadístico de Pearson, en este caso, no es chi-cuadrado, y depende del verdadero valor de  $\theta$ . Lo único que podemos decir, en general, es que los correctos puntos críticos caen entre los puntos críticos de una distribución  $\chi^2_{(r-s-1)}$  y los correspondientes de una  $\chi^2_{(r-1)}$ . Estas cotas hacen que, desde el punto de vista práctico, sea a menudo útil el uso del estadístico de Pearson, con los parámetros estimados por el método de máxima verosimilitud, cuando el número de celdas,  $r$ , es muy grande, y el número de parámetros desconocido,  $s$ , es muy pequeño.

A lo largo del tiempo, encontramos que se han propuesto otros estadísticos del tipo chi-cuadrado de Pearson, como, por ejemplo, el de Freeman-Tukey,

$$FT^2 = 4 \sum_{i=1}^r \{v_i^{1/2} - (np_i)^{1/2}\}^2$$

Cressie y Read (1984), por otra parte, han sistematizado la teoría del procedimiento clásico de la chi-cuadrado, introduciendo una clase de tests estadísticos basados en medidas de divergencia entre distribuciones discretas en  $r$  puntos. Si  $q = (q_1, q_2, \dots, q_r)$  y  $p = (p_1, p_2, \dots, p_r)$  son tales distribuciones discretas, la divergencia directa de orden  $\lambda$  entre  $q$  y  $p$  es:



$$I^{\lambda}(q:p) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^r q_i \left[ \left( \frac{q_i}{p_i} \right)^{\lambda} - 1 \right]$$

Si  $\nu$  es el vector de frecuencias  $\nu_i$ , y  $p(\theta)$  es el vector de probabilidades  $p_i(\theta)$ , los estadísticos de Cressie-Read miden las divergencias entre la distribución empírica  $\nu/n$  y la distribución hipotética  $p(\theta)$ , con la expresión

$$R^{\lambda}(\theta) = 2n I^{\lambda}(\nu/n:p(\theta))$$

Si  $I^{\lambda}$  está definida por continuidad en  $\lambda = 0$ , esta clase de estadísticos incluye a  $\chi^2(\lambda = 1)$ ,  $G^2(\lambda = 0)$  y  $FT^2(\lambda = -1/2)$ .

Aunque los tests de tipo chi-cuadrado son de interés histórico y, como hemos apuntado, están siendo continuamente estudiados y modificados, coincidimos con el profesor D. S. Moore (1986) en no recomendar "estos tipos de tests para el contraste de normalidad cuando la muestra no se encuentra agrupada". En este caso existen otros procedimientos que son bastante más potentes. En el caso en que no esté disponible toda la muestra, o cuando los datos están agrupados en clases, si admitiremos como útiles aquellos procedimientos.

#### Tests basados en la función de distribución empírica.

Básicamente, para la distribución normal, los test basados en la función de distribución empírica implican una medida de la discrepancia entre la función de distribución

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2} dt$$



de la distribución normal y la función de distribución empírica

$$F_n(x) = \frac{\text{número de observaciones } \leq x}{n}$$

de la muestra. Los parámetros  $\mu$  y  $\sigma$ , en la expresión anterior, a menudo no son especificados, siendo reemplazados por la media muestral  $\bar{X}$  y la desviación típica muestral  $S$ , como estimadores de los anteriores.

En el caso en que la hipótesis nula sea simple, (esto es,  $H_0: F(x) = \Phi(x)$ , y  $\mu$  y  $\sigma$  conocidas), presentamos siete estadísticos que están englobados en el grupo de contrastes basados en la función de distribución empírica. Todos ellos están concebidos para medir las diferencias verticales entre la función de distribución empírica y la función de distribución bajo la hipótesis nula. Estos estadísticos suelen ser divididos en dos clases, la *clase del supremo* y la *clase cuadrática*.

Entre los estadísticos de la *clase del supremo*, los dos primeros son  $D^+$  y  $D^-$  que son, respectivamente, la mayor diferencia vertical cuando  $F_n(x)$  es más grande que  $F(x)$ , y la mayor diferencia vertical cuando  $F_n(x)$  es más pequeña que  $F(x)$ . Formalmente,

$$D^+ = \sup_x \{F_n(x) - F(x)\} \quad \text{y} \quad D^- = \sup_x \{F(x) - F_n(x)\}$$

Entre los estadísticos de esta clase, el más conocido es el estadístico  $D$  introducido por Kolmogorov (1933) y Smirnov. Está definido por:

$$D = \sup_x |F_n(x) - F(x)| = \max(D^+, D^-)$$

Un estadístico estrechamente ligado al anterior, es el estadístico  $V$  de Kuiper (1960), muy útil cuando las observaciones están en un círculo, y que está definido por:



$$V = D^+ + D^-$$

Los estadísticos *cuadráticos* constituyen una amplia clase de medidas de discrepancia y vienen dados por la familia de Cramer-von Mises

$$Q = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \psi(x) dF(x)$$

donde  $\psi(x)$  es una función apropiada que pondera las diferencias al cuadrado,  $\{F_n(x) - F(x)\}^2$ . Cuando  $\psi(x) = 1$ , obtenemos el estadístico  $W^2$  de Cramer-von Mises. Cuando  $\psi(x) = [\{F(x)\}\{1-F(x)\}]^{-1}$ , aparece el estadístico  $A^2$  de Anderson-Darling. Una modificación de  $W^2$ , concebido también para el círculo, es el estadístico  $U^2$  de Watson, definido por

$$U^2 = n \int_{-\infty}^{\infty} \left\{ F_n(x) - F(x) - \int_{-\infty}^{\infty} \{F_n(x) - F(x)\} dF(x) \right\}^2 dF(x)$$

Las expresiones de los estadísticos anteriores, tanto en el caso supremo como cuadrático, no auguran facilidad en su cálculo. Veamos como conseguimos obtener expresiones computacionalmente asequibles. Para ello, hacemos uso de la transformación integral de probabilidad,  $Z = F(X)$ . Cuando  $F(x)$  es la verdadera distribución de  $X$ , la nueva variable aleatoria,  $Z$ , está distribuida uniformemente entre 0 y 1. Por ello, la función de distribución de  $Z$  será  $F^*(z) = z$ , para  $0 \leq z \leq 1$ .

Supongamos que una muestra  $X_1, X_2, \dots, X_n$  da valores  $Z_i = F(X_i)$ ,  $i = 1, 2, \dots, n$ , y sea  $F_n^*(z)$  la función de distribución empírica asociada a los  $Z_i$ .

El estadístico correspondiente puede ser ahora calculado comparando  $F_n^*(z)$  con  $F^*(z)$ , dado que puede demostrarse que, para



valores  $z$  y  $x$  relacionados por la expresión  $z = F(x)$ , las diferencias verticales en los diagramas de  $X$  y  $Z$  son iguales; esto es:

$$F_n(x) - F(x) = F_n^*(z) - F^*(z) = F_n^*(z) - z.$$

Esta relación permite establecer las siguientes fórmulas para el cálculo de los citados estadísticos a partir de los valores de  $Z$ .

Consideremos los valores de  $Z$  ordenados de menor a mayor, esto es,

$$Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$$

y sea

$$\bar{Z} = \sum_{i=1}^n \frac{Z_i}{n}$$

con ello,

$$\left. \begin{aligned} D^+ &= \max_i \{i/n - Z_{(i)}\} \\ D^- &= \max_i \{Z_{(i)} - (i-1)/n\} \end{aligned} \right\} D = \max(D^+, D^-) ; V = D^+ + D^-$$

$$W^2 = \sum_i \left\{ Z_{(i)} - \frac{2i-1}{2n} \right\}^2 + \frac{1}{12n} ;$$

$$U^2 = W^2 - n(\bar{Z} - 0.5)^2$$

$$A^2 = -n - \frac{1}{n} \sum_i (2i-1) [\log Z_{(i)} + \log\{1 - Z_{(n+1-i)}\}]$$

o bien,

$$A^2 = -n - \frac{1}{n} \sum_i [(2i-1) \log Z_{(i)} + (2n+1-2i) \log\{1 - Z_{(i)}\}]$$

En las expresiones anteriores, los logaritmos que aparecen son en base  $e$ , y tanto las sumas como los máximos están toma-



dos en  $i$ , para  $1 \leq i \leq n$ .

En resumen, el procedimiento a seguir con estos estadísticos para contrastar la hipótesis de normalidad con parámetros especificados, es:

(a) Ordenar las observaciones  $X_i$ , de la forma

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}.$$

(b) Calcular  $Z_{(i)} = F(X_{(i)})$ ,  $i = 1, 2, \dots, n$ .

(c) Calcular el correspondiente valor del estadístico haciendo uso de las anteriores expresiones de cálculo.

(d) Modificar el test estadístico como a se indica e la Tabla 3.1, y comparar con el correspondiente punto crítico. Si el valor modificado del estadístico ( $T^*$ ) excede al valor dado en la Tabla 3.2,  $H_0$  es rechazada al nivel de confianza  $\alpha$ .

Estadístico (T)	Modificación ( $T^*$ )
$D^+, D^-$	$D^+(\sqrt{n}+0'12+0'11/\sqrt{n})$
D	$D(\sqrt{n}+0'12+0'11/\sqrt{n})$
V	$V(\sqrt{n}+0'155+0'24/\sqrt{n})$
$W^2$	$(W^2-0'4/n+0'6/n^2)(1+1/n)$
$U^2$	$(U^2-0'1/n+0'1/n^2)(1+0'8/n)$
$A^2$	para todo $n \geq 5$

**TABLA 3.1:** Modificaciones de los estadísticos usados en los contrastes basados en la función de distribución empírica. ( $\mu$  y  $\sigma$  conocidos)



Estadístico modificado ( $T^*$ )	nivel de significación $\alpha$							
	.25	.15	.10	.05	.025	.01	.005	.001
$(D^+, D^-)^*$	0.828	0.973	1.073	1.224	1.358	1.518	1.628	1.859
$(D)^*$	1.019	1.138	1.224	1.358	1.480	1.628	1.731	1.950
$(V)^*$	1.420	1.537	1.620	1.747	1.862	2.001	2.098	2.303
$(W^2)^*$	0.209	0.284	0.347	0.461	0.581	0.743	0.869	1.167
$(U^2)^*$	0.105	0.131	0.152	0.187	0.222	0.268	0.304	0.385
$A^2$	1.248	1.610	1.933	2.492	3.070	3.880	4.500	6.000

**TABLA 3.2:** Puntos críticos para los test basados en la función de distribución empírica ( $\mu$  y  $\sigma$  conocidos)

Cuando realizamos un estudio con este tipo de estadísticos para el contraste de normalidad, siendo la hipótesis nula compuesta, distinguimos tres casos:

caso 1: La varianza  $\sigma^2$  es conocida y la media  $\mu$  es desconocida. Estimamos la media por la media muestral  $\bar{X}$ .

caso 2: La media,  $\mu$ , es conocida y la varianza,  $\sigma^2$ , desconocida. Estimamos la varianza por  $s_1^2 = \sum_i (X_i - \mu)^2 / n$ .

caso 3: Tanto la media como la varianza son desconocidas. Las estimamos con  $\bar{X}$  y  $s^2 = \sum_i (X_i - \bar{X})^2 / (n-1)$ , respectivamente.

Algunos de los estadísticos propuestos anteriormente han sido modificados para estos casos. El procedimiento a seguir con estos estadísticos para contrastar la hipótesis de normalidad, en cada uno de los casos anteriores es:



(a) Ordenar las observaciones  $X_i, X_{(1)} < X_{(2)} < \dots < X_{(n)}$ , y determinar

$$W_{(i)} = (X_{(i)} - \bar{X})/\sigma \quad (\text{Caso 1})$$

$$W_{(i)} = (X_{(i)} - \mu)/s_1 \quad (\text{Caso 2})$$

$$W_{(i)} = (X_{(i)} - \bar{X})/s \quad (\text{Caso 3})$$

(b) Calcular  $Z_{(i)} = \Phi(W_{(i)})$ ,  $i = 1, 2, \dots, n$ , donde por  $\Phi(x)$  notamos ahora a la función de distribución de la variable normal estándar.

(c) Calcular el correspondiente valor del estadístico usando las anteriores expresiones de cálculo.

(d) Para los casos 1 y 2, usar la Tabla 3.3 que nos da los puntos críticos para contrastar la normalidad cuando  $\mu$  es desconocida (caso 1), o  $\sigma^2$  es desconocida (caso 2) para los estadísticos  $W^2$  de Cramer-von Mises,  $U^2$  de Watson y  $A^2$  de Anderson-Darling. Para el caso 3, usar la Tabla 3.4 que recoge las modificaciones de los estadísticos  $D$  de Kolmogorov-Smirnov,  $V$  de Kuiper,  $W^2$  de Cramer-von Mises,  $U^2$  de Watson y  $A^2$  de Anderson-Darling, para luego usar la Tabla 3.5 de puntos críticos para estos estadísticos.

Estadístico	Nivel de significación $\alpha$							
	.25	.15	.10	.05	.025	.01	.005	.0025
$W^2$ caso 1	0.094	0.117	0.134	0.165	0.197	0.238	0.270	0.302
$W^2$ caso 2	0.190	0.263	0.327	0.442	0.562	0.725	0.851	0.978
$U^2$ caso 1	0.088	0.110	0.127	0.157	0.187	0.228	0.259	0.291
$U^2$ caso 2	0.085	0.105	0.122	0.151	0.180	0.221	0.252	0.284
$A^2$ caso 1	0.644	0.782	0.894	1.087	1.285	1.551	1.756	1.964
$A^2$ caso 2	1.072	1.430	1.743	2.308	2.898	3.702	4.324	4.954

**TABLA 3.3:** Puntos críticos para los test basados en la función de distribución empírica (caso 1 y caso 2).



Estadístico (T)	Modificación (T*)
D	$D(\sqrt{n}-0'01+0'85/\sqrt{n})$
V	$V(\sqrt{n}+0'05+0'82/\sqrt{n})$
$W^2$	$W^2(1 + 0'5/n)$
$U^2$	$U^2(1 + 0'5/n)$
$A^2$	$A^2(1+0'75/n+2.25/n^2)$

TABLA 3.4: Modificaciones de los estadísticos usados en los contrastes basados en la función de distribución empírica (caso 3)

Estadístico modificado (T*)	nivel de significación $\alpha$							
	.50	.25	.15	.10	.05	.025	.01	.005
(D)*			0.775	0.819	0.895	0.995	1.035	
(V)*			1.320	1.386	1.489	1.585	1.693	
( $W^2$ )*	0.051	0.074	0.091	0.104	0.126	0.148	0.179	0.201
( $U^2$ )*	0.048	0.070	0.085	0.096	0.117	0.136	0.164	0.183
( $A^2$ )*	0.341	0.470	0.561	0.631	0.752	0.873	1.035	1.159

TABLA 3.5: Puntos críticos de los test basados en la función de distribución empírica (caso 3).

Por último, y para acabar este apartado, citemos los recientes trabajos de Ahmad (1993), que usando la norma introducida por Cramer-von Mises, que él nota por  $U(\psi)$  y queda definida como

$$U(\psi) = \int_{-\infty}^{\infty} \{F(x) - F_0(x)\}^2 \psi(F_0(x)) dF_0(x)$$

propone determinar un estimador de  $U(\psi)$ ,  $\hat{U}_n(\psi)$ , tal que  $n^{1/2}\hat{U}_n(\psi)$



sea asintóticamente normal bajo la hipótesis nula, y para cualquier elección de  $\psi$ , de forma que los resultados se podrán aplicar a más elecciones de  $\psi$  de las que actualmente existen. Los resultados por él obtenidos podrán ser aplicados al caso multidimensional.

Tests basados en medidas descriptivas.

A.- ASIMETRIA Y CURTOSIS

Historicamente, el uso de los coeficientes de asimetría y curtosis ha constituido una manera básica de relatar la forma de una distribución de probabilidad. Ya, en 1895, Karl Pearson reconoció que desviaciones de normalidad pueden ser caracterizadas por el tercer y cuarto momento estandarizados de una distribución. Concretando, los tercer y cuarto momentos estandarizados son

$$\sqrt{\beta_1} = \frac{E(X-\mu)^3}{\sigma^3}$$

$$\beta_2 = \frac{E(X-\mu)^4}{\sigma^4}$$

Para la distribución normal,  $\sqrt{\beta_1} = 0$  y  $\beta_2 = 3$ .

El tercer momento  $\sqrt{\beta_1}$  caracteriza la asimetría de una distribución. Si una distribución es simétrica respecto de la media  $\mu$ , como es el caso de la distribución normal,  $\sqrt{\beta_1} = 0$ . Valores de  $\sqrt{\beta_1} \neq 0$  indican asimetría y, por tanto, ausencia de normalidad.

Por su parte, el cuarto momento  $\beta_2$  caracteriza la curtosis de una distribución. Para la distribución normal,  $\beta_2 = 3$ . Valores de  $\beta_2 \neq 3$  indican no normalidad.



Pearson sugirió que en la muestra, los tercer y cuarto momentos estandarizados, dados por

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}}$$

y

$$b_2 = \frac{m_4}{m_2^2}$$

donde

$$m_k = \frac{\sum_i (X_i - \bar{X})^k}{n}, \quad k > 1 \quad \text{y} \quad \bar{X} = \frac{\sum_i X_i}{n}$$

pueden usarse para juzgar alejamientos de la normalidad. De esta forma, las pruebas de normalidad irán dirigidas a comprobar, a partir de  $\sqrt{b_1}$  y  $b_2$ , el valor de  $\sqrt{\beta_1}$  y  $\beta_2$ , que, respectivamente y bajo normalidad, valdrán 0 y 3.

Por tanto, la hipótesis nula será enunciada, en este caso, como

$$H_0: \sqrt{\beta_1} = 0 \quad \text{y/o} \quad H_0: \beta_2 = 3$$

Respecto a la alternativa, ésta puede plantearse en distintas formas,

$$H_1: \sqrt{\beta_1} \neq 0 \quad \text{y/o} \quad H_1: \beta_2 \neq 3$$

$$H_1: \sqrt{\beta_1} < 0 \quad \text{y/o} \quad H_1: \beta_2 < 3$$

$$H_1: \sqrt{\beta_1} > 0 \quad \text{y/o} \quad H_1: \beta_2 > 3$$

apareciendo así contrastes a una cola o a dos colas. (El planteamiento de la hipótesis alternativa se hará en función de la información previa que pueda existir sobre el comportamiento de la distribución bajo estudio).



E.S. Pearson (1930) desarrolló, bajo la hipótesis nula de normalidad, los cuatro primeros momentos de la distribución muestral de  $\sqrt{b_1}$  y  $b_2$  hasta de orden  $n^{-3}$ . Para  $\sqrt{b_1}$  obtuvo

$$E(\sqrt{b_1}) = \beta_2(\sqrt{b_1}) = 0$$

$$\sigma(\sqrt{b_1}) = \sqrt{6/n} \left( 1 - \frac{3}{n} + \frac{6}{n^2} - \frac{15}{n^3} + \dots \right)$$

$$\beta_2(\sqrt{b_1}) = 3 + \frac{36}{n} - \frac{864}{n^2} + \frac{12096}{n^3} - \dots$$

Para  $b_2$  tenemos,

$$E(b_2) = 3 \frac{(n-1)}{(n+1)}$$

$$\sigma(b_2) = \sqrt{24/n} \left( 1 - \frac{15}{2n} + \frac{271}{8n^2} - \frac{2319}{16n^3} + \dots \right)$$

$$\beta_1(b_2) = \frac{216}{n} \left( 1 - \frac{29}{n} + \frac{519}{n^2} - \frac{7637}{n^3} + \dots \right)$$

$$\beta_2(b_2) = 3 + \frac{540}{n} - \frac{20196}{n^2} + \frac{470412}{n^3} - \dots$$

Así,  $\sqrt{b_1}$  tiene una distribución simétrica y leptocúrtica que tiende rápidamente a la distribución normal cuando  $n$  crece. Sin embargo,  $b_2$  es muy asimétrica para  $n = 100$  y fuertemente normal para  $n = 1000$ .

Para las situaciones donde la aproximación a la normalidad no es evidente, K. Pearson desarrollo su sistema de curvas, que puede ser usado para representar la distribución de  $\sqrt{b_1}$  y  $b_2$ , y a partir de ellas determinar puntos de significación para distintos tamaños muestrales.

D'Agostino y Tietjen (1973) y Mulholland (1977), mediante técnicas de simulación, dieron tablas con puntos de significa-



ción de la distribución de  $\sqrt{b_1}$  para distintos niveles y varios tamaños muestrales.

D'Agostino probó que la distribución bajo la hipótesis nula de  $\sqrt{b_1}$  puede considerarse bien aproximada por una curva tipo Johnson  $S_U$ . La aproximación se consigue de la siguiente forma:

(1) Calcular  $\sqrt{b_1}$  para los datos muestrales.

(2) Calcular

$$Y = \sqrt{b_1} \left\{ \frac{(n+1)(n+3)}{6(n-2)} \right\}^{1/2}$$

$$\beta_2 = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$$

$$W^2 = -1 + \{2(\beta_2 - 1)\}^{1/2}$$

$$\delta = 1/\sqrt{\log W}$$

$$\alpha = \{2/(W^2 - 1)\}^{1/2}$$

(3) Calcular

$$Z = \delta \log[Y/\alpha + \{(Y/\alpha)^2 + 1\}^{1/2}]$$

Así obtenida, Z sigue aproximadamente una distribución normal de media cero y varianza la unidad. Esta aproximación es válida para  $n \geq 8$ . La determinación de la región de rechazo (unilateral o bilateral), se hará sin más que referirnos a las tablas de la distribución normal.

De la misma forma, D'Agostino y Tietjen (1973) investigaron una aproximación a la distribución t de Student de la distribución de  $\sqrt{b_1}$ , bajo la hipótesis nula. El procedimiento por ellos propuesto es el siguiente:



Calculamos

$$T = \frac{\sqrt{b_1} \left( \frac{\nu}{\nu-2} \right)^{1/2}}{\sigma(\sqrt{b_1})}$$

donde

$$\nu = \frac{4\beta_2 - 6}{\beta_2 - 3} \quad \text{con} \quad \beta_2 = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$$

y

$$\sigma(\sqrt{b_1}) = \left[ \frac{6(n-2)}{(n+1)(n+3)} \right]^{1/2}$$

Bajo la hipótesis nula de normalidad, T sigue aproximadamente una distribución t de Student con  $\nu$  grados de libertad. Esta aproximación es válida para  $n \geq 8$ , y tan buena como la obtenida anteriormente por la curva de Johnson  $S_U$ .

La aproximación a la distribución t de Student nos sugiere inmediatamente una aproximación a la distribución normal, que está dada por

$$\sqrt{b_1} \left[ \frac{(n+1)(n+3)}{6(n-2)} \right]^{1/2}$$

y que será válida para  $n \geq 150$ .

Respecto a la distribución de  $b_2$ , indiquemos que D'Agostino y Tietjen (1971) y más tarde D'Agostino y Pearson (1973), mediante técnicas de simulación, dieron tablas con puntos de significación para distintos niveles de significación y varios tamaños muestrales.



Anscombe y Glynn (1983) dieron una aproximación de la distribución de  $b_2$  a la normal, obtenida de la siguiente forma:

(1) Calcular  $b_2$  a partir de las observaciones muestrales.

(2) Calcular la media y la varianza de  $b_2$  mediante

$$E(b_2) = \frac{3(n-1)}{n+1}$$

y

$$\text{Var}(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

(3) Calcular el valor normalizado de  $b_2$

$$x = \frac{b_2 - E(b_2)}{\sqrt{\text{Var}(b_2)}}$$

(4) Calcular el tercer momento estandarizado de  $b_2$

$$\sqrt{\beta_1}(b_2) = \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)} \sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}$$

(5) Calcular

$$A = 6 + \frac{8}{\sqrt{\beta_1}(b_2)} \left[ \frac{2}{\sqrt{\beta_1}(b_2)} + \sqrt{1 + \frac{4}{\beta_1(b_2)}} \right]$$



(6) Calcular

$$Z = \frac{\left(1 - \frac{2}{9A}\right) - \left[ \frac{1 - \frac{2}{A}}{1 + x \sqrt{\frac{2}{A-4}}} \right]^{1/3}}{\sqrt{\frac{2}{9A}}}$$

Donde Z se distribuye aproximadamente según una normal de media cero y varianza la unidad. Esta aproximación es válida cuando  $n \geq 20$ .

De la misma forma que ocurría con el coeficiente de asimetría muestral, podemos dar también una aproximación directa a la normal, en la forma

$$\frac{b_2 - E(b_2)}{\sqrt{\text{Var}(b_2)}}$$

que sigue aproximadamente una distribución normal estándar para valores extremadamente grandes de n (valores por encima de 1000, lo que hace que esta aproximación sea desaconsejada).

A partir de los trabajos anteriores se han desarrollado numerosas investigaciones dirigidas a proporcionar un test de normalidad de propósito general, combinando los resultados obtenidos para  $\sqrt{b_1}$  y para  $b_2$ . De destacar algunos, mencionemos el R-test, que, básicamente, consiste en usar los test sobre  $\sqrt{b_1}$  al nivel  $\alpha_1$  y  $b_2$  al nivel  $\alpha_2$ , y rechazar la hipótesis de normalidad si alguno de ellos la rechaza. El nivel de confianza,  $\alpha$ , para la combinación de estos tests, debería ser tal que verificara la desigualdad de Bonferroni, esto es,

$$\alpha \leq \alpha_1 + \alpha_2$$



Pearson, D'Agostino y Bowman (1977) demostraron que si  $\alpha_1 = \alpha_2 = 2\alpha^*$ , una buena aproximación al nivel de significación global es

$$\alpha = 4(\alpha^* - (\alpha^*)^2)$$

que sería exacto si  $\sqrt{b_1}$  y  $b_2$  fuesen independientes. Sin embargo, éstos son incorrelados, pero no independientes, y el uso de ese nivel de significación global produciría test conservadores.

D'Agostino y Pearson (1973) sugirieron el estadístico

$$K^2 = X^2(\sqrt{b_1}) + X^2(b_2)$$

donde  $X(\sqrt{b_1})$  y  $X(b_2)$  son variables normales estándar equivalentes, obtenidas a partir de  $\sqrt{b_1}$  y  $b_2$  como vimos anteriormente. De esta forma,  $K^2$  puede verse como una variable chi-cuadrado con dos grados de libertad. Este test ha sido desarrollado suponiendo independencia entre  $\sqrt{b_1}$  y  $b_2$ , pero como indican Bowman y Shenton (1986), éstas, aún siendo incorreladas, no son independientes, aunque estén próximas a la independencia. Así,  $K^2$  puede verse aproximadamente como una chi-cuadrado, siendo esta aproximación considerablemente buena cuando  $n \geq 100$ .

Bowman y Shenton (1975) sugirieron, de pasada, el estadístico

$$\frac{(\sqrt{b_1})^2}{\sigma_1^2} + \frac{(b_2 - 3)^2}{\sigma_2^2}$$

donde  $\sigma_1^2 = 6/n$  y  $\sigma_2^2 = 24/n$ , (que son las varianzas asintóticas de  $\sqrt{b_1}$  y  $b_2$ , respectivamente). El estadístico así definido está distribuido, bajo la hipótesis de normalidad, asintóticamente según una distribución chi-cuadrado con dos grados de libertad, pero debido a la lenta convergencia de  $b_2$  a normalidad, este estadísti-



co no resulta muy útil.

Otras propuestas en este ambiente son las dadas por Cox y Hinkley (1974), que sugirieron

$$\max (|\sqrt{b_1}|/\sigma_1 ; |b_2 - 3|/\sigma_2)$$

Otra posible es

$$-\log P(\sqrt{b_1}) - \log P(b_2)$$

donde  $P(\sqrt{b_1})$  y  $P(b_2)$  son transformaciones integrales de probabilidad de  $\sqrt{b_1}$  y  $b_2$ , respectivamente. Este estadístico seguirá aproximadamente una distribución chi-cuadrado con cuatro grados de libertad.

#### B. - OTROS CONTRASTES BASADOS EN MEDIDAS DESCRIPTIVAS

Un numerosísimo conjunto de técnicas encaminadas a probar la normalidad basadas en medidas descriptivas (momentos), han sido introducidas a lo largo de la historia. De entre todas, quizá la más destacable es el test de Geary (1935) definido como el cociente de la desviación media y la desviación estándar (supuesto que la media poblacional es conocida e igual a cero):

$$w = \frac{1}{\sqrt{n}} \frac{\sum |x_i|}{(\sum x_i^2)^{1/2}}$$

Bajo la hipótesis nula, Geary obtuvo todos los momentos de  $w$ . Este cociente, bajo normalidad, tiende asintóticamente a  $(2/\pi)^{1/2}$ .



Así mismo, Geary (1935) propuso

$$a = \frac{\sum |x_i - \bar{X}|}{S\sqrt{n}}$$

que es la extensión obvia de  $w$ , para cuando la media y la varianza poblacional son desconocidas. Para éste obtuvo:

$$E(a) = \sqrt{\frac{2}{\pi}} + \frac{0'199471}{n} + \frac{0'024934}{n^2} - \frac{0'031168}{n^3} - \frac{0'008182}{n^4} + \dots$$

$$\text{Var}(a) = \frac{0'045070}{n} - \frac{0'124648}{n^2} + \frac{0'084859}{n^3} + \frac{0'006323}{n^4} + \dots$$

$$\sqrt{\beta_1}(a) = - \frac{1'7618}{\sqrt{n}} \left\{ 1 - \frac{2'3681}{n} - \frac{8'8646}{n^2} + \dots \right\}$$

$$\beta_2(a) = 3 + \frac{5'441}{n} \left\{ 1 - \frac{7'628}{n} + \dots \right\}$$

de forma que bajo  $H_0$ ,  $a$  es asintóticamente normal con media  $(2/\pi)^{1/2}$  y varianza  $(0'045070/n)$ , aproximación válida para  $n \geq 41$ .

A su vez, Geary (1947), consideró la siguiente clase de tests estadísticos:

$$a(c) = \frac{1}{nS^c} \sum_{i=1}^n |x_i - \bar{X}|^c \quad \text{para } c \geq 1$$

y

$$g(d) = \frac{1}{nS^d} \left[ \sum_{x_i \geq \bar{X}} |x_i - \bar{X}|^d - \sum_{x_i < \bar{X}} |x_i - \bar{X}|^d \right] \quad \text{para } d \geq 1$$

Como casos especiales de estos test, tenemos:  $b_2 = a(4)$



y  $\sqrt{b_1} = g(3)$ . Además,  $a(1) = a$ .

Para probar la normalidad asintótica de  $a(c)$ , Geary lo aproximó por  $a_1(c)$ , definido como:

$$a_1(c) = \frac{1}{nS^c} \sum_{i=1}^n |x_i - \mu|^c$$

y obtuvo la distribución asintótica de  $a_1(c)$ . Sin embargo, las cantidades  $a_1(c)$  y  $a(c)$  no siempre tienen la misma distribución límite, ocurriendo sólo cuando la distribución poblacional es simétrica.

David, Hartley y Pearson (1954) presentaron una test de homogeneidad y normalidad basado en el cociente entre el rango y la desviación estándar de una muestra de tamaño  $n$ . El estadístico está definido por:

$$u = \frac{X_{(n)} - X_{(1)}}{S}$$

Dado que  $(x_i - \bar{X})/S$ ,  $i = 1, 2, \dots, n$ , es independiente de  $S$ , podemos escribir

$$E(u^r) = \frac{E\{(X_{(n)} - X_{(1)})^r\}}{E(S^r)}$$

Tomando muestras de distintos tamaños se calcularon los cuatro primeros momentos del numerador y del denominador de  $u$ , y a partir de éstos, usando la expresión anterior, los cuatro primeros momentos de  $u$ . Usando las curvas de Pearson con los apropiados primeros cuatro momentos, se obtuvieron tablas de puntos porcentuales de  $u$  a distintos niveles de confianza. [Pearson y Hartley (1972)].



El test de Shapiro y Wilk y sus modificaciones.

A.- EL TEST W DE SHAPIRO Y WILK

Con la aparición del estadístico W propuesto por Shapiro y Wilk (1965), reaparece el interés por el estudio de los contrastes de normalidad. El procedimiento introducido por estos autores está basado en los estadísticos de orden.

Sea  $\mathbf{c}$  el vector de valores esperados de los  $n$  estadísticos de orden de una distribución  $N(0,1)$ , y sea  $\mathbf{V}$  la correspondiente matriz de covarianzas esperadas. Para una distribución normal con media  $\mu$  y varianza  $\sigma^2$  sabemos que  $E(\mathbf{x})$ , donde  $\mathbf{x}$  es el vector de estadísticos de orden, está dada por

$$E(\mathbf{x}) = \mu \mathbf{1} + \sigma \mathbf{c}$$

donde  $\mathbf{1}$  es un vector columna con  $n$  unos. La matriz de covarianzas está dada por  $\sigma^2 \mathbf{V}$  y, de aquí, el mejor estimador lineal insesgado de  $\sigma$ , a partir del teorema de Gauss-Markov, adopta la expresión

$$\tilde{\sigma}^2 = \frac{\mathbf{c}' \mathbf{V}^{-1} \mathbf{x}}{\mathbf{c}' \mathbf{V}^{-1} \mathbf{c}}$$

Por otra parte, un estimador insesgado de  $\sigma^2$ , con momento de segundo orden finito, es

$$\hat{\sigma}^2 = \frac{\sum_i (x_i - \bar{X})^2}{(n-1)} = S^2$$

De esta forma, un test de normalidad se produce al comparar el cociente  $(\tilde{\sigma} / \hat{\sigma})^2$  con 1, obteniendo, salvo una constante



multiplicativa, el estadístico W.

W está habitualmente definido por

$$W = \frac{(\mathbf{a}'\mathbf{x})^2}{(n-1)S^2} = \frac{(\sum_i a_i x_{(i)})^2}{\sum_i (x_i - \bar{X})^2}$$

donde

$$\mathbf{a}' = \frac{\mathbf{c}'\mathbf{V}^{-1}}{(\mathbf{c}'\mathbf{V}^{-2}\mathbf{c})^{1/2}}$$

Los  $a_i$  son las ponderaciones óptimas para el estimador de mínimos cuadrados ponderado de  $\sigma$ , dado que la población está normalmente distribuida. (Una vez que tratemos el apartado correspondiente a técnicas gráficas para el contraste de normalidad, se puede observar que W puede coincide con el cuadrado del coeficiente de correlación en los gráficos de probabilidad normal).

Los valores exactos de  $a_i$ , para  $n = 3$  hasta 50, fueron dados por Shapiro y Wilk (1965). Para  $n > 50$ , sugirieron la siguiente aproximación:

$$a_i \approx 2m_i c_i \quad i = 2, 3, \dots, n-1$$

y

$$a_1^2 = a_n^2 = \frac{\Gamma\{(n+1)/2\}}{\Gamma\{(n/2)+1\}2^{1/2}}$$

Valores de W próximos a 1 indican normalidad, mientras que valores menores que 1, indican ausencia de normalidad. De esta forma, los valores en la cola baja de la distribución de W, bajo la hipótesis de normalidad, se usarán para determinar la región de rechazo.



B.- EL TEST D DE D'AGOSTINO.

El test de Shapiro-Wilk necesita de un conjunto de ponderaciones,  $a$ , para cada tamaño muestral. D'Agostino (1971) presenta una modificación del test W que no requiere del cálculo de dichas ponderaciones. Este test está dado como:

$$D = \frac{T}{n^2 \sqrt{m_2}} = \frac{T}{n^{3/2} \left\{ \sum_i (x_i - \bar{X})^2 \right\}^{1/2}}$$

donde

$$T = \sum_{i=1}^n \left[ i - \frac{1}{2} (n+1) \right] x_{(i)}$$

La esperanza de D es aproximadamente  $1/(2\sqrt{\pi}) = 0'28209479$ , y la desviación estándar es asintóticamente

$$\left[ \frac{12\sqrt{3} - 27 + 2\pi}{24\pi n} \right] = \frac{0'02998598}{\sqrt{n}}$$

de aquí que una variable aproximadamente estandar sea

$$Y = \frac{\sqrt{n} (D - 0'28209479)}{0'02998598}$$

que bajo la hipótesis de normalidad es asintóticamente  $N(0,1)$ .

Si  $H_0$  no es cierta, el valor de Y tenderá a ser distinto de cero. Un estudio de simulación permitió a D'Agostino indicar que para distribuciones alternativas con menos curtosis que la distribución normal, Y tiende a ser mayor que cero, y para distribuciones con más curtosis, Y tiende a ser menor que cero. De esta forma, para contrastar la normalidad contra la hipótesis alterna-



tiva de "no normalidad", se aconseja usar un contraste a dos colas. Si la hipótesis alternativa está dirigida (esto es  $\beta_2 < 3$  o  $\beta_2 > 3$ ) puede usarse un contraste unilateral.

C.- EL TEST W' DE SHAPIRO Y FRANCIA.

Shapiro y Francia (1972) abordaron el problema de las ponderaciones,  $a$ , del test de Shapiro-Wilk de forma distinta. Observaron que para muestras grandes, los estadísticos de orden pueden ser tratados como si fueran independientes. De esta forma, propusieron reemplazar el estadístico W por

$$W' = \frac{(\mathbf{b}' \mathbf{x})^2}{(n-1)S^2} = \frac{\sum_i (b_i X_{(i)})^2}{\sum_i (x_i - \bar{X})^2}$$

donde

$$\mathbf{b}' = \frac{\mathbf{c}'}{(\mathbf{c}' \mathbf{c})^{1/2}}$$

siendo  $c_i$ , como antes, el valor esperado del  $i$ -ésimo estadístico de orden de una distribución normal estándar.

Shapiro y Francia (1972), proporcionaron las ponderaciones  $b_i$ , y valores críticos de  $W'$  para los tamaños muestrales  $n = 35, 50, 51(2)99$ . Pearson, D'Agostino y Bowman (1977) observaron que esos valores críticos estaban calculados vía simulación, a partir de sólo 1.000 muestras. Ellos reevaluaron los puntos porcentuales para  $n = 99, 100$  y  $125$  basándose en 50.000 simulaciones. Comparando los resultados, concluyeron que los valores obtenidos por Shapiro-Francia en la cola baja eran más grandes que los obtenidos por ellos. Esto les llevó a producir niveles de significación más grandes que los indicados por Shapiro y Francia. Además, indicaron que el test ahora era más potente de lo que era hasta ese momento.



D.- EL TEST  $\tilde{W}'$  DE WEISBERG-BINGHAM.

Weisberg y Bingham (1975) sugirieron sustituir las ponderaciones  $\mathbf{b}$  del test  $W'$  de Shapiro-Francia por  $\mathbf{d}$ , donde

$$\mathbf{d} = \frac{\tilde{\mathbf{c}}}{(\tilde{\mathbf{c}}' \tilde{\mathbf{c}})^{1/2}}$$

estando los elementos del vector  $\tilde{\mathbf{c}}$  definidos por

$$\tilde{c}_i = \Phi^{-1} \left[ \frac{i - 3/8}{n + 1/4} \right] \quad \text{para } i = 1, 2, \dots, n$$

y con  $\Phi^{-1}(p)$  la inversa de la función de distribución normal estándar.

De esta forma, el estadístico se convierte en:

$$\tilde{W}' = \frac{(\tilde{\mathbf{c}}' \mathbf{x}) / (\tilde{\mathbf{c}}' \tilde{\mathbf{c}})}{\sum_i (x_{(i)} - \bar{X})^2}$$

La aproximación a  $c_i$  por  $\tilde{c}_i$  fue sugerida por Blom (1958) y su uso en el estadístico  $\tilde{W}'$  hace que su distribución bajo la hipótesis nula sea muy próxima a la de  $W'$ , al menos para  $n = 5, 20$  y  $35$ , donde los autores realizaron tal comparación. Dada la ventaja de no tener que almacenar constantes, los autores sugirieron usar  $\tilde{W}'$  en lugar de  $W'$ , aún manteniendo el uso de las tablas de valores críticos de  $W'$ .



E.- EL CONTRASTE DEL COEFICIENTE DE CORRELACION EN GRAFICOS DE PROBABILIDAD NORMAL DE FILLIBEN.

Filliben (1975) propone lo que él denomina un coeficiente de correlación en gráficos de probabilidad normal,  $r$ , como un estadístico para contrastar la hipótesis compuesta de normalidad.

Entre las bondades que el autor atribuye a  $r$ , indica:

- (1) Es fácil de entender conceptualmente, dado que combina dos conceptos simples: los gráficos de probabilidad y el coeficiente de correlación.
- (2) Es simple desde el punto de vista computacional, dado que todos los cálculos necesarios son internos y no necesita de almacenamiento de datos.
- (3) Se puede extender al caso en que la hipótesis nula trate de otra distribución distinta a la normal.

La idea básica del contraste de normalidad desarrollado por Filliben es realizar un gráfico de probabilidad de las observaciones ordenadas ( $X_{(i)}$ ) contra las correspondientes medianas de los estadísticos de orden de una distribución normal estándar. Una vez obtenido este gráfico, el estadístico queda definido como el coeficiente de correlación producto-momento de estos valores.

De esta forma, si  $M_i$  es la mediana del  $i$ -ésimo estadístico de orden de una distribución normal estándar, el estadístico coeficiente de correlación en gráficos de probabilidad normal,  $r$ , está definido por:

$$r = \text{corr}(X, M) = \frac{\sum_i (X_{(i)} - \bar{X})(M_i - \bar{M})}{\sqrt{\sum_i (X_{(i)} - \bar{X})^2 \sum_i (M_i - \bar{M})^2}}$$



Filliben propone una tabla de valores críticos para  $r$  obtenida mediante una aproximación a su distribución bajo normalidad, para tamaños muestrales  $n = 3(1)50(5)100$ .

F.- EXTENSION DE LA BRECQUE.

Como se ha indicado anteriormente, el modelo base bajo el que se ha desarrollado el contraste de Shapiro-Wilk era:

$$E(X_{(i)}) = \mu + \sigma E(Y_{(i)}) \quad , \quad i = 1, 2, \dots, n \quad \text{Modelo (A)}$$

donde los  $X_{(i)}$  son los estadísticos de orden de una muestra aleatoria de una población normal de parámetros  $\mu$  y  $\sigma$ , e  $Y_{(i)} = (X_{(i)} - \mu)/\sigma$  son los estadísticos de orden de una distribución normal estándar.

Sean  $\xi' = (\xi_1, \xi_2, \dots, \xi_n)$  y  $\mathbf{V} = (\sigma_{ij})$  el vector de valores esperados y la matriz de varianzas-covarianzas de los estadísticos de orden de la normal estándar, respectivamente.

La Brecque (1977), propone una alternativa al modelo (A), introduciendo uno o dos términos adicionales, expresados en potencias de los  $\xi_i$ , en la forma:

$$E(X_i) = \mu + \sigma \xi_i + \alpha \phi_2(\xi_i) + \beta \phi_3(\xi_i), \quad i = 1, 2, \dots, n \quad \text{Modelo (B)}$$

donde  $\phi_j(\xi_i)$  es un polinomio de grado  $j$  en  $\xi_i$ . Si el modelo (A) es el correcto,  $\alpha = \beta = 0$ . Además, si no fuese el correcto, un ajuste del modelo (B) a los datos será mas apropiado que el modelo (A).

En un estudio de los términos que aparecen en el modelo (B), La Brecque observa que es conveniente el uso de funciones cuadráticas en  $\xi_i$  cuando las alternativas son asimétricas, y el uso de funciones cúbicas cuando las alternativas son simétricas.



Los estimadores de mínimos cuadrados de  $\mu$  y  $\sigma$  en el modelo (A) son

$$\hat{\mu} = C_{\mu}^2 \mathbf{1}' \mathbf{V}^{-1} \mathbf{X}$$

$$\hat{\sigma} = C_{\sigma}^2 \xi' \mathbf{V}^{-1} \mathbf{X}$$

donde  $\mathbf{X}$  es el vector de los estadísticos de orden,  $C_{\mu}^2 = (\mathbf{1}' \mathbf{V}^{-1} \mathbf{1})^{-1}$  y  $C_{\sigma}^2 = (\xi' \mathbf{V}^{-1} \xi)^{-1}$ . Sea  $\xi^{(r)'} = (\xi_1^r, \xi_2^r, \dots, \xi_n^r)$  y definamos  $\Phi_2$  y  $\Phi_3$  como

$$\Phi_2 = \xi^{(2)} - C_{\mu}^2 \mathbf{1}' \mathbf{V}^{-1} \xi^{(2)} \mathbf{1}$$

$$\Phi_3 = \xi^{(3)} - C_{\sigma}^2 \xi' \mathbf{V}^{-1} \xi^{(3)} \xi$$

Para  $\mathbf{A}$ , tal que  $\mathbf{A}' \mathbf{A} = \mathbf{V}^{-1}$ , los vectores  $\mathbf{A}\mathbf{1}$ ,  $\mathbf{A}\xi$ ,  $\mathbf{A}\Phi_2$  y  $\mathbf{A}\Phi_3$  son ortogonales.

Los estimadores de los parámetros asociados con  $\mathbf{1}$  y  $\xi$  son  $\hat{\mu}$  y  $\hat{\sigma}$  dados anteriormente. Los estimadores de  $\alpha$  y  $\beta$  en el modelo (B) son entonces

$$\hat{\alpha} = C_{\alpha}^2 \Phi_2' \mathbf{V}^{-1} \mathbf{X}$$

$$\hat{\beta} = C_{\beta}^2 \Phi_3' \mathbf{V}^{-1} \mathbf{X}$$

donde

$$C_{\alpha}^2 = (\Phi_2' \mathbf{V}^{-1} \Phi_2)^{-1} \quad \text{y} \quad C_{\beta}^2 = (\Phi_3' \mathbf{V}^{-1} \Phi_3)^{-1}.$$



La Brecque considera en su trabajo tres modelos alternativos:

$$(1) E(X_{(i)}) = \mu + \sigma \xi_i + \alpha \phi_2(\xi_i) + \beta \phi_3(\xi_i)$$

$$(2) E(X_{(i)}) = \mu + \sigma \xi_i + \alpha \phi_2(\xi_i)$$

$$(3) E(X_{(i)}) = \mu + \sigma \xi_i + \beta \phi_3(\xi_i)$$

donde  $\phi_j(\xi_i)$  es el  $i$ -ésimo elemento de  $\Phi_j$ . Mientras que (1) es una extensión natural de (2), la introducción de (3) está justificada porque éste es más apropiado para distribuciones simétricas que (1).

La Brecque propone los estadísticos

$$F_j = \frac{D_j}{ks^2} \quad j = 1, 2, 3$$

donde  $D_j$  es la reducción de la suma de cuadrados de residuos para el modelo (j),  $s^2$  es el usual estimador cuadrático para  $\sigma^2$  y  $k$  es el número de parámetros adicionales usados. Los test estadísticos correspondientes a los modelos alternativos (1), (2) y (3) son:

$$F_1 = \frac{\hat{\alpha}^2/C_\alpha^2 + \hat{\beta}^2/C_\beta^2}{2s^2}$$

$$F_2 = \frac{\hat{\alpha}^2/C_\alpha^2}{s^2}$$

$$F_3 = \frac{\hat{\beta}^2/C_\beta^2}{s^2}$$

Los coeficientes necesarios para calcular  $\hat{\alpha}/C_\alpha$  y  $\hat{\beta}/C_\beta$  son



$$\mathbf{a} = C_{\alpha} \mathbf{V}^{-1} \Phi_2$$

$$\mathbf{b} = C_{\beta} \mathbf{V}^{-1} \Phi_3$$

para los que La Brecque, en función del tamaño de la muestra ( $4 \leq n \leq 64$ ), propone una serie de tablas.

Introducidos estos estadísticos, el procedimiento para contrastar la hipótesis de que la muestra proviene de una población normal de media  $\mu$  y varianza  $\sigma^2$  desconocidas, se puede resumir en los siguientes apartados:

- (i) Ordenar la muestra:  $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ .
- (ii) Calcular el valor de  $s^2$ .
- (iii) Para el tamaño de muestra apropiado, determinar  $\mathbf{a}$  y/o  $\mathbf{b}$  para calcular  $\hat{\alpha}/C_{\alpha}$  y/o  $\hat{\beta}/C_{\beta}$ , respectivamente. (Tener en cuenta que  $a_i = a_{n-i+1}$  y  $b_i = b_{n-i+1}$ ).
- (iv) Calcular

$$\hat{\alpha}/C_{\alpha} = \begin{cases} \sum_{i=1}^{n/2} a_i (X_{(i)} - X_{(n-i+1)}) & \text{para } n \text{ par} \\ \sum_{i=1}^{[n/2]} a_i (X_{(i)} - X_{(n-i+1)}) + a_{(n+1)/2} X_{(n+1)/2} & \text{para } n \text{ impar} \end{cases}$$

y

$$\hat{\beta}/C_{\beta} = \sum_{i=1}^{[n/2]} b_i (X_{(i)} - X_{(n-i+1)})$$

(cuando  $n$  es impar,  $b_{(n+1)/2} = 0$ )



- (v) Calcular el estadístico  $F_1$  y/o  $F_2$  y/o  $F_3$ .
- (vi) Los tests son a una cola, y valores grandes de los estadísticos resultan significativos. La Brecque aporta una serie de tablas con puntos críticos para estos test.

G. - UNA GENERALIZACION DE W.

Puri y Rao (1976) consideraron una generalización del test de Shapiro-Wilk. Escribieron el valor esperado del  $i$ -ésimo estadístico de orden en la forma

$$E(X_{(i)}) = \gamma_1 + \gamma_2 c_i + \gamma_3 (c_i^2 - \lambda) + \gamma_4 (c_i^3 - \mu c_i) + \dots$$

donde  $\lambda$  y  $\mu$  son elegidos para proporcionar polinomios ortogonales.

Cuando la distribución es normal,  $\gamma_1 = \mu$ ,  $\gamma_2 = \sigma$ ,  $\gamma_3 = \gamma_4 = \dots = 0$ . El test de Shapiro-Wilk es esencialmente el test  $\gamma_2 = \sigma$ . Puri y Rao investigaron si se puede obtener información adicional combinando tests sobre  $\gamma_3$  y  $\gamma_4$  con el test W.

Bajo la representación por polinomios ortogonales, encontraron que

$$E(\mathbf{x}) = \gamma_1 \mathbf{b}_1 + \gamma_2 \mathbf{b}_2 + \gamma_3 \mathbf{b}_3 + \gamma_4 \mathbf{b}_4$$

donde

$$\mathbf{b}_1 = \mathbf{c}_0, \quad \mathbf{b}_2 = \mathbf{c}_1, \quad \mathbf{b}_3 = \mathbf{c}_2 - \frac{\mathbf{c}_2' \mathbf{V}^{-1} \mathbf{c}_0}{\mathbf{c}_0' \mathbf{V}^{-1} \mathbf{c}_0}, \quad \mathbf{b}_4 = \mathbf{c}_3 - \frac{\mathbf{c}_3' \mathbf{V}^{-1} \mathbf{c}_1}{\mathbf{c}_1' \mathbf{V}^{-1} \mathbf{c}_1}$$

y

$$\mathbf{c}'_j = (c_1^j, c_2^j, \dots, c_n^j), \quad j = 0, 1, 2, \dots$$



El mejor estimador lineal insesgado de  $\gamma_i$  está dado por

$$\hat{\gamma}_i = \frac{\mathbf{b}'_i \mathbf{V}^{-1} \mathbf{x}}{\mathbf{b}'_i \mathbf{V}^{-1} \mathbf{b}_i}$$

con varianza

$$\text{Var}(\hat{\gamma}_i) = \frac{\sigma^2}{\mathbf{b}'_i \mathbf{V}^{-1} \mathbf{b}_i}$$

A partir de aquí, se puede definir el test estadístico para  $\gamma_2 = \sigma$ ,  $\gamma_3 = 0$ ,  $\gamma_4 = 0$  como

$$T_2 = W^{1/2} = \frac{\hat{\gamma}_2 (\mathbf{b}'_2 \mathbf{V}^{-1} \mathbf{b}_2)}{\hat{\sigma} (\mathbf{b}'_2 \mathbf{V}^{-2} \mathbf{b}_2)^{1/2}}$$

$$T_3 = \frac{\hat{\gamma}_3}{\hat{\sigma}} (\mathbf{b}'_3 \mathbf{V}^{-1} \mathbf{b}_3)^{1/2}$$

$$T_4 = \frac{\hat{\gamma}_4}{\hat{\sigma}} (\mathbf{b}'_4 \mathbf{V}^{-1} \mathbf{b}_4)^{1/2}$$

Las investigaciones de Puri y Rao les condujeron a concluir que el uso conjunto de  $W$  y  $T_3$  nos proporciona un test más eficiente para normalidad contra alternativas no simétricas, y similarmente, el uso conjunto de  $W$  y  $T_4$  es más eficiente para medir la normalidad contra alternativas con curtosis no normal.

#### H.- TRANSFORMACION DE LA DISTRIBUCION DE $W$ Y $W'$ A NORMALIDAD.

Royston (1993a) busca una transformación  $g$  de forma que aplicada a cualquier estadístico  $X$  con distribución continua haga



que  $g(X)$  sea (aproximadamente) normal y desarrolla particularmente este caso aplicándolo a los estadísticos  $W$  de Shapiro y Wilk y  $W'$  de Shapiro y Francia. Indica que la forma de definir  $g$  será tal que grandes valores de  $g(X)$  correspondan a valores significativos de  $W$  y  $W'$  (valores pequeños). El  $p$ -valor para  $X$  será entonces  $1-\Phi^{-1}(Z)$ , donde  $Z = (g(X)-\mu)/\sigma$ . En la práctica la distribución de  $X$  es desconocida y es determinada por simulación sobre un conjunto de valores seleccionados de  $n$ , generalmente en el rango  $5 \leq n \leq 2000$ . La transformación es encontrada por el método de prueba y error, a partir de la transformación de Box-Cox  $g(X) = (X^\lambda - 1)/\lambda$  y/o la transformación logarítmica  $g(X) = \ln(X-\gamma)$ . Los parámetros  $\mu$ ,  $\sigma$ ,  $\lambda$  y  $\gamma$ , o simples transformaciones de ellos, son determinados como funciones (normalmente polinómicas) de  $u = f(n)$ , generalmente  $u = \ln(n)$ .

La distribución de  $Z$ , a estas alturas, será aproximadamente normal, pero no estándar, por lo que habrá que determinar su media,  $\mu_z$ , y su varianza,  $\sigma_z^2$ , como funciones de  $n$ . La transformación normalizada de  $X$  (los originales  $W$  o  $W'$ ) será entonces:

$$Z' = \frac{\frac{g(X) - \mu}{\sigma} - \mu_z}{\sigma_z} = \frac{g(X) - \mu - \sigma\mu_z}{\sigma\sigma_z}$$

esto es,  $g(X)$  tiene de media  $\mu + \sigma\mu_z$  y de desviación típica  $\sigma\sigma_z$ .

La siguiente tabla recoge las transformaciones propuestas por Royston y el cálculo de los parámetros involucrados como funciones de  $n$ .



Test	Rango de n	Transformación (g)	u	Parámetros	Coefficientes				
W	4-11	$-\ln\{\gamma - \ln(1-W)\}$	n	$\gamma$	0 -2.273				
					1 0.459				
				$\mu$	0 0.5440				
					1 -0.39978				
					2 0.025054				
					3 -0.0006714				
				$\ln(\sigma)$	0 1.3822				
					1 -0.77857				
					2 0.062767				
					3 -0.0020322				
				W	12-2000 (*)	$\ln(1-W)$	$\ln(n)$	$\mu$	0 -1.5861
									1 -0.31082
2 -0.083751									
$\ln(\sigma)$	3 0.0038915								
	0 -0.4803								
	1 -0.082676								
2 0.0030302									
W'	5-5000	$\ln(1-W')$	$\ln(v) - v$ (**)	$\mu$	0 -1.2725				
					1 1.0521				
			$\ln(v) + 2/v$ (**)	$\sigma$	0 1.0308				
					1 -0.26758				

(\*) Royston (1992) ya aseguró que es válido hasta  $n = 5000$

(\*\*)  $v = \ln(n)$

Por último, indiquemos que Royston (1993b) propone, para el cálculo de las ponderaciones de  $W'$ , el uso de la aproximación de Blom para determinar los valores esperados de los estadísticos de orden de una normal estándar, es decir,

$$\tilde{m}_i = \Phi^{-1}\{(i-3/8)/(n+1/4)\} \quad (i = 1, 2, \dots, n)$$



de forma que las ponderaciones  $\mathbf{b}$  en el test de Shapiro-Francia se tomarán como

$$\mathbf{b} = (\tilde{\mathbf{m}}' \tilde{\mathbf{m}})^{-1/2} \tilde{\mathbf{m}}$$

Así mismo, Royston (1992) proporciona una aproximación  $\tilde{\mathbf{a}}$  para las ponderaciones del test  $W$  de Shapiro-Wilk, que están basadas en  $\tilde{\mathbf{m}}$  y en polinomios en  $u = n^{-1/2}$ , con lo que queda anulada la necesidad de usar la matriz  $\mathbf{V}$ . Para  $n \geq 4$ , define:

$$\tilde{a}_n = b_n + 0.221157u - 0.147981u^2 - 2.071190u^3 + 4.434685u^4 - 2.706056u^5$$

$$\tilde{a}_{n-1} = b_{n-1} + 0.042981u - 0.293762u^2 - 1.752461u^3 + 5.682633u^4 - 3.582633u^5$$

con  $b_n$  y  $b_{n-1}$  último y penúltimo elemento de  $\mathbf{b}$ .

Tomando ahora:

$$\phi = (\tilde{\mathbf{m}}' \tilde{\mathbf{m}} - 2\tilde{m}_n^2) / (1 - 2\tilde{a}_n^2) \quad \text{si } n \leq 5$$

$$= (\tilde{\mathbf{m}}' \tilde{\mathbf{m}} - 2\tilde{m}_n^2 - 2\tilde{m}_{n-1}^2) / (1 - 2\tilde{a}_n^2 - 2\tilde{a}_{n-1}^2) \quad \text{si } n > 5$$

tendremos:

$$\tilde{a}_i = \phi^{-1/2} \tilde{m}_i \quad \text{para } i=2, \dots, n-1 \quad (n \leq 5)$$

$$\quad \quad \quad \text{para } i=3, \dots, n-2 \quad (n > 5)$$

Con este trabajo, Royston corrige los errores en los que, como indica, había incurrido [Royston (1982b, 1982c)] al utilizar aproximaciones incorrectas de las ponderaciones en el test de Shapiro-Wilk.



Contraste de razón de verosimilitudes asociado con transformaciones para incrementar la normalidad.

Un método para asegurar normalidad está asociado al uso de la familia de transformaciones introducida por Box y Cox (1964).

La transformación de las observaciones que proponen estos autores es

$$x_i^{(\xi, \lambda)} = \begin{cases} \frac{(x_i + \xi)^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \ln(x_i + \xi) & \text{para } \lambda = 0 \end{cases} \quad \text{con } x_i > -\xi$$

Un breve análisis sobre  $\xi$  y  $\lambda$  nos permite llevar a cabo las siguientes consideraciones:

- (1) Como sobre  $\xi$  se exige que  $x_i > -\xi$ , si las observaciones son siempre positivas, podremos tomar  $\xi$  igual a cero, y en otro caso, igual al valor absoluto del valor más negativo observado.
- (2) Cuando  $\lambda > 1$ , la transformación produce una mayor separación o dispersión de los valores grandes de  $(x_i + \xi)$ , tanto más acusada cuanto mayor sea el valor de  $\lambda$ . Por el contrario, cuando  $\lambda < 1$ , los valores grandes de  $(x_i + \xi)$  tienden a concentrarse y los valores pequeños  $((x_i + \xi) < 1)$  a dispersarse.

El aspecto básico introducido con estas transformaciones es que:

" $X_1, X_2, \dots, X_n$  son variables aleatorias independientes  $N(\mu, \sigma^2)$ , si, y sólo si,  $\lambda = 1$ ".



Este resultado nos permite inmediatamente plantear un contraste de normalidad en términos de  $\lambda$ , de la siguiente forma;

$$H_0: \lambda = 1 \text{ (normalidad)}$$

$$H_1: \lambda \neq 1 \text{ (ausencia de normalidad)}$$

La relación que existe entre las distribuciones de los datos originales  $x$  y los transformados  $x^{(\xi, \lambda)}$  será

$$f(x) = f(x^{(\xi, \lambda)}) \left| \frac{dx^{(\xi, \lambda)}}{dx} \right|$$

con

$$\frac{dx^{(\xi, \lambda)}}{dx} = \frac{\lambda(x+\xi)^{\lambda-1}}{\lambda} = (x+\xi)^{\lambda-1}$$

(expresión que siempre es mayor que cero dado que  $x > -\xi$ ).

Por tanto, y suponiendo que  $x^{(\xi, \lambda)}$  es  $N(\mu, \sigma^2)$ , la función de densidad de las variables originales será

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} \left( \frac{(x+\xi)^{\lambda} - 1}{\lambda} - \mu \right)^2 \right\} (x+\xi)^{\lambda-1}$$

Así, teniendo en cuenta la independencia de las observaciones, la función de densidad conjunta de  $\mathbf{X}=(X_1, X_2, \dots, X_n)$  será

$$f(\mathbf{X}) = \frac{1}{\sigma^n(\sqrt{2\pi})^n} \left( \prod_{i=1}^n (x_i + \xi)^{\lambda-1} \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( \frac{(x_i + \xi)^{\lambda} - 1}{\lambda} - \mu \right)^2 \right\}$$

y, por ello, la función soporte, o logaritmo de la función de verosimilitud tendrá la expresión



$$L(\xi, \lambda; \mu, \sigma) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi + (\lambda-1) \sum_i \ln(x_i + \xi) - \frac{1}{2\sigma^2} \sum_i \left( \frac{(x_i + \xi)^{\lambda-1}}{\lambda} - \mu \right)^2$$

Si maximizamos el logaritmo de la función de verosimilitud inicialmente respecto a  $\mu$  y  $\sigma^2$ , obtendremos

$$L_{\max}(\xi, \lambda) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \hat{\sigma}^2 + (\lambda-1) \sum_i \ln(x_i + \xi)$$

donde

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left\{ x_i^{(\xi, \lambda)} - \bar{x}^{(\xi, \lambda)} \right\}^2$$

y

$$\bar{x}^{(\xi, \lambda)} = \frac{1}{n} \sum_{i=1}^n x_i^{(\xi, \lambda)}$$

Después de sustituir  $x_i^{(\xi, \lambda)}$  en términos de  $x_i$ , podemos maximizar  $L_{\max}(\xi, \lambda)$  para obtener numéricamente los estimadores de máxima verosimilitud,  $\tilde{\xi}$  y  $\tilde{\lambda}$ , de  $\xi$  y  $\lambda$  respectivamente.

Siguiendo ahora el criterio del contraste de razón de verosimilitud, tenemos que

$$2 \left\{ L_{\max}(\tilde{\xi}, \tilde{\lambda}) - L_{\max}(\tilde{\xi}, 1) \right\}$$

está asintóticamente distribuida como una  $\chi_1^2$ . (Obsérvese que  $L_{\max}(\tilde{\xi}, 1)$ , no depende de  $\xi$ , por lo que cualquier valor de  $\xi$ , incluido  $\tilde{\xi}$ , maximizará  $L_{\max}(\tilde{\xi}, 1)$ ).

#### Técnicas gráficas informales.

A lo largo de la historia de la Estadística se han propuesto una gran variedad de técnicas gráficas, pero no ha sido



hasta mitad de los años 70, coincidiendo con el avance del hardware y el software de las computadoras gráficas, cuando estas técnicas, y otras introducidas posteriormente, han alcanzado su auge.

No pretendemos en este apartado hacer una exposición detallada de las técnicas gráficas más importantes dado que, entre otros motivos, estaríamos totalmente fuera de contexto. Por el contrario queremos mencionar una técnica, que dentro del ambiente de este capítulo, ha sido profusamente utilizada. Se trata de los gráficos de probabilidad (Q-Q plots) usados fundamentalmente para comprobar hipótesis distribucionales establecidas sobre un conjunto de observaciones. En particular, cuando la hipótesis se refiere a la normalidad de los datos, hablaremos del gráfico probabilístico normal (normal Q-Q plots) que será considerado como una herramienta gráfica para el estudio de la hipótesis de normalidad enunciada sobre un conjunto de datos.

Debido a que las conclusiones obtenidas en los gráficos de probabilidad, respecto a las suposiciones distribucionales, están sujetas a cierta subjetividad del observador, éstos han venido a denominarse técnicas gráficas informales, y por tanto es habitual la recomendación de que junto a un gráfico de probabilidad siempre aparezca un contraste formal de normalidad.

No entramos a precisar cómo el gráfico probabilístico normal es usado para contrastar la hipótesis de normalidad, dado que en los capítulos 4 al 6 trataremos con toda profundidad este tema.

#### **Consideraciones finales sobre los contrastes de normalidad univariantes.**

Aunque todas las técnicas expuestas son contrastes de normalidad univariantes, éstos no son todos las existentes. Pode-



mos encontrar otros muchos test introducidos para probar la normalidad en un conjunto de observaciones. La elección particular que hemos hecho de los presentados se ha fundamentado, como indicamos al principio, tanto en la importancia histórica que encontramos en algunos de ellos, como en el uso extenso que de otros observamos en gran cantidad de trabajos de investigación, unos encaminados directamente al estudio de dicha técnica, y otros eligiéndola como la mejor para ser usada como herramienta previa en el análisis de la normalidad de un conjunto de datos que permita validar los posteriores resultados.

Quisiéramos concluir este apartado referido a los contrastes de normalidad univariante haciendo una serie de consideraciones sobre ellos. Desearíamos que estas observaciones concluyeran indicándonos cual de todos es el mejor, de forma que a partir de ese momento lo adoptáramos como técnica única para contrastar la normalidad, abandonando todas las demás o, en su caso, considerándolas en la medida en que pudiesen estar ligadas con la elegida. Sin embargo, cualquier intento que hagamos en este sentido será infructuoso. Incluso un estudio detallado de la potencia de los distintos tests, nos hace calificar a unos más potentes que otros y a los otros más que los unos, en función del enunciado que hagamos de la alternativa a la normalidad. En esta línea, destacamos el trabajo realizado por Shapiro, Wilk y Chen (1968) que puede considerarse como el primer gran estudio de potencia realizado para contrastes de normalidad. En el mismo, utilizaron nueve estadísticos y consideraron doce familias de distribuciones que, con diferentes valores de los parámetros, dieron lugar a 45 alternativas distintas. Los autores establecieron las siguientes conclusiones:

- (1) El test W de Shapiro y Wilk proporciona, generalmente, una mejor medida de la no normalidad.
- (2) Los tests basados en la función de distribución empírica son poco potentes.



- (3) El test  $u$  basado en el rango, tiene buenas propiedades cuando las alternativas son distribuciones simétricas con colas pequeñas, sin embargo, es casi inútil contra asimetrías.
- (4) Una combinación de  $\sqrt{b_1}$  y de  $b_2$  es generalmente bastante potente, aunque está normalmente dominada por  $W$ .

La predominancia que con este estudio otorgaron Shapiro, Wilk y Chen al estadístico  $W$ , ocasionó que surgieran posteriores investigaciones sobre el tema. Así, Dyer (1974) consideró siete test de normalidad y cuatro distribuciones alternativas, distinguiendo en los casos en que los parámetros  $\mu$  y  $\sigma^2$  fuesen desconocidos y el caso en que sólo  $\sigma^2$  era desconocido.

Un estudio de Monte Carlo permitió a Dyer observar que  $W$  y  $A^2$  son, en general, mejores que el resto de los estudiados. Sin embargo, el resultado más interesante que obtuvo, y al que nosotros ya hemos hecho mención, es que la potencia del test es siempre más grande cuando  $\mu$  y  $\sigma^2$  son desconocidos y, por tanto, estimados a partir de la muestra.

Un estudio similar fue realizado por Stephens (1974). En este caso trabajó con nueve test y nueve distribuciones alternativas a la normal. Contrariamente a lo establecido por Shapiro, Wilk y Chen (1968) que indicaron una potencia muy baja en los test basados en la función de distribución empírica, Stephens demostró que estos contrastes tienen una potencia más o menos comparable a la del test  $W$  de Shapiro-Wilk. Como explicación de esta discrepancia, Stephens apuntó que en los estudios previos, se usaban valores críticos erróneos para los estadísticos basados en la función de distribución empírica, dado que éstos estaban calculados suponiendo que la media y la varianza de la población eran conocidas. Para una correcta comparación con  $W$ , concluyó, la media y la varianza deberán ser calculadas a partir de la muestra.

Con independencia al estudio de Dyer, Stephens llegó a



la misma conclusión: cuando se trata de contrastar la normalidad, nunca deben ser consideradas la media y la varianza como conocidas, incluso aunque éstas lo fueran.

Un estudio de Monte Carlo permitió a Stephens indicar que los contrastes de normalidad usando  $W$  cuando el tamaño muestral es menor que 50, y  $W'$  cuando es mayor de 50, se comportan ligeramente mejor que incluso sus mejores competidores basados en la función de distribución empírica.

En la misma línea, Pearson, D'Agostino y Bowman (1977) han producido un extenso estudio dirigido a observar la potencia de los contrastes de normalidad basados en 8 tests con 58 distribuciones alternativas. Concluyeron, entre otras, que para distribuciones alternativas asimétricas,  $W$  es bastante más potente que cualquier otro test de propósito general, mientras que con alternativas simétricas, existen otros mejores. Así mismo, señalan como aconsejable el uso de  $\sqrt{b_1}$  y  $b_2$  si existe alguna información previa sobre las posibles causas de no normalidad ocasionadas por la falta de simetría o por una curtosis no normal.

Mencionemos también los estudios de potencia realizados por Filliben (1975) y La Brecque (1977) encaminados directamente a comprobar la potencia de los contrastes que cada uno de ellos propusieron, obteniendo, por lo general, resultados positivos de sus respectivos contrastes frente al resto.

Los estudios sobre potencia realizados por Landry y Lepage (1992) también mantienen al test de Shapiro-Wilk y al test de propósito general basado en la asimetría y la curtosis como los mejores cuando es desconocida la causa que provoca ausencia de normalidad.

Vistos todos los comentarios anteriores y en un intento de hacer un resumen de los estudios ya mencionados, nos atrevemos



a indicar un conjunto de recomendaciones finales que, en función del ambiente en que nos movamos, pueden dirigir nuestra forma de actuar a la hora de abordar un estudio sobre la posible normalidad de un conjunto de observaciones. Estas son:

- 1.- Realizar un análisis gráfico detallado haciendo especial uso del gráfico de probabilidad normal, que será acompañado de otra técnica formal de estudio de la normalidad.
- 2.- Los tests W de Shapiro-Wilk y sus extensiones, el test  $K^2$  de D'Agostino-Pearson y el test  $A^2$  de Anderson-Darlin parecen ser los mejores tests de propósito general disponibles. Los tests del tipo Shapiro-Wilk, son probablemente los más potentes de todos. Sin embargo, debido a los inconvenientes que se presentan cuando hay empates en las observaciones, el test  $K^2$ , basado en los estadísticos  $\sqrt{b_1}$  y  $b_2$ , puede resultar una alternativa aconsejable.  
Los estadísticos  $\sqrt{b_1}$  y  $b_2$  resultarán muy útiles para indicar posibles causas de ausencia de normalidad.
- 3.- El test D de D'Agostino es un test de propósito general aconsejable y fácil de usar, aunque tiene menos potencia que los señalados en el apartado anterior.
- 4.- Si la dirección de la alternativa a la normalidad es conocida (por ejemplo  $\sqrt{\beta_1} > 0$ , o  $\beta_2 > 3$ ), las versiones direccionales de los tests  $\sqrt{b_1}$  y  $b_2$  y del test D de D'Agostino son recomendados.
- 5.- El test de Kolmogorov-Smirnov tiene una potencia muy pobre en comparación con los anteriores, por lo que debemos considerarlo como interesante sólo desde el punto de vista histórico, pero no será aconsejable su uso.
- 6.- Los contrastes del tipo chi-cuadrado no deberán usarse si se



dispone de la muestra al completo, ya que no tiene mucha potencia comparado con los anteriores. Su uso sólo es aconsejado cuando las observaciones estuviesen agrupadas.



## CAPITULO 4

### EL GRAFICO PROBABILISTICO NORMAL

Como hemos indicado en el capítulo anterior, dentro del conjunto de técnicas gráficas (informales) encaminadas al estudio de las hipótesis distribucionales enunciadas sobre un conjunto de datos, encontramos al denominado gráfico de probabilidad o gráfico probabilístico (Q-Q plots).

De forma muy esquemática, ya que posteriormente profundizaremos en el tema, un gráfico de probabilidad consiste en representar los cuantiles de la muestra contra los correspondientes cuantiles que esperaríamos obtener si nuestras observaciones provinieran de la distribución especificada que, en el caso que en principio nos interesa, es la distribución normal. Cuando los mencionados puntos se encuentren suficientemente próximos a una línea recta, la hipótesis de normalidad podrá mantenerse, mientras que, en caso contrario, ésta será rechazada.

Aunque este procedimiento guarda toda la intuición de lo esperado, está sujeto a la subjetividad del observador, subjetividad que aparece por la falta de precisión con la que se plantea que "los puntos estén suficientemente próximos a una línea recta", y que, por otra parte, justifica el calificativo de "informales" que aparece en el grupo donde encuadramos a esta técnica como prueba de normalidad.

En algunos casos, la falta de precisión mencionada puede



resultarnos ventajosa. Por una parte, por la propia idea que subyace en un contraste de hipótesis, en el que a la pregunta "¿están los datos normalmente distribuidos?", no pretendemos dar una respuesta en sentido exacto, sino que más bien procuramos resolver la cuestión "¿está la distribución empírica suficientemente bien aproximada por una distribución normal?". Por otra parte, y estrechamente relacionado con lo anterior, la falta de precisión nos permitirá la libertad de juzgar ciertos alejamientos de los datos a la distribución normal, como carentes de importancia para el propósito que tenemos en mente, mientras que los mismos pueden considerarse como un serio impedimento para usos distintos que hayamos planeado.

Ahora bien, frente a las ventajas mencionadas, la "informalidad" del procedimiento nos puede llevar a situaciones paradójicas como aquella en la que dos observadores distintos, al enfrentarse al mismo conjunto de datos para los mismos fines, juzguen de forma contraria la aceptabilidad de la hipótesis, planteándonos posibles dudas procedimentales que en algunos casos puedan ser incluso más importantes que la propia normalidad o no de los datos, y que, en resumen, aconsejen el abandono de tales técnicas en beneficio de otros contrastes "más formales".

Por contrario, en lugar de abandonar estas técnicas "informales", pretendemos plantear una "formalización" de ellas que, por una parte, mantengan los beneficios en lo que se refiere a intuición del procedimiento y al coste de su desarrollo, y por otra, minimice la subjetividad del observador y permita poner de relieve elementos que sean los causantes de la falta de normalidad en el conjunto de datos que nos encontremos analizando.

En este capítulo introduciremos el proceso de construcción de un gráfico de probabilidad (Q-Q plot). Comenzaremos con el gráfico de probabilidad empírico (Q-Q plot empírico) que permitirá cargar de intuición al gráfico de probabilidad teórico (Q-Q



plot teórico). Por fin, concluiremos con el análisis de algunos elementos fundamentales que aparecen en el esquema de construcción de este tipo de gráficos.

#### EL GRAFICO DE PROBABILIDAD EMPIRICO (Q-Q PLOT EMPIRICO)

El gráfico de probabilidad empírico constituye un método efectivo para producir una comparación detallada de la distribución de dos conjuntos de datos [Wilk y Gnanadesikan (1968)]. Se construye dibujando los cuantiles de una de las distribuciones empíricas contra los correspondientes cuantiles de la otra.

Formalmente, si los dos conjuntos de datos están representados por  $x_i$  ( $i = 1, 2, \dots, n$ ) e  $y_j$  ( $j = 1, 2, \dots, m$ ), y notamos por  $Q_x(p)$  y  $Q_y(p)$  a los correspondientes cuantiles de orden  $p$  ( $0 < p < 1$ ) de los datos anteriores, respectivamente, se trata de dibujar  $Q_x(p)$  contra  $Q_y(p)$  para un conjunto de valores de  $p$ . (Figura 4.1).

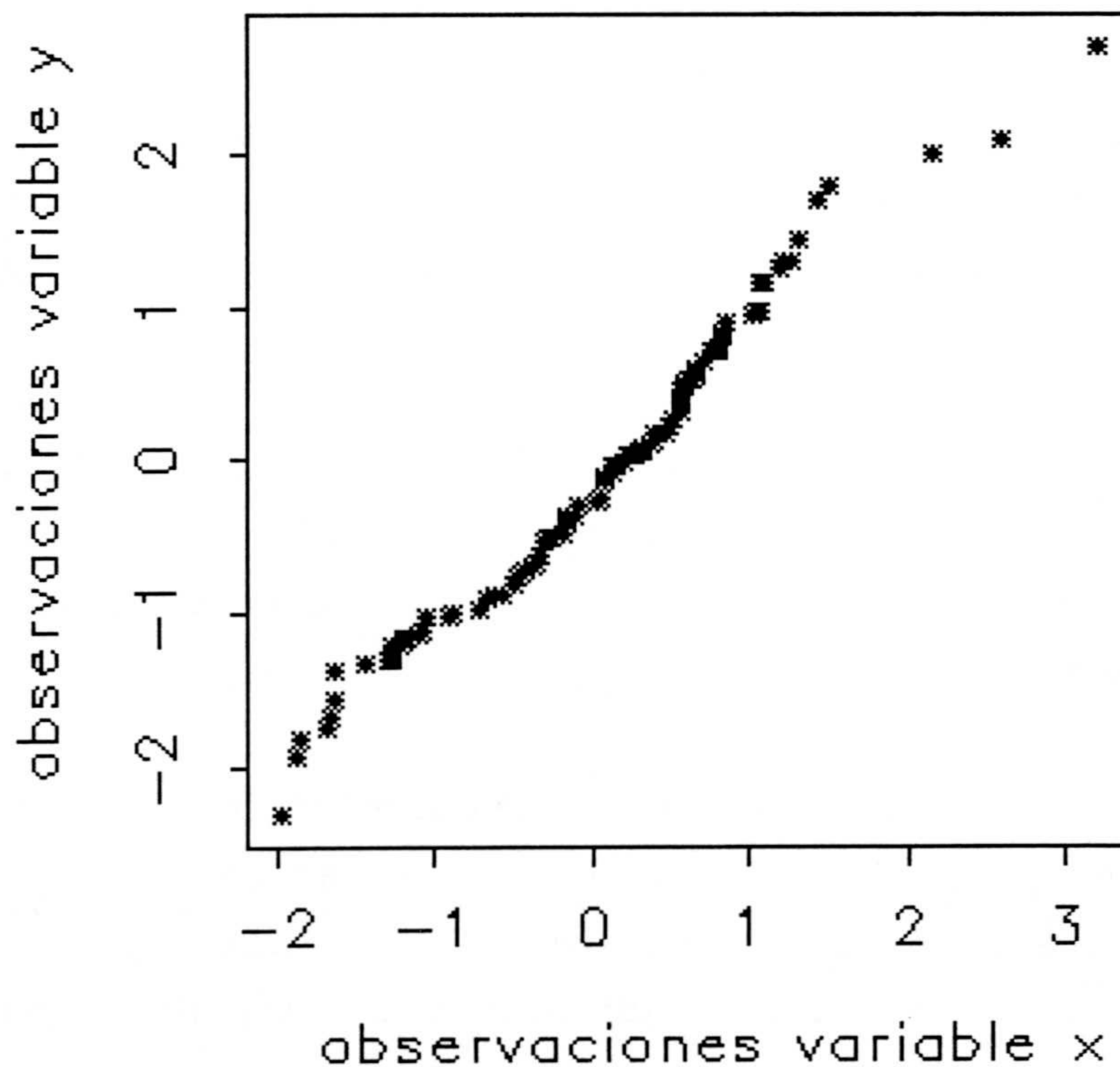


FIGURA 4.1: Gráfico de probabilidad empírico



Seguido el proceso anterior, es evidente que si las dos distribuciones fueran idénticas, los puntos, así dibujados, presentarían una configuración lineal, cayendo sobre la recta  $y = x$ . Una propiedad elemental de los gráficos de probabilidad es que si ambas distribuciones fueran idénticas salvo quizá los parámetros de localización y escala, se mantendría la configuración lineal en el gráfico de probabilidad empírico, sin necesidad de caer los puntos sobre la recta  $y = x$ .

La característica anterior ha hecho que los gráficos de probabilidad se constituyan en una técnica bastante atrayente, dado que la linealidad es una configuración geométrica que el ojo humano percibe fácilmente. Además, cualquier alejamiento de tal configuración puede ser apreciado y evaluado inmediatamente.

Para la elaboración de un Q-Q plot empírico, consideremos  $x_{(i)}$ ,  $i = 1, 2, \dots, n$ , las observaciones  $x_i$  ordenadas de menor a mayor, esto es,  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . De la misma forma, sean  $y_{(j)}$ ,  $j = 1, 2, \dots, m$ , las observaciones  $y_j$  ordenadas de menor a mayor, esto es,  $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(m)}$ . Podemos encontrarnos ante dos situaciones diferentes: que los tamaños muestrales coincidan ( $n = m$ ) o que los tamaños muestrales sean distintos (por ejemplo,  $n > m$ ).

Si los tamaños muestrales coinciden, ( $n = m$ ),  $Q_x(p_i) = x_{(i)}$  y  $Q_y(p_i) = y_{(i)}$ , para  $p_i = (i - 0.5)/n$ , con lo que para generar el Q-Q plot empírico bastará con dibujar los pares  $(x_{(i)}, y_{(i)})$ ,  $i = 1, 2, \dots, n$ .

En el caso en que los tamaños muestrales sean distintos, supongamos que  $n > m$ . Lo que usualmente se hace es partir del conjunto de datos ordenados más pequeño (menos elementos), y a partir de él, se interpolan los correspondientes cuantiles del conjunto más grande (más elementos). Es decir, consideramos



$$Q_y(p_j) = y_{(j)}, \quad \text{con} \quad p_j = \frac{j - 0'5}{m} \quad j = 1, 2, \dots, m$$

Para determinar los correspondientes cuantiles en la distribución de los  $x_i$ , necesitamos encontrar un valor  $\nu$ , tal que:

$$p_\nu = \frac{\nu - 0'5}{n} = \frac{j - 0'5}{m} = p_j$$

o lo que es lo mismo,

$$\nu = \frac{n}{m} (j - 0'5) + 0'5$$

Si al calcular  $\nu$  obtenermos un número entero, el Q-Q plot se obtiene sin más que representar los puntos  $(x_{(\nu)}, y_{(j)})$ ,  $j = 1, 2, \dots, m$ . Si  $\nu$  no es un número entero, éste se podrá escribir como

$$\nu = k + \theta, \quad \text{donde } 0 < \theta < 1, \quad \text{siendo } k \text{ la parte entera de } \nu$$

En este caso, el cuantil interpolado se obtiene como:

$$Q_x\left(\frac{j - 0'5}{m}\right) = (1 - \theta) x_{(k)} + \theta x_{(k+1)}$$

y ya bastará con representar los pares

$$((1 - \theta) x_{(k)} + \theta x_{(k+1)} ; y_{(j)}) , \quad j = 1, 2, \dots, m$$

para así obtener el Q-Q plot empírico.

Para ilustrar este caso, consideremos los dos conjuntos de observaciones recogidos en la Tabla 4.1. Seguido el procedimiento anterior, recogemos en la Tabla 4.2 las coordenadas de los puntos a dibujar en el gráfico de probabilidad empírico.



OBSERVACIONES VARIABLE X	0.5254 , 0.7076 , -0.6281 , 0.1254 , -0.2961
	-0.3694 , -2.1722 , 0.7548 , 1.6541 , -2.0083
	-0.4040 , -1.9125 , -0.4905 , 1.2073 , 0.1123
	0.8987 , -0.1601 , -0.9275 , 0.2722 , -0.5635
	-1.2695 , 0.4325 , -1.5231 , 1.7827 , -0.8202
OBSERVACIONES VARIABLE Y	0.5389 , 0.7109 , -0.5116 , 0.4600 , -0.5696
	-1.7675 , 0.0036 , 0.3762 , 0.7124 , -0.2825
	-0.2100 , 1.1832 , 1.0159 , 1.1332 , 2.3831

TABLA 4.1: Observaciones de las variables X e Y.

EJE X	EJE Y
-2.1722	-1.7675
-1.9399	-0.5996
-1.4145	-0.5116
-0.9122	-0.2825
-0.6555	-0.2100
-0.5218	0.0036
-0.3941	0.3762
-0.2961	0.4600
0.0345	0.5389
0.1883	0.7109
0.4458	0.7124
0.6815	1.0159
0.8370	1.1332
1.3349	1.1832
1.7827	2.3831

TABLA 4.2: Coordenadas de los puntos en el gráfico de probabilidad empírico.

Una vez que las coordenadas de los puntos han sido obtenidas, daremos lugar al gráfico de probabilidad empírico recogido



en la Figura 4.2.

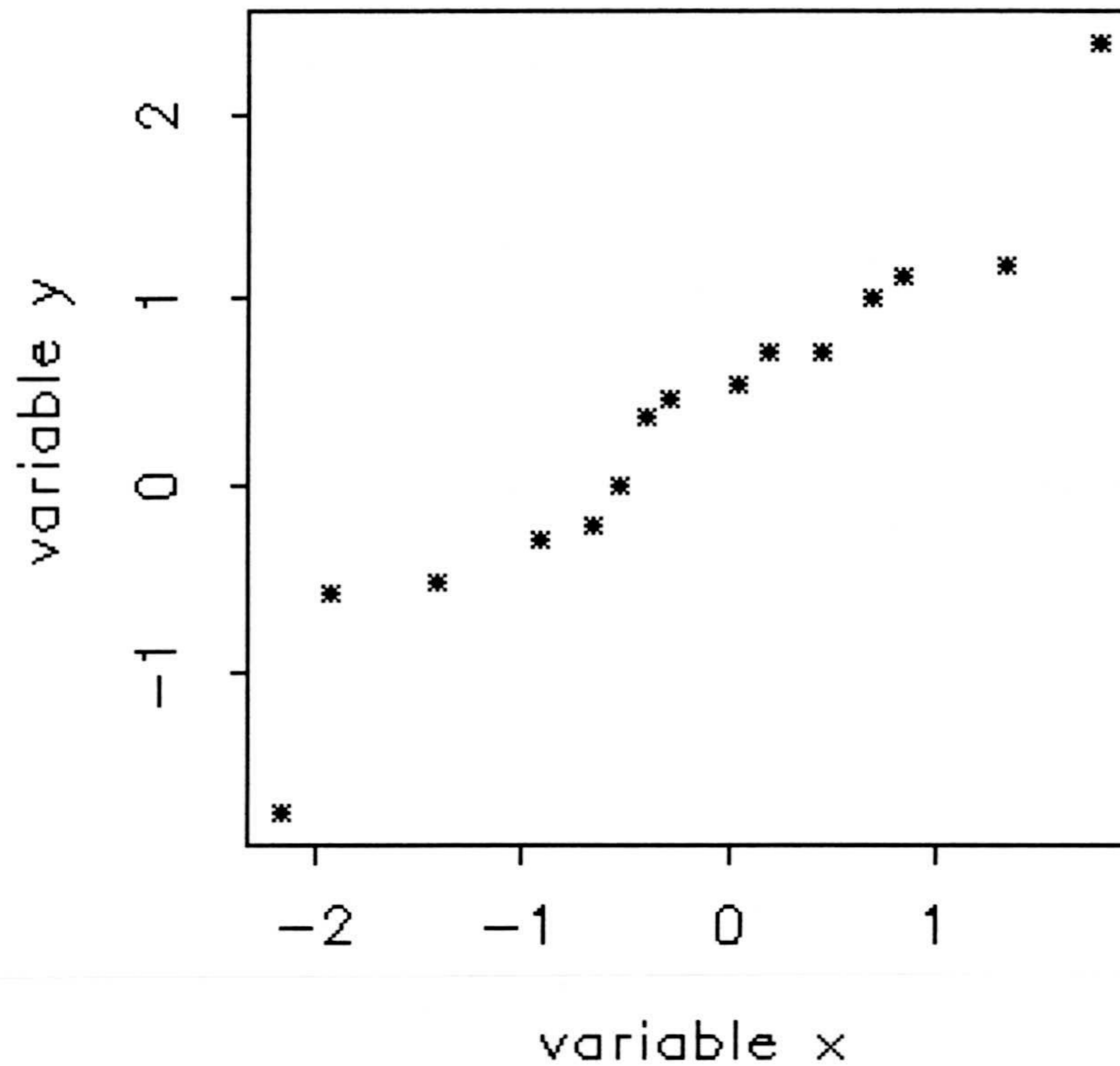
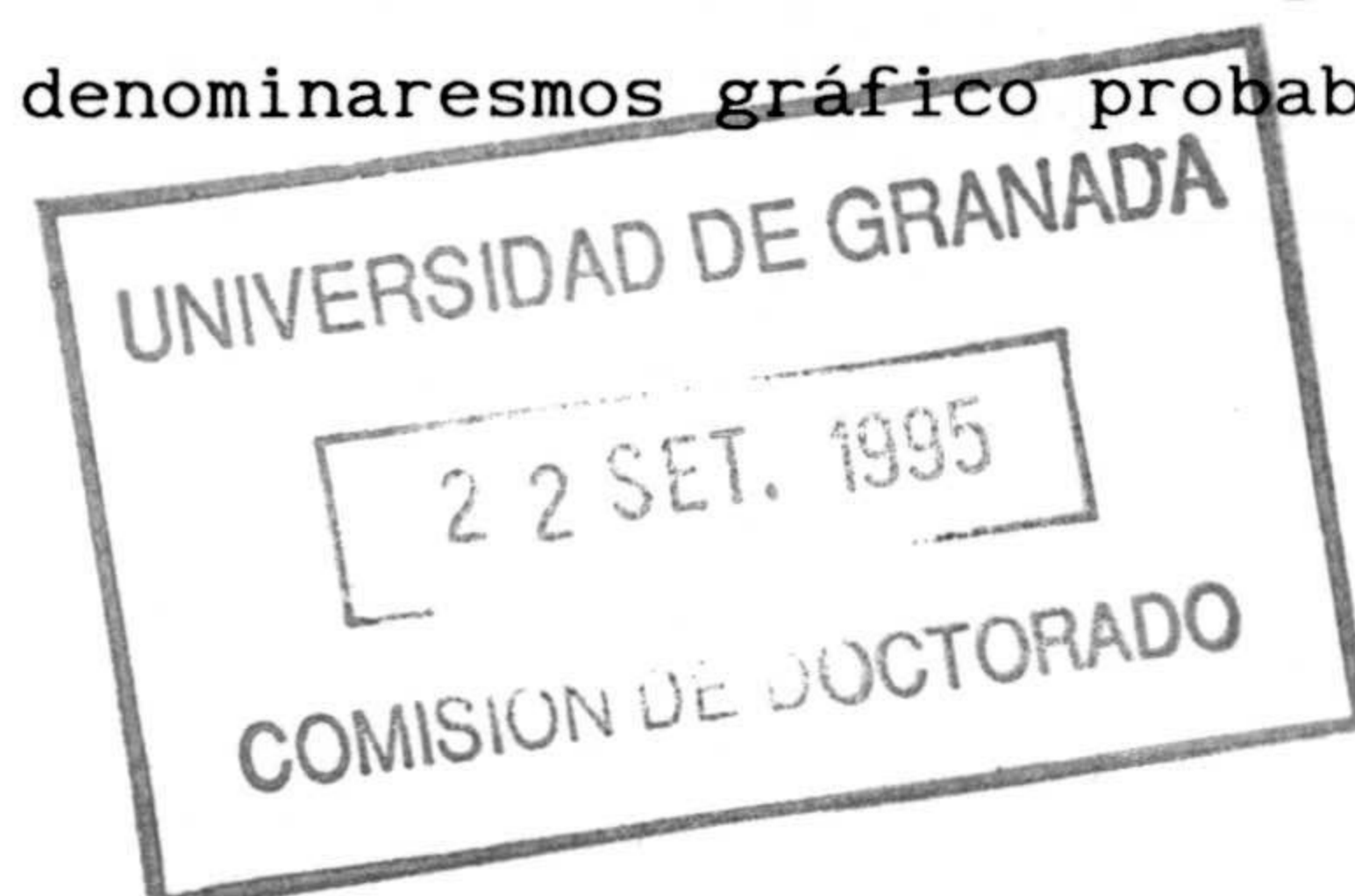


FIGURA 4.2: Gráfico de probabilidad empírico  $n > m$ .

#### EL GRAFICO DE PROBABILIDAD TEORICO (Q-Q PLOT TEORICO)

El gráfico de probabilidad teórico, al que a partir de ahora nos referiremos simplemente como gráfico de probabilidad o Q-Q plot, encuentra su fundamento en el empírico. Lo obtendremos reemplazando una de las distribuciones empíricas por una distribución teórica. Así, dibujaremos los cuantiles de los datos contra los correspondientes cuantiles de la distribución teórica. En el caso particular, que es el que principalmente nos interesa en este trabajo, en el que la distribución teórica sea la normal, al gráfico de probabilidad resultante lo denominaremos gráfico probabilístico normal o Q-Q plot normal.





Para describir con más detalle la construcción del gráfico de probabilidad, supongamos que los datos de los que disponemos son  $x_1, x_2, \dots, x_n$ , que ordenamos de menor a mayor en la forma  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . Por lo establecido anteriormente ocurrirá que

$$x_{(i)} = Q_x(p_i) \quad \text{con} \quad p_i = \frac{i - 0'5}{n} \quad i = 1, 2, \dots, n$$

Supongamos también que  $F(x)$  es la función de distribución de la distribución teórica en cuestión. Si por  $Q_t(p)$  notamos al cuantil de orden  $p$  ( $0 < p < 1$ ) de la distribución  $F$ , se tendrá que

$$F(Q_t(p)) = p$$

o lo que es lo mismo,

$$Q_t(p) = F^{-1}(p)$$

Con lo que ahora bastará con dibujar los pares  $(F^{-1}(p) ; x_{(i)})$ ,  $i = 1, 2, \dots, n$ , para así obtener nuestro gráfico de probabilidad.

Recapitulando, supuesto que disponemos de un conjunto de observaciones  $x_i$  ( $i = 1, 2, \dots, n$ ) y de una distribución teórica dada por su función de distribución  $F(x)$ , la obtención de un gráfico de probabilidad se hará siguiendo las siguientes etapas:

1°) Ordenar las observaciones originales de menor a mayor, con lo que dispondremos de los valores  $x_{(i)}$  ( $i = 1, 2, \dots, n$ ).

2°) Determinar los valores

$$p_i = \frac{i - 0'5}{n} \quad i = 1, 2, \dots, n$$

A estas alturas podemos identificar

$$x_{(i)} = Q_x(p_i) \quad i = 1, 2, \dots, n$$



3°) Determinar los cuantiles de orden  $p_i$  ( $i = 1, 2, \dots, n$ ) de la distribución teórica representada por la función de distribución  $F$ , esto es,

$$Q_t(p_i) = F^{-1}(p_i) \quad i = 1, 2, \dots, n$$

4°) Representar el conjunto de puntos  $(Q_t(p_i); Q_x(p_i))$ ,  $i = 1, 2, \dots, n$ , o lo que es lo mismo, los puntos  $(F^{-1}(p_i); x_{(i)})$ ,  $i = 1, 2, \dots, n$ .

Veamos el siguiente ejemplo que ilustra este procedimiento. Consideramos 26 observaciones que, ordenadas, recogemos en la tabla 4.3. Así mismo, consideremos el valor de los  $p_i$  y el valor de  $F^{-1}(p_i)$ , supuesto que  $F$  es la función de distribución de una  $N(0,1)$ .

OBSERVACIONES ORDENADAS	$p_i = \frac{i-0.5}{n}$	$F^{-1}(p_i)$
-2.1842	0.0192	-2.0706
-1.4330	0.0576	-1.5752
-1.1114	0.0961	-1.3041
-0.8981	0.1346	-1.1049
-0.8896	0.1730	-0.9424
-0.6250	0.2115	-0.8012
-0.5436	0.2500	-0.6745
-0.2301	0.2884	-0.5581
-0.2154	0.3269	-0.4485
-0.1201	0.3653	-0.3443
0.0460	0.4038	-0.2435
0.1220	0.4423	-0.1451
0.3144	0.4807	-0.0484
0.3187	0.5192	0.0484
0.3466	0.5576	0.1451
0.5550	0.5961	0.2435
0.6100	0.6346	0.3443
0.8018	0.6730	0.4485
0.8397	0.7115	0.5581
0.8504	0.7500	0.6745
1.0683	0.7884	0.8012
1.1121	0.8269	0.9424
1.1983	0.8653	1.1049
1.5703	0.9038	1.3041
1.5788	0.9423	1.5752
2.1474	0.9807	2.0706

TABLA 4.3: Coordenadas de un gráfico probabilístico normal.



Dibujadas estas coordenadas daremos lugar al siguiente gráfico probabilístico normal (Figura 4.3).

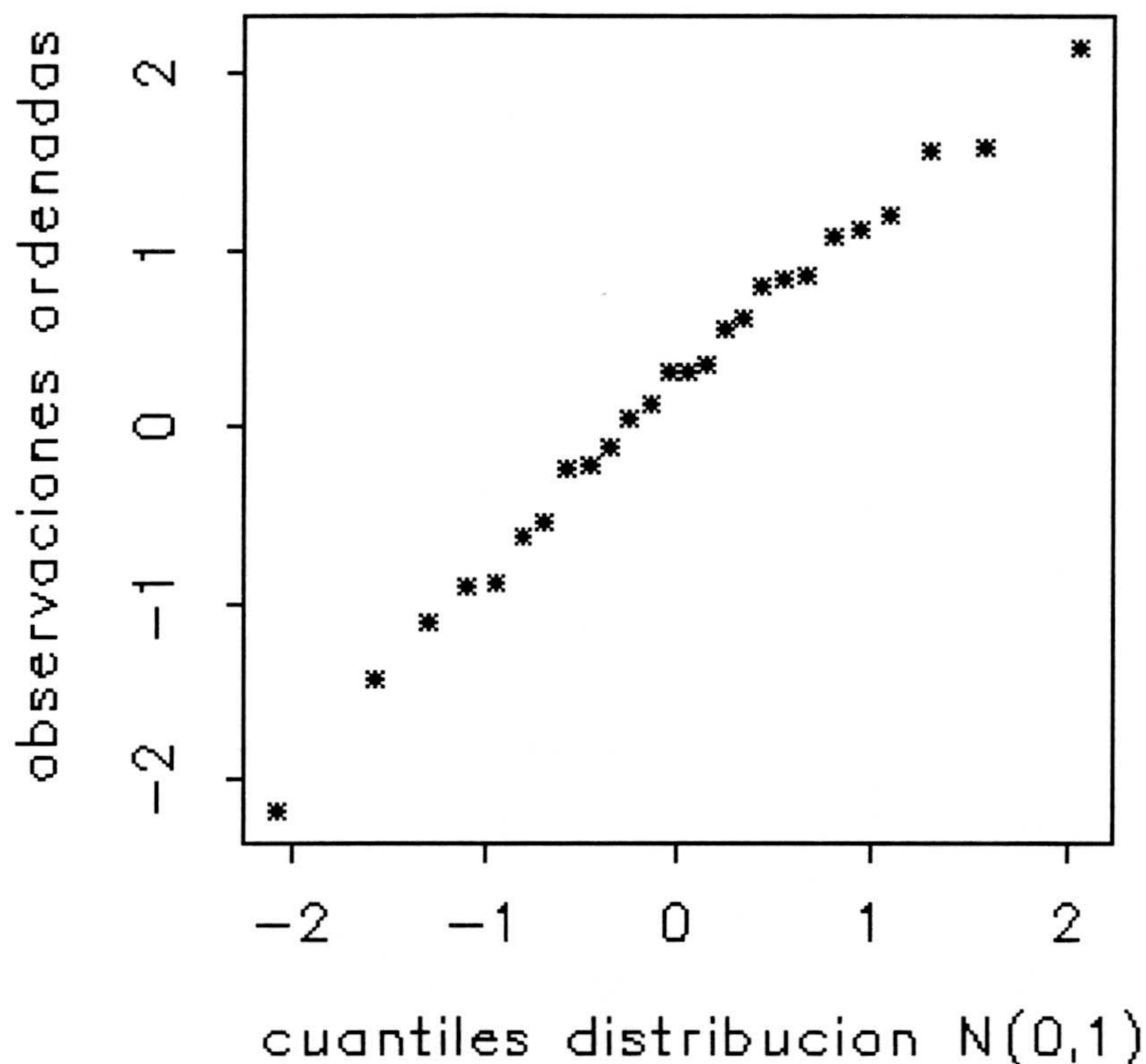


FIGURA 4.3: Gráfico probabilístico normal (n = 26)

**ALGUNOS COMENTARIOS SOBRE LOS PUNTOS DE POSICION GRAFICA ( $p_i$ )**

En la obtención de un gráfico de probabilidad aparecen como elementos básicos los  $p_i$ , que denominamos *puntos de posición gráfica* (plotting position). Por una parte, estos elementos permiten identificar a las observaciones como los cuantiles de la distribución empírica de los datos ( $x_{(i)} = Q_x(p_i)$ ), y por otra, constituyen los elementos fundamentales para determinar los cuantiles de la distribución teórica en curso, representada por la función de distribución  $F(x)$  ( $Q_t(p_i) = F^{-1}(p_i)$ ).

Está claro, a partir del proceso seguido para obtener un gráfico de probabilidad, que los  $p_i$  son plausibles estimadores



empíricos de  $F(x_{(i)})$ . Es por esto que, en principio, nos podía haber resultado más natural definir los  $p_i$  como

$$p_i = \frac{i}{n} \quad i = 1, 2, \dots, n$$

que constituye la más temprana elección que se hizo de los puntos de posición gráfica, y que fue conocida por los hidrólogos como "Método California", pero, este método fue pronto descartado debido a que no era posible dibujar la observación más grande (importante en muchas ocasiones donde aparece el interés de estimar el más alto percencil de la distribución).

Hazen (1930) introdujo la definición

$$p_i = \frac{i - 0'5}{n} \quad i = 1, 2, \dots, n$$

donde el término 0'5 del denominador aparece como un corrector de continuidad, y que permite recoger todas las observaciones en el gráfico. Posteriormente, Weibull (1939) propuso definir los  $p_i$  mediante la expresión

$$p_i = \frac{i}{n + 1} \quad i = 1, 2, \dots, n$$

conformándose estas dos definiciones como las de uso más extendido.

Las dos anteriores propuestas de los  $p_i$  están recogidas en la expresión general

$$p_i = \frac{i - c}{n - 2c + 1} \quad \text{para} \quad 0 \leq c \leq 1$$

donde observamos que si  $c = 0'5$  aparece la propuesta de Hazen, y si  $c = 0$ , obtenemos la propuesta de Weibull. [Barnett (1975) realiza una extensa discusión sobre la selección óptima de  $c$ ].



En la literatura al uso, aparecen otras muchas definiciones distintas de los  $p_i$ . Una idea bastante aceptada y difundida es que los  $p_i$  deben ser determinados a partir de medidas de localización de los estadísticos de orden. Así, si consideramos los estadísticos de orden de la variable transformada en la forma  $(x - \mu)/\sigma$  (donde  $\mu$  y  $\sigma$  no son necesariamente la media y la desviación típica, sino parámetros de localización y escala respectivamente), y notamos por

$$W_{(i)} = \frac{X_{(i)} - \mu}{\sigma} \quad i = 1, 2, \dots, n$$

se propone tomar

$$p_i = F[\text{loc}(W_{(i)})] \quad i = 1, 2, \dots, n$$

donde  $\text{loc}(W_{(i)})$ , simboliza a una medida de localización de  $W_{(i)}$ . En este sentido, Shapiro y Wilk (1965), Shapiro y Francia (1972) y La Breque (1977) propusieron, para los gráficos de probabilidad, y por consiguiente, para los estadísticos que definen para sus respectivos contrastes de normalidad, el uso de

$$p_i = F[E(W_{(i)})] \quad i = 1, 2, \dots, n$$

donde por  $E(W_{(i)})$  denotamos al valor esperado del  $i$ -ésimo estadístico de orden transformado.

Por su parte, Benard y Bos-Levenbach (1953) y Filliben (1975) definieron los  $p_i$  en los términos

$$p_i = F[\text{Med}(W_{(i)})] \quad i = 1, 2, \dots, n$$

donde por  $\text{Med}(W_{(i)})$  se simboliza a la mediana del  $i$ -ésimo estadístico de orden transformado.

Aunque muchos autores han justificado la conveniencia



del uso de las anteriores dos definiciones, en la práctica (refiriéndonos tanto al uso habitual de los gráficos de probabilidad como al procedimiento seguido por aquellos paquetes estadísticos que introducen esta técnica gráfica), raramente son usadas, todo ello en beneficio de las propuestas por Hazen y Weibull.

Vista la dificultad que en algunos casos aparece en la determinación de los  $p_i$  usando las definiciones anteriores (medidas de localización de los estadísticos de orden), algunos autores han propuesto otras definiciones basadas en meras aproximaciones de éstas.

Benard y Bos-Levenbach (1953), con su elección de la mediana como medida de localización más apropiada, demostraron que para el gráfico probabilístico normal, una buena aproximación, para tamaños muestrales intermedios, de los  $p_i$ , podía conseguirse con la expresión:

$$p_i = \frac{i - 0'3}{n + 0'4}$$

Por su parte, Filliben (1975), sugirió el uso de

$$p_i = \begin{cases} 1 - p_n & i = 1 \\ \frac{i - 0'3175}{n + 0'365} & i = 2, 3, \dots, n-1 \\ (0'5)^{(1/n)} & i = n \end{cases}$$

como aproximación a la expresión

$$p_i = F \left[ \text{Med}(W_{(i)}) \right] \quad i = 1, 2, \dots, n$$

Otros autores como Kimball (1960) y Cunnane (1978) han



recomendado el uso de una aproximación a

$$p_i = F\left[E(W_{(i)})\right] \quad i = 1, 2, \dots, n$$

desarrollada por Blom (1958) y usada particularmente cuando la distribución teórica es la distribución normal (gráfico de probabilidad normal). La propuesta consiste en

$$p_i = \frac{i - 3/8}{n + 1/4} \quad i = 1, 2, \dots, n$$

Esta elección de los  $p_i$  ha incrementado su aceptación entre muchos usuarios.

Si observamos, la aproximación dada por Filliben puede obtenerse de la expresión general

$$p_i = \frac{i - c}{n - 2c + 1} \quad \text{para} \quad 0 \leq c \leq 1$$

sin más que elegir  $c = 0.3175$  (para  $i = 2, 3, \dots, n-1$ ). De la misma forma, la aproximación propuesta por Blom se ajusta a la anterior expresión general considerando  $c = 3/8$  ( $i = 1, 2, \dots, n$ ).

Presentamos a continuación una tabla con los distintos valores de los  $p_i$  para las distintas propuestas recogidas anteriormente, y en función de varios tamaños muestrales. (Tabla 4.4).



	i	i/n	$\frac{i-0.5}{n}$	$\frac{i}{n+1}$	$\frac{i-0.3}{n+0.4}$	$\frac{i-3/8}{n+1/4}$	$m_i$	Fillib.
n=2	1	.5	.25	.3333	.2916	.2777	.2928	.2928
	2	1.	.75	.6666	.7083	.7222	.7071	.7071
n=3	1	.333	.166	.25	.2058	.1923	.2062	.2062
	2	.666	.500	.50	.5000	.5000	.5000	.5000
	3	1.	.833	.75	.7941	.8076	.7937	.7937
n=4	1	.25	.125	.20	.1591	.1470	.1591	.1591
	2	.50	.375	.40	.3863	.3823	.3857	.3854
	3	.75	.625	.60	.6136	.6176	.6142	.6145
	4	1.	.875	.80	.8409	.8529	.8408	.8408
n=5	1	.20	.1	.1666	.1296	.1190	.1294	.1294
	2	.40	.3	.3333	.3148	.3095	.3138	.3136
	3	.60	.5	.5000	.5000	.5000	.5000	.5000
	4	.80	.7	.6666	.6851	.6904	.6861	.6863
	5	1.	.9	.8333	.8703	.8809	.8705	.8705
n=10	1	.10	.05	.0909	.0673	.0609	.0669	.0669
	2	.20	.15	.1818	.1634	.1585	.1622	.1623
	3	.30	.25	.2727	.2596	.2560	.2585	.2588
	4	.40	.35	.3636	.3557	.3536	.3550	.3552
	5	.50	.45	.4545	.4519	.4512	.4516	.4517
	6	.60	.55	.5454	.5480	.5487	.5483	.5482
	7	.70	.65	.6363	.6442	.6463	.6449	.6447
	8	.80	.75	.7272	.7403	.7439	.7414	.7411
	9	.90	.85	.8181	.8365	.8414	.8377	.8376
	10	1.	.95	.9090	.9326	.9390	.9330	.9330
n=20	1	.05	.025	.0476	.0343	.0308	.0340	.0340
	2	.10	.075	.0952	.0833	.0802	.0825	.0826
	3	.15	.125	.1428	.1323	.1296	.1314	.1317
	4	.20	.175	.1904	.1813	.1790	.1805	.1808
	5	.25	.225	.2380	.2303	.2283	.2296	.2299
	6	.30	.275	.2857	.2794	.2777	.2788	.2790
	7	.35	.325	.3333	.3284	.3271	.3279	.3281
	8	.40	.375	.3809	.3774	.3765	.3771	.3772
	9	.45	.425	.4285	.4264	.4259	.4262	.4263
	10	.50	.475	.4761	.4754	.4753	.4754	.4754
	11	.55	.525	.5238	.5245	.5246	.5245	.5245
	12	.60	.575	.5714	.5735	.5740	.5737	.5736
	13	.65	.625	.6190	.6225	.6234	.6228	.6227
	14	.70	.675	.6666	.6715	.6728	.6720	.6718
	15	.75	.725	.7142	.7205	.7222	.7211	.7209
	16	.80	.775	.7619	.7696	.7716	.7703	.7700
	17	.85	.825	.8095	.8186	.8209	.8194	.8191
	18	.90	.875	.8571	.8676	.8703	.8685	.8682
	19	.95	.925	.9047	.9166	.9197	.9174	.9173
	20	1.	.975	.9523	.9656	.9691	.9659	.9659

TABLA 4.4: Valores de los  $p_i$ , según distintas definiciones.



	i	i/n	$\frac{i-0.5}{n}$	$\frac{i}{n+1}$	$\frac{i-0.3}{n+0.4}$	$\frac{i-3/8}{n+1/4}$	$m_i$	Fillib.
n=50	1	.02	.01	.0196	.0138	.0124	.0137	.0137
	2	.04	.03	.0392	.0337	.0323	.0333	.0334
	3	.06	.05	.0588	.0535	.0522	.0531	.0532
	4	.08	.07	.0784	.0734	.0721	.0729	.0731
	5	.10	.09	.0980	.0932	.0920	.0927	.0929
	6	.12	.11	.1176	.1130	.1119	.1126	.1128
	7	.14	.13	.1372	.1329	.1318	.1325	.1326
	8	.16	.15	.1568	.1527	.1517	.1523	.1525
	9	.18	.17	.1764	.1726	.1716	.1722	.1723
	10	.20	.19	.1960	.1924	.1915	.1920	.1922
	11	.22	.21	.2156	.2123	.2114	.2119	.2121
	12	.24	.23	.2352	.2321	.2313	.2318	.2319
	13	.26	.25	.2549	.2519	.2512	.2516	.2518
	14	.28	.27	.2745	.2718	.2711	.2715	.2716
	15	.30	.29	.2941	.2916	.2910	.2914	.2915
	16	.32	.31	.3137	.3115	.3109	.3112	.3113
	17	.34	.33	.3333	.3313	.3308	.3311	.3312
	18	.36	.35	.3529	.3511	.3507	.3510	.3510
	19	.38	.37	.3725	.3710	.3706	.3708	.3709
	20	.40	.39	.3921	.3908	.3905	.3907	.3907
	21	.42	.41	.4117	.4107	.4104	.4106	.4106
	22	.44	.43	.4313	.4305	.4303	.4304	.4305
	23	.46	.45	.4509	.4503	.4502	.4503	.4503
	24	.48	.47	.4705	.4702	.4701	.4702	.4702
	25	.50	.49	.4901	.4900	.4900	.4900	.4900
	26	.52	.51	.5098	.5099	.5099	.5099	.5099
	27	.54	.53	.5294	.5297	.5298	.5297	.5297
	28	.56	.55	.5490	.5496	.5497	.5496	.5496
	29	.58	.57	.5686	.5694	.5696	.5695	.5694
	30	.60	.59	.5882	.5892	.5895	.5893	.5893
	31	.62	.61	.6078	.6091	.6094	.6092	.6092
	32	.64	.63	.6274	.6289	.6293	.6291	.6290
	33	.66	.65	.6470	.6488	.6492	.6489	.6489
	34	.68	.67	.6666	.6686	.6691	.6688	.6687
	35	.70	.69	.6862	.6884	.6890	.6887	.6886
	36	.72	.71	.7058	.7083	.7089	.7085	.7084
	37	.74	.73	.7254	.7281	.7288	.7284	.7283
	38	.76	.75	.7450	.7480	.7487	.7483	.7481
	39	.78	.77	.7647	.7678	.7686	.7681	.7680
	40	.80	.79	.7843	.7876	.7885	.7880	.7878
	41	.82	.81	.8039	.8085	.8084	.8079	.8077
	42	.84	.83	.8235	.8273	.8283	.8277	.8276
	43	.86	.85	.8431	.8472	.8482	.8476	.8474
	44	.88	.87	.8627	.8670	.8681	.8674	.8673
	45	.90	.89	.8823	.8869	.8880	.8873	.8871
	46	.92	.91	.9019	.9067	.9079	.9072	.9070
	47	.94	.93	.9215	.9265	.9278	.9270	.9268
	48	.96	.95	.9411	.9464	.9477	.9468	.9467
	49	.98	.97	.9607	.9662	.9676	.9666	.9665
	50	1.	.99	.9803	.9861	.9875	.9862	.9862

92 TABLA 4.4: (continuación)



Observemos que en la tabla se ha incorporado una columna encabezada con  $m_i$  que corresponde a la determinación de los  $p_i$  usando como medida de localización a la mediana de los estadísticos de orden (supuesta normalidad). En el siguiente epígrafe se darán suficientes explicaciones sobre el procedimiento seguido para su cálculo.

En resumen, y ajustándonos a lo mencionado en este epígrafe, indiquemos que entre las distintas definiciones hechas de los  $p_i$ , las aportadas por Hazen, Weibull y Blom, son las que, desde el punto de vista práctico han tenido mayor aceptación a la hora de construir un gráfico de probabilidad.

#### NUESTRA ELECCION DE LOS $p_i$

En este apartado nos vamos a referir al gráfico probabilístico normal o Q-Q plot normal (la distribución teórica es la distribución normal).

Consideraremos la concepción del gráfico de probabilidad como aquel que se obtiene dibujando la  $i$ -ésima observación ordenada de la muestra contra el valor  $F^{-1}(p_i)$ , donde  $p_i$  es obtenido a partir de una medida de localización del  $i$ -ésimo estadístico de orden correspondiente a la distribución teórica en curso (en nuestro caso, la distribución normal estándar).

Como hemos indicado en el apartado anterior, las dos medidas de localización más usadas para el fin que nos interesa son:

$$a) \text{ loc } (X_{(i)}) = E(X_{(i)})$$

$$b) \text{ loc } (X_{(i)}) = \text{Med}(X_{(i)})$$

La medida de localización propuesta en el primer aparta-



do (esperanza del  $i$ -ésimo estadístico de orden), aunque ha sido la usada por Shapiro y Wilk (1965), por Shapiro y Francia (1972) y por La Brecque (1977) para introducir sus correspondientes estadísticos para el contraste de normalidad sobre un conjunto de observaciones, posee una serie de propiedades que podemos calificar de poco deseables, y que son:

- 1°) Las técnicas de integración para calcular  $E(X_{(i)})$  varía drásticamente de una distribución a otra. No existe una técnica uniforme que genere  $E(X_{(i)})$  para todas las distribuciones posibles.
- 2°) Para algunas distribuciones (como por ejemplo la distribución normal), las esperanzas de los estadísticos de orden son difíciles de calcular, consumen mucho tiempo de cálculo y deben ser almacenadas o aproximadas.
- 3°) Para otras distribuciones (por ejemplo la de Cauchy), pueden no estar siempre definidas.

Estas tres propiedades poco deseables son, en general, evitadas abandonando el uso de la esperanza de los estadísticos de orden como medida de localización de éstos, en beneficio de la mediana [ $loc(X_{(i)}) = Med(X_{(i)}) = M_i$ ], opción que a partir de este momento adoptamos como nuestra. Así, nuestro gráfico probabilístico normal quedará definido como aquel que se obtiene representando las observaciones muestrales ordenadas contra las correspondientes medianas,  $M_i$ , de los estadísticos de orden de la distribución normal estándar.



**Determinación de los  $p_i$  a partir de la mediana de los estadísticos de orden de una distribución normal**

En este apartado vamos a determinar el valor de los puntos de posición gráfica,  $p_i$ , a partir de la mediana de los estadísticos de orden de una distribución normal estándar, como medida de localización de éstos, de forma que:

$$p_i = F(\text{Med}(X_{(i)}))$$

o lo que es lo mismo

$$F^{-1}(p_i) = \text{Med}(X_{(i)})$$

Sea, por tanto,  $X_1, X_2, \dots, X_n$  una muestra aleatoria simple tomada de una distribución normal estándar, es decir, la función de distribución adoptará la expresión:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  el correspondiente conjunto de estadísticos de orden, y sea  $M_i = \text{Med}(X_{(i)})$ ,  $i = 1, 2, \dots, n$ .

Para cada  $i$  ( $1 \leq i \leq n$ ),  $M_i$  será aquel valor de la variable  $X_{(i)}$  tal que

$$\Phi_i(M_i) = \frac{1}{2}$$

donde por  $\Phi_i(x)$  notamos a la función de distribución del  $i$ -ésimo estadístico de orden de la distribución normal estándar.

Según se recoge en el apéndice I, y dado que estamos



considerando a la distribución normal estándar como la distribución poblacional, tenemos que

$$\Phi_i(x) = \frac{n!}{(i-1)!(n-i)!} \int_0^{\Phi(x)} t^{i-1} (1-t)^{n-i} dt$$

Por tanto,  $M_i$  será aquel valor tal que:

$$\Phi_i(M_i) = \frac{1}{2} \Leftrightarrow \frac{n!}{(i-1)!(n-i)!} \int_0^{\Phi(M_i)} t^{i-1} (1-t)^{n-i} dt = \frac{1}{2} \quad (4.1)$$

Ahora bien, si consideramos la relación existente entre la distribución de los estadísticos de orden de una distribución normal estándar y los estadísticos de orden de la distribución uniforme estándar (apéndice I), o bien, considerando que si  $m_i$  es la mediana del  $i$ -ésimo estadístico de orden de una distribución uniforme estándar, se tendrá que, dado que la función de distribución de la uniforme es  $F(x) = x$ ,  $m_i$  será aquel valor tal que:

$$\frac{n!}{(i-1)!(n-i)!} \int_0^{m_i} t^{i-1} (1-t)^{n-i} dt = \frac{1}{2} \quad (4.2)$$

De (4.1) y (4.2), obtenemos inmediatamente que  $m_i = \Phi(M_i)$ , o lo que es lo mismo

$$M_i = \Phi^{-1}(m_i)$$

es decir, la mediana del  $i$ -ésimo estadístico de orden de una distribución normal estándar es el valor de la inversa de la función de distribución normal, en  $m_i$ , donde  $m_i$  es la mediana del  $i$ -ésimo estadístico de orden de una distribución uniforme estándar. De esta forma, la determinación de los  $M_i$  pasará por la obtención de



los  $m_i$ . (Queda por tanto establecido que los puntos de posición gráfica,  $p_i$ , con la elección hecha basada en la mediana como medida de localización, no son más que los  $m_i$ )

Para determinar los  $m_i (1 \leq i \leq n)$ , bastará tener en cuenta la distribución del  $i$ -ésimo estadístico de orden de una distribución uniforme estándar. Como recogemos en el apéndice I, su función de distribución adoptará la expresión:

$$\Phi_i(x) = \frac{n!}{(i-1)!(n-i)!} \int_0^x t^{i-1} (1-t)^{n-i} dt$$

que como señalamos en el mismo apéndice, coincide con la función de distribución de una Beta de parámetros  $\alpha = i$  y  $\beta = (n-i+1)$ . Por ello, y dado que  $m_i$  será tal que

$$\Phi_i(m_i) = \frac{1}{2}$$

podemos indicar que  $m_i$  será la mediana de una distribución Beta de parámetros  $\alpha = i$  y  $\beta = (n-i+1)$  y, por tanto,  $m_i$  será aquel valor que verifique la expresión

$$\sum_{k=0}^{n-i} \frac{n!}{(i-1)!(n-i)!} \binom{n-i}{k} (-1)^k \frac{m_i^{i+k}}{(i+k)} = \frac{1}{2}$$

Para cada  $i$ , la expresión anterior es una ecuación polinómica en  $m_i$ , por lo que la determinación de los  $m_i (1 \leq i \leq n)$  para un  $n$  (tamaño muestral) dado, se puede basar en la resolución de  $n$  ecuaciones polinómicas, donde cada uno de los polinomios es de grado  $n$ .

Este proceso de determinación de los  $m_i$  se puede simplificar sin más que tener en cuenta que



$$m_i = 1 - m_{n-i+1} \text{ para } i = 1, 2, \dots, [n/2]$$

donde  $[.]$  simboliza a la parte entera, de manera que si  $n$  es par,  $[n/2] = n/2$ , y la determinación de los  $n/2$  primeros  $m_i$  nos lleva a la obtención de todos ellos. En el caso en que  $n$  sea impar, bastará con tener en cuenta que  $m_{[n/2]+1} = 0.5$ , por lo que la determinación de los  $[n/2]$  primeros  $m_i$  también nos llevará a la obtención de todos ellos.

Al objeto de determinar los  $m_i$ , para un  $n$  fijado, aportamos un programa (apéndice II), escrito en UBASIC, que a la vez que determina los polinomios anteriormente indicados para cada  $i$ , obtiene la solución de éstos. El uso del lenguaje UBASIC está justificado por el hecho de permitirnos una alta precisión, necesaria, a todas luces, debido tanto a los coeficientes que aparecen en los polinomios como a las múltiples operaciones que son necesarias para determinar las raíces de éstos.

A modo de ilustración, consideremos el caso en que  $n = 26$ . Los polinomios a resolver, que los obtenemos como salida del programa mencionado, son:

$$\begin{aligned} & \underline{i = 1} \\ & -0.5 + 26 X - 325 X^2 + 2600 X^3 - 14950 X^4 + 65780 X^5 - 230230 X^6 + \\ & 657800 X^7 - 1562275 X^8 + 3124550 X^9 - 5311735 X^{10} + 7726160 X^{11} - \\ & 9657700 X^{12} + 10400600 X^{13} - 9657700 X^{14} + 7726160 X^{15} - 5311735 X^{16} + \\ & 3124550 X^{17} - 1562275 X^{18} + 657800 X^{19} - 230230 X^{20} + 65780 X^{21} - \\ & 14950 X^{22} + 2600 X^{23} - 325 X^{24} + 26 X^{25} - 1.0 X^{26} = 0 \end{aligned}$$

$$\begin{aligned} & \underline{i = 2} \\ & -0.5 + 325 X^2 - 5200 X^3 + 44850 X^4 - 263120 X^5 + 1151150 X^6 - \\ & 3946800 X^7 + 10935925 X^8 - 24996400 X^9 + 47805615 X^{10} - 77261600 X^{11} + \end{aligned}$$



$$\begin{aligned}
& 106234700 X^{12} - 124807200 X^{13} + 125550100 X^{14} - 108166240 X^{15} + \\
& 79676025 X^{16} - 49992800 X^{17} + 26558675 X^{18} - 11840400 X^{19} + \\
& 4374370 X^{20} - 1315600 X^{21} + 313950 X^{22} - 57200 X^{23} + 7475 X^{24} - \\
& 624 X^{25} + 25 X^{26} = 0
\end{aligned}$$

$$i = 3$$

$$\begin{aligned}
& -0.5 + 2600 X^3 - 44850 X^4 + 394680 X^5 - 2302300 X^6 + 9867000 X^7 - \\
& 32807775 X^8 + 87487400 X^9 - 191222460 X^{10} + 347677200 X^{11} - \\
& 531173500 X^{12} + 686439600 X^{13} - 753300600 X^{14} + 703080560 X^{15} - \\
& 557732175 X^{16} + 374946000 X^{17} - 212469400 X^{18} + 100643400 X^{19} - \\
& 39369330 X^{20} + 12498200 X^{21} - 3139500 X^{22} + 600600 X^{23} - 82225 X^{24} + \\
& 7176 X^{25} - 300 X^{26} = 0
\end{aligned}$$

$$i = 4$$

$$\begin{aligned}
& -0.5 + 14950 X^4 - 263120 X^5 + 2302300 X^6 - 13156000 X^7 + 54679625 X^8 - \\
& 174974800 X^9 + 446185740 X^{10} - 927139200 X^{11} + 1593520500 X^{12} - \\
& 2288132000 X^{13} + 2762102200 X^{14} - 2812322240 X^{15} + 2416839425 X^{16} - \\
& 1749748000 X^{17} + 1062347000 X^{18} - 536764800 X^{19} + 223092870 X^{20} - \\
& 74989200 X^{21} + 19883500 X^{22} - 4004000 X^{23} + 575575 X^{24} - 52624 X^{25} + \\
& 2300 X^{26} = 0
\end{aligned}$$

$$i = 5$$

$$\begin{aligned}
& -0.5 + 65780 X^5 - 1151150 X^6 + 9867000 X^7 - 54679625 X^8 + 218718500 X^9 - \\
& 669278610 X^{10} + 1622493600 X^{11} - 3187041000 X^{12} + 5148297000 X^{13} - \\
& 6905255500 X^{14} + 7733886160 X^{15} - 7250518275 X^{16} + 5686681000 X^{17} - \\
& 3718214500 X^{18} + 2012868000 X^{19} - 892371480 X^{20} + 318704100 X^{21} - \\
& 89475750 X^{22} + 19019000 X^{23} - 2877875 X^{24} + 276276 X^{25} - 12650 X^{26} = 0
\end{aligned}$$

$$i = 6$$

$$\begin{aligned}
& -0.5 + 230230 X^6 - 3946800 X^7 + 32807775 X^8 - 174974800 X^9 + \\
& 669278610 X^{10} - 1946992320 X^{11} + 4461857400 X^{12} - 8237275200 X^{13} +
\end{aligned}$$



$$12429459900 X^{14} - 15467772320 X^{15} + 15951140205 X^{16} - 13648034400 X^{17} + 9667357700 X^{18} - 5636030400 X^{19} + 2677114440 X^{20} - 1019853120 X^{21} + 304217550 X^{22} - 68468400 X^{23} + 10935925 X^{24} - 1105104 X^{25} + 53130 X^{26} = 0$$

$$\underline{i = 7}$$

$$-0.5 + 657800 X^7 - 10935925 X^8 + 87487400 X^9 - 446185740 X^{10} + 1622493600 X^{11} - 4461857400 X^{12} + 9610154400 X^{13} - 16572613200 X^{14} + 23201658480 X^{15} - 26585233675 X^{16} + 25021396400 X^{17} - 19334715400 X^{18} + 12211399200 X^{19} - 6246600360 X^{20} + 2549632800 X^{21} - 811246800 X^{22} + 193993800 X^{23} - 32807775 X^{24} + 3499496 X^{25} - 177100 X^{26} = 0$$

$$\underline{i = 8}$$

$$-0.5 + 1562275 X^8 - 24996400 X^9 + 191222460 X^{10} - 927139200 X^{11} + 3187041000 X^{12} - 8237275200 X^{13} + 16572613200 X^{14} - 26516181120 X^{15} + 34181014725 X^{16} - 35744852000 X^{17} + 30383124200 X^{18} - 20933827200 X^{19} + 11600829240 X^{20} - 5099265600 X^{21} + 1738386000 X^{22} - 443414400 X^{23} + 79676025 X^{24} - 8998704 X^{25} + 480700 X^{26} = 0$$

$$\underline{i = 9}$$

$$-0.5 + 3124550 X^9 - 47805615 X^{10} + 347677200 X^{11} - 1593520500 X^{12} + 5148297000 X^{13} - 12429459900 X^{14} + 23201658480 X^{15} - 34181014725 X^{16} + 40212958500 X^{17} - 37978905250 X^{18} + 28784012400 X^{19} - 17401243860 X^{20} + 8286306600 X^{21} - 3042175500 X^{22} + 831402000 X^{23} - 159352050 X^{24} + 19122246 X^{25} - 1081575 X^{26} = 0$$

$$\underline{i = 10}$$

$$-0.5 + 5311735 X^{10} - 77261600 X^{11} + 531173500 X^{12} - 2288132000 X^{13} + 6905255500 X^{14} - 15467772320 X^{15} + 26585233675 X^{16} - 35744852000 X^{17} + 37978905250 X^{18} - 31982236000 X^{19} + 21268186940 X^{20} - 11048408800 X^{21}$$



$$+ 4394253500 X^{22} - 1293292000 X^{23} + 265586750 X^{24} - 33995104 X^{25} + 2042975 X^{26} = 0$$

$$\underline{i = 11}$$

$$-0.5 + 7726160 X^{11} - 106234700 X^{12} + 686439600 X^{13} - 2762102200 X^{14} + 7733886160 X^{15} - 15951140205 X^{16} + 25021396400 X^{17} - 30383124200 X^{18} + 28784012400 X^{19} - 21268186940 X^{20} + 12153249680 X^{21} - 5273104200 X^{22} + 1681279600 X^{23} - 371821450 X^{24} + 50992656 X^{25} - 3268760 X^{26} = 0$$

$$\underline{i = 12}$$

$$-0.5 + 9657700 X^{12} - 124807200 X^{13} + 753300600 X^{14} - 2812322240 X^{15} + 7250518275 X^{16} - 13648034400 X^{17} + 19334715400 X^{18} - 20933827200 X^{19} + 17401243860 X^{20} - 11048408800 X^{21} + 5273104200 X^{22} - 1834123200 X^{23} + 439425350 X^{24} - 64899744 X^{25} + 4457400 X^{26} = 0$$

$$\underline{i = 13}$$

$$-0.5 + 10400600 X^{13} - 125550100 X^{14} + 703080560 X^{15} - 2416839425 X^{16} + 5686681000 X^{17} - 9667357700 X^{18} + 12211399200 X^{19} - 11600829240 X^{20} + 8286306600 X^{21} - 4394253500 X^{22} + 1681279600 X^{23} - 439425350 X^{24} + 70308056 X^{25} - 5200300 X^{26} = 0$$

$$\underline{i = 14}$$

$$-0.5 + 9657700 X^{14} - 108166240 X^{15} + 557732175 X^{16} - 1749748000 X^{17} + 3718214500 X^{18} - 5636030400 X^{19} + 6246600360 X^{20} - 5099265600 X^{21} + 3042175500 X^{22} - 1293292000 X^{23} + 371821450 X^{24} - 64899744 X^{25} + 5200300 X^{26} = 0$$

$$\underline{i = 15}$$

$$-0.5 + 7726160 X^{15} - 79676025 X^{16} + 374946000 X^{17} - 1062347000 X^{18} + 2012868000 X^{19} - 2677114440 X^{20} + 2549632800 X^{21} - 1738386000 X^{22} +$$



$$831402000 X^{23} - 265586750 X^{24} + 50992656 X^{25} - 4457400 X^{26} = 0$$

$$\underline{i = 16}$$

$$\begin{aligned} & -0.5 + 5311735 X^{16} - 49992800 X^{17} + 212469400 X^{18} - 536764800 X^{19} + \\ & 892371480 X^{20} - 1019853120 X^{21} + 811246800 X^{22} - 443414400 X^{23} + \\ & 159352050 X^{24} - 33995104 X^{25} + 3268760 X^{26} = 0 \end{aligned}$$

$$\underline{i = 17}$$

$$\begin{aligned} & -0.5 + 3124550 X^{17} - 26558675 X^{18} + 100643400 X^{19} - 223092870 X^{20} + \\ & 318704100 X^{21} - 304217550 X^{22} + 193993800 X^{23} - 79676025 X^{24} + \\ & 19122246 X^{25} - 2042975 X^{26} = 0 \end{aligned}$$

$$\underline{i = 18}$$

$$\begin{aligned} & -0.5 + 1562275 X^{18} - 11840400 X^{19} + 39369330 X^{20} - 74989200 X^{21} + \\ & 89475750 X^{22} - 68468400 X^{23} + 32807775 X^{24} - 8998704 X^{25} + 1081575 X^{26} \\ & = 0 \end{aligned}$$

$$\underline{i = 19}$$

$$\begin{aligned} & -0.5 + 657800 X^{19} - 4374370 X^{20} + 12498200 X^{21} - 19883500 X^{22} + \\ & 19019000 X^{23} - 10935925 X^{24} + 3499496 X^{25} - 480700 X^{26} = 0 \end{aligned}$$

$$\underline{i = 20}$$

$$\begin{aligned} & -0.5 + 230230 X^{20} - 1315600 X^{21} + 3139500 X^{22} - 4004000 X^{23} + 2877875 \\ & X^{24} - 1105104 X^{25} + 177100 X^{26} = 0 \end{aligned}$$

$$\underline{i = 21}$$

$$\begin{aligned} & -0.5 + 65780 X^{21} - 313950 X^{22} + 600600 X^{23} - 575575 X^{24} + 276276 X^{25} - \\ & 53130 X^{26} = 0 \end{aligned}$$



$$\underline{i = 22}$$

$$-0.5 + 14950 X^{22} - 57200 X^{23} + 82225 X^{24} - 52624 X^{25} + 12650 X^{26} = 0$$

$$\underline{i = 23}$$

$$-0.5 + 2600 X^{23} - 7475 X^{24} + 7176 X^{25} - 2300 X^{26} = 0$$

$$\underline{i = 24}$$

$$-0.5 + 325 X^{24} - 624 X^{25} + 300 X^{26} = 0$$

$$\underline{i = 25}$$

$$-0.5 + 26 X^{25} - 25 X^{26} = 0$$

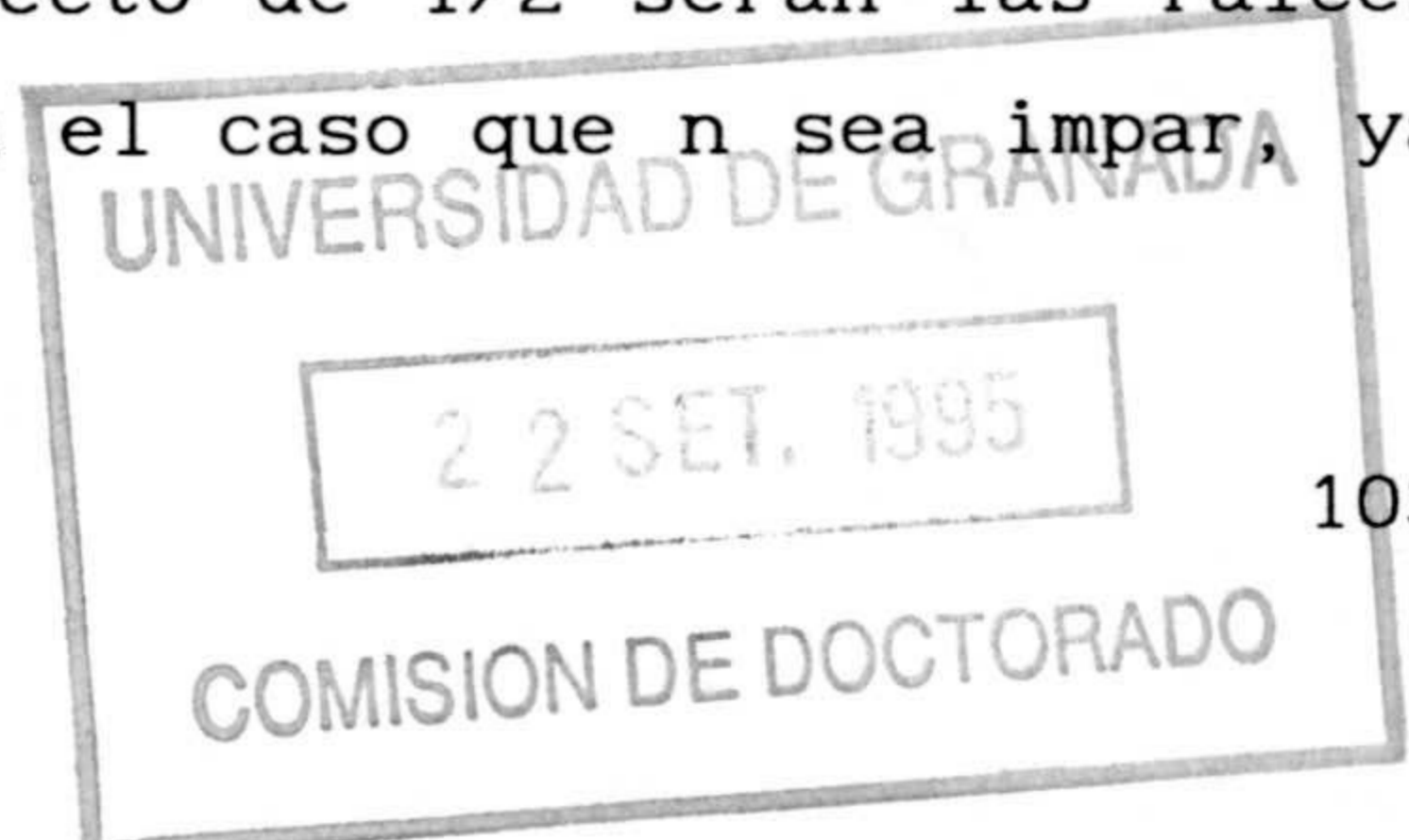
$$\underline{i = 26}$$

$$-0.5 + X^{26} = 0$$

Al escribir todos los polinomios, observamos como propiedad, que establecemos como general, que conforme  $i$  crece, aún manteniendo el correspondiente polinomio grado  $n$ , el número de sumandos de éste disminuye. Por otra parte, los coeficientes que van apareciendo, se hacen algo más manejables. Estas dos características que apreciamos nos inducen a usar la relación

$$m_i = 1 - m_{n-i+1} \quad \text{para } i = 1, 2, \dots, [n/2]$$

en el sentido inverso al anteriormente indicado, es decir, calcularemos las raíces de los polinomios desde  $[n/2] + 1$  hasta  $n$ , y sus correspondientes simétricos respecto de  $1/2$  serán las raíces de los primeros  $[n/2]$  polinomios (en el caso que  $n$  sea impar, ya





hemos indicado que  $m_{\lfloor n/2 \rfloor + 1} = 0.5$ ).

Para determinar las soluciones de los anteriores polinomios, desarrollamos un programa (apéndice II) que, con una altísima precisión, determina las raíces, consiguiendo en una primera etapa una aproximación por el método de bisección, para usar ésta como valor inicial en una segunda etapa donde aplicamos el método de Newton-Raphson.

En una aplicación concreta de este procedimiento, y para  $n = 26$ , obtenemos las siguientes raíces de los polinomios que, por otra parte, corresponderán a las medianas de los estadísticos de orden de una distribución uniforme estándar.

$m_1 = 0.0263072793$	$m_{14} = 0.5189824568$
$m_2 = 0.0637167842$	$m_{15} = 0.5569472181$
$m_3 = 0.1015262003$	$m_{16} = 0.5949115035$
$m_4 = 0.1394224465$	$m_{17} = 0.6328749307$
$m_5 = 0.1773507740$	$m_{18} = 0.6708369998$
$m_6 = 0.2152943069$	$m_{19} = 0.7087970004$
$m_7 = 0.2532461608$	$m_{20} = 0.7467538392$
$m_8 = 0.2912029996$	$m_{21} = 0.7847056931$
$m_9 = 0.3291630002$	$m_{22} = 0.8226492260$
$m_{10} = 0.3671250693$	$m_{23} = 0.8605775535$
$m_{11} = 0.4050884965$	$m_{24} = 0.8984737997$
$m_{12} = 0.4430527819$	$m_{25} = 0.9362832158$
$m_{13} = 0.4810175432$	$m_{26} = 0.9736927207$



Aunque la representación que se recoge de los  $m_i$  es sólo con diez dígitos, el programa nos permite usar una representación de estos elementos de hasta 2.500 dígitos, representación que abandonamos sólo en la presentación que aquí se hace de las soluciones, aunque es mantenida para todos los cálculos internos.

A continuación presentamos una serie de tablas donde se recogen los valores de los  $m_i (= p_i)$  para distintos valores del tamaño muestral. Insistimos que la presentación que hacemos, en estas tablas, sólo con tres dígitos está justificada únicamente por el hecho de conseguir unas tablas suficientemente compactas para su impresión, aunque todos los cálculos internos mantienen tal precisión que hace que las soluciones aportadas puedan ser consideradas como exactas.

En las tablas se representan los  $m_i$  para  $i = 1, 2, \dots, n/2$ , si  $n$  es par, y para  $i = 1, 2, \dots, [n/2] + 1$ , si  $n$  es impar. Para determinar los restantes  $m_i$ , bastará con hacer uso de la relación

$$m_{n-i+1} = 1 - m_i \quad i = 1, 2, \dots, [n/2]$$

$i \backslash n$	2	3	4	5	6	7	8	9	10	11	12	13
1	.293	.206	.159	.129	.109	.094	.083	.074	.067	.061	.056	.052
2		.500	.386	.314	.265	.229	.201	.180	.162	.148	.136	.126
3			.500	.422	.364	.321	.286	.259	.236	.217	.201	
4				.500	.440	.393	.355	.324	.298	.275		
5					.500	.452	.412	.379	.350			
6						.500	.460	.425				
7							.500					.500

TABLA 4.5: Determinación de los  $m_i$  para  $n = 2$  (1) 50



$i \backslash n$	14	15	16	17	18	19	20	21	22	23	24	25
1	.048	.045	.042	.040	.038	.036	.034	.032	.031	.030	.028	.027
2	.117	.109	.103	.097	.091	.087	.082	.079	.075	.072	.069	.066
3	.186	.174	.164	.154	.146	.138	.131	.125	.120	.114	.110	.105
4	.256	.239	.225	.212	.200	.190	.180	.172	.164	.157	.151	.145
5	.326	.304	.286	.269	.255	.241	.230	.219	.209	.200	.192	.184
6	.395	.370	.347	.327	.309	.293	.279	.266	.254	.243	.233	.224
7	.465	.435	.408	.385	.363	.345	.328	.312	.298	.286	.274	.263
8		.500	.469	.442	.418	.397	.377	.359	.343	.329	.315	.303
9			.500	.473	.448	.426	.406	.388	.371	.356	.342	
10				.500	.475	.453	.433	.414	.397	.382		
11					.500	.478	.457	.438	.421			
12						.500	.479	.460				
13												.500

TABLA 4.5: (continuación)



$i \backslash n$	26	27	28	29	30	31	32	33	34	35	36	37
1	.026	.025	.024	.024	.023	.022	.021	.021	.020	.020	.019	.018
2	.064	.061	.059	.057	.055	.053	.052	.050	.049	.047	.046	.045
3	.101	.098	.094	.091	.088	.085	.083	.080	.078	.076	.073	.072
4	.139	.134	.129	.125	.121	.117	.113	.110	.107	.104	.101	.098
5	.177	.171	.165	.159	.154	.149	.144	.140	.136	.132	.128	.125
6	.215	.207	.200	.193	.187	.181	.175	.170	.165	.160	.156	.152
7	.253	.244	.235	.227	.220	.213	.206	.200	.194	.189	.183	.179
8	.291	.280	.271	.261	.253	.245	.237	.230	.223	.217	.211	.205
9	.329	.317	.306	.295	.286	.277	.268	.260	.252	.245	.238	.232
10	.367	.354	.341	.329	.319	.308	.299	.290	.281	.274	.266	.259
11	.405	.390	.376	.364	.352	.340	.330	.320	.311	.302	.294	.286
12	.443	.427	.412	.398	.385	.372	.361	.350	.340	.330	.321	.312
13	.481	.463	.447	.432	.417	.404	.392	.380	.369	.358	.349	.339
14		.500	.482	.466	.450	.436	.423	.410	.398	.387	.376	.366
15			.500	.483	.468	.454	.440	.427	.415	.404	.393	
16				.500	.484	.470	.456	.443	.431	.420		
17					.500	.485	.472	.459	.446			
18						.500	.486	.473				
19												.500

TABLA 4.5: (continuación)



i \ n	38	39	40	41	42	43	44	45	46	47	48	49
1	.018	.018	.017	.017	.016	.016	.016	.015	.015	.015	.014	.014
2	.044	.043	.042	.040	.040	.039	.038	.037	.036	.035	.035	.034
3	.070	.068	.066	.065	.063	.062	.060	.059	.058	.056	.055	.054
4	.096	.093	.091	.089	.087	.085	.083	.081	.079	.077	.076	.074
5	.122	.119	.116	.113	.110	.108	.105	.103	.100	.099	.097	.095
6	.148	.144	.140	.137	.134	.131	.128	.125	.122	.120	.117	.115
7	.174	.169	.165	.161	.157	.154	.150	.147	.144	.141	.138	.135
8	.200	.195	.190	.185	.181	.177	.173	.169	.165	.162	.159	.155
9	.226	.220	.215	.210	.205	.200	.195	.191	.187	.183	.179	.176
10	.252	.246	.240	.234	.228	.223	.218	.213	.209	.204	.200	.196
11	.278	.271	.264	.258	.252	.246	.241	.235	.230	.225	.221	.216
12	.304	.297	.289	.282	.276	.269	.263	.257	.252	.246	.241	.236
13	.330	.322	.314	.306	.299	.292	.286	.279	.273	.268	.262	.257
14	.356	.347	.339	.331	.323	.315	.308	.301	.295	.289	.283	.277
15	.383	.373	.364	.355	.346	.338	.331	.323	.316	.310	.303	.297
16	.409	.398	.388	.379	.370	.361	.353	.346	.338	.331	.324	.317
17	.435	.424	.413	.403	.394	.385	.376	.368	.360	.352	.345	.338
18	.461	.449	.438	.427	.417	.408	.398	.390	.381	.373	.365	.358
19	.487	.474	.463	.452	.441	.431	.421	.412	.403	.394	.386	.379
20		.500	.488	.476	.464	.454	.444	.434	.424	.415	.407	.399
21			.500	.488	.477	.466	.456	.446	.437	.427	.419	
22				.500	.489	.478	.468	.458	.448	.439		
23					.500	.489	.479	.469	.459			
24						.500	.490	.480				
25								.500				

TABLA 4.5: (continuación)



$i \backslash n$	50	55	60	65	70	75	80	85	90	95	100
1	.014	.012	.011	.011	.010	.009	.009	.008	.008	.007	.007
2	.033	.030	.028	.026	.024	.022	.021	.020	.018	.018	.017
3	.053	.048	.044	.041	.038	.035	.033	.031	.030	.028	.027
4	.073	.066	.061	.056	.052	.049	.046	.043	.041	.038	.036
5	.093	.084	.077	.071	.066	.062	.058	.055	.052	.049	.046
6	.113	.102	.094	.087	.081	.075	.070	.066	.063	.059	.056
7	.132	.120	.110	.102	.095	.088	.083	.078	.074	.070	.066
8	.152	.138	.127	.117	.109	.102	.095	.090	.085	.080	.076
9	.172	.157	.144	.133	.123	.115	.108	.101	.096	.091	.086
10	.192	.175	.160	.148	.137	.128	.120	.113	.107	.101	.096
11	.212	.193	.177	.163	.152	.142	.133	.125	.118	.112	.106
12	.232	.211	.193	.178	.166	.155	.145	.137	.129	.122	.116
13	.252	.229	.210	.194	.180	.168	.158	.148	.140	.133	.126
14	.271	.247	.226	.209	.194	.181	.170	.160	.151	.143	.136
15	.291	.265	.243	.224	.208	.195	.182	.172	.162	.154	.146
16	.311	.283	.260	.240	.223	.208	.195	.184	.173	.164	.156
17	.331	.301	.276	.255	.237	.221	.207	.195	.184	.175	.166
18	.351	.319	.293	.270	.251	.234	.220	.207	.195	.185	.176
19	.371	.337	.309	.286	.265	.248	.232	.219	.207	.196	.186
20	.391	.355	.326	.301	.280	.261	.245	.230	.218	.206	.196
21	.410	.373	.342	.316	.294	.274	.257	.242	.229	.217	.206
22	.430	.391	.359	.332	.308	.288	.270	.254	.240	.227	.216
23	.450	.410	.376	.347	.322	.301	.282	.266	.251	.238	.226
24	.470	.428	.392	.362	.336	.314	.295	.277	.262	.248	.236
25	.490	.446	.409	.377	.351	.327	.307	.289	.273	.259	.246

TABLA 4.6: Determinación de los  $m_i$  para  $n = 50$  (5) 100



$i \backslash n$	50	55	60	65	70	75	80	85	90	95	100
26		.464	.425	.393	.365	.341	.319	.301	.284	.269	.256
27		.482	.442	.408	.379	.354	.332	.312	.295	.280	.266
28		.500	.458	.423	.393	.367	.344	.324	.306	.290	.276
29			.475	.439	.407	.380	.357	.336	.317	.301	.286
30			.492	.454	.422	.394	.369	.348	.328	.311	.296
31				.469	.436	.407	.382	.359	.339	.322	.306
32				.485	.450	.420	.394	.371	.350	.332	.316
33				.500	.464	.434	.407	.383	.362	.343	.325
34					.479	.447	.419	.394	.373	.353	.335
35					.493	.460	.431	.406	.384	.364	.345
36						.473	.444	.418	.395	.374	.355
37						.487	.456	.430	.406	.385	.365
38						.500	.469	.441	.417	.395	.375
39							.481	.453	.428	.405	.385
40							.494	.465	.439	.416	.395
41								.476	.450	.426	.405
42								.488	.461	.437	.415
43								.500	.472	.447	.425
44									.483	.458	.435
45									.494	.468	.445
46										.479	.455
47										.489	.465
48										.500	.475
49											.485
50											.495

TABLA 4.6: (continuación )



$i \backslash n$	110	120	130	140	150	160	170	180	190	200
1	.006	.006	.005	.005	.005	.004	.004	.004	.004	.003
2	.015	.014	.013	.012	.011	.010	.010	.009	.009	.008
3	.024	.022	.020	.019	.018	.017	.016	.015	.014	.013
4	.033	.030	.028	.026	.024	.023	.021	.020	.019	.018
5	.042	.039	.036	.033	.031	.029	.027	.026	.024	.023
6	.051	.047	.043	.040	.038	.035	.033	.031	.030	.028
7	.060	.055	.051	.047	.044	.041	.039	.037	.035	.033
8	.069	.064	.059	.055	.051	.048	.045	.042	.040	.038
9	.078	.072	.066	.062	.058	.054	.051	.048	.045	.043
10	.088	.080	.074	.069	.064	.060	.057	.054	.051	.048
11	.097	.089	.082	.076	.071	.066	.063	.059	.056	.053
12	.106	.097	.089	.083	.078	.073	.068	.065	.061	.058
13	.115	.105	.097	.090	.084	.079	.074	.070	.066	.063
14	.124	.113	.105	.097	.091	.085	.080	.076	.072	.068
15	.133	.122	.112	.104	.097	.091	.086	.081	.077	.073
16	.142	.130	.120	.112	.104	.098	.092	.087	.082	.078
17	.151	.138	.128	.119	.111	.104	.098	.092	.087	.083
18	.160	.147	.135	.126	.117	.110	.104	.098	.093	.088
19	.169	.155	.143	.133	.124	.116	.109	.103	.098	.093
20	.178	.163	.151	.140	.131	.123	.115	.109	.103	.098
21	.187	.172	.158	.147	.137	.129	.121	.115	.108	.103
22	.196	.180	.166	.154	.144	.135	.127	.120	.114	.108
23	.205	.188	.174	.161	.151	.141	.133	.126	.119	.113
24	.214	.197	.181	.169	.157	.148	.139	.131	.124	.118
25	.223	.205	.189	.176	.164	.154	.145	.137	.130	.123

TABLA 4.7: Detrminación de los  $m_i$  para  $n = 110 (10) 200$



$i \backslash n$	110	120	130	140	150	160	170	180	190	200
26	.233	.213	.197	.183	.171	.160	.151	.142	.135	.128
27	.242	.222	.205	.190	.177	.166	.156	.148	.140	.133
28	.251	.230	.212	.197	.184	.172	.162	.153	.145	.138
29	.260	.238	.220	.204	.191	.179	.168	.159	.151	.143
30	.269	.246	.228	.211	.197	.185	.174	.164	.156	.148
31	.278	.255	.235	.218	.204	.191	.180	.170	.161	.153
32	.287	.263	.243	.226	.211	.197	.186	.176	.166	.158
33	.296	.271	.251	.233	.217	.204	.192	.181	.172	.163
34	.305	.280	.258	.240	.224	.210	.198	.187	.177	.168
35	.314	.288	.266	.247	.231	.216	.203	.192	.182	.173
36	.323	.296	.274	.254	.237	.222	.209	.198	.187	.178
37	.332	.305	.281	.261	.244	.229	.215	.203	.193	.183
38	.341	.313	.289	.268	.250	.235	.221	.209	.198	.188
39	.350	.321	.297	.275	.257	.241	.227	.214	.203	.193
40	.359	.330	.304	.283	.264	.247	.233	.220	.208	.198
41	.368	.338	.312	.290	.270	.254	.239	.225	.214	.203
42	.378	.346	.320	.297	.277	.260	.245	.231	.219	.208
43	.387	.354	.327	.304	.284	.266	.250	.237	.224	.213
44	.396	.363	.335	.311	.290	.272	.256	.242	.229	.218
45	.405	.371	.343	.318	.297	.278	.262	.248	.235	.223
46	.414	.379	.350	.325	.304	.285	.268	.253	.240	.228
47	.423	.388	.358	.332	.310	.291	.274	.259	.245	.233
48	.432	.396	.366	.340	.317	.297	.280	.264	.250	.238
49	.441	.404	.373	.347	.324	.303	.286	.270	.256	.243
50	.450	.413	.381	.354	.330	.310	.291	.275	.261	.248

TABLA 4.7: (continuación)



$i \backslash n$	110	120	130	140	150	160	170	180	190	200
51	.459	.421	.389	.361	.337	.316	.297	.281	.266	.253
52	.468	.429	.396	.368	.344	.322	.303	.286	.271	.258
53	.477	.438	.404	.375	.350	.328	.309	.292	.277	.263
54	.486	.446	.412	.382	.357	.335	.315	.297	.282	.268
55	.495	.454	.419	.389	.364	.341	.321	.303	.287	.273
56		.463	.427	.397	.370	.347	.327	.309	.292	.278
57		.471	.435	.404	.377	.353	.333	.314	.298	.283
58		.479	.442	.411	.383	.360	.338	.320	.303	.288
59		.487	.450	.418	.390	.366	.344	.325	.308	.293
60		.496	.458	.425	.397	.372	.350	.331	.313	.298
61			.465	.432	.403	.378	.356	.336	.319	.303
62			.473	.439	.410	.385	.362	.342	.324	.308
63			.481	.446	.417	.391	.368	.347	.329	.313
64			.488	.454	.423	.397	.374	.353	.334	.318
65			.496	.461	.430	.403	.380	.358	.340	.323
66				.468	.437	.409	.385	.364	.345	.328
67				.475	.443	.416	.391	.370	.350	.333
68				.482	.450	.422	.397	.375	.355	.338
69				.489	.457	.428	.403	.381	.361	.343
70				.496	.463	.434	.409	.386	.366	.348
71					.470	.441	.415	.392	.371	.353
72					.477	.447	.421	.397	.376	.358
73					.483	.453	.427	.403	.382	.363
74					.490	.459	.432	.408	.387	.368
75					.497	.466	.438	.414	.392	.373

TABLA 4.7: (continuación)



i \ n	110	120	130	140	150	160	170	180	190	200
76						.472	.444	.419	.397	.378
77						.478	.450	.425	.403	.383
78						.484	.456	.431	.408	.388
79						.491	.462	.436	.413	.393
80						.497	.468	.442	.418	.398
81							.473	.447	.424	.403
82							.479	.453	.429	.408
83							.485	.458	.434	.413
84							.491	.464	.439	.418
85							.497	.469	.445	.423
86								.475	.450	.428
87								.480	.455	.433
88								.486	.460	.438
89								.492	.466	.442
90								.497	.471	.447
91									.476	.452
92									.482	.457
93									.487	.462
94									.492	.467
95									.497	.472
96										.477
97										.482
98										.487
99										.492
100										.497

TABLA 4.7: (continuación)



Como comentario final a este apartado, recordemos que aunque tanto Benard y Bos-Levenbach (1953) como Filliben (1975) ya propusieron el uso de la mediana como medida de localización usada para determinar el valor de los pntos de posición,  $p_i$ , éstos abandonaron su cálculo directo en beneficio de las aproximaciones que anteriormente hemos recogido. Aunque estas aproximaciones puedan considerarse bastante buenas (la de Filliben mejor que la de Benard y Bos-Levenbach), incluso para valores pequeños de  $n$ , el procedimiento que aquí proponemos es "casi" exacto, alcanzando, sin un coste elevado de tiempo, una precisión tan grande como queramos. Esta es la razón que nos ha llevado a elegir los puntos de posición gráfica del gráfico probabilístico normal como la mediana del  $i$ -ésimo estadístico de orden de una distribución uniforme estándar, sin necesidad de acudir a aproximaciones.







## CAPITULO 5

### EL GRAFICO PROBABILISTICO NORMAL COMO TECNICA GRAFICA PARA EL ESTUDIO DE NORMALIDAD

En la presentación que hicimos de los gráficos de probabilidad empíricos, resultaba bastante intuitivo que éstos se introducían para comparar las distribuciones de dos conjuntos de datos, aceptándolas tanto más iguales, conforme los puntos en el gráfico se aproximaran más a la recta  $y = x$ .

Al introducir el gráfico de probabilidad teórico a partir del empírico, sin más que sustituir en éste último una de las distribuciones empíricas por una distribución teórica, cabe esperar que la interpretación dada allí sea válida en este ambiente. En efecto, si consideramos la situación en que la distribución teórica constituya una buena aproximación de la distribución empírica, cabe esperar que los cuantiles de los datos estén muy próximos a los de la distribución teórica y, por tanto, los puntos en el gráfico de probabilidad caigan muy cerca de la recta  $y = x$ , conformándose ésta en una línea de referencia en el gráfico. En otras palabras, cuanto más "rectos" queden representados los puntos en el Q-Q plot, tanto más podremos afirmar que la distribución empírica de los datos está bien aproximada por la correspondiente distribución teórica; luego el gráfico de probabilidad nos permite comparar el comportamiento distribucional de un conjunto de datos respecto de una distribución teórica especificada. De esta forma, en el caso en particular que la distribución teórica sea la nor-



mal, el gráfico probabilístico normal nos permitirá comparar la distribución empírica de los datos con la distribución normal, y por tanto, el gráfico probabilístico normal puede considerarse como una técnica gráfica para la prueba de normalidad sobre un conjunto de datos.

Para reafirmarnos en esta idea, consideremos el caso de comparar el comportamiento distribucional de un conjunto de datos, recogido en la tabla 5.1, con la distribución normal.

OBSERVACIONES	OBSERVACIONES ORDENADAS
0.167	-1.669
-0.348	-0.757
0.783	-0.669
0.081	-0.471
-0.006	-0.348
-0.005	-0.238
0.608	-0.120
1.192	-0.118
-1.669	-0.006
1.211	-0.005
0.580	0.081
0.649	0.167
-0.238	0.219
0.219	0.260
0.490	0.490
-0.120	0.580
0.607	0.607
0.893	0.608
-0.669	0.649
-0.471	0.755
0.917	0.783
-0.757	0.893
0.755	0.917
-0.118	1.192
0.260	1.211
1.350	1.350

FIGURA 5.1: Observaciones de una variable (n = 26)

Con el fin de comprender cómo el gráfico de probabilidad produce una comparación entre la distribución empírica de los da-



tos y la distribución teórica, consideremos la siguiente figura 5.1

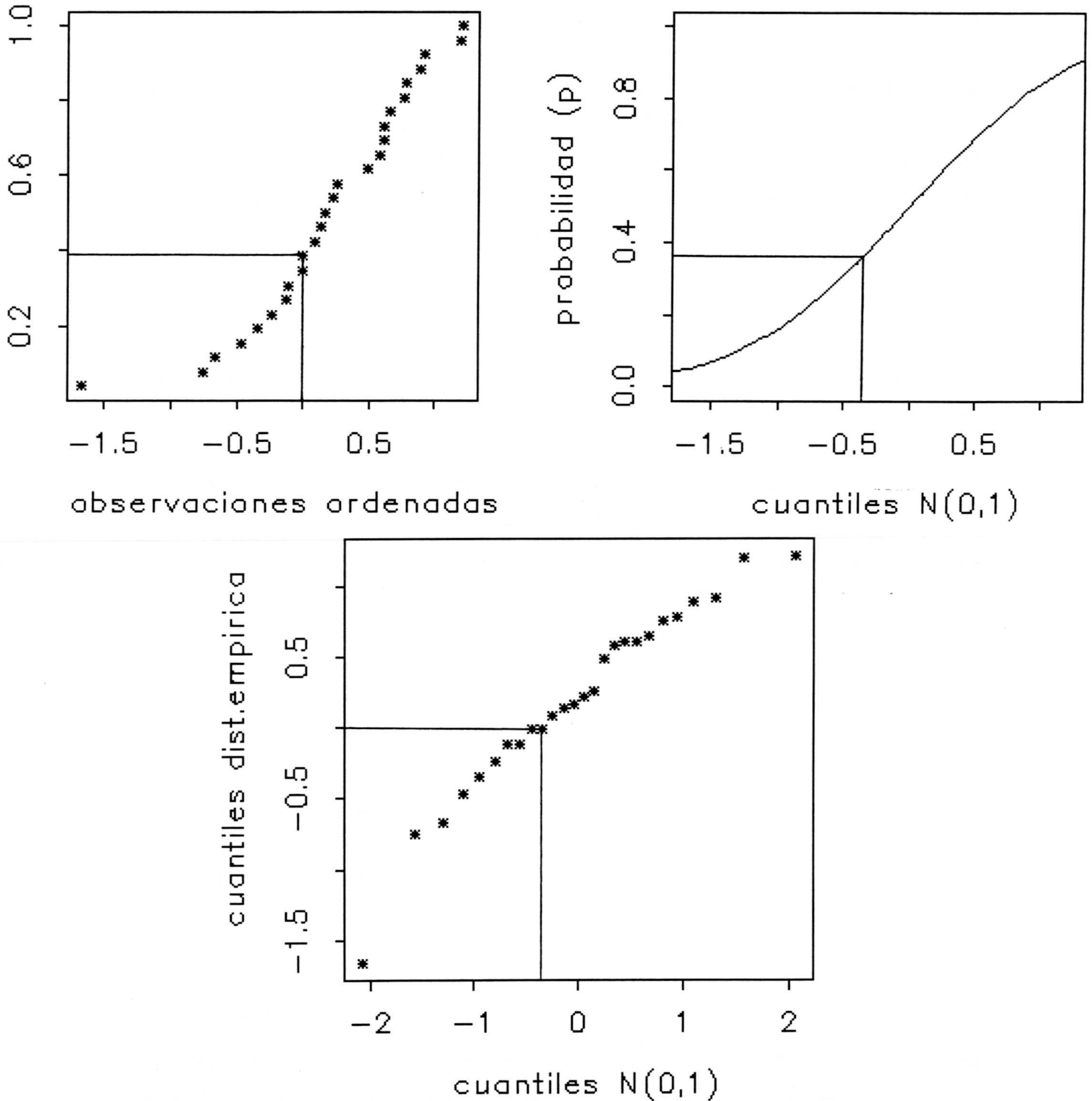


FIGURA 5.1: (a) Frecuencia acumulada variable  $x$ . (b) Función de distribución de la  $N(0,1)$ . (c) Gráfico probabilístico normal.



Para proceder a comparar la distribución empírica con la distribución normal, podríamos observar conjuntamente las figuras 5.1 (a) y 5.1 (b), que nos daría una idea aproximada del posible comportamiento normal de los datos. Sin embargo, con el gráfico probabilístico normal se propone una conjunción o combinación de los anteriores, obteniendo así un gráfico como el que aparece en la figura 5.1 (c), donde los puntos que se representan son tomados de los anteriores, de manera que las coordenadas de un punto en el gráfico de probabilidad serán las abscisas correspondientes de los anteriores gráficos (figuras 5.1 (a) y (b)) para una cantidad de probabilidad fijada  $p_i$ . (Observemos los puntos señalados en las figuras 5.1 (a) y 5.1 (b), para un mismo valor de  $p$ . Sus correspondientes abscisas constituirán las coordenadas del punto marcado en la figura 5.1 (c)).

La propia definición de los gráficos de probabilidad nos permite concebirlos como una "herramienta informal" para comprobar las hipótesis distribucionales establecidas sobre el conjunto de datos bajo estudio, donde juzgaremos éstas como "aceptables" o "inaceptables" si los puntos en dicho gráfico se disponen "próximos" o "lejanos" a la recta  $y = x$ , respectivamente. Sin embargo, es necesario hacer un breve comentario sobre las posibles fluctuaciones que puedan existir de los puntos respecto de la recta.

Algunas fluctuaciones son debidas no más que al carácter aleatorio de los datos, causando que los puntos en el gráfico difieran de la recta, pero permaneciendo "razonablemente" cerca de ella, por lo que seguiremos calificando como correcta la hipótesis distribucional.

En otros casos puede ocurrir que los puntos no estén próximos a la recta  $y = x$ , pero que su configuración se aproxime a una recta paralela a ésta, o incluso a otra que ni siquiera sea paralela a  $y = x$ . En este caso podemos encontrar las correspondientes constantes que produzcan un cambio de localización y esca-



la en los datos de forma que los valores resultantes nos lleven, en el gráfico de probabilidad, a una disposición de los puntos suficientemente próxima a la recta  $y = x$ . Concluiremos, en este caso, que la distribución empírica es compatible con la distribución teórica, salvo en los parámetros de localización y escala.

En la línea de lo indicado, cuando el gráfico de probabilidad es usado para verificar la normalidad de un conjunto de datos, su construcción se hará siguiendo el esquema ya presentado, de forma que aprovechando la propiedad de invarianza frente a cambios de localización y escala, siempre produciremos una comparación con la distribución normal estándar (media cero y varianza uno). De esta manera, bastará con calcular los cuantiles de la distribución  $N(0;1)$ , aceptando la hipótesis de normalidad de los datos siempre que los puntos en el gráfico tengan un comportamiento "suficientemente rectilíneo", no necesariamente sobre la recta  $y = x$ .

La propiedad de invarianza en el procedimiento, frente a cambios de localización y escala, nos permite usarlo en sentido inverso, y obtener así estimadores de los parámetros poblacionales de localización y escala, sin más que trazar la línea a la que mejor se ajustan los puntos en el gráfico y determinar su intersección e inclinación como estimadores respectivos de los anteriores parámetros. En particular, en el caso normal, supuesto que consideramos compatible la distribución empírica y la teórica, podemos producir una estimación de la media y la desviación típica de la población, como parámetros de localización y escala en la distribución normal.

A la hora de juzgar el comportamiento rectilíneo de los datos que nos permita aceptar la hipótesis distribucional establecida sobre ellos, resultará útil trazar la línea que mejor se ajuste a los puntos en el gráfico de probabilidad. Esto lo podemos



hacer bien "a ojo" (con lo que garantizamos un ahorro en lo que se refiere al tiempo necesario para llevar a cabo una serie de cálculos), o bien mediante métodos numéricos (con lo que, aunque esto exija dedicar tiempo a los cálculos, conseguiremos aumentar la precisión en nuestros posteriores análisis). Dado que más tarde trataremos en particular la determinación analítica de la recta que mejor se ajusta al conjunto de puntos en el gráfico de probabilidad, trabajaremos hasta entonces con una recta determinada bajo cualquier procedimiento, y que será la que mejor explica el comportamiento de los puntos en el Q-Q plot.

No todas las fluctuaciones, respecto de la recta, que podamos encontrar en el gráfico de probabilidad son debidas a las causas citadas anteriormente, dado que si así ocurriera, aceptaríamos como válida cualquier suposición distribucional que establecieramos sobre cualquier conjunto de datos. Por el contrario, pueden existir otras muchas, incluso de carácter sistemático, que nos abocarán a rechazar la hipótesis de normalidad enunciada sobre el conjunto de datos. A estas alturas, ponemos de relieve una segunda característica crucial de los gráficos de probabilidad, y es que, a parte de indicarnos un pobre ajuste de la distribución empírica a la distribución teórica, nos pueden orientar sobre la naturaleza y causa de tal alejamiento.

Las principales causas que pueden dar lugar a un alejamiento de los puntos en el Q-Q plot respecto de la recta ajustada, son las siguientes:

- Existencia de outliers.
- Curvatura en ambos extremos, indicando la existencia de colas grandes o pequeñas respecto de la distribución teórica.
- Curvatura cóncava o convexa poniendo de relieve una falta de



simetría.

- Existencia de segmentos horizontales, grupos de puntos o discontinuidades.

### **Causas de no normalidad apreciables en el gráfico probabilístico normal**

Citadas las anteriores causas que ponen de relieve la ausencia de normalidad mediante el gráfico probabilístico normal, entremos en una discusión detallada de cada una de ellas. Esto nos permitirá además establecer una especie de patrón de comportamiento en los gráficos de probabilidad que nos sirva para rechazar la hipótesis de normalidad de un conjunto de datos.

#### **A.- EXISTENCIA DE OUTLIERS.**

Un hecho que desafortunadamente ocurre de manera habitual en la estadística al trabajar con muestras, es que aparezcan datos que podemos calificar como "contaminados" y que proporcionan valores raros dentro del conjunto del resto de observaciones. Estos elementos, que son denominados outliers, son detectados por los gráficos de probabilidad como puntos que salen del rango de comportamiento del resto.

Los outliers aparecerán siempre, si es que existen, en las colas de la distribución de los datos. Precisamente por ello es por lo que su identificación se ha de hacer con el suficiente cuidado, ya que incluso en muestras con buen comportamiento existen observaciones en las colas que, aunque parezca que tienen un comportamiento anómalo, no deben ser interpretadas como outliers.

En los casos en que algún outlier sea detectado, debemos, si es posible, volver a la fuente de donde han sido obtenidos



los datos, dado que tales observaciones anómalas pueden provenir de errores de medición que, en su caso, procederemos a corregir. Supuesto que lo anterior no es posible, o concluimos que las observaciones son correctas, eliminaremos los outliers (siempre que éstos no constituyan un número muy elevado o sean juzgados como observaciones fundamentales) y construiremos un gráfico de probabilidad con el resto de elementos muestrales, encontrando pequeños cambios en la forma del dibujo respecto del anterior, cambios que rara vez harán cambiar nuestro juicio sobre el resto de los datos, por lo que la producción del nuevo gráfico de probabilidad sólo será necesaria en el caso en que los outliers fueran tan extremos que hubieran forzado a una severa compresión del resto de los datos, o cuando se haga uso del gráfico de probabilidad para producir una estimación de los parámetros poblacionales.

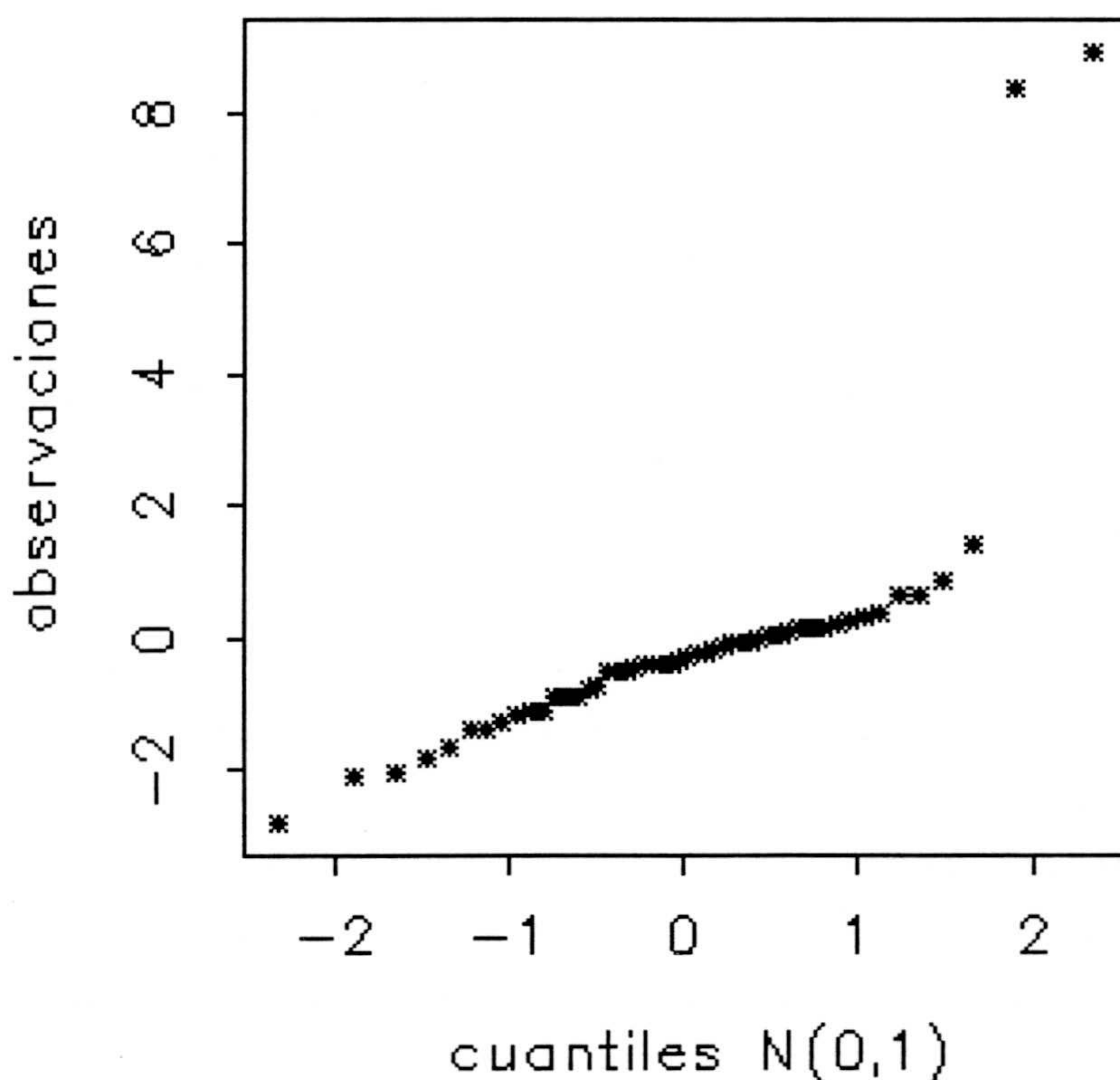


FIGURA 5.2: Gráfico probabilístico normal con outliers.



En la figura 5.2, observamos la existencia de dos outliers. Si eliminamos estos valores, y redibujamos el gráfico probabilístico normal, su configuración quedará como aparece en la figura 5.3.

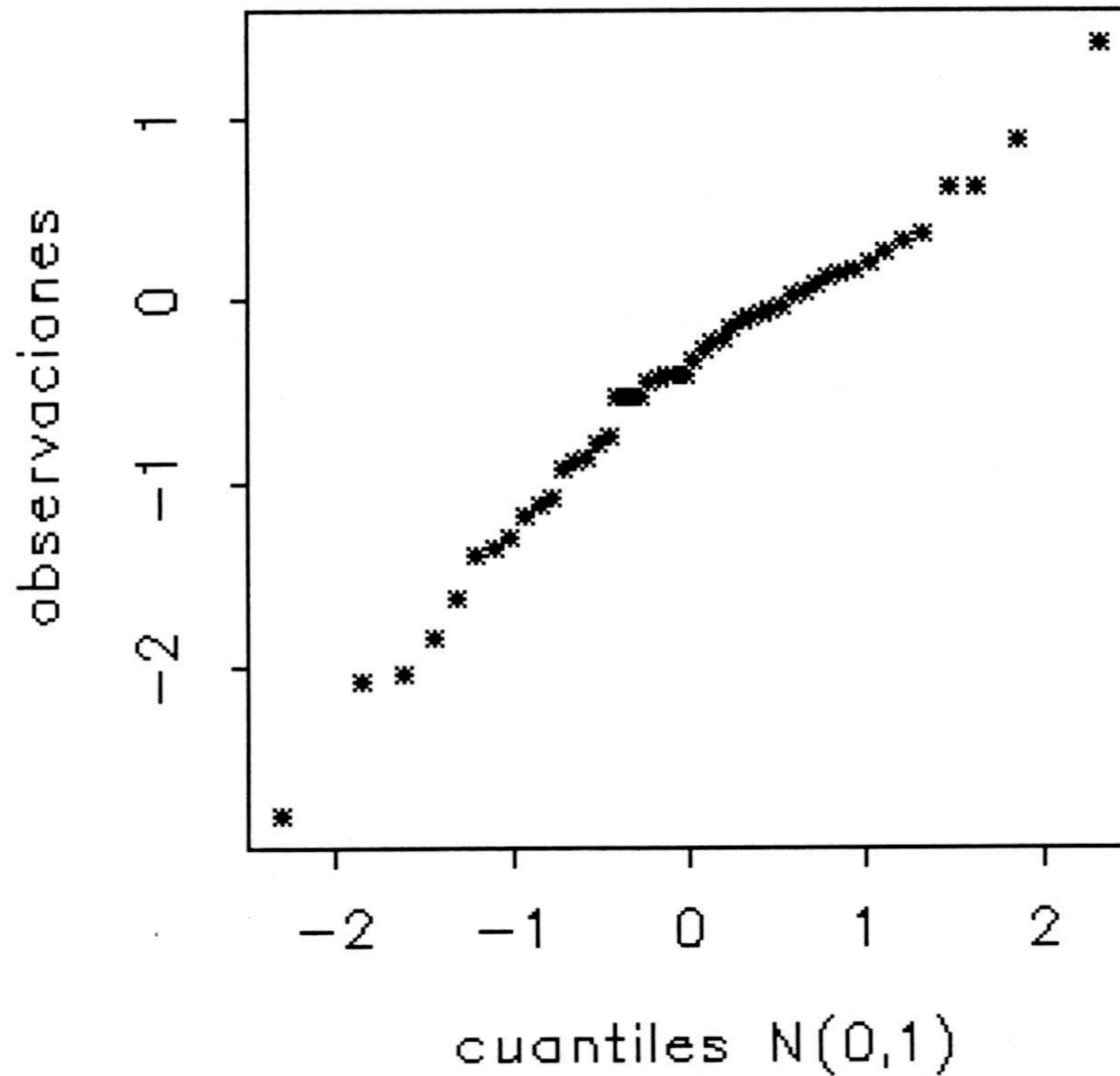


FIGURA 5.3: Gráfico probabilístico normal, eliminados los outliers.

#### B.- EXISTENCIA DE PEQUEÑAS O GRANDES COLAS.

Una segunda razón que nos aleja de la linealidad en los gráficos de probabilidad y que, por tanto, pone de relieve la falta de ajuste de la distribución de los datos a la distribución teórica enunciada como hipótesis, es la existencia en la distribución empírica de colas más grandes, o bien más pequeñas que las que presenta la distribución teórica.



En el caso particular que la distribución teórica sea la distribución normal, la existencia de colas más grandes respecto de la distribución normal se manifiesta al aparecer, en la izquierda del gráfico de probabilidad, un considerable número de puntos por debajo de la línea ajustada, y por encima, a la derecha. La situación inversa a la anterior (a la izquierda por encima y a la derecha por debajo), indica la existencia de colas más pequeñas que la distribución normal.

Al ocurrir cualquiera de las anteriores circunstancias, rechazaremos la normalidad de los datos bajo estudio.

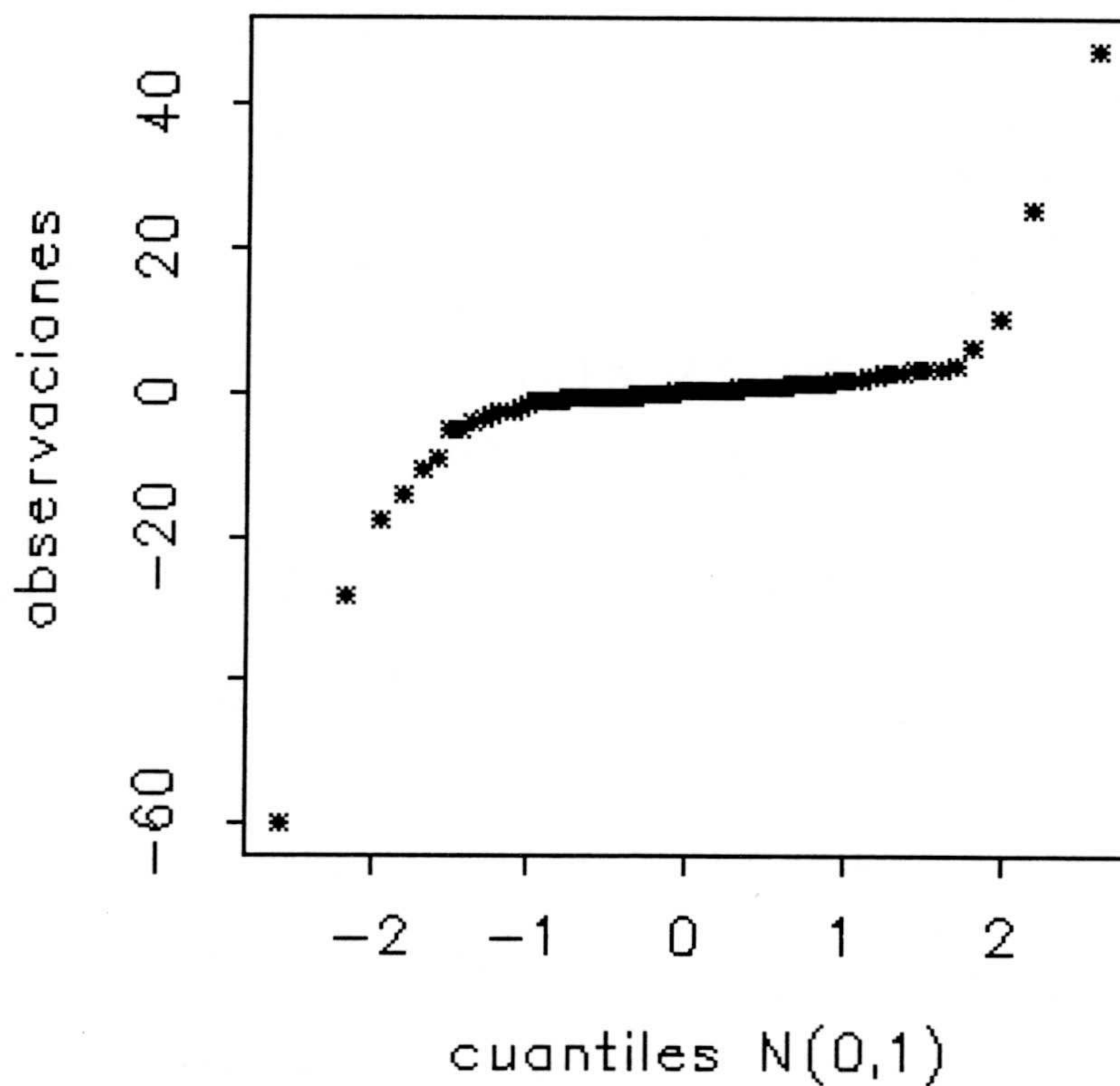


FIGURA 5.4: Distribución con colas más grandes que la normal.



### C.- ASIMETRIA.

Otra posibilidad de desacuerdo entre la distribución de los datos y la distribución teórica se manifiesta cuando ésta es simétrica y la de aquellos no lo es. Los gráficos de probabilidad permitirán también descubrir tal situación.

La evidencia de una asimetría a la derecha en la distribución de los datos bajo estudio se pondrá de relieve con una, más o menos acentuada, curvatura convexa en la disposición de los puntos en el gráfico de probabilidad. Dicho de otra manera, si nos movemos hacia el final del conjunto de observaciones ordenadas, encontramos que los cuantiles por encima de la mediana se encuentran más lejos de ésta que los situados antes de ella.

En el caso contrario, que se manifiesta con una curvatura cóncava en la disposición de los puntos en el gráfico de probabilidad, con análogos comentarios con respecto a la localización de los datos ordenados frente a la mediana, nos lleva a concluir que en la distribución empírica existe una, más o menos acentuada, asimetría a la izquierda.

En todo caso, la falta de simetría en la distribución de los datos, que será puesta de manifiesto en los gráficos de probabilidad, nos llevará a desaconsejar la hipótesis distribucional en el caso en que ésta recoja a una distribución teórica que sea simétrica (como es el caso en que el gráfico de probabilidad sea usado como prueba de normalidad).

En la figura 5.5 recogemos los casos anteriormente relatados, donde en la figura 5.5 (a) aparece una distribución con una marcada asimetría a la derecha, y en la figura 5.5 (b), con una asimetría a la izquierda.



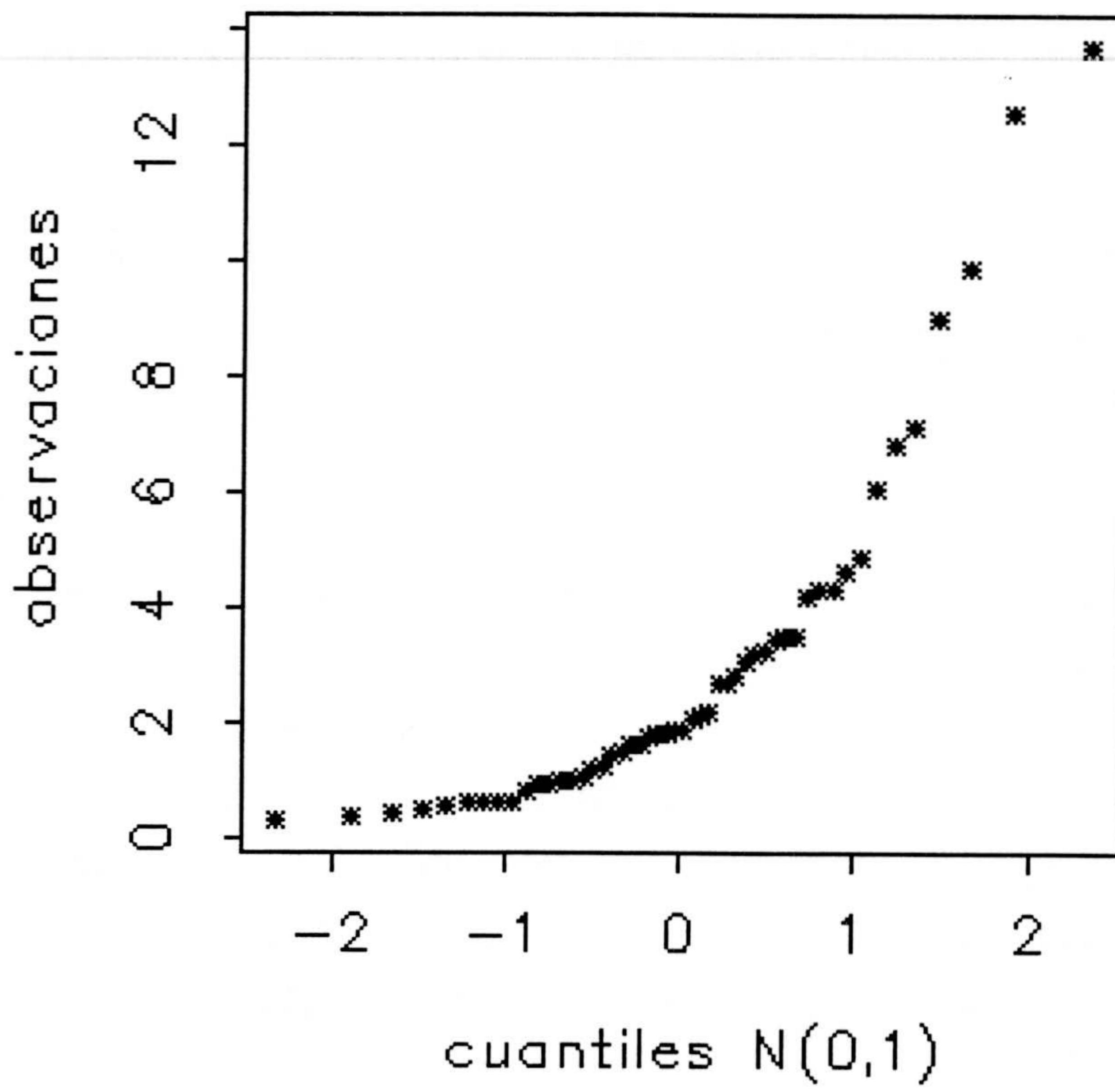


FIGURA 5.5 (a): Distribución con asimetría a la derecha.

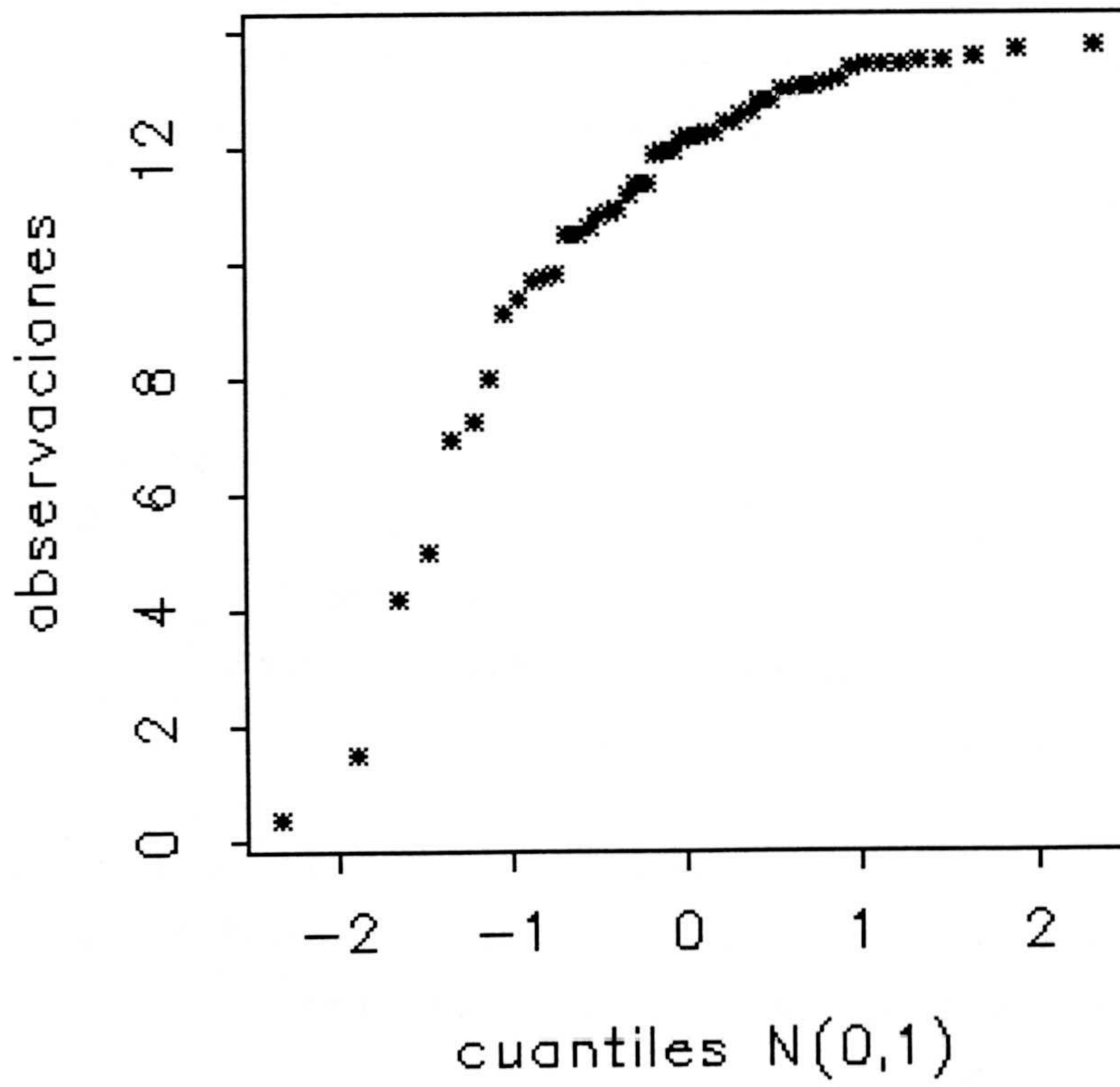


FIGURA 5.5 (b): Distribución con asimetría a la izquierda.



D.- EXISTENCIA DE SEGMENTOS HORIZONTALES, GRUPOS DE PUNTOS Y DISCONTINUIDADES.

Otras situaciones que nos llevarán a desechar las hipótesis distribucionales enunciadas sobre el conjunto de datos, se refieren a aquellas en las que encontramos en los gráficos de probabilidad situaciones sospechosas como ocurre al ver los datos dispuestos en segmentos horizontales (Figura 5.6) o puntos agrupados en dos o más bloques (Figura 5.7) o bien al observar claras discontinuidades (Figura 5.8).

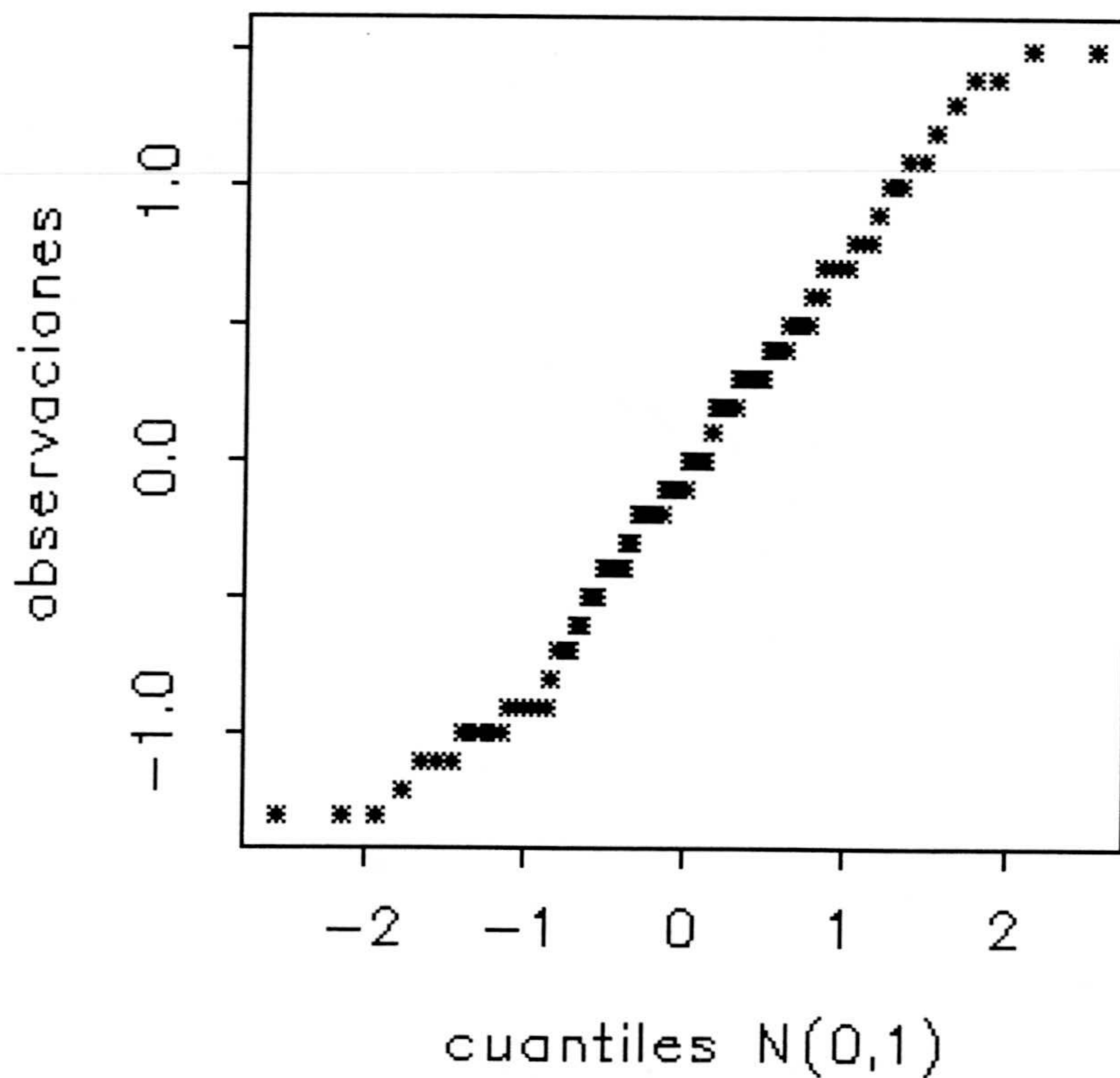


FIGURA 5.6: Gráfico probabilístico normal con observaciones dispuestas en segmentos horizontales



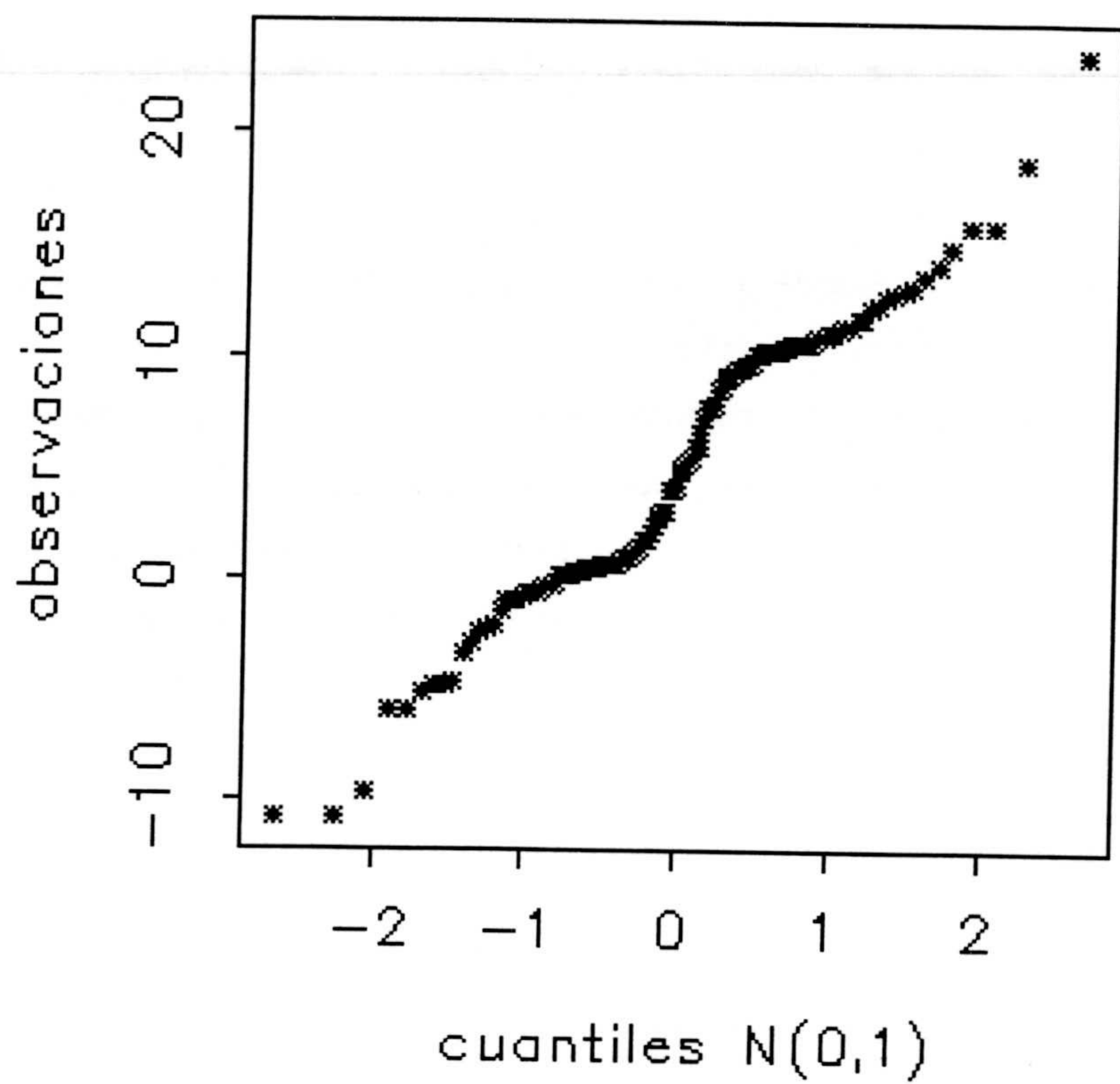


FIGURA 5.7: Gráfico probabilístico normal con dos bloques.

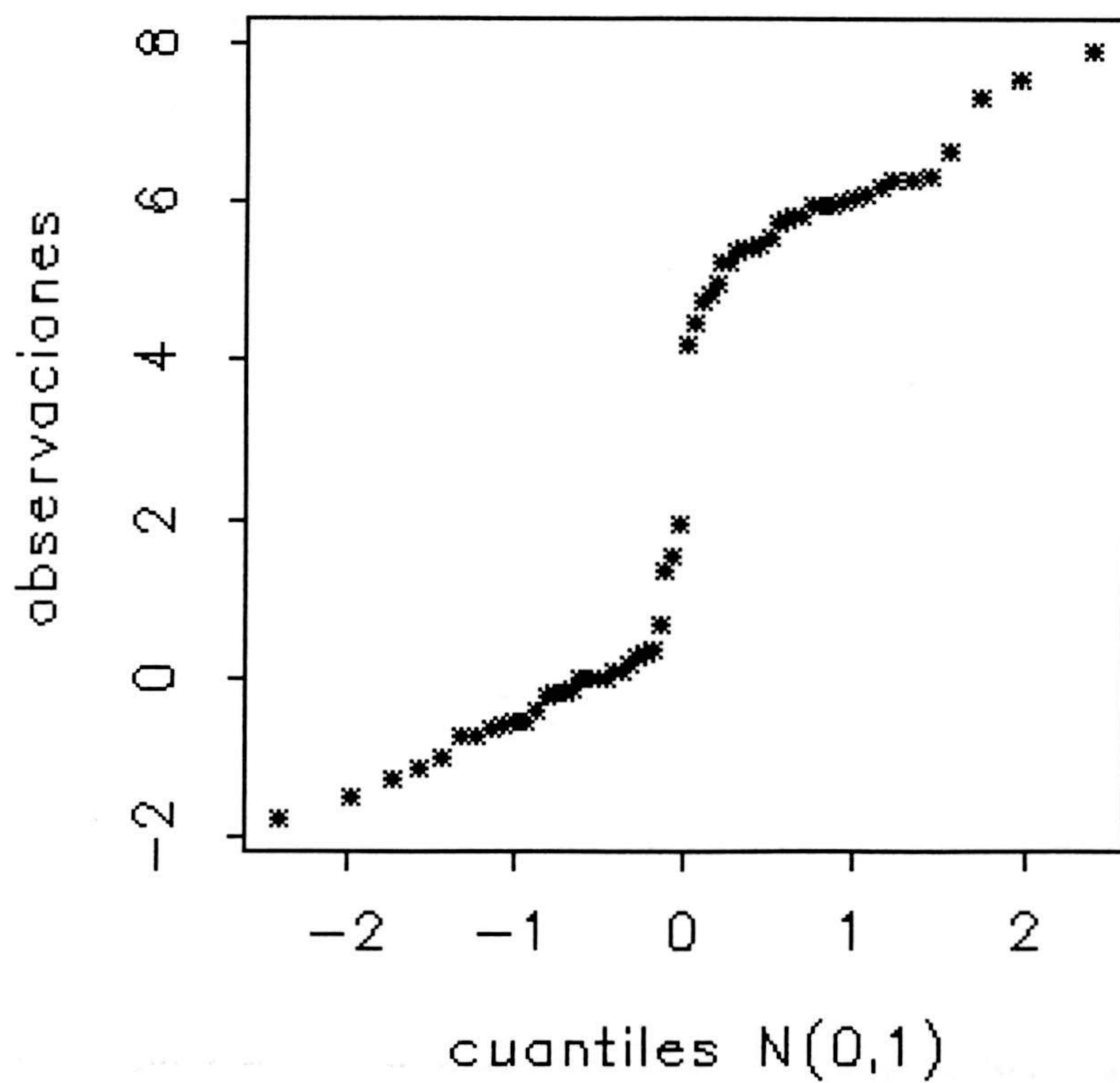


FIGURA 5.8: Gráfico probabilístico normal con discontinuidad.



Cuando, en particular, pretendemos probar la normalidad de los datos y nos encontramos ante alguna de las circunstancias anteriores, se evidencia un claro rechazo de tal hipótesis.

Algunas de las anteriores situaciones no son propias de la naturaleza de los datos, sino que son provocadas por posibles redondeos producidos en la toma de éstos. En tal caso, y con todas las precauciones posibles, la no compatibilidad con la distribución hipotética puede ser juzgada como carente de importancia debido a las actuaciones futuras previstas con el conjunto de datos.

#### **Ajuste de una recta al gráfico de probabilidad.**

Una vez que los puntos ya han sido dibujados en el gráfico de probabilidad, el paso siguiente será proceder a discutir la bondad del ajuste de los datos observados a la distribución hipotética. Esto, como ya ha sido indicado en reiteradas ocasiones, se hará atendiendo a la presentación "más o menos" rectilínea de los datos en el Q-Q plot. Con el objetivo de ayudar en tal discusión, se determina una recta que "mejor se ajuste" a los puntos en el gráfico de probabilidad. Además, el valor de la recta en el punto de abscisas 0, proporcionará una estimación de la media poblacional, y el valor de la recta en el punto de abscisas 1, nos dará el valor de la suma de la media y la desviación típica, lo que usaremos, junto con lo anterior, para producir una estimación de la desviación típica poblacional.

La determinación de la línea que mejor se ajusta a los puntos en el gráfico de probabilidad, se puede llevar a cabo mediante distintos procedimientos:

- 1º) El procedimiento más simple que podemos encontrar consiste en trazar la línea que mejor se ajusta a los puntos "a ojo". Este procedimiento tiene la ventaja de no consumir tiempo en cálculos.

22 SET. 1995

131  
COMISION DE DOCTORADO



los auxiliares, aunque posee la desventaja de producir estimadores "dudosos" de los parámetros poblacionales. Sin embargo, aconsejaremos este método cuando lo que pretendamos es observar de forma muy general el comportamiento distribucional de los datos, y estemos dispuestos a asumir, más que normalidad, proximidad a la normalidad.

- 2°) Una forma de formalizar el procedimiento de ajuste de la recta obtenida "a ojo", es localizar un punto en el gráfico correspondiente, aproximadamente, al décimo percentil ( $p_i \approx 0'1$ ) y otro próximo al percentil de orden 90 ( $p_i \approx 0'9$ ) y unir esos dos puntos. La recta así determinada será la que mejor se ajuste a los puntos en el gráfico (la elección del décimo percentil y del de orden 90, puede variar, y así elegir, por ejemplo, los percentiles de orden 20 y 80, respectivamente).
- 3°) Un tercer procedimiento consiste en tener en cuenta que los parámetros de intersección en el origen e inclinación de la recta, producirán estimadores de los parámetros poblacionales media y desviación típica. Dado que la media muestral,  $\bar{x}$ , y la desviación típica muestral,  $s$ , producen buenos estimadores de estos parámetros poblacionales, este método propone ajustar a los puntos del gráfico la recta

$$x = \bar{x} + s z$$

Este procedimiento posee la desventaja de no obtener estimadores de los parámetros poblacionales como consecuencia del ajuste de la recta, sino que estos serán determinados previamente, y usados para la determinación de aquella.

- 4°) Un cuarto procedimiento para obtener tanto la línea ajustada como los estimadores de los parámetros poblacionales, consiste en tener en cuenta que la línea deseada tendrá la forma



$$x = \mu + \sigma z$$

y la estimación de  $\mu$  y  $\sigma$  será obtenida usando mínimos cuadrados no ponderados (regresión lineal simple). La solución general en este caso consistirá en:

$$\tilde{\sigma} = \frac{\sum_i (z_i - \bar{z}) x_i}{\sum_i (z_i - \bar{z})^2}$$

y

$$\tilde{\mu} = \bar{x} - \tilde{\sigma} \bar{z}$$

que en nuestro caso (distribución normal), dado que  $\sum_i z_i = 0$ , se convierte en

$$\tilde{\mu} = \bar{x}$$

y

$$\tilde{\sigma} = \frac{\sum_i z_i x_i}{\sum_i z_i^2}$$

Así, la recta ajustada tendrá la forma

$$x = \tilde{\mu} + \tilde{\sigma} z$$

En este caso cabe el mismo comentario que el realizado en el procedimiento desarrollado en el apartado 3°).

En un comentario general que compare la bondad de los cuatro métodos propuestos, encontramos que cada uno de ellos posee sus ventajas e inconvenientes frente a los restantes. El primero, frente a los otros tres, presenta la ventaja de no consumir tiempo en cálculos para la determinación de los parámetros de la recta,



pero a su vez, esto puede ser considerado una desventaja, dado que producirá estimadores "dudosos" de los parámetros de localización y escala de la distribución teórica.

El segundo, al introducirse como una pequeña formalización del primero, poseerá las ventajas de éste, incrementadas, además, en el hecho de producir estimadores de los parámetros poblacionales que no se verán afectados por observaciones extremas. Sin embargo, esta ventaja puede ser, a su vez, un inconveniente en todos aquellos casos en los que los valores extremos sean calificados como fundamentales dentro del conjunto de todas las observaciones.

Los procedimientos propuestos en los apartados tercero y cuarto, producen, por lo general, buenas estimaciones de los parámetros poblacionales, aunque éstas pueden verse muy influidas por observaciones extremas. Además, estos dos últimos procedimientos presentan una clara independencia entre la estimación de los parámetros y la validación del modelo teórico con el uso del gráfico probabilístico normal.



## CAPITULO 6

### ANALISIS DE LA VARIABILIDAD EN EL GRAFICO

#### PROBABILISTICO NORMAL

Como ya indicábamos en el capítulo anterior, donde hemos introducido al gráfico probabilístico normal como técnica gráfica para el análisis de la posible normalidad de un conjunto de datos, existen una serie de fluctuaciones de los puntos del gráfico respecto de su comportamiento rectilíneo que están justificadas solamente por el carácter aleatorio de las observaciones. Por contra, pueden aparecer también otra serie de fluctuaciones, de tendencia sistemática, que nos informan sobre la no normalidad de las observaciones.

En el presente capítulo nos vamos a situar en el límite de los tipos de fluctuaciones mencionados anteriormente. Es decir, en aquellos casos donde no tengamos evidencia directa para calificar las fluctuaciones como sistemáticas y, sin embargo, el gráfico probabilístico normal nos haga sospechar de cierta tendencia no normal de los datos. O dicho de otra forma, pretendemos responder a la pregunta ¿son las fluctuaciones en el gráfico probabilístico normal consecuencia del carácter aleatorio de las observaciones, o, por el contrario, ponen de relieve una ausencia de normalidad en estos?.

Pretendemos introducir una metodología que responda a la



pregunta anterior, de manera que, la posible aceptación de la linealidad de los puntos en el gráfico de probabilidad normal, no se vea afectada por la subjetividad del observador. Con ello, conseguiremos que dos observadores enfrentados ante el mismo conjunto de datos, y haciendo uso de los gráficos de probabilidad, concluyan siempre en idéntica manera respecto de la normalidad de éstos.

Propondremos, por tanto, una "formalización" en el uso de los gráficos de probabilidad encaminados al análisis de la normalidad de un conjunto de observaciones, que, entre otras, nos permita recalificar a éstos dentro del conjunto de técnicas gráficas para el estudio de normalidad, donde eliminaremos el calificativo de "informales" que estamos acostumbrados a encontrar en la bibliografía al uso.

Para realizar un análisis de la variabilidad que estamos dispuestos a admitir en los gráficos de probabilidad para calificar las posibles fluctuaciones respecto de la recta como simple consecuencia de la aleatoriedad de las observaciones y, por tanto, aceptar la hipótesis de normalidad de los datos, estudiaremos la variabilidad de los puntos en el gráfico probabilístico normal respecto de la recta ajustada. Proponemos tres procedimientos para llevar a cabo este estudio: el primero consiste en analizar la variabilidad basándonos en la varianza de los estadísticos de orden de una distribución normal. El segundo, hará uso de la distribución de los estadísticos de orden de una distribución normal. Y, por último, el tercero, tendrá en cuenta el comportamiento asintótico de los estadísticos de orden de una distribución normal. En éste último epígrafe introduciremos un criterio, que estimamos oportuno, para clasificar los estadísticos de orden en las categorías de estadísticos de orden extremos, intermedios y centrales.



**ANALISIS DE LA VARIABILIDAD EN EL GRAFICO PROBABILISTICO NORMAL  
A PARTIR DE LA VARIANZA DE LOS ESTADISTICOS DE ORDEN.**

Un primer camino que proponemos para medir la variabilidad admisible en los gráficos de probabilidad es, para cada punto, dibujar, verticalmente, un intervalo centrado en el correspondiente punto de la recta ajustada, con radio una desviación típica. Sin embargo, este procedimiento presenta ciertas dificultades técnicas. Para calcular la mencionada desviación típica hay que considerar que los puntos dibujados tienen como coordenadas a estadísticos de orden, y la desviación típica de los estadísticos de orden depende de la distribución poblacional, que debe, por tanto, ser asumida antes de llevar a cabo este procedimiento.

Bajo la hipótesis de normalidad, tenderemos entonces que calcular la desviación típica de los estadísticos de orden procedentes de tal distribución.

En el apéndice I referido a los estadísticos de orden hemos establecido que, en general, la varianza del  $i$ -ésimo estadístico de orden tiene la expresión

$$\sigma_{(i)}^2 = \mu_{(i)}^{(2)} - \mu_{(i)}^2 \quad i = 1, 2, \dots, n$$

donde

$$\mu_{(i)}^{(m)} = \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x^m [F(x)]^{i-1} [1-F(x)]^{n-i} f(x) dx$$

con la notación  $\mu_{(i)} = \mu_{(i)}^{(1)}$ , y siendo  $F(x)$  y  $f(x)$  la función de distribución y la función de densidad, respectivamente, de la variable bajo estudio.

Haciendo uso de resultados establecidos en el citado



apéndice, podemos escribir

$$\mu_{(i)}^{(m)} = \frac{n!}{(i-1)!(n-i)!} \int_0^1 [F^{-1}(u)]^m u^{i-1} (1-u)^{n-i} du$$

en donde, en nuestro caso en particular, supuesta la normalidad,

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2} dt$$

Nos encontramos a estas alturas con un gran inconveniente: el cálculo de  $\mu_{(i)}^{(2)}$  y  $\mu_{(i)}$  necesario para determinar la desviación típica de cada estadístico de orden a partir de la expresión general de  $\mu_{(i)}^{(m)}$ , hace presagiar un cálculo complicado, tanto más si tenemos en cuenta que al ser  $F(x) = \Phi(x)$ , tendremos que usar  $\Phi^{-1}$ , y  $\Phi^{-1}$  no tiene una expresión conocida.

Por todo lo anterior, vamos a optar por realizar el cálculo de la varianza de los estadísticos de orden mediante el procedimiento que desarrollamos a continuación.

Como apreciamos en el apéndice I,

$$X_{(i)} = F^{-1}(U_{(i)})$$

donde  $X_{(i)}$  es el  $i$ -ésimo estadístico de orden de una población distribuida según una función de distribución  $F$ , y  $U_{(i)}$  es el  $i$ -ésimo estadístico de orden de una distribución uniforme estándar.

Desarrollando  $F^{-1}(U_{(i)})$  en serie de Taylor alrededor del punto  $E[U_{(i)}] = i/(n+1) = e_i$ , obtenemos, a la vez, un desarrollo de  $X_{(i)}$  con la expresión



$$X_{(i)} = F^{-1}(e_i) + F^{-1(1)}(e_i)(U_{(i)} - e_i) + \frac{1}{2}F^{-1(2)}(e_i)(U_{(i)} - e_i)^2 + \\ + \frac{1}{6}F^{-1(3)}(e_i)(U_{(i)} - e_i)^3 + \frac{1}{24}F^{-1(4)}(U_{(i)} - e_i)^4 + \dots$$

donde por  $F^{-1(1)}(e_i), F^{-1(2)}(e_i), F^{-1(3)}(e_i), F^{-1(4)}(e_i), \dots$ , notamos a las derivadas sucesivas de  $F^{-1}(u)$  evaluadas en el punto  $u = e_i$ .

Si tomamos esperanzas en ambos miembros, teniendo en cuenta la expresión de los momentos de los estadísticos de orden de una distribución uniforme estándar, y reordenando en potencias inversas de  $(n+2)$  [David y Johnson (1954)], podemos escribir:

$$\mu_{(i)} \cong F^{-1}(e_i) + \frac{e_i d_i}{2(n+2)} F^{-1(2)}(e_i) + \\ + \frac{e_i d_i}{(n+2)^2} \left[ \frac{1}{3} (d_i - e_i) F^{-1(3)}(e_i) + \frac{1}{8} e_i d_i F^{-1(4)}(e_i) \right] + \\ + \frac{e_i d_i}{(n+2)^3} \left[ -\frac{1}{3} (d_i - e_i) F^{-1(3)}(e_i) + \frac{1}{4} \{ (d_i - e_i)^2 - d_i e_i \} F^{-1(4)}(e_i) \right. \\ \left. + \frac{1}{6} e_i d_i (d_i - e_i) F^{-1(5)}(e_i) + \frac{1}{48} e_i^2 d_i^2 F^{-1(6)}(e_i) \right]$$

donde  $d_i = (1 - e_i) = \frac{n-i+1}{n+1}$ .

De forma similar, tomando esperanzas en los dos miembros del desarrollo análogo de  $X_{(i)}^2$ , y restándole a esta expresión el valor aproximado de  $\mu_{(i)}^2$ , obtendremos un valor aproximado de la varianza del  $i$ -ésimo estadístico de orden  $X_{(i)}$ , recogido en la expresión



$$\begin{aligned} \sigma_{(i)}^2 \cong & \frac{e_i d_i}{n+2} \{F^{-1(1)}(e_i)\}^2 + \frac{e_i d_i}{(n+2)^2} \left[ 2(d_i - e_i) F^{-1(1)}(e_i) F^{-1(2)}(e_i) \right. \\ & \left. + e_i d_i \left[ F^{-1(1)}(e_i) F^{-1(3)}(e_i) + \frac{1}{2} \{F^{-1(2)}(e_i)\}^2 \right] \right] \\ & + \frac{e_i d_i}{(n+2)^3} \left[ -2(d_i - e_i) F^{-1(1)}(e_i) F^{-1(2)}(e_i) \right. \\ & \left. + \{(d_i - e_i)^2 - e_i d_i\} \left[ 2F^{-1(1)}(e_i) F^{-1(3)}(e_i) + \frac{3}{2} \{F^{-1(2)}(e_i)\}^2 \right] \right] \\ & + e_i d_i (d_i - e_i) \left\{ \frac{5}{3} F^{-1(1)}(e_i) F^{-1(4)}(e_i) + 3F^{-1(2)}(e_i) F^{-1(3)}(e_i) \right\} \\ & + \frac{1}{4} e_i^2 d_i^2 \left[ F^{-1(1)}(e_i) F^{-1(5)}(e_i) + 2F^{-1(2)}(e_i) F^{-1(4)}(e_i) \right. \\ & \left. + \frac{5}{3} \{F^{-1(3)}(e_i)\}^2 \right] \end{aligned}$$

En esta expresión aparecen las derivadas sucesivas de  $F^{-1}(u)$ , por lo que, para su cálculo tengamos en cuenta que

$$F^{-1(1)}(u) = \frac{d}{du} F^{-1}(u) = \frac{dx}{du} = \frac{1}{(du/dx)} = \frac{1}{f(x)} = \frac{1}{f[F^{-1}(u)]} \quad (6.1)$$

de manera que las siguientes derivadas se obtendrán derivando esta expresión.

Dado que nos encontramos en el caso particular en que  $F(x) = \Phi(x)$ , con  $\Phi$  la función de distribución de una normal de media  $\mu$  y desviación típica  $\sigma$ , la expresión que aproxima el valor de la varianza del  $i$ -ésimo estadístico de orden adoptará la forma



$$\begin{aligned}
 \sigma_{(i)}^2 \cong & \frac{e_i d_i}{n+2} \{\Phi^{-1(1)}(e_i)\}^2 + \frac{e_i d_i}{(n+2)^2} \left[ 2(d_i - e_i) \Phi^{-1(1)}(e_i) \Phi^{-1(2)}(e_i) \right. \\
 & + e_i d_i \left[ \Phi^{-1(1)}(e_i) \Phi^{-1(3)}(e_i) + \frac{1}{2} \{\Phi^{-1(2)}(e_i)\}^2 \right] \\
 & + \frac{e_i d_i}{(n+2)^3} \left[ -2(d_i - e_i) \Phi^{-1(1)}(e_i) \Phi^{-1(2)}(e_i) \right. \\
 & + \{(d_i - e_i)^2 - e_i d_i\} \left[ 2\Phi^{-1(1)}(e_i) \Phi^{-1(3)}(e_i) + \frac{3}{2} \{\Phi^{-1(2)}(e_i)\}^2 \right] \\
 & + e_i d_i (d_i - e_i) \left\{ \frac{5}{3} \Phi^{-1(1)}(e_i) \Phi^{-1(4)}(e_i) + 3\Phi^{-1(2)}(e_i) \Phi^{-1(3)}(e_i) \right\} \\
 & + \frac{1}{4} e_i^2 d_i^2 \left[ \Phi^{-1(1)}(e_i) \Phi^{-1(5)}(e_i) + 2\Phi^{-1(2)}(e_i) \Phi^{-1(4)}(e_i) \right. \\
 & \left. + \frac{5}{3} \{\Phi^{-1(3)}(e_i)\}^2 \right] \tag{6.2}
 \end{aligned}$$

Si hacemos uso de la relación

$$\frac{d}{dx} \phi(x) = -\left( \frac{x-\mu}{\sigma} \right) \phi(x)$$

donde por  $\phi(x)$  representamos a la función de densidad de la anterior distribución normal, y teniendo en cuenta (6.1), podemos escribir:

$$\Phi^{-1(1)}(u) = \frac{1}{\phi[\Phi^{-1}(u)]}$$

$$\Phi^{-1(2)}(u) = \frac{\frac{\Phi^{-1}(u) - \mu}{\sigma}}{\{\phi[\Phi^{-1}(u)]\}^2}$$



$$\Phi^{-1(3)}(u) = \frac{1 + 2 \left\{ \frac{\Phi^{-1}(u) - \mu}{\sigma} \right\}^2}{\{\phi[\Phi^{-1}(u)]\}^3}$$

$$\Phi^{-1(4)}(u) = \frac{\left\{ \frac{\Phi^{-1}(u) - \mu}{\sigma} \right\} \left[ 7 + 6 \left\{ \frac{\Phi^{-1}(u) - \mu}{\sigma} \right\}^2 \right]}{\{\phi[\Phi^{-1}(u)]\}^4}$$

$$\Phi^{-1(5)}(u) = \frac{7 + 46 \left\{ \frac{\Phi^{-1}(u) - \mu}{\sigma} \right\}^2 + 24 \left\{ \frac{\Phi^{-1}(u) - \mu}{\sigma} \right\}^4}{\{\phi[\Phi^{-1}(u)]\}^5}$$

Sustituyendo estos valores en (6.2), podemos reescribir el valor aproximado de la varianza del  $i$ -ésimo estadístico de orden de una distribución normal de parámetros  $\mu$  y  $\sigma$ , como:

$$\begin{aligned} \sigma_{(i)}^2 \cong & \frac{e_i d_i}{n+2} \frac{1}{\{\phi[\Phi^{-1}(e_i)]\}^2} + \\ & + \frac{e_i d_i}{(n+2)^2} \left[ 2(d_i - e_i) \frac{\left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\}}{\{\phi[\Phi^{-1}(e_i)]\}^3} + e_i d_i \frac{1 + \frac{5}{2} \left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\}^2}{\{\phi[\Phi^{-1}(e_i)]\}^4} \right] \\ & + \frac{e_i d_i}{(n+2)^3} \left[ -2(d_i - e_i) \frac{\left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\}}{\{\phi[\Phi^{-1}(e_i)]\}^3} + 2\{(d_i - e_i)^2 - e_i d_i\} \frac{1 + \frac{11}{4} \left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\}^2}{\{\phi[\Phi^{-1}(e_i)]\}^4} \right] \\ & + \frac{1}{3} e_i d_i (d_i - e_i) \frac{44 \left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\} + 48 \left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\}^2}{\{\phi[\Phi^{-1}(e_i)]\}^5} \end{aligned}$$



$$+ \frac{1}{12} e_i^2 d_i^2 \left[ \frac{26+200 \left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\}^2 + 128 \left\{ \frac{\Phi^{-1}(e_i) - \mu}{\sigma} \right\}^4}{\{\phi[\Phi^{-1}(e_i)]\}^6} \right]$$

Esta aproximación, aunque pueda parecer algo intratable, constituye una sencilla expresión desde el punto de vista computacional, por lo que será usada para determinar el valor aproximado de la varianza del  $i$ -ésimo estadístico de orden en una distribución normal de parámetros  $\mu$  y  $\sigma$ .

Por tanto, el procedimiento a seguir para dibujar en el gráfico probabilístico normal la variabilidad admitida basada en la varianza de los estadísticos de orden, lo recogemos en las siguientes etapas:

- 1°) Supuesto que se dispone de las observaciones  $x_i$ , realizar el correspondiente gráfico de probabilidad normal.
- 2°) Ajustar una recta al gráfico probabilístico normal. Para ello podrá usarse cualquiera de los métodos anteriormente introducidos, con las observaciones que en cada uno de ellos hemos realizado.
- 3°) La recta ajustada en el gráfico de probabilidad nos proporciona una estimación de los parámetros  $\mu$  y  $\sigma$  de la distribución normal bajo la hipótesis nula (como ya se ha indicado serán los puntos de intersección e inclinación de la recta ajustada)
- 4°) Determinar, para cada  $i$  ( $1 \leq i \leq n$ ), el valor aproximado de la desviación típica del  $i$ -ésimo estadístico de orden usando la expresión aportada anteriormente. Para su cálculo, usar, allí donde es necesario, los valores estimados de la media y la desviación típica poblacional a los que nos hemos referido en



el apartado anterior.

5°) Para cada  $i$ , dibujar verticalmente un intervalo centrado en el correspondiente punto de la recta ajustada, con radio una vez la desviación típica del  $i$ -ésimo estadístico de orden.

Una vez concluido este procedimiento, daremos lugar a un gráfico como el que se aporta en la figura 6.1.

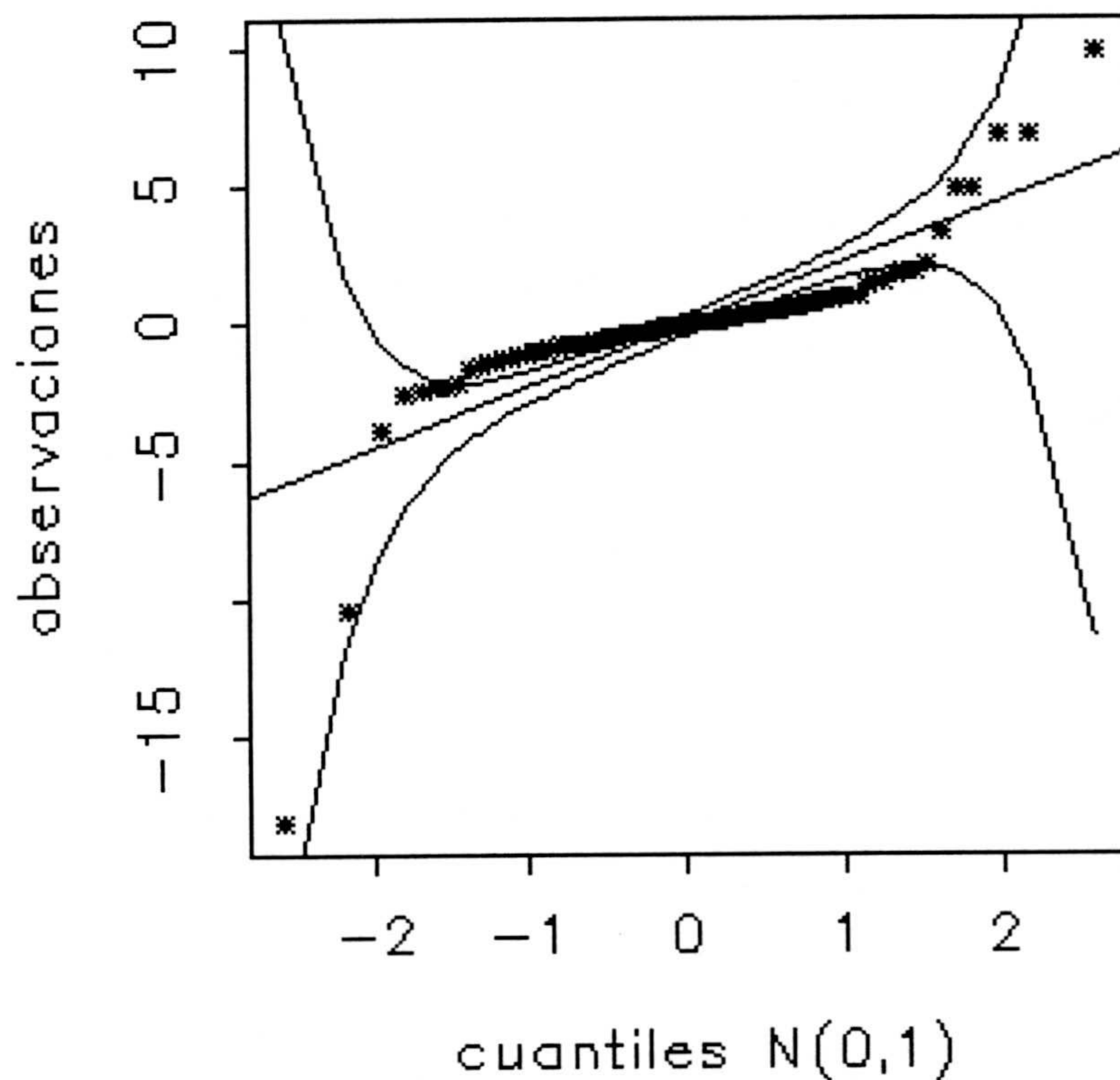


FIGURA 6.1: Región de variabilidad admitida haciendo uso de la varianza de los estadísticos de orden.

Aunque el método así introducido presenta la ventaja de ser sencillo de realizar, encontramos en el algunos inconvenientes. En primer lugar, indiquemos que como en la expresión aproximada de la varianza del  $i$ -ésimo estadístico de orden aparecen los



parámetros poblacionales  $\mu$  y  $\sigma$ , deberemos hacer uso de sus valores estimados. Estos valores los determinaremos, como ha quedado dicho, a partir de la recta ajustada, y por tanto, variarán en función del procedimiento que sigamos para determinarla.

En segundo lugar, la varianza de los estadísticos de orden de una distribución normal es tanto más grande conforme el correspondiente estadístico sea más extremo. Esto es, la varianza es más grande para los primeros y los últimos estadísticos de orden, por lo que las bandas determinadas en función de éstas presentarán una región más grande en los extremos del gráfico de probabilidad (véase la figura 6.1), cerrándose en el centro del gráfico. Esto nos llevará en muchas ocasiones a recoger dentro de las bandas de variabilidad a puntos que en otras condiciones deberían ser calificados como outliers. Este inconveniente se ve acentuado cuando el método seguido para la determinación de la recta ajustada, y por tanto para la determinación de los estimadores de los parámetros  $\mu$  y  $\sigma$ , se encuentra muy afectado por las observaciones extremas, por lo que sería recomendable hacer uso de alguno de los dos primeros métodos introducidos en el capítulo anterior para determinar la recta que mejor se ajusta al gráfico de probabilidad normal (es decir, bien "a ojo", bien uniendo los puntos del gráfico de probabilidad correspondientes a cuantiles de orden prefijado), con las deficiencias que ello pueda llevar en lo que se refiere a las estimaciones de los parámetros poblacionales.

Por todo lo anterior, entendemos que la no aceptación de la hipótesis de normalidad haciendo uso de estas regiones introducidas en el gráfico deberá basarse más en la observación del comportamiento en el centro del gráfico que en lo que ocurra en sus extremos.

Para la obtención de estos gráficos hemos desarrollado una aplicación en S-PLUS que recogemos en el apéndice II.



Por último, y para acabar con los comentarios sobre el estudio de la variabilidad en los gráficos de probabilidad normal basado en la varianza de los estadísticos de orden, indiquemos que Chambers, Cleveland, Kleiner y Tukey (1983) introducen un método análogo al que nosotros hemos usado, con una gran diferencia producida por la expresión que usan en el cálculo aproximado de la varianza de los estadísticos de orden. Ellos proponen usar como valor aproximado de la varianza, el desarrollado por Kendall y Stuart (1977), que puede obtenerse también a partir del desarrollo de  $X_{(i)}$ , pero cortando en el primer término. (El razonamiento seguido por Kendall y Stuart difiere ampliamente del que nosotros acabamos de señalar pero, dado que también puede obtenerse de esta manera, nos parece más conveniente hacer el comentario en este sentido para poder comparar la precisión de ambas aproximaciones al valor de la varianza de los estadísticos de orden). Entendemos que el procedimiento que aquí se propone, es mejor que el propuesto por Chambers, Cleveland, Kleiner y Tukey (1983), ya que aunque ambos se basen en un cálculo aproximado de la varianza del  $i$ -ésimo estadístico de orden, la aproximación que nosotros aportamos es más precisa. (El cortar el desarrollo de  $X_{(i)}$  en el primer término es usado en este trabajo, por conveniencia, para obtener, en el apéndice I, la distribución asintótica de los estadísticos de orden centrales).

**ANALISIS DE LA VARIABILIDAD EN EL GRAFICO PROBABILISTICO NORMAL  
A PARTIR DE LA DISTRIBUCION EXACTA DE LOS ESTADISTICOS DE ORDEN.**

Siguiendo la idea ya introducida anteriormente al dibujar una "banda de confianza" que nos delimite la región aceptable de variabilidad de los puntos en el gráfico probabilístico normal, proponemos a continuación un procedimiento que entendemos guardará más precisión que el anterior a la hora de detectar fluctuaciones que nos induzcan a aceptar o rechazar la hipótesis de normalidad



enunciada sobre el conjunto de observaciones.

La base de este procedimiento, coincidiendo con el anterior, consistirá en, una vez que en el gráfico probabilístico normal se ha dibujado la recta que mejor se ajusta a los puntos, determinar sobre ella una "banda de confianza" que nos delimite la región donde pueden moverse los puntos para concluir si las fluctuaciones sobre la recta son mera consecuencia de la aleatoriedad de las observaciones, de forma que si un cierto número de puntos cayera fuera de esta región, pondremos en duda la cusa de tales fluctuaciones y, por tanto, la hipótesis de normalidad. Además, tal y como ocurre con un contraste formal, pretendemos que la determinación de esa "región de confianza" dependa de un "nivel de confianza",  $\alpha$ , de manera que aceptemos o rechacemos la hipótesis de normalidad a dicho nivel  $\alpha$ .

Para el propósito anterior proponemos hacer uso de la distribución de los estadísticos de orden. Como recogemos en el apéndice I, la función de distribución del  $i$ -ésimo estadístico de orden adopta la expresión

$$\begin{aligned} \Phi_i(x) &= \sum_{m=i}^n \frac{n!}{m!(n-m)!} [F(x)]^m [1-F(x)]^{n-m} = \\ &= \frac{n!}{(i-1)!(n-i)!} \int_0^{F(x)} t^{i-1} (1-t)^{n-i} dt \end{aligned}$$

donde por  $F(x)$  representamos a la función de distribución de la población bajo estudio. Supuesta la hipótesis de normalidad, es decir,  $X_{(i)}$  es el  $i$ -ésimo estadístico de orden de una distribución normal de media  $\mu$  y desviación típica  $\sigma$ , la expresión anterior adoptará la forma



$$\begin{aligned} \Phi_i(x) &= \sum_{m=i}^n \frac{n!}{m!(n-m)!} [\Phi(x)]^m [1-\Phi(x)]^{n-m} = \\ &= \frac{n!}{(i-1)!(n-i)!} \int_0^{\Phi(x)} t^{i-1} (1-t)^{n-i} dt \end{aligned}$$

donde ahora

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2} dt$$

Por tanto, fijado un nivel de confianza  $\alpha$ , y para cada  $i$  ( $1 \leq i \leq n$ ), podemos determinar sendos puntos porcentuales,  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$ , asociados a la distribución de  $X_{(i)}$ , tales que,

$$P \left[ p_1^{(i)}(\alpha) \leq X_{(i)} \leq p_2^{(i)}(\alpha) \right] = 1 - \alpha$$

Dado que  $\Phi_i(x)$  es la función de distribución de  $X_{(i)}$ , podemos escribir

$$P \left[ p_1^{(i)}(\alpha) \leq X_{(i)} \leq p_2^{(i)}(\alpha) \right] = \Phi_i(p_2^{(i)}(\alpha)) - \Phi_i(p_1^{(i)}(\alpha))$$

y haciendo uso de la expresión de  $\Phi_i$ , bajo normalidad,



$$\begin{aligned} \Phi_i(p_2^{(i)}(\alpha)) - \Phi_i(p_1^{(i)}(\alpha)) &= \frac{n!}{(i-1)!(n-i)!} \int_0^{\Phi(p_2^{(i)}(\alpha))} t^{i-1}(1-t)^{n-i} dt - \\ &- \frac{n!}{(i-1)!(n-i)!} \int_0^{\Phi(p_1^{(i)}(\alpha))} t^{i-1}(1-t)^{n-i} dt = \\ &= \frac{n!}{(i-1)!(n-i)!} \int_{\Phi(p_1^{(i)}(\alpha))}^{\Phi(p_2^{(i)}(\alpha))} t^{i-1}(1-t)^{n-i} dt \end{aligned}$$

Por lo que

$$P \left[ p_1^{(i)}(\alpha) \leq X_{(i)} \leq p_2^{(i)}(\alpha) \right] = 1 - \alpha$$

es equivalente a

$$\frac{n!}{(i-1)!(n-i)!} \int_{\Phi(p_1^{(i)}(\alpha))}^{\Phi(p_2^{(i)}(\alpha))} t^{i-1}(1-t)^{n-i} dt = 1 - \alpha$$

Si tenemos en cuenta la función beta ( $B(\alpha, \beta)$ ), y de ella la propiedad

$$B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)},$$

la igualdad anterior se puede escribir como



$$\frac{1}{B(i, n-i+1)} \int_{\Phi(p_1^{(i)}(\alpha))}^{\Phi(p_2^{(i)}(\alpha))} t^{i-1}(1-t)^{n-i} dt = 1 - \alpha$$

donde, si observamos, en el primer miembro recogemos la probabilidad asociada a una distribución Beta de parámetros  $i$  y  $(n-i+1)$  entre los puntos  $\Phi(p_1^{(i)}(\alpha))$  y  $\Phi(p_2^{(i)}(\alpha))$ , o dicho de forma más conveniente para nuestro propósito,  $\Phi(p_1^{(i)}(\alpha))$  y  $\Phi(p_2^{(i)}(\alpha))$  son los puntos asociados a una variable distribuida según una Beta de parámetros  $i$  y  $n-i+1$ , que encierran una probabilidad de  $(1-\alpha)$ .

La discusión sobre la determinación de  $\Phi(p_1^{(i)}(\alpha))$  y  $\Phi(p_2^{(i)}(\alpha))$  puede plantearse de forma análoga a como se obtienen los extremos de un intervalo de confianza. Es decir, con las consideraciones finales hechas, queremos determinar  $\Phi(p_1^{(i)}(\alpha))$  y  $\Phi(p_2^{(i)}(\alpha))$  de forma que el intervalo  $[\Phi(p_1^{(i)}(\alpha)) ; \Phi(p_2^{(i)}(\alpha))]$  tenga una probabilidad de  $(1-\alpha)$ , y, a su vez, sea de longitud mínima.

En la estimación por intervalos, siempre que la distribución de la variable pivote sea simétrica y unimodal, la solución a la determinación de los extremos del intervalo se obtiene tomando éstos de forma simétrica. En este ambiente, y aún en el caso de ausencia de simetría, por simplicidad, también se toman los extremos de los intervalos de forma simétrica.

En nuestro caso, aunque no podemos garantizar la simetría de la distribución de la que dependen nuestros extremos, también por simplicidad, los tomaremos forma simétrica. De esta manera,  $\Phi(p_1^{(i)}(\alpha))$  será el valor asociado a una variable Beta( $i, n-i+1$ ) que deja por debajo una probabilidad de  $(\alpha/2)$ , y  $\Phi(p_2^{(i)}(\alpha))$  será, igualmente, el valor asociado a una variable Beta( $i, n-i+1$ ) que deja por encima una probabilidad de  $(\alpha/2)$ . O, dicho de otra mane-



ra,  $\Phi(p_1^{(i)}(\alpha))$  y  $\Phi(p_2^{(i)}(\alpha))$  serán los cuantiles de orden  $(\alpha/2)$  y  $(1-\alpha/2)$ , respectivamente, asociados a una distribución Beta( $i, n-i+1$ ).

Una vez determinados los cuantiles anteriores, bastará con calcular el valor de  $\Phi^{-1}$ , en esos puntos para hallar  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$  que usaremos para dibujar nuestra "banda de confianza".

En resumen, el procedimiento a seguir que proponemos para analizar la variabilidad admisible en el gráfico de probabilidad normal, haciendo uso de la distribución de los estadísticos de orden de una distribución normal, se puede ordenar en las siguientes etapas:

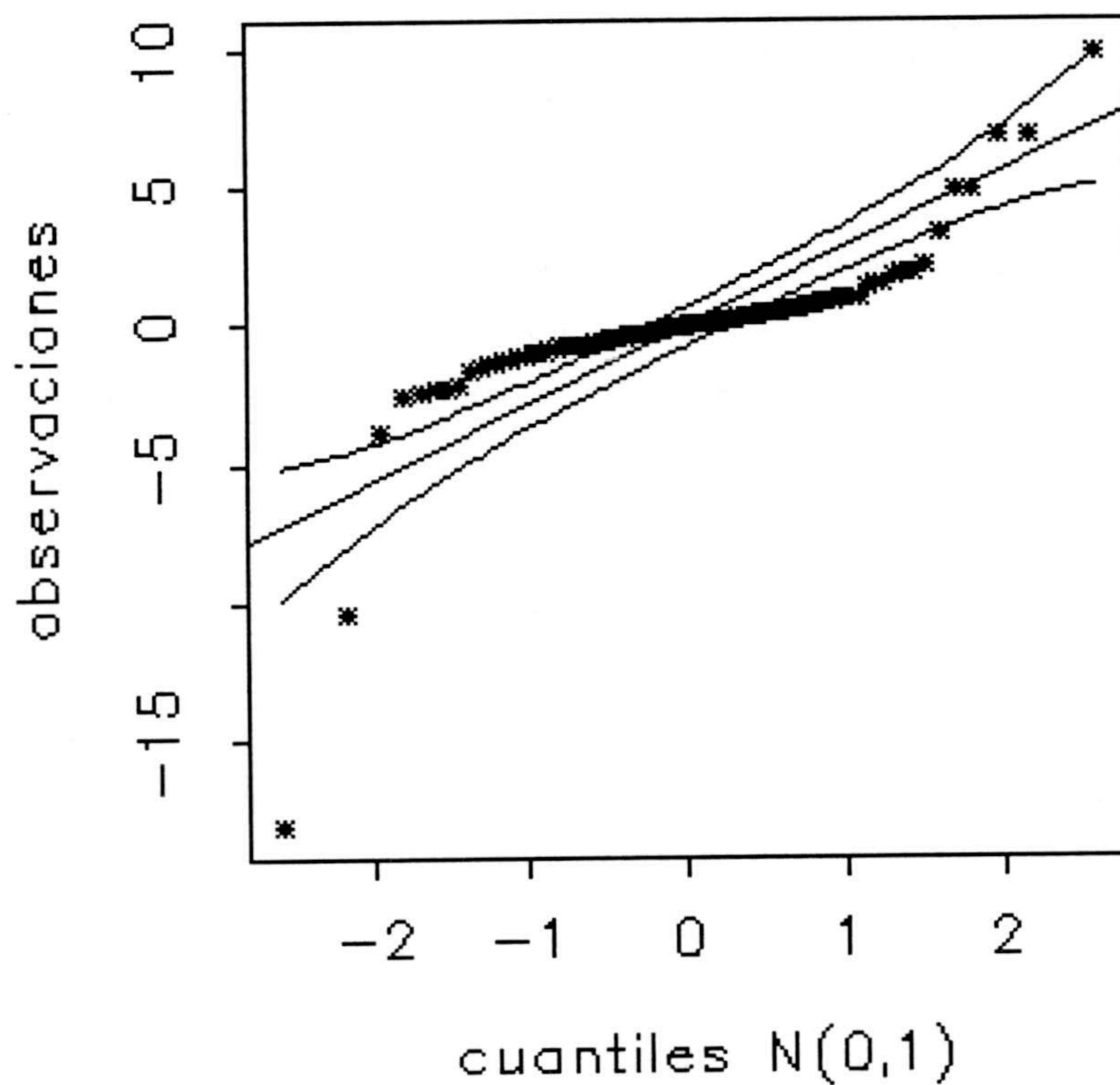
- 1°) Supongamos que disponemos de las observaciones  $x_i$ . Fijemos el nivel de confianza  $\alpha$ .
- 2°) Ajustar una recta al gráfico probabilístico normal. Para ello podrá usarse cualquiera de los métodos anteriormente introducidos, con las observaciones que en cada uno de ellos hemos realizado.
- 3°) La recta ajustada en el gráfico de probabilidad nos proporciona una estimación de los parámetros  $\mu$  y  $\sigma$  de la distribución normal bajo la hipótesis nula (como ya se ha indicado serán los puntos de intersección e inclinación de la recta ajustada)
- 4°) Determinar, para cada  $i$ , ( $1 \leq i \leq n$ ), los valores  $\Phi(p_1^{(i)}(\alpha))$  y  $\Phi(p_2^{(i)}(\alpha))$  definidos como los cuantiles de orden  $(\alpha/2)$  y  $(1-\alpha/2)$  de una distribución Beta( $i, n-i+1$ ).
- 5°) Para cada  $i$ , determinar los valores  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$  obtenidos como el valor de  $\Phi^{-1}$  en los cuantiles calculados en el apartado anterior, y donde  $\Phi$  será la función de distribución de una normal de parámetros  $\mu$  y  $\sigma$  enunciada en la hipótesis



nula. (Para el valor de los parámetros  $\mu$  y  $\sigma$  haremos uso de las estimaciones obtenidas en el apartado 3°)).

6°) Para cada  $i$ , dibujar verticalmente un intervalo centrado en el correspondiente punto de la recta ajustada, con extremo inferior igual a  $p_1^{(i)}(\alpha)$  y extremo superior igual a  $p_2^{(i)}(\alpha)$ .

Una vez concluido este procedimiento, daremos lugar a un gráfico como el que aparece en la figura 6.2.



**FIGURA 6.2:** Banda de confianza en el gráfico probabilístico normal, haciendo uso de la distribución de los estadísticos de orden.



Para concluir la presentación de este procedimiento quedará por determinar cuándo aceptaremos la hipótesis de normalidad, es decir, cuando aceptaremos que las fluctuaciones de los puntos en el gráfico de probabilidad son debidas únicamente a la aleatoriedad de las observaciones. Proponemos, en este caso, un simple conteo de las observaciones que caen fuera de las bandas de confianza. Cuando el número de observaciones que caen fuera de nuestra región es superior al  $100\alpha\%$  del número total de observaciones, rechazaremos la hipótesis de normalidad, en otro caso, no tendremos evidencias suficientes para rechazarla.

#### ANALISIS DE LA VARIABILIDAD EN EL GRAFICO PROBABILISTICO NORMAL A PARTIR DEL COMPORTAMIENTO ASINTOTICO DE LOS ESTADISTICOS DE ORDEN.

Para concluir con el estudio de la variabilidad en los gráficos de probabilidad normal, y en la misma línea de los epígrafes anteriores, pretendemos determinar las que hemos denominado "bandas de confianza" basándonos en el comportamiento asintótico de los estadísticos de orden.

La idea básica consiste en localizar, una vez fijado el nivel de confianza  $\alpha$ , para cada  $i$  ( $1 \leq i \leq n$ ), sendos valores  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$  tales que

$$P \left[ p_1^{(i)}(\alpha) \leq X_{(i)} \leq p_2^{(i)}(\alpha) \right] = 1 - \alpha$$

considerando, en esta ocasión, el comportamiento asintótico de los estadísticos de orden  $X_{(i)}$ , supuesto que provienen de una distribución normal de media  $\mu$  y desviación típica  $\sigma$ .



El porqué trabajar ahora con el comportamiento asintótico de los estadísticos de orden habiendo ya trabajado con su comportamiento exacto, queda justificado por la convergencia a normalidad que obtenemos en muchos de ellos y, por tanto, por las ventajas que esto nos reporta.

Según recopilamos en el apartado dedicado al estudio del comportamiento asintótico de los estadísticos de orden en el apéndice I, éste queda determinado en función del tipo de estadístico a que nos estemos refiriendo, donde hemos distinguido entre "comportamiento asintótico de los estadísticos de orden centrales", "comportamiento asintótico de los estadísticos de orden intermedios" y "comportamiento asintótico de los estadísticos de orden extremos", destacando en este último grupo como casos especiales el "comportamiento asintótico del máximo muestral" y el "comportamiento asintótico del mínimo muestral" (los criterios teóricos específicos que se usan para encuadrar a cada estadístico de orden en cada uno de esos grupos, están recogidos en el citado apéndice).

Bajo la hipótesis de normalidad, se obtienen los resultados distribucionales asintóticos que presentamos en la tabla 6.1, donde distinguimos en función del tipo de estadístico de orden de que se trate.



ESTADISTICO	COMPORTAMIENTO ASINTOTICO	CONSTANTES DE NORMALIZACION
ESTADISTICOS CENTRALES $X_{(i)}$	$\sqrt{n} \phi[\Phi^{-1}(p)] \frac{\frac{X_{(i)} - \mu}{\sigma} - \Phi^{-1}(p)}{\sqrt{p(p-1)}} \longrightarrow N(0, 1)$	
ESTADISTICOS INTERMEDIOS SUPERIORES $X_{(n-i+1)}$	$\frac{\frac{X_{(n-i+1)} - \mu}{\sigma} - a_n}{b_n} \longrightarrow N(0, 1)$	$a_n = \Phi^{-1}\left(1 - \frac{i}{n}\right)$ $b_n = \frac{\sqrt{i}}{n\phi(a_n)}$
ESTADISTICOS INTERMEDIOS INFERIORES $X_{(i)}$	$\frac{\frac{X_{(i)} - \mu}{\sigma} - a_n^*}{b_n^*} \longrightarrow N(0, 1)$	$a_n^* = \Phi^{-1}\left(\frac{i}{n}\right)$ $b_n^* = \frac{\sqrt{i}}{n\phi(a_n^*)}$
ESTADISTICOS EXTREMOS SUPERIORES $X_{(n-i+1)}$	$\frac{\frac{X_{(n-i+1)} - \mu}{\sigma} - a_n}{b_n} \longrightarrow \exp\{-\exp(-x)\} \sum_{r=0}^{i-1} \frac{\exp(-rx)}{r!}$	$a_n = \Phi^{-1}\left(1 - \frac{1}{n}\right)$ $b_n = \{n\phi(a_n)\}^{-1}$
ESTADISTICOS EXTREMOS INFERIORES $X_{(i)}$	$\frac{\frac{X_{(i)} - \mu}{\sigma} - a_n^*}{b_n^*} \longrightarrow 1 - \exp\{-\exp(x)\} \sum_{r=0}^{i-1} \frac{\exp(rx)}{r!}$	$a_n^* = \Phi^{-1}\left(\frac{1}{n}\right)$ $b_n^* = \{n\phi(a_n^*)\}^{-1}$
MAXIMO MUESTRAL $X_{(n)}$	$\frac{\frac{X_{(n)} - \mu}{\sigma} - a_n}{b_n} \longrightarrow \exp\{-\exp(-x)\}$	$a_n = \Phi^{-1}\left(1 - \frac{1}{n}\right)$ $b_n = \{n\phi(a_n)\}^{-1}$
MINIMO MUESTRAL $X_{(1)}$	$\frac{\frac{X_{(1)} - \mu}{\sigma} - a_n^*}{b_n^*} \longrightarrow 1 - \exp\{-\exp(x)\}$	$a_n^* = \Phi^{-1}\left(\frac{1}{n}\right)$ $b_n^* = \{n\phi(a_n^*)\}^{-1}$

TABLA 6.1: Comportamiento asintótico de los estadísticos de orden de una distribución normal.



El primer problema que nos planteamos, una vez que el comportamiento asintótico de los estadísticos de orden ha sido estudiado, y recogido en la tabla anterior, es hacer que la clasificación que de ellos proponemos, sea exhaustiva, dado que los criterios teóricos aportados en el apéndice I no establecen con claridad dónde situar las fronteras de los tres grupos.

Proponemos aprovechar, de una parte, el comportamiento simétrico de los estadísticos de orden de una distribución normal, y de otra, el concepto de intervalo de tolerancia introducido en el apéndice dedicado al estudio de los estadísticos de orden, en concreto, en el epígrafe donde se estudian algunos usos de los estadísticos de orden.

Recordamos que un intervalo de tolerancia para una distribución  $F$  (que suponemos continua), con coeficiente de tolerancia  $\gamma$ , es un intervalo aleatorio tal que  $\gamma$  es la probabilidad de que el intervalo cubra al menos un porcentaje especificado ( $100p$ ) de la distribución.

En el caso en que los extremos del intervalo de tolerancia sean dos estadísticos de orden  $X_{(r)}$  y  $X_{(s)}$  ( $1 \leq r < s \leq n$ ), podremos escribir

$$P \left\{ P[X_{(r)} < X < X_{(s)}] \geq p \right\} = \gamma$$

Aprovechando la relación existente entre los estadísticos de orden  $X_{(i)}$  y los estadísticos de orden de una distribución uniforme estándar, con referencia al apéndice I, podemos escribir

$$\gamma = \sum_{i=0}^{s-r-1} \binom{n}{i} p^i (1-p)^{n-i}$$



de manera que, fijados  $\gamma$  y  $p$ , podemos determinar el valor de  $(s-r-1)$  sin más que tener en cuenta la distribución Binomial de parámetros  $n$  y  $p$ .

La determinación de  $(s-r-1)$  siguiendo el procedimiento anterior, no siempre nos reportará una solución única de  $r$  y  $s$ , por lo que nosotros, dado que trabajaremos bajo la hipótesis de normalidad, y aprovechando la simetría de tal distribución, exigiremos que  $r$  y  $s$  sean índices simétricos, esto es,

$$s = n - r + 1$$

Con ello,

$$s - r - 1 = (n - r + 1) - r - 1 = n - 2r$$

y la expresión anterior la podremos escribir como

$$\gamma = \sum_{i=0}^{n-2r} \binom{n}{i} p^i (1-p)^{n-i}$$

Por último, a partir de esta expresión, proponemos seguir el siguiente criterio de clasificación de los estadísticos de orden:

#### ESTADISTICOS DE ORDEN CENTRALES

los comprendidos entre  $X_{(r)}$  y  $X_{(n-r+1)}$  tales que

$$P \left\{ P[X_{(r)} < X < X_{(n-r+1)}] \geq 0,4 \right\} = \gamma$$

#### ESTADISTICOS DE ORDEN INTERMEDIOS

los comprendidos entre  $X_{(r)}$  y  $X_{(n-r+1)}$ , excepto los del grupo an-



terior, tales que

$$P \left\{ P[X_{(r)} < X < X_{(n-r+1)}] \geq 0'7 \right\} = \gamma$$

### ESTADISTICOS DE ORDEN EXTREMOS

los no recogidos en los apartados anteriores, excepto

**MAXIMO MUESTRAL,  $X_{(n)}$**

**MINIMO MUESTRAL,  $X_{(1)}$**

Los valores de  $p$  que hemos tomado, 0'4 y 0'7, son susceptibles de ser cambiados por otros. La elección particular que nosotros hemos hecho de estos valores la podemos justificar desde el momento en que en una distribución normal, la probabilidad asociada al intervalo centrado en la media y de radio una desviación típica, es de aproximadamente 0'4, y de 0'7 si el radio es de dos veces la desviación típica).

Nos queda indicar cómo iremos fijando el coeficiente de tolerancia  $\gamma$ . Dado que el objetivo final es obtener, como se hizo en el epígrafe anterior, una "banda de confianza" en el gráfico probabilístico normal, a un nivel de significación,  $\alpha$ , fijado, proponemos que el coeficiente de tolerancia sea determinado en cada caso a partir del nivel de significación, mediante la relación

$$\gamma = 1 - \alpha$$

Recogemos en la tabla 6.2 ejemplos de la clasificación de los estadísticos de orden en función de varios tamaños muestrales y de distintos niveles de confianza.



NIVEL CONFI	COEFIC TOLER.	TAMAÑO DE LA MUESTRA	ESTADISTICOS CENTRALES		ESTADISTICOS INTERMED. (*)		ESTADISTICOS EXTREMOS (*)	
			desde	hasta	desde	hasta	desde	hasta
0.05	0.95	n = 20	X <sub>(4)</sub>	X <sub>(17)</sub>	X <sub>(2)</sub>	X <sub>(19)</sub>	X <sub>(1)</sub>	X <sub>(20)</sub>
		n = 30	X <sub>(7)</sub>	X <sub>(24)</sub>	X <sub>(2)</sub>	X <sub>(29)</sub>	X <sub>(1)</sub>	X <sub>(30)</sub>
		n = 40	X <sub>(9)</sub>	X <sub>(32)</sub>	X <sub>(3)</sub>	X <sub>(38)</sub>	X <sub>(1)</sub>	X <sub>(40)</sub>
		n = 50	X <sub>(12)</sub>	X <sub>(39)</sub>	X <sub>(5)</sub>	X <sub>(46)</sub>	X <sub>(1)</sub>	X <sub>(50)</sub>
		n = 60	X <sub>(15)</sub>	X <sub>(46)</sub>	X <sub>(6)</sub>	X <sub>(55)</sub>	X <sub>(1)</sub>	X <sub>(60)</sub>
		n = 70	X <sub>(17)</sub>	X <sub>(54)</sub>	X <sub>(7)</sub>	X <sub>(64)</sub>	X <sub>(1)</sub>	X <sub>(70)</sub>
		n = 80	X <sub>(20)</sub>	X <sub>(61)</sub>	X <sub>(8)</sub>	X <sub>(73)</sub>	X <sub>(1)</sub>	X <sub>(80)</sub>
		n = 90	X <sub>(23)</sub>	X <sub>(68)</sub>	X <sub>(10)</sub>	X <sub>(81)</sub>	X <sub>(1)</sub>	X <sub>(90)</sub>
		n = 100	X <sub>(31)</sub>	X <sub>(70)</sub>	X <sub>(11)</sub>	X <sub>(90)</sub>	X <sub>(1)</sub>	X <sub>(100)</sub>
		0.01	0.99	n = 20	X <sub>(3)</sub>	X <sub>(18)</sub>	X <sub>(2)</sub>	X <sub>(19)</sub>
n = 30	X <sub>(6)</sub>			X <sub>(25)</sub>	X <sub>(2)</sub>	X <sub>(29)</sub>	X <sub>(1)</sub>	X <sub>(30)</sub>
n = 40	X <sub>(8)</sub>			X <sub>(33)</sub>	X <sub>(3)</sub>	X <sub>(38)</sub>	X <sub>(1)</sub>	X <sub>(40)</sub>
n = 50	X <sub>(11)</sub>			X <sub>(40)</sub>	X <sub>(4)</sub>	X <sub>(47)</sub>	X <sub>(1)</sub>	X <sub>(50)</sub>
n = 60	X <sub>(13)</sub>			X <sub>(48)</sub>	X <sub>(5)</sub>	X <sub>(56)</sub>	X <sub>(1)</sub>	X <sub>(60)</sub>
n = 70	X <sub>(16)</sub>			X <sub>(55)</sub>	X <sub>(6)</sub>	X <sub>(65)</sub>	X <sub>(1)</sub>	X <sub>(70)</sub>
n = 80	X <sub>(19)</sub>			X <sub>(62)</sub>	X <sub>(7)</sub>	X <sub>(74)</sub>	X <sub>(1)</sub>	X <sub>(80)</sub>
n = 90	X <sub>(21)</sub>			X <sub>(70)</sub>	X <sub>(8)</sub>	X <sub>(83)</sub>	X <sub>(1)</sub>	X <sub>(90)</sub>
n = 100	X <sub>(24)</sub>			X <sub>(77)</sub>	X <sub>(10)</sub>	X <sub>(91)</sub>	X <sub>(1)</sub>	X <sub>(100)</sub>
0.005	0.995			n = 20	X <sub>(3)</sub>	X <sub>(18)</sub>	X <sub>(2)</sub>	X <sub>(19)</sub>
		n = 30	X <sub>(5)</sub>	X <sub>(26)</sub>	X <sub>(2)</sub>	X <sub>(29)</sub>	X <sub>(1)</sub>	X <sub>(30)</sub>
		n = 40	X <sub>(8)</sub>	X <sub>(33)</sub>	X <sub>(2)</sub>	X <sub>(39)</sub>	X <sub>(1)</sub>	X <sub>(40)</sub>
		n = 50	X <sub>(10)</sub>	X <sub>(41)</sub>	X <sub>(3)</sub>	X <sub>(48)</sub>	X <sub>(1)</sub>	X <sub>(50)</sub>
		n = 60	X <sub>(13)</sub>	X <sub>(48)</sub>	X <sub>(4)</sub>	X <sub>(57)</sub>	X <sub>(1)</sub>	X <sub>(60)</sub>
		n = 70	X <sub>(15)</sub>	X <sub>(56)</sub>	X <sub>(6)</sub>	X <sub>(65)</sub>	X <sub>(1)</sub>	X <sub>(70)</sub>
		n = 80	X <sub>(18)</sub>	X <sub>(63)</sub>	X <sub>(7)</sub>	X <sub>(74)</sub>	X <sub>(1)</sub>	X <sub>(80)</sub>
		n = 90	X <sub>(21)</sub>	X <sub>(70)</sub>	X <sub>(8)</sub>	X <sub>(83)</sub>	X <sub>(1)</sub>	X <sub>(90)</sub>
		n = 100	X <sub>(23)</sub>	X <sub>(78)</sub>	X <sub>(9)</sub>	X <sub>(92)</sub>	X <sub>(1)</sub>	X <sub>(100)</sub>

(\*) Excepto los clasificados en los grupos anteriores.

TABLA 6.2: Clasificación de los estadísticos de orden para n = 20 (10) 100, y  $\alpha = 0.05, 0.01$  y  $0.005$ .



Como comentario a la tabla, tengamos en cuenta que en muchas ocasiones, el valor de  $(n-2r)$  se obtiene de forma que  $r$  no es un número entero. En estos casos, debido a que el intervalo debe recoger "al menos" una probabilidad  $p$ , hemos optado por tomar como  $r$  a su parte entera. Esta es la razón por la que en muchos casos, para distintos niveles de confianza, se obtenga la misma clasificación de los estadísticos de orden. Además, observemos que en el grupo de estadísticos extremos hemos incorporado al máximo y al mínimo muestral, a sabiendas que éstos tienen un comportamiento asintótico especial.

Una vez que hemos recordado el comportamiento asintótico de los estadísticos de orden en función de la clase a la que pertenecen, y han quedado totalmente delimitadas tales clases, retomamos el principio de este epígrafe, recordando que nuestro interés estaba en determinar para cada  $i$  ( $1 \leq i \leq n$ ), los valores  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$  tales que

$$P \left[ p_1^{(i)}(\alpha) \leq X_{(i)} \leq p_2^{(i)}(\alpha) \right] = 1 - \alpha$$

En el caso en que  $i$  sea tal que  $X_{(i)}$  sea un estadístico de orden central o intermedio, la determinación de los extremos del intervalo pasará por la localización de cuantiles de la distribución  $N(0,1)$ , y el procedimiento podrá ser automatizado de forma muy sencilla (asumiremos, como en el epígrafe anterior, simetría en los extremos de los intervalos).

El problema lo podemos encontrar en el caso en que  $i$  sea tal que  $X_{(i)}$  sea un estadístico de orden extremo. En este caso, la expresión que adopta la distribución límite de  $X_{(i)}$  nos exige que, para calcular  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$ , ésta sea tabulada o simulada.



Basándonos en un proceso de simulación, en la tabla 6.3 se obtienen los valores de  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$  para los estadísticos de orden extremos en muestras de distintos tamaños. El nivel de confianza ha sido fijado en  $\alpha = 0.05$ . En el proceso de simulación que hemos llevado a cabo, se han considerado 20.000 muestras para cada tamaño muestral.

TAMAÑO MUESTRA	EXTREMOS SUPERIORES	$p_1^{(i)}(\alpha)$	$p_2^{(i)}(\alpha)$	EXTREMOS INFERIORES	$p_1^{(i)}(\alpha)$	$p_2^{(i)}(\alpha)$
n = 50	$X_{(47)}$	0.9702	2.0192	$X_{(2)}$	-2.5773	-1.2475
	$X_{(48)}$	1.0872	2.2550	$X_{(3)}$	-2.2550	-1.0872
	$X_{(49)}$	1.2475	2.5773	$X_{(4)}$	-2.0192	-0.9702
n = 80	$X_{(74)}$	1.0114	1.8008	$X_{(2)}$	-2.7429	-1.4923
	$X_{(75)}$	1.0840	1.9124	$X_{(3)}$	-2.4184	-1.3585
	$X_{(76)}$	1.1651	2.0404	$X_{(4)}$	-2.1976	-1.2495
	$X_{(77)}$	1.2495	2.1976	$X_{(5)}$	-2.0404	-1.1651
	$X_{(78)}$	1.3585	2.4184	$X_{(6)}$	-1.9124	-1.0840
	$X_{(79)}$	1.4923	2.7429	$X_{(7)}$	-1.8008	-1.0114
n = 100	$X_{(91)}$	0.9738	1.6542	$X_{(2)}$	-2.8031	-1.6071
	$X_{(92)}$	1.0274	1.7277	$X_{(3)}$	-2.4965	-1.4725
	$X_{(93)}$	1.0830	1.8042	$X_{(4)}$	-2.2813	-1.3749
	$X_{(94)}$	1.1436	1.9021	$X_{(5)}$	-2.1302	-1.2837
	$X_{(95)}$	1.2094	2.0163	$X_{(6)}$	-2.0163	-1.2094
	$X_{(96)}$	1.2837	2.1302	$X_{(7)}$	-1.9021	-1.1436
	$X_{(97)}$	1.3749	2.2813	$X_{(8)}$	-1.8042	-1.0830
	$X_{(98)}$	1.4725	2.4965	$X_{(9)}$	-1.7277	-1.0274
	$X_{(99)}$	1.6071	2.8031	$X_{(10)}$	-1.6542	-0.9738

TABLA 6.3: Valores de  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$  para  $\alpha = 0.05$

Para la elaboración de esta tabla se ha tenido en cuenta la relación existente entre la distribución asintótica del  $X_{(n-i+1)}$  estadístico extremo superior y la del  $X_{(i)}$  estadístico



extremo inferior, que podemos indicar que es:

$$\Phi_{n-i+1}(-x) = 1 - \Phi_i(x)$$

Determinados los puntos porcentuales de los estadísticos de orden extremos, el procedimiento que proponemos seguir para analizar la variabilidad de en los gráficos de probabilidad normal basado en el comportamiento asintótico de los estadísticos de orden, se puede resumir y ordenar en las siguientes etapas:

- 1°) Supongamos que disponemos de las observaciones  $x_i$ . Fijemos el nivel de confianza  $\alpha$ .
- 2°) Ajustar una recta al gráfico probabilístico normal. Para ello podrá usarse cualquiera de los métodos anteriormente introducidos, con las observaciones que en cada uno de ellos hemos realizado.
- 3°) La recta ajustada en el gráfico de probabilidad nos proporciona una estimación de los parámetros  $\mu$  y  $\sigma$  de la distribución normal bajo la hipótesis nula (como ya se ha indicado serán los puntos de intersección e inclinación de la recta ajustada)
- 4°) Conocido el tamaño muestral, clasificar los estadísticos de orden en las categorías de estadísticos de orden centrales, intermedios y extremos como se muestra en la tabla 6.2.
- 5°) Para cada  $i$ , determinar los valores  $p_1^{(i)}(\alpha)$  y  $p_2^{(i)}(\alpha)$  a partir de su comportamiento asintótico (tabla 6.1), una vez que  $X_{(i)}$  ha sido clasificado. Para el caso en que  $i$  sea tal que  $X_{(i)}$  sea un estadístico de orden extremo, tener en cuenta la tabla 6.3 (u otra similar para los correspondientes valores de  $n$  y  $\alpha$ ). (El valor de los parámetros  $\mu$  y  $\sigma$  será sustituido por sus estimadores obtenidos en el apartado 3°)).



6°) Para cada  $i$ , dibujar verticalmente un intervalo centrado en el correspondiente punto de la recta ajustada, con extremo inferior igual a  $p_1^{(i)}(\alpha)$  y extremo superior igual a  $p_2^{(i)}(\alpha)$ .

Una vez concluido este procedimiento, daremos lugar a un gráfico como el que aparece en la figura 6.3.

**FIGURA 6.3:** Banda de confianza en el gráfico probabilístico normal, haciendo uso del comportamiento asintótico de los estadísticos de orden.

Obtenido el gráfico en la forma propuesta, y como se razonó en el epígrafe anterior, aceptaremos la hipótesis de normalidad cuando el número de observaciones que caen fuera de la re-



gión determinada por la "banda de confianza" no sea superior al  $100\alpha\%$  del total de las observaciones.



## APENDICE I

### ESTADISTICOS DE ORDEN

Como indican Arnold, Balakrishnan y Nagaraja (1992) en la introducción al libro "A First Course in Order Statistics", "hace unos años, el estudio de los estadísticos de orden resultaba una tarea curiosa. Desde la aparición del volumen editado por Sarhan y Greenberg en 1962, y el tratado realizado por H. A. David en 1970 y su siguiente revisión en 1981, todo ha cambiado", y continúan diciendo, "actualmente, y aunque sea de forma somera, cualquier estudio serio en estadística precisará de una introducción al tema".

La justificación de la incorporación de este apéndice al trabajo no es sólo como respuesta al consejo que dan los autores anteriormente mencionados, sino que, a su vez, responde al uso directo e indirecto que en toda la memoria se hace de los estadísticos de orden y de algunos resultados obtenidos a partir de ellos. En lo referido al uso directo, éste es evidente. Cuando hablamos de uso indirecto, queremos referirnos al que implícitamente damos a los estadísticos de orden cuando hablamos de técnicas de bondad de ajuste, dado que, a menudo, en ellas el interés recae en las desviaciones entre la función empírica de cuantiles y la función de cuantiles bajo la hipótesis nula (inversa de la función de distribución).



DEFINICION DE ESTADISTICO DE ORDEN

Sea  $(X_1, X_2, \dots, X_n)$  un vector aleatorio n-dimensional, y  $(x_1, x_2, \dots, x_n)$  un n-upla de valores de  $(X_1, X_2, \dots, X_n)$ . Supongamos que reordenamos  $x_1, x_2, \dots, x_n$  de forma creciente, de tal manera que

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

donde  $x_{(1)} = \min \{x_1, x_2, \dots, x_n\}$ ,  $x_{(2)}$  es el segundo valor más pequeño, y así sucesivamente,  $x_{(n)} = \max \{x_1, x_2, \dots, x_n\}$ . (Si cualesquiera dos  $x_i, x_j$ , son iguales, su orden es irrelevante).

**DEFINICION A.I.1.** La función  $X_{(k)}$  de  $(X_1, X_2, \dots, X_n)$  que toma el valor  $x_{(k)}$  para cada posible secuencia  $(x_1, x_2, \dots, x_n)$  de valores de  $(X_1, X_2, \dots, X_n)$  se conoce con el nombre de k-ésimo estadístico de orden, o estadístico de orden k.

Para cada n dado, y asociados a  $(X_1, X_2, \dots, X_n)$ , se pueden formar los n estadísticos de orden  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  que suelen denominarse conjunto de estadísticos de orden para  $(X_1, X_2, \dots, X_n)$ .

**Ejemplo.** Sean  $X_1, X_2$  y  $X_3$  tres variables aleatorias de tipo discreto, donde suponemos que  $X_1$  y  $X_3$  pueden tomar valores 0, 1, y  $X_2$  puede tomar valores 1, 2, 3. Entonces, el vector aleatorio  $(X_1, X_2, X_3)$  puede tomar cualquiera de los siguientes valores: (0,1,0), (0,1,1), (0,2,0), (0,2,1), (0,3,0), (0,3,1), (1,1,0), (1,1,1), (1,2,0), (1,2,1), (1,3,0), (1,3,1). De esta forma,  $X_{(1)}$  puede tomar los valores 0, 1;  $X_{(2)}$  puede tomar los valores 0, 1; y  $X_{(3)}$  puede tomar los valores 1, 2, 3.

La primera consideración importante que debemos hacer es que para cualquiera que sea  $(X_1, X_2, \dots, X_n)$  un vector aleatorio, el k-ésimo estadístico de orden  $X_{(k)}$  ( $1 \leq k \leq n$ ), es una variable aleatoria.



**FUNCION DE DISTRIBUCION EMPIRICA Y FUNCION DE DISTRIBUCION DE  $X_{(k)}$**

Sea  $(X_1, X_2, \dots, X_n)$  un vector aleatorio, y sea  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  su correspondiente conjunto de estadísticos de orden, con valores  $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ .

**DEFINICION A.I.2.** La función de  $x$  ( $-\infty < x < +\infty$ ) igual a cero para  $x \leq x_{(1)}$ , e igual a  $m/n$  ( $m = 1, 2, \dots, n$ ), para  $x > x_{(1)}$ , donde  $m$  es el índice más grande para el que  $x_{(m)} < x$ , se le denomina *Función de distribución empírica*, y la notaremos por  $F_n(x)$ .

De esta forma definida,  $n F_n(x)$  será el número de elementos en la muestra que son menores o iguales que  $x$ .

La función  $F_n(x)$  puede tomar valores entre cero y uno y es una función no decreciente de  $x$ . Se puede probar que  $F_n(x)$  es continua a la izquierda y satisface todas las propiedades de una función de distribución. Precisamente esto justifica que haya sido denominada función de distribución empírica. De cualquier manera,  $F_n(x)$  es una variable aleatoria para cada valor de  $x$ .

Supongamos que las variables aleatorias  $X_r$  ( $r = 1, 2, \dots, n$ ), son independientes e idénticamente distribuidas (iid) según una función de distribución continua  $F(x)$ . Sean  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  la colección de valores observados de las variables aleatorias  $X_1, X_2, \dots, X_n$ , ordenados de menor a mayor. Al suponer que la función de distribución  $F(x)$  es continua, y que las variables  $X_1, X_2, \dots, X_n$  son independientes, podemos asegurar que la probabilidad de que aparezcan dos valores  $x_{(r)}$  y  $x_{(r+1)}$  iguales, es cero. Por ello, asumimos que

$$x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

pudiendo, por tanto, representar la función de distribución empí-



rica como

$$F_n(x) = \begin{cases} 0 & \text{para } x \leq x_{(1)} \\ \frac{k}{n} & \text{para } x_{(k)} < x \leq x_{(k+1)} ; k=1,2,\dots,n-1 \\ 1 & \text{para } x > x_{(n)} \end{cases}$$

Supongamos ahora que  $X_1, X_2, \dots, X_n$  son independientes y que tiene la misma función de distribución  $F(x)$ . Esto nos permite afirmar que para cualquier  $x$  fijo, tendremos

$$P(X_r < x) = F(x) = p = \text{constante } (r = 1, 2, \dots, n)$$

Por ello, para cualquier  $x$  fijo,  $F_n(x)$  es la frecuencia de éxito en el esquema de Bernoulli y, por tanto, la frecuencia de ocurrencia de algún suceso aleatorio en  $n$  ensayos independientes, donde la probabilidad de ocurrencia de dicho evento, en cada ensayo en particular, es constante e igual a  $p = F(x)$ . Así,

$$P\left[F_n(x) = \frac{m}{n}\right] = \frac{n!}{m!(n-m)!} [F(x)]^m [1-F(x)]^{n-m} \quad (m = 0, 1, \dots, n)$$

Notemos a la función de distribución de la variable aleatoria  $X_{(k)}$  por  $\Phi_k(x)$ . Teniendo en cuenta que

$$\Phi_k(x) = P[X_{(k)} < x]$$

determinemos dicha función de distribución en términos de  $F_n(x)$ .

El suceso  $[X_{(k)} < x]$  ocurre cuando la  $k$ -ésima observación ordenada es más pequeña que  $x$ , o lo que es lo mismo, cuando al menos  $k$  observaciones (de entre las  $n$ ) son más pequeñas que  $x$ . Esto es equivalente al suceso de que la frecuencia observada  $F_n(x)$



no sea más pequeña que  $k/n$ , es decir:

$$\Phi_k(x) = P[X_{(k)} < x] = P\left[F_n(x) \geq \frac{k}{n}\right] = \sum_{m=k}^n P\left[F_n(x) = \frac{m}{n}\right]$$

que teniendo en cuenta el desarrollo anterior, podemos escribir

$$\Phi_k(x) = \sum_{m=k}^n \frac{n!}{m!(n-m)!} [F(x)]^m [1-F(x)]^{n-m}$$

que es la expresión de la función de distribución del  $k$ -ésimo estadístico de orden.

$\Phi_k(x)$  admite también la siguiente representación:

$$\Phi_k(x) = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} t^{k-1} (1-t)^{n-k} dt$$

Veamos a continuación cómo esta última expresión se reduce a la anterior sin más que integrar convenientemente por partes; esto es, si tomamos

$$u = (1-t)^{n-k} \quad \text{y} \quad dv = t^{k-1} dt$$

tendremos

$$\begin{aligned} \Phi_k(x) &= \frac{n!}{(k-1)!(n-k)!} \left\{ \left[ \frac{1}{k} t^k (1-t)^{n-k} \right]_0^{F(x)} + \int_0^{F(x)} \frac{1}{k} t^k (1-t)^{n-k-1} (n-k) dt \right\} = \\ &= \frac{n!}{(k-1)!(n-k)!} \left\{ \frac{1}{k} [F(x)]^k [1-F(x)]^{n-k} + \frac{(n-k)}{k} \int_0^{F(x)} t^k (1-t)^{n-k-1} dt \right\} \end{aligned}$$



De nuevo, tomando

$$u = (1-t)^{n-k-1} \quad y \quad dv = t^k dt$$

tendremos

$$\begin{aligned} \Phi_k(x) = \frac{n!}{(k-1)!(n-k)!} & \left\{ \frac{1}{k} [F(x)]^k [1-F(x)]^{n-k} + \frac{(n-k)}{k(k+1)} [F(x)]^{k+1} [1-F(x)]^{n-k-1} + \right. \\ & \left. + \frac{(n-k)(n-k-1)}{k(k+1)} \int_0^{F(x)} t^{k+1} (1-t)^{n-k-2} dt \right\} \end{aligned}$$

Si repetimos este proceso de integración por partes hasta un total de (n-k) veces, obtendremos:

$$\begin{aligned} \Phi_k(x) = \frac{n!}{(k-1)!(n-k)!} & \left\{ \frac{1}{k} [F(x)]^k [1-F(x)]^{n-k} + \frac{(n-k)}{k(k+1)} [F(x)]^{k+1} [1-F(x)]^{n-k-1} + \right. \\ & + \dots + \frac{(n-k)(n-k-1)\dots 1}{k(k+1)\dots(n-1)n} [F(x)]^n \left. \right\} = \\ & = \sum_{m=k}^n \frac{n!}{m!(n-m)!} [F(x)]^m [1-F(x)]^{n-m} \end{aligned}$$

que coincide con la expresión que habíamos propuesto.

El procedimiento seguido es válido para  $k = 2, \dots, (n-1)$ , pero veamos que en el caso en que  $k = 1$  y  $k = n$ , también se mantiene la igualdad en las expresiones, ya que:



Si  $k = 1$

$$\Phi_1(x) = \frac{n!}{(n-1)!} \int_0^{F(x)} (1-t)^{n-1} dt = 1 - [1-F(x)]^n$$

Por otra parte,

$$\begin{aligned} \Phi_1(x) &= \sum_{m=1}^n \frac{n!}{m!(n-m)!} [F(x)]^m [1-F(x)]^{n-m} = \\ &= [F(x) + 1 - F(x)]^n - \binom{n}{0} [F(x)]^0 [1-F(x)]^{n-0} = 1 - [1-F(x)]^n \end{aligned}$$

Con lo que se mantiene la igualdad.

Si  $k = n$

$$\Phi_n(x) = \frac{n!}{(n-1)!} \int_0^{F(x)} t^{n-1} dt = [F(x)]^n$$

Por otra parte,

$$\Phi_n(x) = \sum_{m=n}^n \frac{n!}{m!(n-m)!} [F(x)]^m [1-F(x)]^{n-m} = \binom{n}{n} [F(x)]^n = [F(x)]^n$$

Con lo que, de la misma forma, también se mantiene la igualdad.

Por todo ello, y tal y como proponíamos, tanto la prime-



ra como la segunda expresión son iguales y, por tanto, ambas serán usadas, indistintamente, como la función de distribución del  $k$ -ésimo estadístico de orden.

Supongamos que la función de densidad  $f(x) = F'(x)$  existe. Entonces existe la función de densidad de  $X_{(k)}$  ( $k = 1, 2, \dots, n$ ), que notaremos por  $\varphi_k(x)$  y que la obtenemos directamente diferenciando la función de distribución  $\Phi_k(x)$  respecto de  $x$ , obteniendo:

$$\varphi_k(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1-F(x)]^{n-k} f(x)$$

**Ejemplo.** Sean  $X_1, X_2, \dots, X_n$ ,  $n$  variables aleatorias independientes e idénticamente distribuidas según una distribución uniforme estándar. Sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  el correspondiente conjunto de estadísticos de orden. Calculemos la distribución (función de distribución y función de densidad) de cada  $X_{(k)}$ ,  $k = 1, 2, \dots, n$ .

Sin más que tener en cuenta que

$$f(x) = \begin{cases} 1 & \text{para } 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

y que

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

tendremos que, siendo  $X_{(k)}$  el  $k$ -ésimo estadístico de orden,



$$\Phi_k(x) = \frac{n!}{(k-1)! (n-k)!} \int_0^x t^{k-1} (1-t)^{n-k} dt$$

y a su vez,

$$\varphi_k(x) = \frac{n!}{(k-1)! (n-k)!} x^{k-1} (1-x)^{n-k}$$

Observamos que la expresión que tiene esta distribución de probabilidad coincide con la de una variable aleatoria Beta con parámetros  $k$  y  $(n-k+1)$ . Por ello podemos establecer el siguiente resultado, que nos será de gran utilidad:

"Sean  $X_1, X_2, \dots, X_n$   $n$  variables aleatorias independientes e idénticamente distribuidas según una distribución uniforme estándar, y sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  el correspondiente conjunto de estadísticos de orden. Para cualquier  $k$ , con  $1 \leq k \leq n$ ,  $X_{(k)}$  sigue una distribución Beta ( $k ; n-k+1$ )".

#### DISTRIBUCION CONJUNTA DE LOS ESTADÍSTICOS DE ORDEN

Dedicaremos este apartado al estudio de la distribución conjunta de los estadísticos de orden. Una vez determinada ésta, sin más que considerar las distribuciones marginales, obtendremos en primer lugar la distribución de un estadístico de orden, seguidamente la distribución conjunta de dos estadísticos de orden, para acabar con la distribución conjunta de  $k$  estadísticos de orden. Para facilitar el trabajo, y visto el ambiente de esta memoria, obtendremos los anteriores resultados para el caso en que  $F$  represente a una distribución continua.

Supongamos que  $X_1, X_2, \dots, X_n$  son  $n$  variables aleatorias independientes e idénticamente distribuidas según una función de



distribución F. Supongamos que  $f$  existe y que  $f(x) = F'(x)$ . Sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ , el conjunto de estadísticos de orden asociados a  $X_1, X_2, \dots, X_n$ . Dado que se ha supuesto la continuidad de las variables aleatorias  $X_i$ , podemos afirmar que, con probabilidad 1,

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}$$

**TEOREMA A.I.1.** Bajo el ambiente anteriormente enunciado, la función de densidad conjunta de  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  viene dada por:

$$\varphi(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = \begin{cases} n! \prod_{i=1}^n f(x_{(i)}) & x_{(1)} < x_{(2)} < \dots < x_{(n)} \\ 0 & \text{en otro caso} \end{cases}$$

Demostración.

Dada la independencia de los  $X_i$ , es obvio que

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$$

Por otra parte, la transformación de  $X_1, X_2, \dots, X_n$  a  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  es uno a uno. De hecho, hay un total de  $n!$  posibles ordenaciones de  $x_1, x_2, \dots, x_n$  en orden creciente. Por lo tanto, hay  $n!$  inversas de las transformaciones. El Jacobiano de cada una de las transformaciones es el determinante de una matriz identidad de dimension  $n \times n$ , con las filas reordenadas, dado que cada  $x_{(i)}$  es igual a uno sólo de los  $x_1, x_2, \dots, x_n$ . Por lo tanto,  $J = \pm 1$ , y la función de densidad asociada a cada una de esas posibles ordenaciones es

$$|J| \prod_{i=1}^n f(x_{(i)}) \quad x_{(1)} < x_{(2)} < \dots < x_{(n)}$$



para un total de  $n!$  ordenaciones posibles de los  $x_1, x_2, \dots, x_n$ .

Por último, y teniendo en cuenta los resultados sobre transformaciones de variables aleatorias [pgs. 135 a 137, Rohatgi (1976)], podemos escribir la función de densidad conjunta de  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  por la expresión

$$\varphi(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = \sum_{n! \text{ inversas}} \prod_{i=1}^n f(x_{(i)}) = n! \prod_{i=1}^n f(x_{(i)})$$

$$\text{para } x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

■

Como consecuencia de este resultado, podemos obtener de forma sencilla la función de densidad de cada estadístico de orden, expresión que recogemos en el siguiente resultado.

**TEOREMA A.I.2.** En el ambiente dado al principio del epígrafe, la función de densidad de  $X_{(k)}$  ( $1 \leq k \leq n$ ) adopta la forma

$$\varphi_k(y_k) = \frac{n!}{(k-1)! (n-k)!} [F(y_k)]^{k-1} [1-F(y_k)]^{n-k} f(y_k)$$

Demostración.

$\varphi_k(y_k)$  no es más que la función de densidad marginal de  $X_{(k)}$ , y la obtenemos a partir de la expresión de  $\varphi(x_{(1)}, x_{(2)}, \dots, x_{(n)})$  dada en el teorema anterior.

$$\varphi_k(y_k) = n! f(y_k) \int_{-\infty}^{y_k} \int_{-\infty}^{y_{k-1}} \dots \int_{-\infty}^{y_2} \int_{y_k}^{\infty} \int_{y_{k+1}}^{\infty} \dots \int_{y_{n-1}}^{\infty} \prod_{i \neq k} f(y_i) dy_n \dots dy_{k+1} dy_1 \dots dy_{k-1}$$



$$\begin{aligned}
 &= n! f(y_k) \frac{[1-F(y_k)]^{n-k}}{(n-k)!} \int_{-\infty}^{y_k} \dots \int_{-\infty}^{y_2} \prod_{i=1}^{k-1} f(y_i) dy_1 dy_2 \dots dy_{k-1} \\
 &= n! f(y_k) \frac{[1-F(y_k)]^{n-k}}{(n-k)!} \frac{[F(y_k)]^{k-1}}{(k-1)!}
 \end{aligned}$$

■

Observemos que esta expresión coincide con la que obtuvimos al final del apartado anterior, al diferenciar la función de distribución de  $X_{(k)}$ .

Sin necesidad de realizar un esfuerzo mucho más grande, podemos obtener la distribución conjunta de dos estadísticos de orden, expresión que recogemos en el siguiente resultado:

**TEOREMA A.I.3.** En el mismo ambiente enunciado al comienzo del epígrafe, la función de densidad conjunta de  $X_{(j)}$  y  $X_{(k)}$ , con  $1 \leq j < k \leq n$ , está dada por

$$\varphi_{j,k}(y_j, y_k) = \begin{cases} \frac{n!}{(j-1)!(k-j-1)!(n-k)!} [F(y_j)]^{j-1} [F(y_k)-F(y_j)]^{k-j-1} \times \\ \quad \times [1-F(y_k)]^{n-k} f(y_j) f(y_k) & \text{si } y_j < y_k \\ 0 & \text{en otro caso} \end{cases}$$



Demostración.

$$\begin{aligned} \varphi_{j,k}(y_j, y_k) &= \int_{-\infty}^{y_j} \dots \int_{-\infty}^{y_2} \int_{y_j}^{y_k} \dots \int_{y_{k-2}}^{y_k} \int_{y_k}^{\infty} \dots \int_{y_{n-1}}^{\infty} n! \prod_{i=1}^n f(y_i) dy_n \dots dy_{k+1} \\ &\quad dy_{k-1} \dots dy_{j+1} dy_1 \dots dy_{j-1} \\ &= n! \int_{-\infty}^{y_j} \dots \int_{-\infty}^{y_2} \int_{y_j}^{y_k} \dots \int_{y_{k-2}}^{y_k} \frac{[1-F(y_k)]^{n-k}}{(n-k)!} \prod_{i=1}^k f(y_i) dy_{k-1} \dots dy_{j+1} dy_1 \dots dy_{j-1} \\ &= n! \frac{[1-F(y_k)]^{n-k}}{(n-k)!} f(y_k) \int_{-\infty}^{y_j} \dots \int_{-\infty}^{y_2} \frac{[F(y_k)-F(y_j)]^{k-j-1}}{(k-j-1)!} \prod_{i=1}^j f(y_i) dy_1 \dots dy_{j-1} \\ &= n! \frac{[1-F(y_k)]^{n-k}}{(n-k)!} \frac{[F(y_k)-F(y_j)]^{k-j-1}}{(k-j-1)!} f(y_k) f(y_j) \frac{[F(y_j)]^{j-1}}{(j-1)!} \\ &\quad \text{para } y_j < y_k. \end{aligned}$$

■

De manera similar, obtendríamos la función de densidad conjunta de  $X_{(j_1)}, X_{(j_2)}, \dots, X_{(j_k)}$ , con  $1 \leq j_1 < j_2 < \dots < j_k \leq n$ ,  $1 \leq k \leq n$ , que adopta la expresión:

$$\begin{aligned} \varphi_{j_1, j_2, \dots, j_k}(y_1, y_2, \dots, y_k) &= \frac{n!}{(j_1-1)! (j_2-j_1-1)! \dots (n-j_k)!} \times \\ &\quad \times [F(y_1)]^{j_1-1} f(y_1) [F(y_2)-F(y_1)]^{j_2-j_1-1} f(y_2) \times \dots \times \\ &\quad \times [1-F(y_k)]^{n-j_k} f(y_k) \quad \text{para } y_1 < y_2 < \dots < y_k \end{aligned}$$

y con valor 0 en cualquier otro caso.



### MOMENTOS DE LOS ESTADISTICOS DE ORDEN

En este apartado obtendremos las expresiones de los momentos de los estadísticos de orden, pero antes de entrar de lleno en este apartado avanzaremos el siguiente resultado:

Sea  $U_1, U_2, \dots, U_n$  una muestra aleatoria de tamaño  $n$  de una distribución uniforme estándar, y sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una población con función de distribución  $F(x)$ . Consideremos además  $U_{(1)} < U_{(2)} < \dots < U_{(n)}$  y  $X_{(1)} < X_{(2)} < \dots < X_{(n)}$  los respectivos conjuntos de estadísticos de orden asociados a cada una de las muestras.

Cuando  $F(x)$  es continua, la transformación  $U = F(X)$  produce una distribución uniforme estándar. Así, cuando  $F(x)$  es continua, tendremos que

$$F(X_{(i)}) \stackrel{d}{=} U_{(i)} \quad i = 1, 2, \dots, n \quad (I.1)$$

(donde  $\stackrel{d}{=}$  significa que tiene la misma distribución). Además, si consideramos la función cuantil (inversa de la función de distribución) definida como

$$F^{-1}(y) = \sup \{x : F(x) \leq y\}$$

es bien conocido que

$$F^{-1}(U_i) \stackrel{d}{=} X_i \quad i = 1, 2, \dots, n$$

para cualquiera que sea  $F$ . Dado, además, que  $F^{-1}(\cdot)$  preserva el orden, tendremos que

$$F^{-1}(U_{(i)}) \stackrel{d}{=} X_{(i)} \quad i = 1, 2, \dots, n \quad (I.2)$$



Los resultados (I.1) y (I.2), que fueron originalmente observados por Scheffé y Tukey (1945), serán de gran utilidad tanto en la determinación de los momentos de los estadísticos de orden, como en la obtención de algunos resultados que, sobre el comportamiento asintótico de los estadísticos de orden, veremos más adelante.

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de tamaño  $n$  tomada de una población con función de distribución  $F(x)$  y función de densidad  $f(x)$ , y sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  el correspondiente conjunto de estadísticos de orden. A partir de la función de densidad de  $X_{(i)}$ , tenemos que, para  $1 \leq i \leq n$  y  $m = 1, 2, \dots$ ,

$$\begin{aligned} \mu_{(i)}^{(m)} = E\{X_{(i)}^m\} &= \int_{-\infty}^{\infty} x^m \varphi_i(x) dx = \\ &= \frac{n!}{(i-1)! (n-1)!} \int_{-\infty}^{\infty} x^m [F(x)]^{i-1} [1-F(x)]^{n-i} f(x) dx \end{aligned}$$

(por conveniencia en la notación, escribiremos  $\mu_{(i)} = \mu_{(i)}^{(1)}$ ).

Supuesto que existen los dos primeros momentos de  $X_{(i)}$ , su varianza quedará determinada por

$$\sigma_{(i,i)} = \sigma_{(i)}^{(2)} = \mu_{(i)}^{(2)} - \mu_{(i)}^2 \quad 1 \leq i \leq n$$

Similarmente, a partir de la distribución conjunta de  $X_{(i)}$  y  $X_{(j)}$ , para  $1 \leq i < j \leq n$  y  $m_i, m_j = 1, 2, \dots$ , tendremos



$$\begin{aligned} \mu_{(i,j)}^{(m_i, m_j)} &= E\{X_{(i)}^{m_i} X_{(j)}^{m_j}\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i^{m_i} x_j^{m_j} \varphi_{i,j}(x_i, x_j) dx_i dx_j = \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i^{m_i} x_j^{m_j} [F(x_i)]^{i-1} [F(x_j)-F(x_i)]^{j-i-1} \times \\ &\quad \times [1-F(x_j)]^{n-j} f(x_i) f(x_j) dx_i dx_j \end{aligned}$$

De nuevo, por conveniencia, escribiremos  $\mu_{(i,j)} = \mu_{(i,j)}^{(1,1)}$ .

La covarianza de  $X_{(i)}$  y  $X_{(j)}$  será determinada por la expresión

$$\sigma_{(i,j)} = \text{Cov}(X_{(i)}, X_{(j)}) = \mu_{(i,j)} - \mu_{(i)} \mu_{(j)}$$

A estas alturas, si consideramos la definición de función cuantil y las relaciones distribucionales establecidas al principio del epígrafe, podemos escribir, para  $1 \leq i \leq n$ , y  $m = 1, 2, \dots$ ,

$$\mu_{(i)}^{(m)} = E\{X_{(i)}^m\} = \frac{n!}{(i-1)!(n-i)!} \int_0^1 [F^{-1}(u)]^m u^{i-1} (1-u)^{n-i} du$$

y de la misma forma, para  $1 \leq i < j \leq n$ , y  $m_i, m_j = 1, 2, \dots$ ,



$$\mu_{(i,j)}^{(m_i, m_j)} = E\{X_{(i)}^{m_i} X_{(j)}^{m_j}\} = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int \int_{0 < u_i < u_j < 1} [F^{-1}(u_i)]^{m_i} \times \\ \times [F^{-1}(u_j)]^{m_j} u_i^{i-1} (u_j - u_i)^{j-i-1} (1-u_j)^{n-j} du_i du_j$$

ALGUNOS USOS DE LOS ESTADÍSTICOS DE ORDEN

En esta sección vamos a considerar algunas aplicaciones de los estadísticos de orden. En particular, desarrollaremos tres aplicaciones, a saber: *intervalos de tolerancia para la distribución, alcance (o extensión) e intervalos de confianza estimados para cuantiles.*

**DEFINICION A.I.3** Sea  $F$  una función de distribución continua. Un intervalo de tolerancia para  $F$ , con un coeficiente de tolerancia  $\gamma$ , es un intervalo aleatorio tal que  $\gamma$  es la probabilidad de que el intervalo cubra al menos un especificado porcentaje ( $100p$ ) de la distribución.

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de tamaño  $n$  tomada de una población cuya función de distribución es  $F$ , y sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  el correspondiente conjunto de estadísticos de orden. Si los extremos del intervalo de tolerancia son dos estadísticos de orden,  $X_{(r)}$  y  $X_{(s)}$ ,  $1 \leq r < s \leq n$ , tendremos:

$$P\{P[X_{(r)} < X < X_{(s)}] \geq p\} = \gamma \tag{I.3}$$

Dado que  $F$  es continua,  $F(X)$  es uniforme estándar, y tendremos



$$\begin{aligned}
 P\{X_{(r)} < X < X_{(s)}\} &= P\{X < X_{(s)}\} - P\{X < X_{(r)}\} = \\
 &= F(X_{(s)}) - F(X_{(r)}) = U_{(s)} - U_{(r)} \quad (I.4)
 \end{aligned}$$

donde  $U_{(r)}$  y  $U_{(s)}$  son estadísticos de orden de la distribución uniforme. Así,

$$P\{P[X_{(r)} < X < X_{(s)}] \geq p\} = \gamma \Leftrightarrow P\{[U_{(s)} - U_{(r)}] \geq p\} = \gamma \quad (I.5)$$

Usando la distribución conjunta de  $U_{(r)}$  y  $U_{(s)}$ , y a partir de (I.3), podemos escribir

$$\gamma = \int_p^1 \int_0^{y-p} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} x^{r-1} (y-x)^{s-r-1} (1-y)^{n-s} dx dy \quad (I.6)$$

para todo  $0 < p < 1$  y  $r < s$ . Por desgracia, (I.6) no es fácil de resolver. En su lugar, es conveniente considerar  $U = U_{(s)} - U_{(r)}$  y  $V = U_{(s)}$ .

La función de densidad conjunta de  $U$  y  $V$  está dada por

$$f_{U,V}(u,v) = \begin{cases} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} (v-u)^{r-1} (u)^{s-r-1} (1-v)^{n-s} & 0 < u < v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Así, la función de densidad marginal de  $U$  tiene la expresión



$$\begin{aligned}
f_U(u) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} u^{s-r-1} \int_u^1 (v-u)^{r-1} (1-v)^{n-s} dv = \\
&= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} u^{s-r-1} \int_0^1 t^{r-1} [1-u-t(1-u)]^{n-s} (1-u)^r dt = \\
&= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} u^{s-r-1} (1-u)^{n-s+r} \beta(r, n-s+1) = \\
&= \frac{n!}{(s-r-1)!(n-s+r)!} u^{s-r-1} (1-u)^{n-s+r} \quad 0 < u < 1 \quad (I.7)
\end{aligned}$$

Usando (I.7) en (I.5), reescribimos (I.5) como:

$$\gamma = \int_p^1 n \binom{n-1}{s-r-1} u^{s-r-1} (1-u)^{n-s+r} du \quad (I.8)$$

Vista la expresión (I.8), y teniendo en cuenta la relación existente entre la distribución Beta y la Binomial, obtenemos

$$\gamma = \sum_{i=0}^{s-r-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \quad (I.9)$$

que es mucho más fácil de usar que la expresión que habíamos obtenido en (I.6). Aún así, supuesto que los valores de  $n, p$  y  $\gamma$  son conocidos, no siempre será posible encontrar  $(s-r)$  que dé exactamente un coeficiente de tolerancia igual a  $\gamma$ .



**Ejemplo.** Supongamos que  $s = n$  y  $r = 1$ . Entonces

$$\gamma = \sum_{i=0}^{n-2} \binom{n}{i} p^i (1-p)^{n-i} = 1 - p^n - np^{n-1}(1-p).$$

Si  $p = 0'8$ ,  $n = 5$ ,  $r = 1$ , entonces:

$$\gamma = 1 - (0'8)^5 - 5 (0'8)^4(0'2) = 0'263$$

De esta forma, el intervalo  $(X_{(1)}, X_{(5)})$  determina en este caso a un intervalo de tolerancia al 26% para una probabilidad de 0'8 bajo la distribución (de X).

**Ejemplo.** Sea  $X_1, X_2, X_3, X_4, X_5$  una muestra aleatoria de tamaño 5 de una población cuya función de distribución continua es F. Pretendemos encontrar  $r$  y  $s$ ,  $r < s$ , tales que  $(X_{(r)}, X_{(s)})$  sea un intervalo de tolerancia al 90% para una probabilidad de 0'5 bajo F. Esto es,

$$0'9 = P\{U \geq 0'5\} = \sum_{i=0}^{s-r-1} \binom{5}{i} \left(\frac{1}{2}\right)^5$$

Si elegimos  $s-r = 4$ , entonces  $\gamma = 0'81$ , y si elegimos  $s-r = 5$ , entonces  $\gamma = 0'969$ . En este caso deberemos tender por un intervalo de tolerancia al 96'9%, excediendo el valor deseado (90%).

En general, dado  $p$ ,  $0 < p < 1$ , será posible elegir un tamaño de muestra  $n$ , suficientemente grande, y un valor correspondiente de  $s-r$  tal que, con probabilidad  $\geq \gamma$ , un intervalo de la forma  $(X_{(r)}, X_{(s)})$  cubra al menos el 100p por ciento de la distri-



bución. Si  $s-r$  está dado en función de  $n$ , elegiremos el menor tamaño de la muestra posible.

**Ejemplo.** Sea  $p = 3/4$  y  $\gamma = 0.75$ , supongamos que queremos determinar el tamaño muestral más pequeño requerido tal que  $(X_{(2)}, X_{(n)})$  cubra al menos el 75% de la distribución. Dicho de otra forma, determinaremos el valor de  $n$  más pequeño posible que satisfaga

$$0.75 \leq \sum_{i=0}^{n-2} \binom{n}{i} \left(\frac{3}{4}\right)^i \left(\frac{1}{4}\right)^{n-i}$$

De la distribución binomial, o usando la aproximación a la distribución normal, obtenemos que  $n = 14$ .

El estadístico introducido anteriormente  $U = U_{(s)} - U_{(r)}$ , para  $1 \leq r < s \leq n$ , suele denominarse amplitud o extensión del intervalo  $(X_{(r)}, X_{(s)})$ . Más precisamente, las diferencias

$$\left\{ \begin{array}{l} U_1 = F(X_{(1)}) = U_{(1)} \\ U_2 = F(X_{(2)}) - F(X_{(1)}) = U_{(2)} - U_{(1)} \\ \vdots \\ U_n = F(X_{(n)}) - F(X_{(n-1)}) = U_{(n)} - U_{(n-1)} \\ U_{n+1} = 1 - F(X_{(n)}) = 1 - U_{(n)} \end{array} \right.$$

son denominadas extensiones. Dicho en otros términos

$$U_{(s)} - U_{(r)} = F(X_{(s)}) - F(X_{(r)})$$



es denominada extensión o amplitud del intervalo  $(X_{(r)}, X_{(s)})$ . Así, la extensión de un intervalo aleatorio (basado en estadísticos de orden) es la cantidad de probabilidad de la distribución contenida en dicho intervalo aleatorio.

En (I.7) hemos desarrollado la expresión de la función de densidad de  $U_{(s)} - U_{(r)}$ . Si tenemos en cuenta que

$$U_{(s)} - U_{(r)} = U_{r+1} + U_{r+2} + \dots + U_s \quad 1 \leq r < s \leq n$$

bastará con tomar  $r = i-1$  y  $s = i$ , para obtener la función de densidad marginal de  $U_i$ , con la expresión

$$f_{U_i}(u) = \begin{cases} n(1-u)^{n-1} & 0 < u < 1 \\ 0 & \text{en otro caso} \end{cases}$$

A partir de esta expresión, es inmediato que  $E\{U_i\} = 1/(n+1)$ , que puede ser interpretado como sigue: "Los estadísticos de orden  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  dividen el área bajo la curva definida por la función de densidad en  $(n+1)$  partes, de tal forma que cada parte tiene el mismo área esperada".

La función de densidad conjunta de  $U_1, U_2, \dots, U_n$  es fácil de determinar. Teniendo en cuenta que las transformaciones inversas son

$$\begin{aligned} U_{(1)} &= U_1 \\ U_{(2)} &= U_1 + U_2 \\ &\vdots \\ U_{(n)} &= U_1 + U_2 + \dots + U_n \end{aligned}$$



el Jacobiano de la transformación será 1. Además,  $u \geq 0$  y  $\sum_1^n u_i < 1$ . Por otra parte, ya es conocido que la función de densidad conjunta de  $U_{(1)}, U_{(2)}, \dots, U_{(n)}$  es

$$\varphi(x_1, x_2, \dots, x_n) = \begin{cases} n! & \text{si } 0 < x_1 < x_2 < \dots < x_n < 1 \\ 0 & \text{en otro caso} \end{cases}$$

por lo que la función de densidad de  $U_1, U_2, \dots, U_n$ , será

$$h(x_1, x_2, \dots, x_n) = \begin{cases} n! & \text{si } u_i \geq 0 \text{ y } \sum_1^n u_i < 1 \\ 0 & \text{en otro caso} \end{cases}$$

De otro lado,  $h$  es simétrica en  $U_1, U_2, \dots, U_n$ , así que la suma de cualesquiera  $r$  ( $r < n$ ) extensiones, es la misma, y además coincide con la distribución de  $U_{(r)} = U_1 + U_2 + \dots + U_r$ . De esta forma, la función de densidad de  $\sum_1^r U_i (= U_{(r)})$  viene dada por

$$f_{U_{(r)}}(x) = \begin{cases} n \binom{n-1}{r-1} x^{r-1} (1-x)^{n-r} & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Además, es inmediato que

$$E\{U_{(r)}\} = \sum_{i=1}^r E\{U_i\} = \frac{r}{n+1}$$

A menudo, a la suma de  $r$  cualesquiera extensiones sucesivas se le suele denominar  $r$ -extensión. Para ella tenemos que



$$U_{i+1} + U_{i+2} + \dots + U_{i+r} = U_{(i+r)} - U_{(i)} \quad \text{para } (i+r) \leq n$$

Consideremos ahora el uso de los estadísticos de orden en la construcción de intervalos de confianza para los cuantiles poblacionales. Sea  $X$  una variable aleatoria continua con función de distribución  $F$ . Sea  $0 < p < 1$ . Entonces, el cuantil de orden  $p$ ,  $K_p$ , de la distribución satisface

$$F(K_p) = p$$

Sean  $X_1, X_2, \dots, X_n$   $n$  observaciones independientes de la variable aleatoria  $X$ . El número de  $X_i < K_p$ , es una variable aleatoria que sigue una distribución Binomial con parámetros  $n$  y  $p$ . Similarmente, el número de  $X_i$  que son al menos  $K_p$ , sigue también una distribución Binomial de parámetros  $n$  y  $(1-p)$ .

Sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  el conjunto de estadísticos de orden asociados a la muestra. Entonces,

$$\begin{aligned} P\{X_{(r)} \leq K_p\} &= P\{\text{al menos } r \text{ de los } X_i \text{ sean } \leq K_p\} = \\ &= \sum_{i=r}^n \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

Similarmente,

$$\begin{aligned} P\{X_{(s)} \geq K_p\} &= P\{\text{al menos } n-s+1 \text{ de los } X_i \text{ sean } \geq K_p\} \\ &= P\{\text{al menos } s-1 \text{ de los } X_i \text{ sean } < K_p\} \\ &= \sum_{i=0}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

De las dos expresiones anteriores, tendremos que:



$$\begin{aligned}
P\{X_{(r)} \leq K_p \leq X_{(s)}\} &= P\{X_{(s)} \geq K_p\} - P\{X_{(r)} > K_p\} = \\
&= P\{X_{(r)} \leq K_p\} - 1 + P\{X_{(s)} \geq K_p\} = \\
&= \sum_{i=r}^n \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=0}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} - 1 = \\
&= \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}
\end{aligned}$$

De esta última expresión es fácil determinar un intervalo de confianza para  $K_p$ , una vez que el nivel de confianza ha sido fijado.

En la práctica, uno determina  $r$  y  $s$  tales que  $s-r$  sea lo más pequeño posible, sujeto a la condición de que el nivel es  $1-\alpha$ .

**Ejemplo.** Supongamos que queremos determinar un intervalo de confianza para la mediana ( $p = 1/2$ ) basado en una muestra de tamaño 7, con un nivel de confianza de 0'9. Para ello es suficiente encontrar  $r$  y  $s$  ( $r < s$ ), tales que

$$\sum_{i=r}^{s-1} \binom{7}{i} \left(\frac{1}{2}\right)^7 \geq 0'9$$

Por tanteo, usando la distribución Binomial de parámetros  $n = 7$  y  $p = 1/2$ , vemos que podemos elegir  $s = 7$  y  $r = 2$  o bien  $s = 6$  y  $r = 1$ . En ambos casos,  $s-r$  ( $= 5$ ) es mínimo, y el nivel de confianza es al menos de 0'92.



## COMPORTAMIENTO ASINTOTICO DE LOS ESTADISTICOS DE ORDEN

Como viene siendo habitual, comencemos con una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una población cuya distribución viene dada por una función de distribución  $F$ . Para cada estadístico de orden  $X_{(i)}$ , hemos determinado su distribución exacta. Seguidamente queremos establecer cuál será su comportamiento asintótico.

En general, el comportamiento asintótico de  $X_{(i)}$  dependerá de la posición que ocupe  $i$  respecto de  $n$ , y todo ello teniendo en cuenta la función de distribución  $F$ . Por ejemplo, el comportamiento en muestras grandes del máximo muestral es muy distinto del que tiene el mínimo muestral, o incluso del de la mediana muestral.

Con el propósito de presentar resultados generales de forma sistemática, consideremos la siguiente clasificación que, sin ser exhaustiva, nos permite encuadrar a los distintos estadísticos de orden en tres grupos:

- (i) ESTADISTICOS DE ORDEN EXTREMOS, cuando  $i$  o  $(n-i)$  están fijados y el tamaño muestral  $n \rightarrow \infty$ .
- (ii) ESTADISTICOS DE ORDEN CENTRALES, cuando  $i = [np] + 1$ ,  $0 < p < 1$ , donde  $[.]$  es la parte entera.
- (iii) ESTADISTICOS DE ORDEN INTERMEDIOS, cuando tanto  $i$  como  $(n-i)$  tienden a infinito, pero  $i/n \rightarrow 0$  o  $1$ .

Una advertencia importante que debemos tener presente sobre el comportamiento de los estadísticos de orden extremos, es que si su distribución límite existe, no será normal, y dependerá directamente del comportamiento en las colas de la distribución  $F$ .



### Máximo y Mínimo muestral.

Vamos a presentar una discusión detallada de las posibles distribuciones no degeneradas que podemos encontrar en el límite para  $X_{(n)}$ , para luego tratar, brevemente, con resultados paralelos, el caso de  $X_{(1)}$ . Técnicamente, será suficiente considerar sólo el máximo, dado que el mínimo muestral para una función de distribución  $F$ , tendrá la misma distribución que el opuesto del máximo muestral de la función de distribución  $F^*$ , donde  $F^*(x) = 1-F(-x)$ .

Según hemos obtenido ya anteriormente, la función de distribución de  $X_{(n)}$  es

$$\Phi_n(x) = [F(x)]^n,$$

por lo que  $X_{(n)} \xrightarrow{P} F^{-1}(1)$ , que es el límite superior del soporte de  $F$ . Con la esperanza de obtener una distribución límite no degenerada, trabajaremos teniendo en cuenta normalizaciones o estandarizaciones de  $X_{(n)}$ . Es decir, estudiaremos la distribución de la secuencia  $\{(X_{(n)} - a_n)/b_n; n \geq 1\}$ , donde  $a_n$  representa una variación producida en la localización de  $X_{(n)}$ , y  $b_n > 0$  representa un cambio en la escala. La función de distribución exacta de  $X_{(n)}$  una vez estandarizada, será  $F^n(a_n + b_n x)$ .

El problema que tenemos planteado, tras su nueva reformulación, estará resuelto una vez que hayamos dado respuesta a las siguientes cuestiones:

- (i) ¿Es posible encontrar  $a_n$  y  $b_n > 0$  tales que  $F^n(a_n + b_n x) \rightarrow G(x)$  en todos los puntos de continuidad de una función de distribución  $G(x)$  no degenerada?.
- (ii) ¿Qué clase de función de distribución,  $G$ , puede aparecer como límite?.



(iii) ¿Cómo estarán F y G relacionadas?. Esto es, dada F, ¿podremos identificar G?.

(iv) ¿Cómo elegiremos apropiadamente  $a_n$  y  $b_n$  de (i)?.

Con la idea de responder a las anteriores cuestiones de forma precisa y facilitar la posterior discusión, introducimos las siguientes dos definiciones:

**DEFINICION A.I.4** (*Dominio de máxima atracción*). Una función de distribución, F (discreta o absolutamente continua), se dice que pertenece al "dominio de máxima atracción" de otra función de distribución G no degenerada, si existen sendas sucesiones  $\{a_n\}$  y  $\{b_n > 0\}$ , tales que

$$\lim_{n \rightarrow \infty} F^n(a_n + b_n x) = G(x) \quad (I.10)$$

en todos los puntos de continuidad de G(x). Si lo anterior se mantiene, escribiremos  $F \in \mathcal{D}(G)$ .

Sea W una variable aleatoria cuya función de distribución sea G. Al afirmar que  $F \in \mathcal{D}(G)$ , estamos indicando que  $(X_{(n)} - a_n)/b_n \xrightarrow{d} W$ .

**DEFINICION A.I.5.** Dos funciones de distribución,  $F_1$  y  $F_2$ , se dice que son "del mismo tipo", si existen constantes  $a_0$  y  $b_0 > 0$  tales que

$$F_1(a_0 + b_0 x) = F_2(x)$$

Si dos variables aleatorias  $W_1$  y  $W_2$ , con funciones de



distribución respectivas  $F_1$  y  $F_2$ , están linealmente relacionadas, entonces  $F_1$  y  $F_2$  son del mismo tipo.

Supongamos que  $(X_{(n)} - a_n)/b_n \xrightarrow{d} W$ , siendo  $G$  la función de distribución de  $W$ . Sean  $\{\alpha_n\}$  y  $\{\beta_n > 0\}$  dos sucesiones de números reales tales que  $\beta_n/b_n \rightarrow b_0$  y  $(\alpha_n - a_n)/b_n \rightarrow a_0$  cuando  $n \rightarrow \infty$ . Entonces  $(X_{(n)} - \alpha_n)/\beta_n \xrightarrow{d} (W - a_0)/b_0$ , donde la función de distribución de  $(W - a_0)/b_0$  es  $G(a_0 + b_0 x)$ . Es decir, las distribuciones límite de  $(X_{(n)} - a_n)/b_n$  y  $(X_{(n)} - \alpha_n)/\beta_n$  son del mismo tipo. A la inversa, se puede probar que si  $F^n(a_n + b_n x) \rightarrow G(x)$  y  $F^n(\alpha_n + \beta_n x) \rightarrow G_0(x)$ , entonces existen constantes  $a_0$  y  $b_0 > 0$  tales que  $\beta_n/b_n \rightarrow b_0$  y  $(\alpha_n - a_n)/b_n \rightarrow a_0$  y  $G_0(x) = G(a_0 + b_0 x)$ . (Para una demostración formal de este resultado, se puede consultar Galambos (1987)).

La discusión informal realizada hasta el momento, justifica las siguientes observaciones: Primero, la elección de las constantes de normalización no tiene por qué ser única. Segundo, una función de distribución,  $F$ , no puede estar en el dominio de máxima atracción de más de un tipo de funciones de distribución. O expresado con otras palabras, podemos encontrar varios conjuntos de constantes normalizadoras que den lugar a diversas funciones de distribución límite, pero todas ellas serán del mismo tipo.

Por lo dicho, a partir de este momento dirigiremos nuestra atención a identificar los tipos de funciones  $G$  que pueden ser elegidas para conformarse en límite de funciones de distribución como en (I.10).

Mediante un ingenioso argumento, Fisher y Toppett (1928) identificaron la clase de tales funciones de distribución. Primero, sea  $n = mr$ , donde  $m$  y  $r$  son enteros positivos. Cuando  $m \rightarrow \infty$ , mientras  $F^n(a_n + b_n x) \rightarrow G(x)$ , de la discusión anterior se obtiene que  $[F^m(a_{mr} + b_{mr} x)]^r \rightarrow [G(a_r^0 + b_r^0 x)]^r$ , para ciertas



constantes  $a_r^0$  y  $b_r^0 > 0$ . Así,  $G$  debe ser de tal forma que

$$G^n(a_n^0 + b_n^0 x) = G(x) \quad (I.11)$$

para algunas constantes  $a_n^0$  y  $b_n^0 > 0$ , para todo  $x$  y  $n \geq 1$ .

Las funciones de distribución que satisfacen (I.11), son denominadas funciones de distribución *max-estables*. La solución de (I.11) depende de si  $b_n^0 > 1$ ,  $b_n^0 < 1$ , o  $b_n^0 = 1$ . Frechét (1927) identificó la solución en el primer caso y, poco después, Fisher y Tippett (1928) caracterizaron todas las posibles soluciones de (I.11). Los resultados los recogemos en el siguiente teorema:

**TEOREMA A.I.4.** Si se satisface (I.10), la función de distribución límite,  $G$ , del máximo muestral, convenientemente normalizado, es una de las siguientes:

$$G_1(x; \alpha) = \begin{cases} 0 & x \leq 0 \\ \exp \{-x^{-\alpha}\} & x > 0 \quad ; \quad \alpha > 0 \end{cases}$$

$$G_2(x; \alpha) = \begin{cases} \exp \{-(-x)^\alpha\} & x < 0 \quad ; \quad \alpha > 0 \\ 1 & x \geq 0 \end{cases}$$

$$G_3(x) = \exp \{-\exp(-x)\} \quad -\infty < x < \infty$$

Las dos primeras funciones de distribución límite anteriores,  $G_1$  y  $G_2$ , involucran a un parámetro adicional,  $\alpha$ , que está estrechamente ligado con el comportamiento en las colas de la función de distribución poblacional  $F$ . Las tres funciones de distribución  $G_1$ ,  $G_2$  y  $G_3$  son conocidas con distintos nombres. A la pri-



mera se le suele referir como la de *tipo Frechét*;  $G_2$  es conocida como *tipo Weibull*; y  $G_3$  es a menudo denominada como la función de distribución del *valor extremo*. [Observese que la opuesta de una variable aleatoria de Weibull con parámetro de forma  $\alpha$ , tiene como función de distribución a  $G_2(x; \alpha)$ ].

Aunque sean parecidas a simple vista, las tres funciones de distribución recogidas en el teorema anterior no parecen tener una relación directa. Sin embargo, pueden verse como miembros de una misma familia de funciones de distribución. Es decir, si consideramos la función de distribución  $G(x; \theta)$  definida como:

$$G(x; \theta) = \exp \{-(1 + x\theta^{-1})^{-\theta}\} \quad ; \quad 1 + x\theta^{-1} > 0 \quad ; \quad -\infty < \theta < \infty$$

obtendremos que, para  $\theta > 0$ ,  $G(x; \theta)$  y  $G_1(x; \theta)$  son del mismo tipo; cuando  $\theta < 0$ ,  $G(x; \theta)$  y  $G_2(x; \theta)$  son del mismo tipo; y cuando  $\theta \rightarrow \pm\infty$ ,  $G(x; \theta) \rightarrow G_3(x)$ .

La función de distribución  $G(x; \theta)$  es conocida con el nombre de *función de distribución generalizada del valor extremo*, o bien como *función de distribución del valor extremo en la forma de von Mises*.

Hasta ahora, hemos respondido a la cuestión (ii) que se había propuesto al principio de esta sección, al probar que si  $F \in \mathcal{D}(G)$ , entonces  $G$  es o  $G_1$ , o  $G_2$ , o  $G_3$ . A continuación, pretendemos precisar la respuesta de la cuestión (i), presentando un conjunto de condiciones necesarias y suficientes, impuestas sobre  $F$ , para garantizar que  $F \in \mathcal{D}(G_i)$ ,  $i = 1, 2, 3$ . A su vez, nos plantearemos responder también a la cuestión (iii).



**TEOREMA A.I.5.** (Condiciones necesarias y suficientes de convergencia débil).

- (i)  $F \in \mathcal{D}(G_1)$  si, y sólo si,  $F^{-1}(1) = +\infty$  y existe una constante  $\alpha > 0$ , tal que

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha} [= -\log G_1(x; \alpha)] \quad (I.12)$$

para todo  $x > 0$ . Además, si (I.12) se satisface,  $G_1(x) = G_1(x; \alpha)$ , y escribiremos  $F \in \mathcal{D}(G_1(x; \alpha))$ .

- (ii)  $F \in \mathcal{D}(G_2)$  si, y sólo si,  $F^{-1}(1)$  es finito y existe una constante  $\alpha > 0$  tal que para todo  $x > 0$ ,

$$\lim_{\varepsilon \rightarrow 0^+} \frac{1 - F(F^{-1}(1) - \varepsilon x)}{1 - F(F^{-1}(1) - \varepsilon)} = x^\alpha [= -\log G_2(-x; \alpha)] \quad (I.13)$$

Si (I.13) se satisface,  $G_2(x) = G_2(x; \alpha)$ , y escribiremos  $F \in \mathcal{D}(G_2(x; \alpha))$ .

- (iii)  $F \in \mathcal{D}(G_3)$  si, y sólo si,  $E(X/X > c)$  es finita para algún  $c < F^{-1}(1)$ , y para todo real  $x$ ,

$$\lim_{t \rightarrow F^{-1}(1)} \frac{1 - F(t + xE(X-t/X > t))}{1 - F(t)} = \exp(-x) [= -\log G_3(x)] \quad (I.14)$$

donde  $X$  representa a la variable aleatoria poblacional cuya función de distribución es  $F$ .



La esperanza condicionada,  $E(X-t/X>t)$ , que aparece en (I.14), es conocida en el ambiente de la teoría de la fiabilidad como "vida media residual".

El resultado recogido en el teorema A.I.5, no es más que una adaptación de los resultados dados en los teoremas 2.1.1-2.1.3 y teorema 2.4.1 de Galambos (1987).

Las condiciones necesarias y suficientes dadas en el teorema anterior, son equivalentes a las condiciones originales usadas por Gnedenko (1943) para probar los resultados sobre dominios de atracción. Adviértase que cuando  $G = G_1$ , el límite superior del soporte de  $F$  será infinito, y cuando  $G = G_2$ , este será necesariamente finito. Este simple resultado será en muchas ocasiones útil para descartar, con una rápida comprobación, a  $G_1$  o  $G_2$  como posible distribución límite.

Las condiciones necesarias y suficientes recogidas en el teorema A.I.5 son, a menudo, de muy difícil comprobación. Es por esta razón por lo que presentamos a continuación un conjunto de condiciones suficientes debidas a von Mises (1936), que sin ser muy restrictivas, son de fácil verificación. Advertamos que este conjunto de condiciones son sólo aplicables al caso de distribuciones poblacionales absolutamente continuas.

**TEOREMA A.I.6** (condiciones suficientes de von Mises para convergencia débil).

Sea  $F$  una función de distribución absolutamente continua, y sea  $h(x) = f(x)/\{1 - F(x)\}$ .

(i) Si  $h(x) > 0$  para  $x$  grande, y para algún  $\alpha > 0$

$$\lim_{x \rightarrow \infty} xh(x) = \alpha \quad (\text{I.15})$$



entonces,  $F \in \mathcal{D}(G_1(x; \alpha))$ .

(ii) Si  $F^{-1}(1) < \infty$  y para algún  $\alpha > 0$ ,

$$\lim_{x \rightarrow F^{-1}(1)} (F^{-1}(1) - x) h(x) = \alpha \quad (\text{I.16})$$

entonces,  $F \in \mathcal{D}(G_2(x; \alpha))$ .

(iii) Supongamos que  $h(x)$  es distinta de cero y diferenciable para  $x$  próximo a  $F^{-1}(1)$  (o para  $x$  grande si  $F^{-1}(1) = \infty$ ). Entonces,  $F \in \mathcal{D}(G_3)$  si

$$\lim_{x \rightarrow F^{-1}(1)} \frac{d}{dx} \left\{ \frac{1}{h(x)} \right\} = 0 \quad (\text{I.17})$$

Una simple demostración de estos resultados se recoge en de Haan (1976).

Finalmente, respondamos a la última cuestión determinando las posibles elecciones que podemos hacer de las constantes normalizadoras  $a_n$  y  $b_n$ . Tan importante es determinar sus valores, como probar su existencia. Estas constantes, mientras no sean únicas, dependerán del tipo de función de distribución que sea  $G$ . En el siguiente teorema indicamos, en general, cuáles son las elecciones convenientes de las constantes normalizadoras.

**TEOREMA A.I.7.** (Constantes de normalización).

Para conseguir que  $F^n(a_n + b_n x) \longrightarrow G(x)$ , en función de quien sea  $G$ , podremos elegir  $a_n$  y  $b_n > 0$ , como:



(i)  $a_n = 0$  ;  $b_n = F^{-1}(1 - n^{-1})$ , si  $G = G_1$ .

(ii)  $a_n = F^{-1}(1)$  ;  $b_n = F^{-1}(1) - F^{-1}(1 - n^{-1})$ , si  $G = G_2$ .

(iii)  $a_n = F^{-1}(1 - n^{-1})$  ;  $b_n = E(X - a_n / X > a_n)$  o

$$b_n = F^{-1}(1 - (ne)^{-1}) - F^{-1}(1 - n^{-1}) \text{ si } G = G_3.$$

Además, si se satisface la condición (I.17), podemos tomar  $b_n = \{nf(a_n)\}^{-1}$ .

Concluido el estudio del comportamiento asintótico del máximo muestral, presentamos, de forma breve debido al paralelismo ya mencionado, los correspondientes resultados para el mínimo muestral. El siguiente teorema presentará una clase de funciones de distribución límite para el mínimo muestral, una vez que éste ha sido normalizado. Continuaremos dando una serie de condiciones necesarias y suficientes impuestas sobre  $F$  para que ésta pertenezca al "dominio de máxima atracción", a la vez que se sugieren algunas elecciones convenientes para las constantes normalizadoras.

**TEOREMA A.I.8.** Supongamos que existen constantes  $a_n^*$  y  $b_n^* > 0$ , y una variable aleatoria no degenerada  $W^*$  tales que  $(X_{(1)} - a_n^*)/b_n^* \xrightarrow{d} W^*$ . Entonces, la función de distribución,  $G^*$ , de la variable aleatoria  $W^*$ , deberá ser una de las siguientes:

(i)  $G_1^*(x; \alpha) = 1 - G_1(-x; \alpha)$ .

(ii)  $G_2^*(x; \alpha) = 1 - G_2(-x; \alpha)$ .

(iii)  $G_3^*(x) = 1 - G_3(-x)$ ,



donde  $G_1$ ,  $G_2$  y  $G_3$  han sido dadas anteriormente. [Adviértase que  $G_2^*(x; \alpha)$  no es más que la función de distribución de Weibull con parámetro de forma  $\alpha$ ].

**TEOREMA A.I.9.** (condiciones de convergencia y constantes de normalización).

- (i)  $F \in \mathcal{D}(G_1^*(x; \alpha))$ , es decir, la función de distribución de  $W^*$  es  $G_1^*(x; \alpha)$ , si, y sólo si,  $F^{-1}(0) = -\infty$ , y

$$\lim_{t \rightarrow -\infty} \frac{F(tx)}{F(t)} = x^{-\alpha}$$

para todo  $x > 0$ . Se pueden elegir  $a_n^* = 0$  y  $b_n^*$  el valor absoluto de  $F^{-1}(1/n)$ .

- (ii)  $F \in \mathcal{D}(G_2^*(x; \alpha))$ , si, y sólo si,  $F^{-1}(0)$  es finito y

$$\lim_{\varepsilon \rightarrow 0^+} \frac{F(F^{-1}(0) + \varepsilon x)}{F(F^{-1}(0) + \varepsilon)} = x^\alpha$$

para todo  $x > 0$ . En este caso, podremos elegir  $a_n^* = F^{-1}(0)$  y  $b_n^* = F^{-1}(1/n) - F^{-1}(0)$ .

- (iii)  $F \in \mathcal{D}(G_3^*(x))$ , si, y sólo si,  $E(X/X \leq c)$  es finita para algún  $c > F^{-1}(0)$  y

$$\lim_{t \rightarrow F^{-1}(0)} \frac{F(t + xE(t-X/X \leq t))}{F(t)} = \exp(x)$$



para todo  $x$  real, donde la variable aleatoria  $X$  tiene a  $F$  como función de distribución. Si  $F \in \mathcal{D}(G_3^*(x))$ , tomamos como elecciones convenientes de las constantes de normalización a  $a_n^* = F^{-1}(1/n)$  y  $b_n^* = E(a_n^* - X/X \leq a_n^*)$ .

**Ejemplo.** (Distribución normal estándar).

Consideremos el caso en que

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2}t^2\right) dt$$

es la función de distribución normal estándar. Dado que  $F^{-1}(1) = +\infty$ , afirmamos que  $\Phi \notin \mathcal{D}(G_2)$ .

Por otra parte, y haciendo uso de la regla de l'Hospital, tendremos que, para  $x > 0$ ,

$$\lim_{t \rightarrow \infty} \frac{1 - \Phi(tx)}{1 - \Phi(t)} = \lim_{t \rightarrow \infty} \frac{x \phi(tx)}{\phi(t)} = \lim_{t \rightarrow \infty} \frac{x (1/\sqrt{2\pi}) \exp\{-(1/2)(tx)^2\}}{(1/\sqrt{2\pi}) \exp\{-(1/2)t^2\}} =$$

$$\lim_{t \rightarrow \infty} \frac{x \exp\{-(1/2)(tx)^2\}}{\exp\{-(1/2)t^2\}} = \lim_{t \rightarrow \infty} x \exp\{-(1/2)t^2(x^2 - 1)\} = \begin{cases} \infty & \text{para } x < 1 \\ 1 & \text{para } x = 1 \\ 0 & \text{para } x > 1 \end{cases}$$

con lo que no se satisface la condición (I.12), y por tanto podremos afirmar que  $\Phi \notin \mathcal{D}(G_1)$ . La única posibilidad que nos queda es  $G_3$ .

La condición necesaria y suficiente (I.14) es difícil de probar, por lo que intentaremos ver si se satisface la condición



suficiente (I.17) de von Mises.

$$h(x) = \frac{\phi(x)}{1-\Phi(x)} \Rightarrow \frac{1}{h(x)} = \frac{1-\Phi(x)}{\phi(x)} \Rightarrow \frac{d}{dx} \left\{ \frac{1-\Phi(x)}{\phi(x)} \right\} =$$

$$= \frac{-[\phi(x)]^2 - (1-\Phi(x)) \phi'(x)}{[\phi(x)]^2}$$

Teniendo en cuenta que en la distribución normal estándar,  
 $\phi'(x) = -x \phi(x)$ ,

$$\frac{d}{dx} \left\{ \frac{1-\Phi(x)}{\phi(x)} \right\} = \frac{x (1-\Phi(x)) \phi(x)}{[\phi(x)]^2} - 1 = \frac{x (1-\Phi(x))}{\phi(x)} - 1 \quad (I.18)$$

De esta forma,

$$\lim_{x \rightarrow F^{-1}(1)} \frac{d}{dx} \left\{ \frac{1}{h(x)} \right\} = \lim_{x \rightarrow +\infty} \left\{ \frac{x (1-\Phi(x))}{\phi(x)} - 1 \right\} = \left\{ \lim_{x \rightarrow +\infty} \frac{x (1-\Phi(x))}{\phi(x)} \right\} - 1$$

Aplicando dos veces la regla de l'Hospital, obtendremos:

$$\lim_{x \rightarrow F^{-1}(1)} \frac{d}{dx} \left\{ \frac{1}{h(x)} \right\} = \left\{ \lim_{x \rightarrow +\infty} \frac{(1-\Phi(x)) - x \phi(x)}{-x \phi(x)} \right\} - 1 = \lim_{x \rightarrow +\infty} \frac{\Phi(x)-1}{x \phi(x)} =$$

$$= \lim_{x \rightarrow +\infty} \frac{\phi(x)}{\phi(x) - x^2 \phi(x)} = \lim_{x \rightarrow +\infty} \frac{1}{1 - x^2} = 0$$

Que como recordamos, es la condición suficiente de von Mises, y por tanto, podemos afirmar que  $\Phi \in \mathcal{D}(G_3)$ .

A su vez, por la simetría de la distribución normal es-



tándar, es claro que  $\Phi \in \mathcal{D}(G_3^*)$ .

Respecto de las constantes de normalización digamos que, dado que la función cuantil  $\Phi^{-1}(u)$  no tiene una expresión conocida,  $a_n = \Phi^{-1}(1-n^{-1})$  no puede ser dada de forma explícita, aunque para cada  $n$  dado puede ser calculada de forma aproximada. En diversos trabajos se ha sugerido que se pueden tomar:

$$a_n = \sqrt{2 \log(n)} - \frac{1}{2} \frac{\log(4\pi \log(n))}{\sqrt{2 \log(n)}}$$

$$b_n = \frac{1}{\sqrt{2 \log(n)}}$$

Estas constantes de normalización no son únicas, ni tan siquiera las mejores. Pero, ¿cómo podemos garantizar que unas constantes elegidas son las idóneas de entre todas las existentes?. A este respecto, Hall (1979) ha probado que se obtiene un mejor grado de convergencia de  $\sup_{-\infty < x < \infty} |\Phi^n(a_n + b_n x) - G_3(x)|$  cuando tanto  $a_n$  y  $b_n$  son elegidas bajo la condición  $n \varphi(a_n) = a_n$  y  $b_n = a_n^{-1}$ . Veamos que las elecciones sugeridas en el teorema A.I.7 verifican, aproximadamente, esta relación.

A partir de (I.18), y visto que la condición (I.17) se mantiene, podemos inferir que para  $x$  grande,

$$1 - \Phi(x) \approx \phi(x)/x.$$

Así, en lugar de usar  $a_n = \Phi^{-1}(1-n^{-1})$  o  $\Phi(a_n) = 1-(1/n)$ , podremos usar  $\phi(a_n)/a_n = 1/n$ , o, equivalentemente,  $n \phi(a_n) = a_n$ . Además, dado que se mantiene la condición suficiente de von Mises, se nos permite tomar  $b_n = \{n \phi(a_n)\}^{-1}$ , lo que significa que



$b_n = a_n^{-1}$ , con la elección que se ha hecho de  $a_n$ .

Hall probó que incluso con esta elección óptima, el grado de convergencia es bastante lento (del orden de  $\log n$ ). Fisher y Tippett (1928) fueron los primeros en notar la lentitud en la convergencia de  $\Phi^n(a_n + b_n x)$  a  $G_3(x)$ . Probaron que los cuatro primeros momentos de la función de distribución  $\Phi^n(a_n + b_n x)$  estaban más próximos a los de  $G_2(x; \alpha)$ , para una conveniente elección de  $\alpha$ , que aquellos de  $G_3(x)$ .

#### otros estadísticos de orden extremos

En este epígrafe abordaremos la cuestión de determinar la posible distribución límite no degenerada para el  $i$ -ésimo estadístico de orden extremo por arriba, y para el  $i$ -ésimo estadístico de orden extremo por abajo. Estudiaremos también el comportamiento asintótico conjunto de los  $i$ -ésimos estadísticos de orden extremos (por arriba y por abajo).

Comencemos recordando un resultado elemental en cálculo que será extremadamente útil en nuestra discusión.

**Lema.** Sea  $\{c_n ; n \geq 1\}$  una sucesión de números reales.  $c_n \rightarrow c$  cuando  $n \rightarrow \infty$ , si, y sólo si,  $(1 - c_n/n)^n \rightarrow \exp(-c)$ .

Para  $x$  fijo, definimos

$$c_n \equiv n \{1 - F(a_n + b_n x)\}$$

donde  $a_n$  y  $b_n > 0$ , son constantes de normalización verificando (I.10), esto es,  $F^n(a_n + b_n x) \rightarrow G(x)$ .



Lo anterior puede ser expresado en términos de  $c_n$  como

$$(1 - c_n/n)^n \longrightarrow \exp(\log(G(x)))$$

Entonces, en virtud del lema anterior, podemos decir que (I.10) se mantiene, si, y sólo si,  $c_n \longrightarrow (-\log(G(x)))$ . De esta forma, (I.10) es equivalente a que  $c_n$ , como ha quedado definido, converja a  $-\log(G(x))$  en todos los puntos de continuidad de  $G(x)$ , cuando  $n \rightarrow \infty$ .

Como ya ha sido determinada la función de distribución exacta del estadístico de orden  $X_{(n-i+1)}$ , escribamos:

$$\begin{aligned} P(X_{(n-i+1)} \leq a_n + b_n x) &= \sum_{r=n-i+1}^n \binom{n}{r} [F(a_n + b_n x)]^r [1-F(a_n + b_n x)]^{n-r} = \\ &= \sum_{s=0}^{i-1} \binom{n}{s} (c_n/n)^s (1-c_n/n)^{n-s} . \end{aligned}$$

Como observamos claramente, se trata de la función de distribución de una variable aleatoria Binomial  $(n; c_n/n)$ , evaluada en  $(i-1)$ , que, a su vez, convergerá a la función de distribución de una variable aleatoria de Poisson de parámetro  $(-\log(G(x)))$ , si, y sólo si, se satisface (I.10). (Para probar este resultado basta con tener en cuenta las condiciones de convergencia de la distribución Binomial a la distribución de Poisson). Este argumento tan simple permite establecer el siguiente resultado:

**TEOREMA A.I.10.** (Distribución asintótica de un estadístico de orden extremo).

Para cualquier número real  $x$ ,  $P\{X_{(n)} \leq a_n + b_n x\} \longrightarrow G(x)$  cuando



$n \rightarrow \infty$ , si, y sólo si, para cualquier  $i$  fijo, con  $i > 1$  se tiene que

$$P\{X_{(n-i+1)} \leq a_n + b_n x\} \longrightarrow \sum_{r=0}^{i-1} G(x) \{-\log(G(x))\}^r / r! \quad (I.19)$$

O dicho de otra manera, (I.10) se satisface, o bien  $F \in \mathcal{D}(G)$ , si, y sólo si, para algún  $i$  finito, se verifica (I.19) para todo  $x$ .

Este teorema establece una fuerte conexión entre el comportamiento asintótico de  $X_{(n)}$  y de  $X_{(n-i+1)}$ . Las constantes de normalización serán las mismas, y la función  $G(x)$  en (I.10) y en (I.19) coinciden. Así,  $(X_{(n-i+1)} - a_n) / b_n$  convergerá en distribución a una variable aleatoria cuya función de distribución está recogida en (I.19), con  $G = G_1$ , o  $G_2$ , o  $G_3$ . Por ello, el teorema A.I.5 proporciona condiciones necesarias y suficientes, y el teorema A.I.6, condiciones suficientes para esta convergencia. Las constantes de normalización que usaremos en (I.19) serán las proporcionadas en el teorema A.I.7.

Por ejemplo, cuando  $F \in \mathcal{D}(G_3)$ , como es el caso de la distribución normal,

$$\lim_{n \rightarrow \infty} P[(X_{(n-i+1)} - a_n) / b_n \leq x] = \exp \{-\exp(-x)\} \sum_{r=0}^{i-1} \frac{\exp(-rx)}{r!}$$

donde  $a_n = F^{-1}(1-n^{-1})$  y  $b_n = E(X - a_n / X > a_n)$ .

Los resultados para el  $i$ -ésimo estadístico de orden (por abajo), son similares. Cuando  $F \in \mathcal{D}(G^*)$ , para cualquier  $i > 1$ ,



$$P[X_{(i)} > a_n^* + b_n^* x] \longrightarrow \sum_{r=0}^{i-1} \{1-G^*(x)\} \{-\log[1-G^*(x)]\}^r / r!$$

Por último, para concluir el estudio del comportamiento asintótico de los estadísticos de orden extremos, indiquemos que, mediante un análisis del comportamiento asintótico conjunto de los estadísticos de orden extremos, cosa que excede de la intención de este trabajo, podemos obtener como distribuciones marginales las funciones de densidad de la distribución límite de los estadísticos de orden extremos. Así, la función de densidad de la distribución límite de  $(X_{(n-i+1)} - a_n) / b_n$  es

$$g_{(i)}(x_i) = \frac{\{-\log G(x_i)\}^{i-1}}{(i-1)!} g(x_i)$$

y, similarmente, para  $(X_{(i)} - a_n^*) / b_n^*$ , la función de densidad límite será

$$g_{(i)}^*(x_i) = \frac{\{-\log [1 - G^*(x_i)]\}^{i-1}}{(i-1)!} g(x_i)$$

#### estadísticos de orden centrales e intermedios

Para  $0 < p < 1$ , sea  $i = [np] + 1$ . Entonces  $X_{(i)}$  representa al  $p$ -ésimo cuantil muestral, y será un estadístico de orden central.

Cuando la función de distribución  $F$  es absolutamente



continua, con función de densidad finita y positiva en  $F^{-1}(p)$ ,  $X_{(i)}$ , después de una conveniente normalización, será asintóticamente normal. Este es el resultado básico que informa del comportamiento asintótico de los estadísticos de orden centrales e intermedios, y que enunciaremos formalmente y demostraremos a continuación.

**TEOREMA A.I.11.** (Distribución asintótica de un estadístico de orden central).

Para  $0 < p < 1$ , sea  $F$  absolutamente continua, con función de densidad  $f$  que suponemos positiva y continua en  $F^{-1}(p)$ . Para  $i = [np] + 1$ , cuando  $n \rightarrow \infty$ , se tiene que

$$\sqrt{n} f[F^{-1}(p)] \frac{(X_{(i)} - F^{-1}(p))}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1) \quad (I.20)$$

Demostración.

Comenzaremos probando que (I.20) es cierto cuando  $F$  es la función de distribución de una variable aleatoria uniforme en el intervalo  $(0,1)$ , para luego usar la transformación integral inversa  $X_{(i)} \stackrel{d}{=} F^{-1}(U_{(i)})$ , que nos permitirá probar el resultado para cualquiera que sea  $F$  función de distribución que verifique las condiciones establecidas en las hipótesis del teorema.

Para probar que

$$\sqrt{n} (U_{(i)} - p) \xrightarrow{d} N(0;p(1-p)) \quad (I.21)$$

cuando  $n$ , y por tanto  $i$ , se aproxima a infinito, recordemos que  $U_{(i)}$  se distribuye según una variable aleatoria Beta  $(i;n-i+1)$ .



Por esto,  $U_{(i)}$  podrá expresarse como

$$U_{(i)} = \frac{A_n}{A_n + B_n} \quad (I.22)$$

con  $A_n = \sum_{r=1}^i Z_r$  y  $B_n = \sum_{r=i+1}^{n+1} Z_r$ , donde las  $Z_r$  son variables aleatorias

independientes e idénticamente distribuidas según una  $\text{Exp}(1)$ .

Dado que  $E(Z_r) = 1$  y  $\text{Var}(Z_r) = 1$ , aplicando el teorema central del límite se obtiene que

$$\{(A_n - i)/\sqrt{i}\} \xrightarrow{d} N(0,1)$$

Si recordamos que  $i = [np] + 1$ , obtendremos

$$\{(A_n - i)/\sqrt{n}\} \xrightarrow{d} A$$

donde  $A$  es una variable aleatoria  $N(0,p)$ . Similarmente podemos concluir que

$$\{(B_n - (n-i+1))/\sqrt{n}\} \xrightarrow{d} B$$

donde  $B$  es una variable aleatoria  $N(0,1-p)$ .

Dado que  $A_n$  y  $B_n$  son independientes para todo  $n$ , las variables límite,  $A$  y  $B$ , son también independientes. Esto nos lleva a afirmar que

$$C_n = \frac{1}{\sqrt{n}} \{(1-p)(A_n - i) - p(B_n - (n-i+1))\}$$

converge en distribución a una variable aleatoria  $N(0,p(1-p))$ .



Ahora, usando (I.22), podemos escribir

$$\sqrt{n} (U_{(i)} - p) = \frac{C_n - \{(i-np-1)/\sqrt{n}\}}{\{(A_n + B_n)/n\}}$$

El numerador de esta expresión converge en distribución a una variable aleatoria  $N(0, p(1-p))$ . Por la ley débil de los grandes números, el denominador converge en probabilidad a 1. Por tanto, y en virtud del teorema de Slutsky, se verifica (I.21).

Para una función de distribución,  $F$ , arbitraria, teniendo en cuenta que  $X_{(i)} \stackrel{d}{=} F^{-1}(U_{(i)})$ , y haciendo uso del desarrollo en series de Taylor para  $X_{(i)}$ , podremos escribir:

$$X_{(i)} \stackrel{d}{=} F^{-1}(p) + (U_{(i)} - p) \{f[F^{-1}(D_n)]\}^{-1}$$

donde  $D_n$  es una variable aleatoria que está entre  $U_{(i)}$  y  $p$ .

Esta expresión permite ser reordenada en los siguientes términos:

$$\sqrt{n} \{X_{(i)} - F^{-1}(p)\} \stackrel{d}{=} \sqrt{n} (U_{(i)} - p) \{f[F^{-1}(D_n)]\}^{-1} \quad (I.23)$$

Dado que  $f$  es continua en  $F^{-1}(p)$ , se tiene que  $f[F^{-1}(D_n)] \xrightarrow{P} f[F^{-1}(p)]$  cuando  $n \rightarrow \infty$ . Si aplicamos ahora (I.21) y el teorema de Slutsky a (I.23), concluiremos la demostración de la normalidad asintótica de  $X_{([np]+1)}$  como la recogida en el enunciado del teorema. ■

El teorema anterior permite afirmar que el cuantil muestral  $X_{(i)}$  es un estimador consistente de  $F^{-1}(p)$ . Además, el teore-



ma puede usarse para obtener un intervalo de confianza aproximado para  $F^{-1}(p)$  siempre que, o bien  $f$  sea totalmente conocida alrededor de  $F^{-1}(p)$ , o esté disponible un buen estimador de  $f(F^{-1}(p))$ .

Por la aproximación de David-Johnson, sabemos también que

$$E(X_{(i)}) \approx F^{-1}(p)$$

y que

$$\text{Var}(X_{(i)}) \approx \frac{p(1-p)}{n[f(F^{-1}(p))]^2}$$

Por lo que concluiremos que  $X_{(i)}$  es un estimador de  $F^{-1}(p)$  asintóticamente insesgado, consistente y asintóticamente normal. [Obsérvese que este resultado, usando la aproximación de David-Johnson, coincide con el desarrollado en el capítulo referente al análisis de la variabilidad en el gráfico probabilístico normal, salvo que aquí se ha truncado en el primer término].

Por último, cerraremos esta sección con una breve discusión sobre el comportamiento asintótico de los estadísticos de orden intermedios. Hemos llamado a  $X_{(n-i+1)}$  un estadístico de orden intermedio superior, si  $i \rightarrow \infty$  pero  $i/n \rightarrow 0$  cuando  $n \rightarrow \infty$ . La distribución límite de  $X_{(n-i+1)}$  depende del comportamiento de  $i$  respecto de  $n$ , y podrá ser normal o no normal.

Presentamos a continuación un interesante resultado debido a Falk (1989), que establece la normalidad asintótica de  $X_{(n-i+1)}$ .



**TEOREMA A.I.12.** Sea  $F$  una función de distribución absolutamente continua satisfaciendo una de las condiciones del teorema de von Mises. Supongamos que  $i \rightarrow \infty$  y  $i/n \rightarrow 0$  cuando  $n \rightarrow \infty$ . Existen entonces constantes normalizadoras  $a_n$  y  $b_n > 0$  tales que

$$(X_{(n-i+1)} - a_n)/b_n \xrightarrow{d} N(0,1)$$

El teorema propone elegir

$$a_n = F^{-1}(1-i/n)$$

y

$$b_n = \frac{\sqrt{i}}{\{n f(a_n)\}}$$

Un resultado similar se obtiene para  $X_{(i)}$  cuando  $i \rightarrow \infty$  y  $i/n \rightarrow 0$ .



## APENDICE II

### APLICACIONES INFORMATICAS

En el presente apéndice se recogen una serie de aplicaciones informáticas que se han desarrollado durante la elaboración de la memoria, y que han sido usadas para obtener algunos de los resultados que en ésta aparecen.

Para la obtención de tales aplicaciones, se ha hecho uso de dos lenguajes: el S-PLUS y el UBASIC.

El S-PLUS es un lenguaje y un entorno de programación interactivo orientado al análisis de datos y obtención de gráficos.

El UBASIC es un lenguaje de programación (traductor) de precisión infinita. Sus raíces las encuentra en el clásico BASIC, lo que hace de este lenguaje que sea sencillo de entender y de fácil manejo.



PROGRAMAS EN S-PLUS

cálculo aproximado de la varianza de los estadísticos de orden de una distribución normal.

```
function(x)
{
  n <- length(x)
  m <- mean(x)
  sd <- sqrt(var(x))
  e <- c((1:n)/(n + 1))
  d <- 1 - e
  q <- qnorm(e, m, sd)
  p <- dnorm(q, m, sd)
  q <- (q - m)/sd
  a <- 2 * (d - e) * (q/p^3) + (e * d) * ((1 + (5/2) *
    q^2)/p^4)
  b <- - a + 2 * ((d - e)^2 - (e * d)) * ((1 + (11/4) *
    q^2)/p^4)
  b <- b + (e * d) * (d - e) * (((44/3) * q + (48/3) *
    q^3)/p^5)
  b <- b + (1/12) * (e^2 * d^2) * ((26 + 200 * q^2 + 128 *
    q^4)/p^6)
  va <- ((e * d)/(n + 2)) * (1/p^2) + ((e * d)/(n + 2)^2)
    * a + ((e * d)/(n + 2)^3) * b
  va <- sqrt(va)
  va
}
```



determinación de la variabilidad de los puntos sobre la recta del grafico probabilístico normal basado en la varianza de los estadísticos de orden

```
function(x)
{
  x <- sort(x)
  zz <- qqnorm(x, plot = F)
  rec <- lsfit(zz$x, zz$y)
  i <- rec$coef[2] * zz$x
  i <- rec$coef[1] + i
  su <- i + banda1(x)
  il <- i - banda1(x)
  xlim <- range(zz$x)
  ylim <- range(zz$y)
  qqnorm(x, xlim = xlim, ylim = ylim, xlab = "cuantiles
      N(0,1)", ylab = "observaciones")
  par(new = T)
  abline(lsfit(zz$x, zz$y))
  par(new = T)
  plot(zz$x, su, xlim = xlim, ylim = ylim, xlab = "", ylab
      = "", type = "l")
  par(new = T)
  plot(zz$x, il, xlim = xlim, ylim = ylim, xlab = "", ylab
      = "", type = "l")
}
```



determinación de la banda de confianza en el grafico probabilístico normal, basada en la distribución exacta de los estadísticos de orden de una distribución normal.

```
function(x)
{
  x <- sort(x)
  m <- mean(x)
  sd <- sqrt(var(x))
  n <- length(x)
  for(i in 1:n) {
    a[i] <- qbeta(0.025, i, n - i + 1)
    b[i] <- qbeta(0.975, i, n - i + 1)
    a[i] <- qnorm(a[i])
    b[i] <- qnorm(b[i])
  }
  zz <- qqnorm(x, plot = F)
  ss <- lsfit(zz$x, zz$y)
  infe <- a * ss$coef[2]
  infe <- infe + ss$coef[1]
  supe <- b * ss$coef[2]
  supe <- supe + ss$coef[1]
  xlim <- range(zz$x)
  ylim <- range(zz$y, supe, infe)
  qqnorm(x, xlim = xlim, ylim = ylim)
  par(new = T)
  abline(m, sd)
```



```

par(new = T)
plot(zz$x, infe, type = "l", xlim = xlim, ylim = ylim,
      xlab = "", ylab = "")
par(new = T)
plot(zz$x, supe, type = "l", xlim = xlim, ylim = ylim,
      xlab = "", ylab = "")
}

```

#### PROGRAMAS EN UBASIC

**Determinación de los polinomios que definen a las medianas de los estadísticos de orden de una uniforme estándar, y de las raíces de éstos**

```

10 point 30:X0=1/2
20 cls:cls
30 input "TAMAÑO DE LA MUESTRA";N
40 rem determinación de cada uno de los polinomios.
50 for K=int(N/2)+1 to N
60     clr Cons
70     dim Co(N-K+1)
80     Co(0)=-0.5
90     print K;".....";Co(0);
100    Cons=! (N)/(! (K-1)*! (N-K))
110    for I=1 to N-K+1
120        Co(I)=Cons*combi(N-K,I-1)*(-1)^(I-1)/(K+I-1)
130        if sgn(Co(I))=+1 then print "+";
140        print Co(I);"*X^";K+I-1;

```



```

150     next I
160     print:print "PULSE UNA TECLA"
170     Pul=input$(1)
180     ' APLICACION DEL METODO DE BISECCION
190     A=X0:B=1
200     gosub *Valorp1(Co(),N,K,A,&BX)
210     PA=BX
220     if PA=0 then X0=A:goto *Fin
230     gosub *Valorp1(Co(),N,K,B,&BX)
240     Pb=BX
250     while ((B-A)>0.00000000001 or abs(BX)>0.00000000001)
260         X0=A+(B-A)/2
270         gosub *Valorp1(Co(),N,K,X0,&BX)
280         P0=BX
290         if P0=0 goto *Fin
300         if P0*PA<0 then B=X0:Pb=P0
310             :else A=X0:PA=P0
320         endif
330     wend
340     X0=A:' +(B-A)/2
350     gosub *Valorp2(Co(),N,K,X0,&BX,&Cx)
360     while abs(BX)>0.000000000001
370         X0=X0-(BX/Cx)
380         gosub *Valorp2(Co(),N,K,X0,&BX,&Cx)
390     wend
400     *Fin:print:print "raiz en el intervalo (0,1)=";X0
410     erase Co()
420     print:print "PULSE UNA TECLA"
430     Pul=input$(1)
440 next K
450 end
460 *Valorp1(Co(),N,K,X0,&BX)
470     local I
480     BX=Co(N-K+1)
490     for I=(N-K) to 1 step (-1)

```



```

500             BX=BX*X0+Co(I)
510         next I
520         BX=BX*X0^K+Co(0)
530     return
540 *Valorp2(Co(),N,K,X0,&BX,&Cx)
550         local I,H
560         BX=Co(N-K+1)
570         Cx=Co(N-K+1)
580         for I=(N-K) to 1 step (-1)
590             BX=BX*X0+Co(I)
600             Cx=Cx*X0+BX
610         next I
620         for H=1 to K-1
630             BX=BX*X0
640             Cx=Cx*X0+Cx
650         next H
660         BX=BX*X0+Co(0)
670     return

```







## BIBLIOGRAFIA

- Ahmad, I.A. (1993), Modification of some goodness-of-fit statistics to yield asymptotically normal null distributions, *Biometrika*. **80** (2), 466-472.
- Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley & sons. New York (segunda ed.)
- Andrews, D.F., Gnanadesikan, R. y Warner, J.L. (1973), Methods for Assessing Multivariate Normality, en *Multivariate Analysis III*, P.R. Krishnaiah, ed. Academic Press, New York. 95-116.
- Anscombe, F.J. y Glynn, W.J. (1983), Distribution of the kurtosis  $b_2$  for normal statistics, *Biometrika*. **70**, 227-234.
- Arnold, B.C., Balakrishnan, N. y Nagaraja, H.N. (1992). *A first course in order statistics*. John Wiley & Sons, Inc.
- Barnett, V. (1975), Probability Plotting Methods and Order Statistics. *Applied Statistics*. **24** (1), 95-108.
- Barnett, V. (1976), Convenient Probability Plotting Positions for the Normal Distribution, *Applied Statistics*, **25** (1), 47-50.



- Basak, I., Balch, W.R. y Basak, P (1992), Skewness: asymptotic critical values for a test related to Pearson's measure, *Journal of Applied Statistics*, **19** (4), 479-487.
- Becker, R.A., Chambers, J.M. y Wilks, A.R. (1988). *The new S Language*. Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.
- Bernard, A y Bos-Levenbach, E.C. (1953), Het uitzetten van waarneminger op waarschiyn-lijkheidspapier. *Statistica*. **7**, 163-173.
- Blom, G (1958). *Statistical Estimates and Transformed Beta-Variables*. Wiley. New York.
- Bowman, A.W. y Foster, P.J. (1993), Adaptive Smoothing and Density-Based Test of Multivariate Normality, *Journal of th American Statistical Association*, **88** (422), 529-537.
- Bowman, K.O. y Shenton, B.R. (1975), Omnibus test contours for departures from normality based on  $\sqrt{b_1}$  and  $b_2$ . *Biometrika*. **62**, 243-250.
- Bowman, K.O. y Shenton, L.R. (1986), Moment ( $\sqrt{b_1}$ ,  $b_2$ ) Techniques, en *Goodness-of-fit techniques*, D'Agostino, R.B. y Stephens, M.A. ed. Marcel Dekker. Inc. (279-329)
- Box, G.E.P. y Cox, D.R. (1964), An analysis of transformations, *J. Roy. Statist. Soc. B*. **26**, 211-252.
- Burden, R.L. y Faires, J.D. (1985). *Análisis Numérico*. Grupo Editorial Iberoamérica. México, D.F.
- Chambers, J.M., Cleveland, W.S., Kleiner, B. y Tukey, P.A. (1983) *Graphical Methods for Data Analysis*. Wadsworth International



Group. Belmont, California.

Chernoff, H y Lehmann, E.L. (1954), The use of maximum-likelihood estimates in  $\chi^2$  test for goodness of fit, *Ann. Math. Statist.* **25**, 579-686.

Cox, D.R. y Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall. London.

Cox, D.R. y Wermuth, N. (1994), Tests of Linearity, Multivariate Normality and the Adequacy of Linear Scores, *Applied Statistics*, **43** (2), 347-355.

Cressie, N. y Read, T.R.C. (1984), Multinomial goodness-of-fit tests, *J. Roy. Statist. Soc. B.* **46**, 440-464.

Csörgó, S. (1986), Testing for Normality in Arbitrary Dimension, *The Annals of Statistics*, **14** (2), 708-723.

Cunnane, C. (1978), Unbiased Plotting Positions- A Review, *Journal of Hydrology*. **37**, 205-222.

D'Agostino, R.B. (1971), An omnibus test of normality for moderate and large sample sizes, *Biometrika*. **58**, 341-348.

D'Agostino, R.B. y Pearson, E.S. (1973), Testing for departures from normality. I. Fuller empirical results for the distribution of  $b_2$  and  $\sqrt{b_1}$ , *Biometrika*. **60**, 613-622.

D'Agostino, R.B. y Tietjen, G.L. (1971), Simulation probability points for  $b_2$  for small samples, *Biometrika*. **58**, 669-672.

D'Agostino R.B. y Tietjen, G.L. (1973), Approaches to the null distribution of  $\sqrt{b_1}$ , *Biometrika*. **60**, 169-173.



- David, H.A., Hartley, H.O. y Pearson, E.S. (1954), The distribution of the ratio in a single normal sample of range to standard deviation, *Biometrika*. **41**, 482-493.
- David, F.N. y Johnson, N.L. (1954), Statistical treatment of censored data, I. Fundamental formulae, *Biometrika*. **41**, 228-240.
- De Haan, L. (1976), Sample extremes: An elementary introduction, *Statist. Neerland*. **30**, 161-172.
- Doksum, K.A. y Sievers, G.L. (1976), Plotting with confidence: Graphical comparisons of two populations, *Biometrika*, **63** (3), 421-434.
- Dyer, A.R. (1974), Comparisons of tests for normality with a cautionary note, *Biometrika*. **61**, 185-189.
- Everitt, B.S. (1994). *Statistical Analyses using S-Plus*. Chapman & Hall. London.
- Falk, M. (1989), A note on uniform asymptotic normality of intermediate order statistics, *Ann. Inst. Statist. Math.* **41**, 19-29.
- Filliben, J.J. (1975), The Probability Plot Correlation Coefficient Test for Normality, *Technometrics*. **17**, 111-117.
- Fisher, R.A. y Tippett, L.H.C. (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proc. Cambridge Philos. Soc.* **24**, 180-190.
- Fisz, M. (1963). *Probability Theory and Mathematical Statistics*. John Wiley & sons. New York.
- Frechet, M. (1927), Sur le loi de probabilité de l'écart maximum, *Ann. Soc. Polonaise Math.* **6**, 92-116.



- Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Kreiger, FL.
- Geary, R.C. (1935), The ratio of the mean deviation to the standard deviation as a test of normality, *Biometrika*. **27**, 310-332.
- Geary, R.C. (1947), Testing for normality, *Biometrika*. **34**, 209-242.
- Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une serie aleatoire, *Ann. Math.* **44**, 423-453.
- Hall, P. (1979), On the rate of convergence of normal extremes, *J. Appl. Probab.* **16**, 433-439.
- Hazen, A (1914), Storage to Be Provided in the Impounding Reservoirs for Municipal Water Supply, *Transactions of the American Society of Civil Engineers*. **77**, 1547-1550.
- Hazen, A (1930). *Flood Flows. A Study of Frequencies and Magnitudes*. Wiley, New York.
- Horswell, R.L. y Looney, S.W. (1993), Diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality, *Commun.Statist.: Simula.*, **22** (2), 437-459.
- Johnson, N.L. y Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions, vol 1 y 2*. John Wiley & sons. New York.
- Johnson, N.L. y Kotz, S. (1972). *Distributions in statistics: Continuous multivariate distributions*, John Wiley & sons. New York.



- Johnson, R.A. y Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall International, Inc. (tercera ed.)
- Kimball, B.F. (1960), On the Choice of Plotting Positions on Probability Paper, *Journal of the American Statistical Association*. **55**, 546-560.
- Kolmogorov, A.N. (1933), Sulla determinazione empirica di una legge di distribuzione, *Giorna. Ist. Attuari*. **4**, 83-91.
- Koziol, J.A. (1982), A class of invariant procedures for assessing multivariate normality, *Biometrika*, **69** (2), 423-427.
- Koziol, J.A. (1983), On Assessing Multivariate Normality, *Journal of the Royal Statistical Society, B*. **45** (3), 358-361.
- Koziol, J.A. (1986), Assessing Multivariate Normality: A Compendium, *Communications in Statistics: Theory and Methods*. **15** (9), 2763-2783.
- Koziol, J.A. (1993), Probability plots for assessing multivariate normality, *The Statistician*, **42**, 161-173.
- Krouse, D.P. (1994), The power of D'Agostino's test of normality against a normal mixture alternative, *Commun. Statist: Theory Meth*, **23** (1), 47-57.
- Kuiper, N.H. (1960), Tests concerning random points on a circle, *Proc. Koninkl. Neder. Akad. van Wetenschappen, A*. **63**, 38-47.
- La Brecque, J. (1977), Goodness-of-fit tests based on nonlinearity in probability plots, *Technometrics*. **19** (3), 293-306.



- Landry, L. Lepage, Y. (1992), Empirical behavior of some test for normality, *Communications in Statistics: Simula. and Comput.* **21** (4), 971-999.
- Looney, S.W. y Gullledge, T.R. Jr. (1985), Use of Correlation Coefficient With Normal Probability Plots, *The American Statistician*, **39** (1), 75-79.
- Looney, S.W. (1995), How to Use Test for Univariate Normality to Assess Multivariate Normality, *The American Statistician*, **49** (1), 64-70.
- Mardia, K.V. (1980), Tests of Univariate and Multivariate Normality, en *Handbook of Statistics, vol I*, P.R. Krishnaiah, ed. North-Holland, New York.
- Moore, D.S. (1986), Tests of Chi-Squared Type, en *Goodness-of-fit techniques*, D'Agostino, R.B. y Stephens, M.A., ed. Marcel Dekker, Inc.
- Mudholkar, G.S, McDermott, M y Srivastava, D.K. (1992), A test of p-variate normality, *Biometrika*, **79** (4), 850-854.
- Mulholland, H.P. (1977), On the null distribution of  $\sqrt{b_1}$  for samples of size at most 25, with tables, *Biometrika*. **64**, 401-409.
- Pearson, E.S. (1930), A further development of test for normality, *Biometrika*. **22**, 239.
- Pearson, E.S., D'Agostino, R.B. y Bowman, K.O. (1977), Tests for departure from normality: Comparison of powers, *Biometrika.*, **64**, 231-246.
- Pearson, E.S. y Hartley, H.O. (1972). *Biometrika Tables for Statisticians, vol I y II*. Cambridge University Press. Cambridge.



- Peña Sánchez de Rivera, D y Peña Sánchez de Rivera, J.I. (1986), Un contraste de normalidad basado en la transformación Box-Cox, *Estadística Española*. **110**, 33-46.
- Pollard, J.H. (1977). *A Handbook of Numerical and Statistical Techniques*. Cambridge University Press. Cambridge.
- Puri, M.L. y Rao, C.R. (1976), Augmenting Shapiro-Wilk Test for Normality, *Contributions to Applied Statistics*, Birkhauser (Grossohaus), Berlin. 129-139.
- Quiroz, A.J. y Dudley, R.M. (1991), Some tests for multivariate normality. *Probability Theory and Related Fields*. **87**, 521-546.
- Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley & sons. Mew York. (segunda ed.)
- Rohatgi, V.K. (1976). *An Introduction Probability Theory and Mathematical Statistics*. John Wiley and Sons.
- Royston, J.P. (1982a), Algorithm AS177. Expected normal order statistics (exact and approximate), *Applied Statistics*. **31**, 161-165.
- Royston, J.P. (1982b), An extension of Shapiro and Wilk W test for normality to large samples, *Applied Statistics*. **31** (2), 115-124.
- Royston, J.P. (1982c), Algorithm AS 181. The W test for normality, *Applied Statistics*. **31**, 176-180.
- Royston, J.P. (1983), Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. *Applied Statistics*, **32** (2), 121-133.



- Royston, P. (1992), Approximating the Shapiro-Wilk W-test for non-normality, *Statistics and Computing*. **2**, 117-119.
- Royston, P. (1993a), A toolkit for testing for non-normality in complete and censored samples, *The Statistician*. **42**, 37-43.
- Royston, P. (1993b), A pocket-calculator algorithm for the Shapiro-Francia test for non-normality: an application to medicine, *Statistics in Medicine*. **12**, 181-184.
- Shapiro, S.S. y Francia, R.S. (1972), An approximate analysis of variance test for normality, *Journal of the American Statistical Association*. **67**, 215-216.
- Shapiro, S.S. y Wilk, M.B. (1965), An analysis of variance test for normality (complete samples), *Biometrika*. **52** (3,4), 591-611.
- Shapiro, S.S., Wilk, M.B. y Chen, H.J. (1968), A comparative study of various tests for normality, *Journal of the American Statistical Association*. **63**, 1343-1372.
- Scheffé, H. y Tukey, J.W. (1945), Non-parametric estimation-I. Validation of order statistics, *Ann. Math. Statist.* **16**, 187-192.
- Small, N.J.H. (1978), Plotting squared radii, *Biometrika*, **65** (3), 657-658.
- Small, N.J.H. (1980), Marginal Skewness and Kurtosis in Testing Multivariate Normality, *Applied Statistics*, **29** (1), 85-87.
- Stephens, M.A. (1974), EDF statistics for goodness of fit and some comparisons, *Journal of the American Statistical Association*. **69**, 730-737.



- Stephens, M.A. (1986), Tests Based on EDF Statistics, en *Goodness-of-fit Techniques*, D'Agostino, R.B. y Stephens, M.A., ed. Marcel Dekker, Inc.
- Stuart, A. y Ord, J.K. (1993). *Kendall's, Advanced Theory of Statistics. Volume 1: Distribution Theory*. Edward Arnold, ed. London. (sexta ed.).
- Stuart, A. y Ord, J.K. (1991). *Kendall's, Advanced Theory of Statistics. Volume 1: Classical Inference and Relationship*. Edward Arnold, ed. London. (quinta ed.).
- Sugiura, N (1993), Assessing Multinormality and ordered covariance matrices in classical data, *Commun. Statist: Simula.*, 22 (3), 643-654.
- Ubasic for IBM-PC versión 8.74. *user's manual for UBHELP system*.
- Venables, W.N. y Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus*. Springer-Verlag. New-York
- Verrill, S y Johnson, R.A. (1988), Tables and Large-Sample Distribution Theory for Censored-Data Correlation Statistics for Testing Normality, *Journal of the American Statistical Association*, 83 (404), 1192-1197.
- Viollaz, A.J. y Bru de Labanda, E. (1992), Una paradoja en el test de bondad de ajuste, *Estadística Española*. 34 (130), 247-260.
- Von Mises, R. (1936), La distribution de la plus grande de  $n$  valeurs, *Rev. Math. Union Interbalcanique*. 1, 141-160. Reproducido en Selected Papers of Richard von Mises, *Am. Math. Soc. II*. (1964), 271-294.



Weibull, W (1939), The Phenomenon of Rupture in Solids, *Ingeniors Vetenskaps Akademien Handlingar*. 153, 17.

Weisber, S. y Bingham, C. (1975), An approximate analysis of variance test for non-normality suitable for machine calculation, *Technometrics*. 17, 133-134.

Wilk, M.B. y Gnanadesikan, R. (1968), Probability plotting methos for the analysis of data, *Biometrika*. 55 (1), 1-17.