ESTADÍSTICA ESPAÑOLA
Núm. 110, 1986, págs. 33 a 46

Un contraste de normalidad basado en la transformación Box-Cox

por DANIEL PEÑA SANCHEZ DE RIVERA
ETSII, Universidad Politécnica de Madrid

JUAN IGNACIO PEÑA SANCHEZ DE RIVERA
Dpto. de Econometría, Universidad Autónoma de Madrid

RESUMEN

En este trabajo se presenta un estudio de las posibilidades de utilizar las transformaciones Box-Cox para testar la Normalidad de una variable aleatoria. Se derivan los tests de Razón de Verosimilitud y Multiplicador de Lagrange correspondientes y se presenta un estudio de simulación de la potencia comparada del test LR frente a otros tipos de «omnibus» y direccionales.

Palabras Clave: Transformación Box-Cox, Test Omnibus y Direccionales, LR y LM test, Potencia Comparada.

1. INTRODUCCION

Los contrastes de Normalidad pueden clasificarse en cuatro grandes grupos. El primero de ellos mide si la distancia entre el valor de frecuencias observadas y el vector de frecuencias teóricas es lo suficientemente pequeña.

El contraste más popular con esta filosofía es el de Pearson, que utiliza como estadístico la distancia de Mahalanobis:

$$D_1 = (x-\mu)^T \Sigma_x (x-\mu) \tag{1.1}$$

donde x es el vector de frecuencias observadas, μ el vector de frecuencias teóricas si las observaciones provienen de una distribución normal y Σ_x es la matriz de varianzas y covarianzas de la variable multinomial definida por las clases. Un problema central de este contraste es como construir el vector x, es decir, como agrupar las frecuencias en clases; este tema ha sido objeto de numerosa literatura (Dahiya y Gurland (1973), Hutchinson (1979)).

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

$$D_2 = \text{máx. } |F_n(x) - F(x)| \tag{1.2}$$

Cramer y Von Mises han sugerido definir la distancia entre estas funciones por:

$$D_3 = \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \, dF(x) \tag{1.3}$$

Kendall y Stuart (1969) revisan estos y otros estadísticos basados en la misma idea. 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 los datos ordenados y sus valores esperados una medida de este ajuste. Los tests de Shapiro y Wilks (1965) y Shapiro y Francia (1972) responden a este principio.

El cuarto grupo se centra en el análisis de los coeficientes de asimetría y curtosis, solos o conjuntamente (Pearson et al. (1977)).

El alcance y generalidad de estos cuatro tipos de pruebas es muy distinto. El primer grupo puede aplicarse al ajuste de cualquier distribución, el segundo sólo para distribuciones continuas y los dos últimos son específicos para situaciones gaussianas. Por otro lado, los tres primeros son contrastes generales, mientras que el último va dirigido específicamente hacia la asimetría y/o la curtosis.

Los estudios relativos a la potencia de estos contrastes (véase Pearson et al. (1977) y Shapiro et al. (1968)) no son concluyentes ya que la potencia depende de la distribución alternativa, pero podemos señalar que el contraste W de Shapiro-Wilks, parece, en la mayoría de los casos, dominar el resto de los contrastes, especialmente en pequeñas muestras.

Un problema importante en la utilización de estos contrastes desde un punto de vista práctico, es que si la hipótesis de normalidad es rechazada, ninguno de ellos proporciona una guía de cómo proceder a continuación.

Por tanto, sería deseable conocer qué tipo de transformación de los datos observados puede hacerlos aproximadamente normales, de manera que los métodos estándar sean aplicables.

Un camino para evitar los inconvenientes anteriores es utilizar el enfoque de Box y Cox y construir un test de normalidad basándose en la estimación de la transformación que convierte los datos en gaussianos. En la sección dos se deduce un test basado en este principio, en la sección tres se presentan los resultados de un estudio de simulación de sus propiedades y finalmente, en la sección cuatro, se comentan los resultados obtenidos.

2. DEDUCCION DEL CONTRASTE DE NORMALIDAD

Dada una variable aleatoria, y, Box y Cox (1964) han propuesto la familia de transformaciones:

$$y(\lambda)=\begin{cases}\dfrac{(y+m)^{\lambda}-1}{\lambda}; & \lambda\neq 0\\ \ln(y+m); & \lambda=0\end{cases} \tag{2.1}$$

donde λ y m son parámetros a estimar. La familia es contínua en λ y ha sido extensamente utilizada para obtener homocedasticidad, linealidad y normalidad en modelos estadísticos.

En este trabajo supondremos que y es un vector de variables aleatorias con componentes:

$$y_i(\lambda)=\mu+u_i \qquad u_i\sim N(0,\ \sigma^2)$$

donde $y_i^{(\lambda)}$ viene dado por (2.1). La función soporte de las observaciones será:

$$L(\lambda,\ \sigma^2,\ \mu\,|\,y)=-\frac{n}{2}\ln\ \sigma^2-\frac{1}{2\sigma^2}\ \Sigma(y_i(\lambda)-\mu)^2-(\lambda-1)\ \Sigma\ \ln\ y_i \tag{2.2}$$

Podemos simplificar la función soporte haciendo el Jacobiano de la transformación independiente de λ. Si en lugar de (2.1) tomamos

$$Z_i(\lambda)=\begin{cases}\dfrac{\left(\dfrac{y_i}{\dot{y}}\right)^{\lambda}-1}{\lambda} & \lambda\neq 0\\ \ln(y/\dot{y}) & \lambda=0\end{cases} \tag{2.3}$$

donde $\dot{y}=\left(\prod_i y_i\right)^{1/n}$ es la media geométrica de las observaciones, la función soporte se reduce a (Apéndice III)

$$L(\lambda,\ \sigma^2,\ \mu'|z)=-\frac{n}{2}\ln\ \sigma^2-\frac{1}{2\sigma^2}\ \sum_i\ (Z_i(\lambda)-\mu')^2 \tag{2.4}$$

que concentrada con respecto a σ^2 y μ' queda finalmente

$$L(\lambda)=-\frac{n}{2}\ln\ \sum_i\ (Z_i(\lambda)-\overline{Z}(\lambda))^2=-\frac{n}{2}\ln\ M(\lambda) \tag{2.5}$$

donde $\overline{Z}(\lambda)$ es la media aritmética de los valores transformados $Z_i(\lambda)$.

El máximo de esta función será:

$$\frac{\partial\ L(\lambda)}{\partial\lambda}=0=-\frac{n}{2}\ \frac{\partial\ M(\lambda)}{\partial\lambda}\cdot\frac{1}{M(\lambda)} \tag{2.6}$$

donde

$$\frac{\partial\ M(\lambda)}{\partial\lambda}=2\ \Sigma\ e_i(\lambda)\cdot x_i(\lambda) \tag{2.7}$$

siendo

$$e_i(\lambda)=Z_i(\lambda)-\overline{Z}(\lambda)$$

$$x_i(\lambda)=\frac{-\partial e_i(\lambda)}{\partial\lambda}=-\left[\frac{Z_i(\lambda)(\lambda\ln\ Z_i-1)}{\lambda^2}-\frac{1}{n}\Sigma\ \frac{Z_i(\lambda)(\lambda\ln\ Z_i-1)}{\lambda^2}\right] \tag{2.8}$$

Por tanto, el máximo de $L(\lambda)$ se produce cuando la covarianza entre las variables $e_i(\lambda)$ y $x_i(\lambda)$ es cero y el valor de la derivada (2.6) es precisamente la estimación del coeficiente de regresión entre las variables $e_i(\lambda)$ y $x_i(\lambda)$.

La estimación óptima de λ puede llevarse a la práctica fácilmente utilizando un algoritmo tipo Gauss-Newton, de acuerdo con el esquema iterativo:

$$\hat{\lambda}_n = \hat{\lambda}_{n-1} + (\Sigma x_i^2(\hat{\lambda}_{n-1}))^{-1} \Sigma x_i e_i(\hat{\lambda}_{n-1}) \tag{2.9}$$

Con este desarrollo puede plantearse a continuación un test de normalidad, en el cual la hipótesis nula será $\lambda_0 = 1.0$. Supongamos que mediante iteraciones tipo (2.9) se ha obtenido $\hat{\lambda}_1$ como valor maximoverosímil de λ.

El test de razón de verosimilitud de la hipótesis nula de datos gaussianos, está definido por la región crítica

$$\theta = \frac{\text{máx } L(\lambda_0)}{\text{máx } L(\hat{\lambda}_1)} < k \tag{2.10}$$

que es equivalente a

$$-2 \ln \theta \sim n \ln \frac{\Sigma e_i^2(\lambda_0)}{\Sigma e_i^2(\hat{\lambda}_1)} \sim \chi_1^2 \tag{2.11}$$

y obtenemos un test asintótico inmediato, que además nos proporciona un intervalo de confianza aproximado para el valor estimado $\hat{\lambda}_1$.

Una forma alternativa de proceder sería emplear el test del «scoring» o LM (Rao (1947), Silvey (1959)) que en su forma más general toma la expresión

$$\text{LM} = G(\phi_0)^T I(\phi_0)^{-1} G(\phi_0) \sim \chi_m^2 \tag{2.12}$$

donde $G(\phi_0)$ es el vector gradiente evaluado en la hipótesis nula y $I(\phi_0)$ es la matriz de información. El vector de parámetros ϕ es de dimensión m. La distribución asintótica es χ^2 con m grados de libertad.

En nuestro caso tendríamos

$$\text{LM} = G(\lambda_0)^T I(\lambda_0)^{-1} G(\lambda_0) \sim \chi_1^2 \tag{2.13}$$

donde

$$G(\lambda_0) = \left. \frac{\partial L(\lambda)}{\partial \lambda} \right|_{\lambda = \lambda_0} \tag{2.14}$$

Es fácil ver que

$$G(\lambda_0) = n \frac{\Sigma\, e_i(\lambda_0)\ x_i(\lambda_0)}{\Sigma e_i^2(\lambda_0)} \tag{2.15}$$

$$I(\lambda_0) \simeq n \frac{\Sigma\, x_i^2(\lambda_0)}{\Sigma\, e_i^2(\lambda_0)} \tag{2.16}$$

por tanto, el test adopta la forma

$$\mathrm{LM} = n \frac{\Sigma(e_i(\lambda_0)\ x_i(\lambda_0))^2}{\Sigma\, e_i^2(\lambda_0)\ \Sigma\, x_i^2(\lambda_0)} \tag{2.17}$$

que es precisamente n veces el coeficiente de determinación en la regresión entre $e_i(\lambda_0)$ y $x_i(\lambda_0)$. Esta es la misma expresión dada por Harvey (1981) en el contexto de análisis de restricciones en el modelo lineal

$$\mathrm{LM} = nR^2 \sim \chi_1^2 \tag{2.18}$$

Es bien sabido que los tests LM y LR son asintóticamente equivalentes. El inconveniente básico del test LM es que no se estima el parámetro λ y por tanto sólo sabemos si la muestra es aproximadamente gaussiana o no, pero desconocemos cómo transformarla. Sin embargo, desde un punto de vista computacional es mucho más sencillo el LR, ya que sólo es necesario efectuar una sencilla regresión entre variables fácilmente construibles, $e_i(1.0)$, $x_i(1.0)$.

Sin embargo, en este trabajo se ha optado por construir el test más general, esto es el de razón de verosimilitud, con objeto de conocer el valor óptimo de λ para las distribuciones escogidas.

3. PRINCIPALES RESULTADOS

Para obtener resultados comparables a Pearson et al. (1977) se ha utilizado el mismo procedimiento de simulación seguido por estos autores: generar 200 muestras para cinco de las distribuciones que aparecen en su trabajo, con tamaños muestrales n = 20, 50, 100, y calcular el test para un nivel de significación $\alpha = 0{,}005$. En este caso, al distribuirse el test como χ_1^2, el valor es 3,85.

Las cinco distribuciones estudiadas han sido:

a) χ^2_{10} (Chi-cuadrado con 10 g de l. de l.)

b) $\beta_{2,2}$ (Beta con 2,2 g. de l.)

c) t_{10} (t de Studnet con 10 g. de l.)

d) log N(0,1) (Log normal $\gamma=0$, $\delta=1$, $\beta=0$)

e) CS(0,2; 3) (Normal contaminada de escala 0,2; 3)

correspondientes a las distribuciones 7, 14, 19, 46 y 58 del mencionado artículo.

En los cuadros 1, 2 y 3 se compara la potencia del contraste propuesto (columna BC) con los resultados de Pearson et al. (1979) para $n=20$, 50, 100, dando además el valor medio estimado para λ, y su derivación típica.

Las conclusiones más importantes son: el test BC supera a todos sus rivales en el caso de distribuciones claramente asimétricas (a) y (d) para todos los tamaños muestrales. Se muestra, por tanto, como la alternativa más potente en estos casos, superior incluso a los tests direccionales.

En el caso de la distribución (b), platicúrtica, su potencia es muy baja, superada por todos los otros tests para cualquier tamaño de muestra. En el caso de la distribución (c), leptocúrtica, su potencia es comparable a la de los otros tests «omnibus» sobre todo en $n=20$, 50 y ligeramente inferior a los direccionales. Finalmente, en el caso de la normal contaminada (e) su potencia es inferior en todos los casos y todos los tamaños de muestra a los otros tests.

Por tanto, puede decir que el BC es un excelente test direccional en el caso de asimetría, un discreto test en el caso de apuntamiento y una floja alternativa en los casos de distribuciones contaminadas.

Sin embargo, una ventaja adicional de BC y que no poseen sus oponentes, es que facilita una transformación de los datos que permite tratar a éstos como aproximadamente normales en la mayoría de los casos, con las ventajas que ello lleva consigo. Nótese que los otros tests señalan la ausencia o no de normalidad, pero no indican cómo modificar esta situación. Esta es la gran ventaja que, a nuestro juicio, presenta el uso del test BC y por ello pensamos que es una alternativa interesante a la hora de contrastar la Normalidad de una distribución.

$n=20$

	BC	K^2	W	b_2	$\sqrt{b_1}$	Y	Valor medio de λ en BC
χ^2_{10}	58(*)	27	27	—	36	36	0,24(.03)
$\beta_{2\cdot 2}$	0	14 (*)	2	11	—	11	0,80(.03)
t_{10}	11	13	10	16 (*)	—	14	0,41(0.1)
log $N(0,1)$	100 (*)	82	83	—	91	90	0,02(.01)
CS(0.2, 3)	24	40	38	49 (*)	—	47	0,38(.16)

Fig. 1

Porcentajes de veces que los contrastes indicados rechazan la hipótesis de normalidad al nivel $\alpha=0{,}05$. BC test de Box-Cox. Véanse definiciones de los tests en el apéndice II. Entre paréntesis en la columna de λ su desviación típica.

$n=50$

	BC	k^2	W	b_2	$\sqrt{b_1}$	Y	Valor medio de λ en BC
χ^2_{10}	75 (*)	46	53	—	57	57	0,26(.02)
$\beta_{2.2}$	2	18	23	53 (*)	—	47	0,86(.02)
t_{10}	12	17	12	23 (*)	—	26	0,83(0,06)
log $N(0,1)$	100 (*)	99	100 (*)	—	100	100	−0,002(.009)
CS(0.2, 3)	40	74 (*)	57	74 (*)	—	80	1,08(.14)

Fig. 2

$n = 100$

	BC	k^2	W	b_2	$\sqrt{b_1}$	Y	Valor medio de λ en BC
χ^2_{10}	98 (*)	79	—	—	—	38	0,29 (.01)
$\beta_{2,2}$	2	56	—	—	—	60 (*)	0,9 (.01)
t_{10}	11	26	—	—	—	30 (*)	0,99 (.04)
log $N(0,1)$	100	100	—	—	100	100	0,008 (.006)
CS (0.2, 3)	46	92 (*)	—	—	—	90	1,23 (.25)

Fig. 3

La casilla que contiene un asterisco (*) indica que ese es el test de mayor potencia para la distribución correspondiente.

4. DISCUSION

De los apartados anteriores se deduce que el test BC puede ser una alternativa razonable para testar la normalidad de una variable aleatoria. En particular su potencia es superior a la mayoría de los tests comúnmente usados, en el caso de las distribuciones asimétricas aquí presentadas. Además, proporciona un método de inducir normalidad en una muestra que no lo es, con las ventajas que ello proporciona, cara a su ulterior tratamiento estadístico.

Sin embargo, es importante conocer no solamente si la hipótesis de normalidad se rechaza o acepta, si no, cual ha sido la causa de un posible rechazo y, en particular, si éste ha sido debido a observaciones atípicas (outliers) que han podido distorsionar los resultados. Este análisis aunque reconocido como importante en la literatura, no ha sido objeto de estudio sistemático.

Resulta interesante discutir la interpretación del test BC efectuado; particularicemos el gradiente (2.15) para $\lambda_0 = 1.0$. Entonces, siguiendo a Box (1980), el estadístico se reduce a:

$$G(1) = \frac{-\Sigma(y_i - \bar{y})\ x_i(1)}{S_y^2} = -S_y^{-1}\ \Sigma\ \frac{y_i - \bar{y}}{S_y}\ x_i(1) \tag{4.1}$$

donde, particularizando en (2.8):

$$-x_i(1) = (y_i/\dot{y})\ \left(1 - \ln(y_i/\dot{y})\right) = G(y) \tag{4.2}$$

se observa que $G(1)$ es proporcional a $\Sigma\ r_i x_i(1)$ siendo $r_i = (y_i - \bar{y})/S_y$, los valores estandarizados de y_i.

Desarrollando en serie la función $G(y)$ hasta términos de segundo orden:

$$G(y) \cong G(\dot{y}) + (y - \dot{y})\ G'(\dot{y}) + (y - \dot{y})^2\ G''(\dot{y})/2 \tag{4.3}$$

Como $G(\dot{y}) = 1$ y $G'(\dot{y})$ es cero, entonces:

$$G(y) \simeq 1 + B(y - y)^2 \tag{4.4}$$

siendo $B = G''(\dot{y})/2$, una constante. Sustituyendo en la expresión del gradiente este valor de $G(y)$, tenemos:

$$G(1) = S_y^{-1}\ \Sigma\ r_i\left(1 + B(y - \dot{y})^2\right) = B\ \Sigma\ r_i(y - \dot{y})^2/S_y \tag{4.5}$$

como: $(y - \dot{y})/S_y = (y - \bar{y})/S_y + (\bar{y} - \dot{y})/S_y = r_i + d$, donde d es positivo, entonces tenemos que:

$$G(1) = B \cdot S_y(\Sigma r_i^3 + 2d\ \Sigma\ r_i^2) \tag{4.6}$$

es decir,

$$G(1) = \mathrm{BS}_y\ n(2d + \sqrt{b_1}) \tag{4.7}$$

siendo b_1 el coeficiente de asimetría de los datos. Por tanto, el estadístico $G(y)$ mide, básicamente, la asimetría de la distribución original. Si hubiésemos tomado términos de tercer orden en el desarrollo, $G(1)$, dependería también de la curtosis, aunque con un peso muy pequeño de ésta.

Estos resultados teóricos confirman los obtenidos por los ejercicios de simulación efectuados en los apartados anteriores. El test BC es básicamente una herramienta para detectar problemas de asimetría, y dentro de los tests usualmente empleados, se presenta como la alternativa más potente.

APENDICE I

METODOS DE COMPUTACION

El trabajo de simulación se llevó a cabo en el ordenador VAX-11/780 de la Universidad Autónoma de Madrid. Para investigar la potencia del test BC, se generaron 250 muestras aleatorias independientes $n = 20$, 50 y 100 para cada una de las cuatro distribuciones no normales, mediante el uso de las rutinas IMSL (1977); la normal contaminada se generó de la forma siguiente: se generó una Normal (0,1) y a continuación una uniforme (0,1) independiente de la anterior. Si esta nueva variable tenía valores menores o iguales que el valor γ de contaminación, la normal (0,1) se ajustaba multiplicando el término correspondiente por el valor de escala θ de contaminación. A continuación se estimó el λ óptimo mediante un algoritmo no-lineal tipo Levenberg-Marquardt y se efectuó el test LR para $H_0: \lambda = 1{,}0$, con nivel de significación del 5 %.

APENDICE II

Definiciones de los estadísticos utilizados como comparaciones. ($\sqrt{b_1}$ y b_2 son los habituales estadísticos para asimetría y curtosis, m_3/s^3 y $m_4/s^4 - 3$, donde $m_i = \Sigma (x_i - \bar{x})^i/n$; $s^2 = m_2$.)

El test k^2

Si $\sqrt{b_1}$ y b_2 fuesen independientes, en la hipótesis nula de Normalidad

$$k^2 = (\sqrt{b_1}/ST_1)^2 + (b_2/ST_2)^2$$

donde ST_1 y ST_2 son sus errores estándar se distribuiría como una χ^2 con 2 grados de libertad. En realidad no lo son, y los percentiles de esta distribución se han tabulado mediante simulación. (D'Agostino y Pearson (1973)).

El test W

El test W de Shapiro y Wilk (1965) se calcula

$$W = \sum_{i=1}^{h} a_{in}(x_{(n-i+1)} - x_i)^2 \quad \sum_{i=1}^{n} (x_i - \bar{x})^2$$

siendo x_i los valores ordenados de la muestra y a_{in} coeficientes obtenidos en las tablas correspondientes. Bajos valores de W indican no normalidad.

El test Y

Este test se desarrolla en D'Agostino (1971) y consiste básicamente en un test de Kurtosis. Para más detalles ver D'Agostino (1971, p. 343 ss.).

APENDICE III

UN COMENTARIO SOBRE LA FUNCION DE VEROSIMILITUD

Si tenemos

$$y^{(\lambda)}=\frac{y^{\lambda}-1}{\lambda}$$

la función de verosimilitud es

$$f(y)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}(y^{(\lambda)}-\mu)^2/\sigma^2}\cdot\left|\frac{dy(\lambda)}{dy}\right|\;;\;\frac{dy(\lambda)}{dy}=y^{\lambda-1}$$

$$\log f(\lambda|y_1,\ \dots,\ y_n)=-\frac{n}{2}\ln\sigma^2-\frac{1}{2}\Sigma\left(\frac{y_i^{(\lambda)}-\mu}{\sigma}\right)^2-(\lambda-1)\Sigma\ln\ y_i$$

llamando $\dot{y}=\prod(y_1,\ \dots,\ y_n)^{1/n}$; $n\ \log\ \dot{y}=\Sigma\log y_i$

$$L(\lambda)=-\frac{n}{2}\ln\sigma^2-\frac{1}{2}\Sigma\left(\frac{y_i^{(\lambda)}-\mu}{\sigma}\right)^2-(\lambda-1)\ \ln\dot{y} \tag{1}$$

como $L(\lambda,\ \sigma^2,\ \mu)$ es lineal en $(\sigma^2,\ \mu)$ respecto a λ, es decir, tiene la forma

$$L(\lambda,\ \sigma^2,\ \mu)=g(\sigma^2,\ \lambda,\ \mu)+h(\lambda)$$

Resolviendo respecto a σ^2 y μ, para λ constante

$$\hat{\sigma}^2=\frac{1}{n}\Sigma(y_i^{(\lambda)}-\bar{y}^{(\lambda)})^2\ ;\ \bar{y}^{(\lambda)}=\frac{1}{n}\Sigma y_i^{(\lambda)}=\hat{\mu}$$

y la función de verosimilitud concentrada es:

$$L\ \lambda;\ \hat{\sigma}^2(\lambda),\ \hat{\mu}(\lambda))=-\frac{n}{2}\ln\ \hat{\sigma}^2(\lambda)-\frac{n}{2}-(\lambda-1)\ \ln\ \dot{y}\ ;\ A=(\lambda-1)\ln\dot{y}$$

para que el término A desaparezca y el soporte dependa sólo de la varianza, la función (1) no debe incluirlo, es decir el Jacobiano de la transformación debe ser constante respecto a λ.

Esto puede conseguirse:

a) Como proponen Box y Cox (1964), haciendo:

$$z(\lambda)=(y^{\lambda}-1)/\lambda\,\dot{y}^{\lambda-1}$$

y se demuestra que el jacobiano de la transformación es constante.

b) Haciendo $Z_i(\lambda)=\dfrac{(y/\dot{y})^{\lambda}-1}{\lambda}$

entonces

$$\frac{dZ_i^{(\lambda)}}{dy_i}=\frac{1}{\lambda}\lambda\left(\frac{y_1}{\dot{y}}\right)^{\lambda-1}\cdot\left(\frac{\dot{y}-\frac{1}{n}\frac{\dot{y}}{y_1}\cdot y_1}{\dot{y}^2}\right)=\left(\frac{y_1}{\dot{y}}\right)^{\lambda-1}\left[\frac{n-1}{n\dot{y}}\right]$$

$$\text{Jacobiano}=\left(\frac{n-1}{n}\right)^n\cdot\frac{(y_1,\ \dots,\ y_n)^{\lambda-1}}{(\dot{y}^n)^{\lambda-1}}\cdot\frac{1}{\dot{y}^n}=\left(\frac{n-1}{n}\right)^n\cdot\frac{1}{\dot{y}^n}$$

y el Jacobiano no depende de λ y es una constante que puede eliminarse. Por tanto

$$L(\lambda,\ \sigma^2,\ \mu)=cte-\frac{n}{2}\ \ln\ \sigma^2\frac{1}{2\sigma^2}\Sigma(Z_i(\lambda)-\mu)^2$$

y por tanto basta minimizar $\Sigma(Z_i(\lambda)-\mu)^2$.

BIBLIOGRAFIA

ATKINSON, A. C. (1973): *Testing transformations to normality*. *JRSS*, B, 35, 473-479.

BOX, G. E. P. (1980): *Sampling and Bayes Inference in Scientific Modelling and Robustness*. *JRSS*, A, 143, 4, 390-430.

BOX, G. E. P. y COX, D. R. (1964): *An Analysis of Transformations*. *JRSS*, B, 211-243.

COOK, R. D. (1977): *Detection of Influential observations in Regression*. *Technometrics*, 19, 15-18.

D'AGOSTINO, R. B. (1971): *An omnibus test of normality for moderate and large size samples*. *Biométrika*, 58, 341-348.

D'AGOSTINO, R. B. y PEARSON, E. S. (1973): *Test for Departure from normality fuller empirical results for the distributions of b_2 and b_1*. **Biométrika**, 60, 613-22.

DAHIYA, R. C. y GURLAND, J. (1973): *How many classes in the Pearson Chi-Square test?*. *JASA*, 68, 343, 707-712.

DRAPER, N. y SMITH, H. (1981): *Applied Regression Analysis*. Wiley. New York.

IMSL (1977): *User's Manual*.

KENDALL y STUART (1968): *The Advanced Theory of Statistics*. 2.ª ed. Charles Griffin & Co. London.

HATCHINSON, T. P. (1979): *The validity of the Chi-squared test when expected frequencies are small: a list of recent research references*. *Comm. STA*, 8, 327-336.

MADDALA, G. S. (1977): *Econometrics*. McGraw Hill.

PEARSON, E. S., D'AGOSTINO, R. B. y BOWMAN, K. O. (1977): *Test for departure from normality: Comparison of powers*. **Biométrika**, 64, 231-46.

RAO, C. R. (1947): *Large Sample test of Statistical Hypothesis concerning several parameters with applications to problems of estimation*. *Proc. Camb. Phil. Soc*, 44, 50-57.

SHAPIRO, S. S., WILK, M. B. y CMEN, H. J. (1968): *A comparative study of various tests of normality*. *JASA*, 63, 1343-72.

SHAPIRO, S. S. y WILK, M. B. (1965): *An analysis of variance test of normality*. **Biométrika**, 52, 591-611.

SHAPIRO, S. S. y FRANCIA, R. S. (1972): *An approximate analysis of variance test for normality*. *JASA*, 67, 215-216.

SILVEY, S. D. (1959): *The Lagrangian Multiplier Test*. *Ann. of Math. Stat.*, 30, 389-407.

SUMMARY

A NORMALITY TEST BASED ON BOX-COX TRANSFORMATION

In this paper we show how to use the family of transformations by Box and Cox as a test for normality. Likelihood ratio and lagrange multiplier test are derivated, and it is showed the results of a Monte Carlo simulation about the power of LR test faced with some others alternatives.

Key words: Box-Cox transformation, Omnibus and directional test, LR and LM test, comparative power.

AMS, 1980. Subject classification 65B10.