# Una mirada geométrica a la transformación de Box-Cox

**Francisco J. Cepeda Coronado**
*Profesor Universidad Pedagógica Nacional*
*Bogotá D.C, Colombia*
fcepedacoronado@yahoo.com

**Fabio A. Fajardo Molinares**
*Profesor Universidad Nacional de Colombia*
*Bogotá D.C, Colombia*
ffajardo@unal.edu.co

## 1. Planteamiento del Problema

En el análisis de los datos, cuando la normalidad no es un supuesto viable, una alternativa es ignorar los resultados del chequeo de normalidad y proceder como si los datos estuvieran normalmente distribuidos. Este procedimiento no es recomendable puesto que, en la mayoría de los casos, lleva a conclusiones incorrectas.

Una segunda alternativa consiste en convertir los datos no normales en unos que tengan más apariencia de normales considerando transformaciones de los datos. Estas transformaciones son simplemente reexpresiones de los datos en diferentes unidades, frecuentemente sucede que las nuevas unidades proporcionan expresiones más naturales a las características que se están estudiando.

Para datos normalmente distribuidos, las *transformaciones lineales* de los datos son comunmente usadas para crear una variable con distribución normal estándar.

La ecuación $z = (x - \mu)/\sigma$ es una transformación lineal de la variable $x$ en la variable $z$. Como resultado de esta transformación, la media cambia de $\mu$ a 0 y de $\sigma^2$ a 1. La forma de la distribución, sin embargo, no cambia como resultado de esta transformación lineal.

Para cambiar la forma de la distribución se requiere una *transformación no lineal*. Teniendo en cuenta que la forma de una distribución puede ser analizada en términos de sesgo y empleando la distribución normal como referencia o estándar, se puede buscar obtener "casi normalidad"por una transformación no lineal que obtenga simetría.

Es sencillo imaginar ejemplos de transformaciones que pueden ser usadas para reducir o comprimir la escala. Para valores positivos de $X$, la transformación

$Y = X^2$ reduce la escala para $X > 1$ y comprime la escala para $X < 1$. El inverso de estas propiedades es cierto para la transformación $Y = X^{1/2}$.

Luego, dada una distribución sesgada con valores $X$ positivos, una transformación de la forma $Y = X^k$ puede ser util en la eliminación del sesgo, valores de $k$ mayores que 1 pueden ser usados para eliminar el sesgo negativo, mientras que valores de $k$ en el rango $0 < k < 1$ pueden ser usados para eliminar el sesgo positivo. Esta transformación es un caso especial de una familia de transformaciones conocidas como *transformaciones de potencia*.

## 2. Transformación de Box-Cox

Una forma más general de las transformaciones de potencia es la de Box-Cox dada por

$$Y = \begin{cases} \dfrac{X^\lambda - 1}{\lambda} & si \quad \lambda \neq 0 \\ \ln X & si \quad \lambda = 0; \quad X > 0 \end{cases}$$

La expresión $Y = \ln X$ simplemente refleja el límite de la formula de la transformación cuando $X$ es positivo y $\lambda$ se aproxima a cero.

Dado un conjunto de datos observados $(x_1, \ldots, x_n)$ el parámetro de la transformación, $\lambda$, debe ser estimado. Las aproximaciones para estimar $\lambda$ son basadas usualmente en el supuesto de que los valores transformados $y_1, \ldots, y_n$ están normalmente distribuidos.

Usando una aproximación de Máxima Verosimilitud, Box y Cox desarrollaron una técnica para estimar $\lambda$. Ellos mostraron que la función de máxima verosimilitud es

$$L_m(\lambda) = -\frac{1}{2n} \ln \tilde{\sigma}_z^2$$

donde

$$\tilde{\sigma}_z^2 = \sum_{i=1}^{n} \frac{(z_i - \bar{z})^2}{n}; \qquad z_i = \frac{x_i^\lambda}{\tilde{x}_G^\lambda} \quad \text{y} \quad \tilde{x}_G = (x_1 \cdot x_2 \cdots x_n)^{1/n}$$

Las inferencias para $\lambda$ pueden ser llevadas a cabo usando una aproximación basada en una distribución $\chi^2$ con un grado de libertad. Un intervalo confidencial de

$100(1-\alpha)\,\%$ para $\lambda$ está dado por el conjunto de todos los $\lambda$ que satisfacen $[L_m(\hat{\lambda}) - L_m(\lambda)] \leq \frac{1}{2}x^2_{\alpha,1}$ donde $\hat{\lambda}$ es el estimador máximo verosímil. Para probar $H_0 : \lambda = \lambda_0$ la prueba estadística $2[L_m(\hat{\lambda}) - L_m(\lambda_0)]$ se compara con $\chi^2_{\alpha,1}$.

Una aproximación común de ensayo y error para la determinación del estimador máximo verosímil de $\lambda$ envuelve la determinación de $L_m(\lambda)$ para un rango de valores de $\lambda$. Entonces se emplea un proceso iterativo para hallar el valor de $\lambda$ que maximiza $L_m(\lambda)$.

Una solución para $\lambda$ es usualmente un valor entero tal como 2, 3 o 4; o si $\lambda < 1$, las fracciones comunes son 1/2, 1/3 o 1/4. Si $\lambda$ es cercano a 0 y $X$ es positivo, $\ln X$ es frecuentemente empleado. Una vez se ha determinado el valor de $\lambda$, no hay garantía de que la solución resultará en una distribución aceptablemente similar a la normal. Entonces, la distribución de los datos transformados debe ser cuidadosamente estudiada.

## Ejemplo

La Tabla 1 presenta la producción total anual de crudo para 50 paises productores en el año 1976. Una breve inspección de los datos revela un fuerte sesgo positivo. El rango de la distribución es $[1{,}1, 520]$, la mediana es 10.5, la media es 57.1, la desviación estándar es 113.2 y el sesgo tiene un valor de 2.9, lo cual hace que la prueba estadística para esta medida resulte altamente significativa y por consiguiente se rechace el supuesto de normalidad de los datos. El sesgo positivo en esta distribución podría sugerir una transformación de Box-Cox con un valor pequeño de $\lambda$, o una transformación logarítmica.

Empleando un método iterativo, la fución de verosimilitud $L_m(\lambda)$ es maximizada en $\lambda = -0{,}13$. Un intervalo confidencial de 95 % para $\lambda$ está dado por $(-0{,}30, 0{,}04)$. Como este intervalo contiene el valor $\lambda = 0$, la transformación logarítmica parece razonable.

La Tabla 1 también contiene los valores transformados $Y = \ln X$. La mediana de la distribución transformada tiene un valor de 1.02, la media es 1.13, la desviación estándar es 0.75 y el sesgo es 0.42. Con estos valores se encuentra que el sesgo no es estadísticamente significativo y las pruebas de significancia para el supuesto de normalidad son significativas.

| País | Producción | Log. | País | Producción | Log. |
|---|---|---|---|---|---|
| Francia | 1.1 | 0.04 | Trinidad | 11.0 | 1.04 |
| Chile | 1.1 | 0.04 | Gabon | 11.4 | 1.06 |
| Italia | 1.1 | 0.04 | Reino Unido | 12.0 | 1.08 |
| Burma | 1.2 | 0.08 | Noruega | 13.7 | 1.14 |
| Países Bajos | 1.4 | 0.15 | Rumania | 14.7 | 1.17 |
| Bolivia | 2.0 | 0.30 | Egipto | 16.7 | 1.22 |
| España | 2.0 | 0.30 | Oman | 18.3 | 1.26 |
| Congo | 2.0 | 0.30 | Australia | 20.5 | 1.31 |
| Austria | 2.0 | 0.30 | Argentina | 20.9 | 1.32 |
| Hungría | 2.1 | 0.32 | Qatar | 23.5 | 1.37 |
| Albania | 2.3 | 0.36 | México | 40.8 | 1.61 |
| Turquía | 2.6 | 0.41 | Algeria | 50.0 | 1.70 |
| Bahrain | 2.9 | 0.46 | Canadá | 64.1 | 1.81 |
| Peru | 3.7 | 0.57 | Indonesia | 74.0 | 1.87 |
| Tunez | 3.7 | 0.57 | China | 85.0 | 1.93 |
| Yugoslavia | 3.9 | 0.59 | Libia | 92.8 | 1.97 |
| Alemania Occ. | 5.5 | 0.74 | Emiratos Arabes | 93.3 | 1.97 |
| Angola | 6.3 | 0.80 | Nigeria | 102.7 | 2.01 |
| Colombia | 7.5 | 0.88 | Kuwait | 108.6 | 2.04 |
| Malasia | 8.0 | 0.90 | Iraq | 112.4 | 2.05 |
| Brasil | 8.5 | 0.93 | Venezuela | 119.8 | 2.08 |
| India | 8.6 | 0.93 | Iran | 296.5 | 2.47 |
| Brunei | 8.6 | 0.93 | USA | 401.6 | 2.60 |
| Ecuador | 9.5 | 0.98 | Arabia | 424.2 | 2.63 |
| Siria | 10.0 | 1.00 | Unión Sovietica | 520.0 | 2.72 |

Tabla 1: Producción Mundial de Crudo de 50 Naciones en 1976