

# Transformaciones de datos

---

Alberto Benavides  
20 de octubre de 2020

## 1. INTRODUCCIÓN

En esta práctica se ha desarrollado una metodología computacional que permite comparar las correlaciones de ciertas funciones después de transformar sus variables de entrada y salida mediante las transformadas de Tukey y de Box-Cox.

## 2. TRANSFORMADAS

Las *transformadas de Tukey* [1] parten de la idea de transformar variables independientes  $X$  o dependientes  $Y = f(X)$  en potencias de dichas variables  $X^\lambda, Y^\lambda$  a partir de los distintos valores que pueda tomar  $\lambda$ , tal que

$$z_\lambda = \begin{cases} x^\lambda, & \text{si } \lambda > 0, \\ \log(x), & \text{si } \lambda = 0, \\ -(x^\lambda), & \text{si } \lambda < 0. \end{cases} \quad (2.1)$$

Por otro lado, las *transformadas de Box-Cox* [2], por su cuenta, se realizan a partir de la ecuación

$$z_\lambda = \frac{x^\lambda - 1}{\lambda}. \quad (2.2)$$

## 3. CORRELACIONES

Ambas transformadas suelen usarse para mejorar la correlación de las funciones resultantes con respecto a la función original. Para ello, es posible transformar sólo  $X$ ,  $Y$

o ambas al mismo tiempo y luego calcular las correlaciones entre dichas variables. Así, una manera de encontrar la mejor transformada para una determinada función sería definir algunos valores de  $\lambda$ , transformarla mediante ambas transformadas y graficar las correlaciones de las funciones con transformaciones en  $X$ ,  $Y$  o ambas variables.

## 4. DISEÑO DE EXPERIMENTOS

Se realiza un diseño de experimentos para comparar las correlaciones de las funciones transformadas a partir de las transformaciones de Tukey y Box–Cox. Como ejemplo, se utilizan las funciones

- $1/x$ ,
- $x^2$ ,
- $x^3$ ,
- $\log(x)$ ,
- $e^x$ ,
- $\sin(x \cdot \frac{180}{\pi})$ ,
- $\cos(x \cdot \frac{180}{\pi})$ ,
- $\tan(x \cdot \frac{180}{\pi})$ .

La variable  $\lambda = [-3.0, -2.5, -2.0, \dots, 2.0, 2.5, 3.0]$ , mientras que las transformaciones se aplican sobre sólo  $X$ , sólo  $Y$  y ambas simultáneamente. Se generan mil valores de  $X$  a partir de una distribución uniforme  $\mathcal{U}(-100, 100)$  y se calculan  $Y$  para cada función. Luego, se grafican las funciones de las transformadas de Tukey y Box–Cox para cada una de estas variantes. Por último, se grafican las correlaciones con los distintos  $\lambda$  tanto para las transformaciones de Tukey como las de Box–Cox.

## 5. RESULTADOS

Se muestran algunos ejemplos de esta práctica. Primero  $1/x$ . La función se muestra en la figura 5.1 (p. 3). El despliegue de correlaciones para las transformaciones de  $X$ ,  $Y$  y ambas se muestra en la figura 5.2 (p. 4). En dicha figura se puede apreciar que los valores de las correlaciones para ambas transformadas con los mismos valores  $\lambda$  son iguales. Además, una animación de las diferentes transformaciones a lo largo de los cambios en  $\lambda$  puede consultarse en <https://tinyurl.com/yxjepzdf>, <https://tinyurl.com/y468ba9o> y <https://tinyurl.com/y2e6a6w2> para las transformaciones de  $X$ ,  $Y$  y ambas, respectivamente. El resto de las funciones, comparaciones entre correlaciones, animaciones y el código se hallan en <https://github.com/jbenavidesv87/probabilidad/tree/master/tema7>.

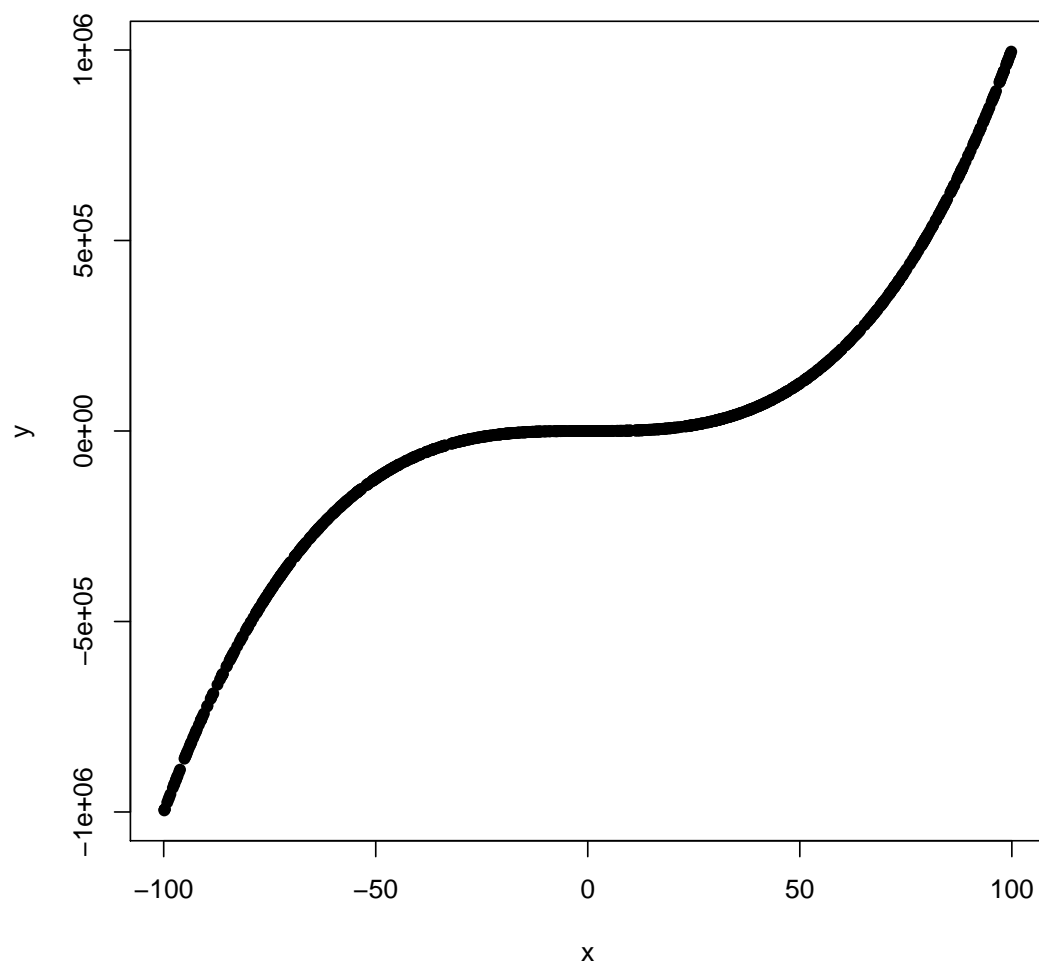
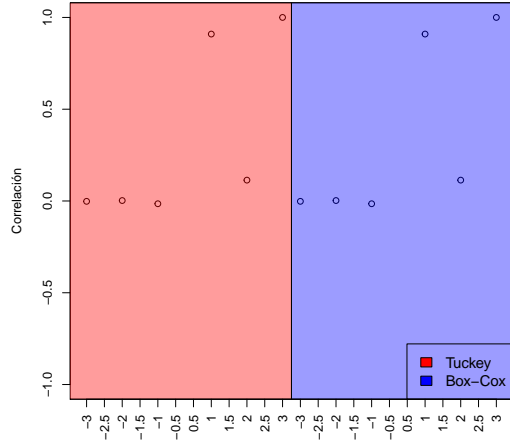
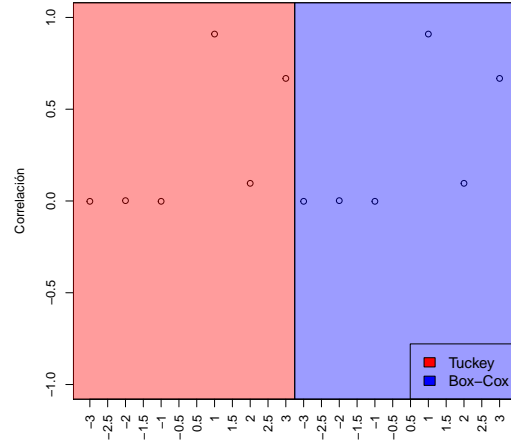


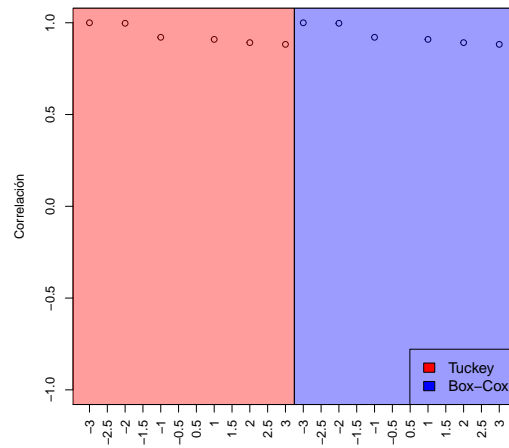
Figura 5.1: Función  $1/x$  a partir de mil valores de  $X$  desde una distribución uniforme con valores  $[-100, 100]$ .



(a)  $X$  transformada.



(b)  $Y$  transformada.



(c)  $X$  y  $Y$  transformadas.

Figura 5.2: Gráficas de las correlaciones para la función  $1/x$  con transformaciones para  $X$ ,  $Y$  y ambas. En el eje horizontal se grafican los valores de  $\lambda$ , la mitad roja corresponde a transformación de Tukey, mientras que la azul a la de Box-Cox.

## REFERENCIAS

- [1] Elisa Schaeffer. Modelos Probabilísticos Aplicados – Curso en Línea. <https://elisa.dyndns-web.com/teaching/prob/pisis/prob.html>, 2020.
- [2] David Scott. Box-Cox Transformations. <http://onlinestatbook.com/2/transformations/box-cox.html>, 2020.