Transformaciones

En muchas ocasiones se quiere transformar los datos originales para que la distribución de la variable transformada tenga mejores propiedades de simetría etc., o para simplificar el análisis.

Es interesante saber cómo cambian las características de la muestra como la media y desviación típica.

En general, no existe una fórmula sencilla para hallar la media de los datos transformadas, salvo en el caso de que la transformación sea lineal.

Transformaciones lineales

Supongamos que tenemos una muestra $x_1, \ldots, x_n$ con media $\bar{x}$ y desviación típica s_x y que hacemos una transformación lineal de los datos

$$y_i = \alpha + \beta x_i \quad \text{para } i = 1, \ldots, n$$

Entonces, tenemos el siguiente teorema.

Teorema 3 *La media, varianza y desviación típica de la muestra $y_1, \ldots, y_n$ son*

$$\begin{aligned} \bar{y} &= \alpha + \beta \bar{x} \\ s_y^2 &= \beta^2 s_x^2 \\ s_y &= \beta s_x \end{aligned}$$

Demostración

$$
\begin{aligned}
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \textit{por definición de la media} \\
&= \frac{1}{n}\sum_{i=1}^{n}(\alpha + \beta x_i) \qquad \textit{transformando} \\
&= \frac{1}{n}\sum_{i=1}^{n}\alpha + \frac{1}{n}\sum_{i=1}^{n}\beta x_i \\
&= \frac{1}{n}n\alpha + \beta\frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \alpha + \beta\bar{x} \\
s_y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \textit{por definición de la varianza} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\alpha + \beta x_i - (\alpha + \beta\bar{x})\right)^2 \quad \textit{por el resultado anterior} \\
&= \frac{1}{n}\sum_{i=1}^{n}\beta^2(x_i - \bar{x})^2 \\
&= \beta^2\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \beta^2 s_x^2 \qquad \diamond
\end{aligned}
$$

Ejemplo 53 *Volvemos al Ejemplo 32 sobre los ratoncitos.*

Un cientifico quien quiere hacer experimentos con animales paga a los padres una cantidad de 10 *euros por ratoncito criado. ?Cuál es la cantidad media pagada a unos padres que dejan sus críos al cientifico?*

Si $x_i =$ *el número de críos por familia y* $y_i =$ *pago, se tiene* $y_i = 10x_i$.

Hemos visto en el Ejemplo 35, que el número medio de críos por pareja de ratones era de $5{,}33\dot{3}$. *Entonces el pago medio es de* 53,33 *euros por familia.*

Tipificando las observaciones

Teorema 4 *Dada la muestra $x_1, \ldots, x_n$, con media $\bar{x}$ y varianza s_x^2, la distribución de las variables típificados*

$$y_i = \frac{x_i - \bar{x}}{s_x} \quad \textit{para } i = 1, \ldots, n$$

tiene media 0 *y varianza y desviación típica* 1.

Demostración

Se tiene $y_i = -\frac{\bar{x}}{s_x} + \frac{1}{s_x} x_i$ y aplicando el Teorema 3 con $\alpha = -\frac{\bar{x}}{s_x}$ y $\beta = \frac{1}{s_x}$, implica que

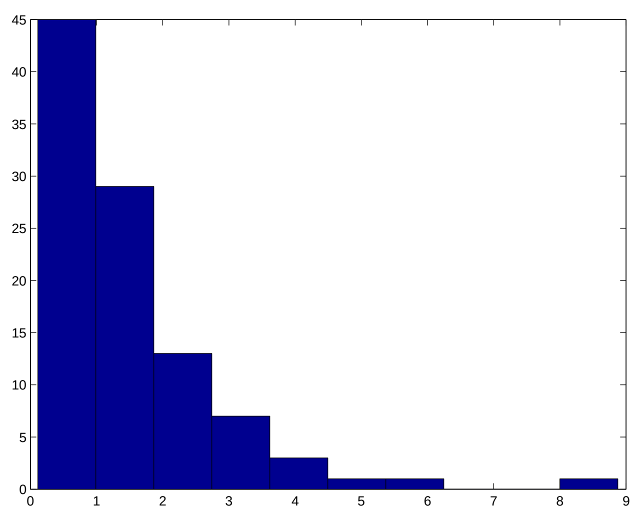
$$\begin{aligned} \bar{y} &= -\frac{\bar{x}}{s_x} + \frac{1}{s_x}\bar{x} \\ &= 0 \\ s_y^2 &= \left(\frac{1}{s_x}\right)^2 s_x^2 \\ &= 1 \end{aligned}$$

$\diamond$

Transformaciones no lineales

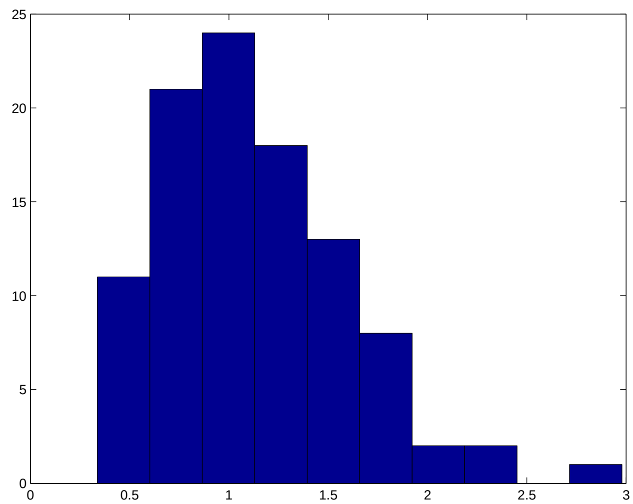
Se puede usar una transformación no lineal para convertir una muestra asimétrica en una muestra mucho más simétrica.

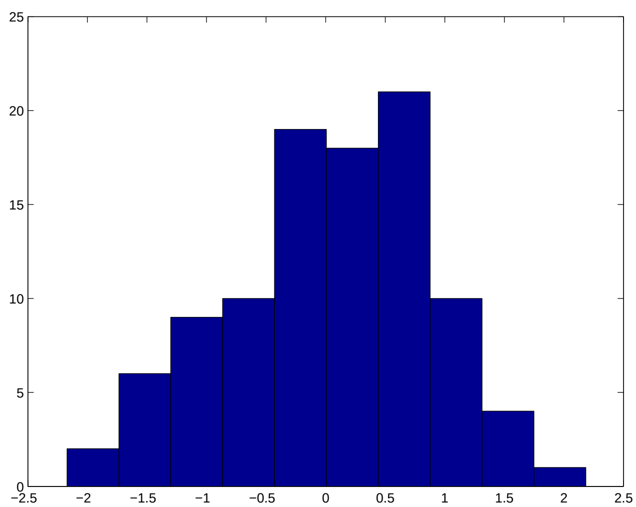
Ejemplo 54 *Los datos ilustrados en el histograma son los tiempos de funcionamiento de 100 piezas electrónicas.*



El histograma es muy asimétrica a la derecha.

Los siguientes histogramas ilustran los efectos de las transformaciones $y = \sqrt{x}$ e $y = \log x$ respectivamente.





Los resultados son mucho menos asimétricas.