

## Cambio de origen y escala de una variable estadística x.-

Consideremos una variable  $x$ , cuya distribución venga dada por la tabla adjunta y sean la media  $\bar{x} = \frac{1}{N} \sum_{i=1}^r x_i n_i$  y la varianza  $S_x^2 = \frac{1}{N} \sum_{i=1}^r (x_i - \bar{x})^2 n_i$ .

Efectuemos un cambio de origen y de escala sobre la variable  $x$ , es decir, construyamos otra variable  $y = ax + b$ , siendo  $a > 0$  y  $b$  constantes (multiplicar  $x$  por una constante es efectuar un cambio de escala y sumarle a  $x$  una constante es efectuar un cambio de origen). Esto quiere decir que para cada  $x_i$  hay un  $y_i = ax_i + b$  con su misma frecuencia  $n_i$ . La tabla de la variable  $y$  será:

$x_i$	$n_i$
$x_1$	$n_1$
$x_2$	$n_2$
$x_3$	$n_3$
$x_4$	$n_4$
....	....
$x_r$	$n_r$
	$N$

$y_i$	$n_i$
$y_1$	$n_1$
$y_2$	$n_2$
$y_3$	$n_3$
$y_4$	$n_4$
....	....
$y_r$	$n_r$
	$N$

Entonces, la media

$$\bar{y} = \frac{1}{N} \sum_{i=1}^r y_i n_i = \frac{1}{N} \sum_{i=1}^r (ax_i + b) n_i = a \cdot \frac{1}{N} \sum_{i=1}^r x_i n_i + b \cdot \frac{1}{N} \sum_{i=1}^r n_i = a\bar{x} + b$$

y la varianza

$$S_y^2 = \frac{1}{N} \sum_{i=1}^r (y_i - \bar{y})^2 n_i = \frac{1}{N} \sum_{i=1}^r (ax_i + b - a\bar{x} - b)^2 n_i = a^2 \frac{1}{N} \sum_{i=1}^r (x_i - \bar{x})^2 n_i = a^2 S_x^2$$

Es decir, **la media se ve afectada por el mismo cambio de origen y de escala efectuado sobre la variable**, mientras que **la varianza no se ve afectada por el cambio de origen pero se ve afectada por el cuadrado del cambio de escala efectuado sobre la variable**.

Para el coeficiente de variación se tiene:  $CV_y = \frac{S_y}{\bar{y}} = \frac{aS_x}{a\bar{x} + b}$ . Por tanto, si sólo efectuamos un

cambio de escala (es decir, si  $b = 0$ ), entonces  $CV_y = \frac{S_y}{\bar{y}} = \frac{aS_x}{a\bar{x}} = \frac{S_x}{\bar{x}} = CV_x$ , es decir, el **cambio de escala no le afecta al coeficiente de variación**.

Si sólo efectuamos un cambio de origen (es decir, si  $a = 1$ ) entonces  $CV_y = \frac{S_y}{\bar{y}} = \frac{S_x}{\bar{x} + b} \neq CV_x$ , es decir, **sí que le afecta el cambio de origen**.

## Cambios de origen y escala en una distribución bidimensional.-

Consideremos una distribución bidimensional  $(x_i, y_i)$ ,  $i = 1, 2, \dots, N$ .

Supóngase que hacemos un cambio de origen y de escala, es decir, introducimos otras variables  $x'_i$  e  $y'_i$ , relacionadas con las anteriores de la siguiente forma:

$$\begin{aligned} x'_i &= mx_i + n \\ y'_i &= py_i + q \end{aligned}$$

sean:

$a_{10}$ ,  $a_{01}$ ,  $m_{11}$ ,  $m_{20}$  y  $m_{02}$  los momentos referidos a la variable  $(x_i, y_i)$  y

$a'_{10}$ ,  $a'_{01}$ ,  $m'_{11}$ ,  $m'_{20}$  y  $m'_{02}$  los referidos a la variable  $(x'_i, y'_i)$ .

Se tiene:

$$a'_{10} = \frac{1}{N} \sum_{i=1}^N x'_i = \frac{1}{N} \sum_{i=1}^N (mx_i + n) = m \cdot \frac{1}{N} \sum_{i=1}^N x_i + \frac{1}{N} \sum_{i=1}^N n = ma_{10} + n$$

Análogamente  $a'_{01} = pa_{01} + q$ .

(es decir, **las medias se ven afectadas por el mismo cambio de origen y de escala efectuado en la variable**)

Por lo tanto:

$$m'_{11} = \frac{1}{N} \sum_{i=1}^N (x'_i - a'_{10})(y'_i - a'_{01}) = \frac{1}{N} \sum_{i=1}^N (mx_i - ma_{10})(py_i - pa_{01}) = mp \frac{1}{N} \sum_{i=1}^N (x_i - a_{10})(y_i - a_{01}) =$$

$$= mp \cdot m_{11}$$

(es decir, **la covarianza es invariante ante un cambio de origen pero no ante un cambio de escala**)

$$m'_{20} = \frac{1}{N} \sum_{i=1}^N (x'_i - a'_{10})^2 = \frac{1}{N} \sum_{i=1}^N (mx_i - ma_{10})^2 = m^2 \frac{1}{N} \sum_{i=1}^N (x_i - a_{10})^2 = m^2 \cdot m_{20}.$$

Análogamente  $m'_{02} = p^2 \cdot m_{02}$ .

(es decir, **las varianzas son invariantes ante un cambio de origen pero no ante un cambio de escala**)

Sean ahora  $b_{Y/X}$  y  $b_{X/Y}$  los coeficientes de regresión de las rectas  $Y/X$  y  $X/Y$  respectivamente y correspondientemente  $b'_{Y'/X'}$  y  $b'_{X'/Y'}$  los coeficientes de regresión de las rectas  $Y'/X'$  y  $X'/Y'$  respectivamente. Se tendrá:

$$b'_{Y'/X'} = \frac{m'_{11}}{m'_{20}} = \frac{mp \cdot m_{11}}{m^2 \cdot m_{20}} = \frac{p}{m} b_{Y/X}.$$

$$\text{Análogamente } b'_{X'/Y'} = \frac{m}{p} b_{X/Y}.$$

Es decir, **los coeficientes de regresión son invariantes ante un cambio de origen pero no ante un cambio de escala.**

$$\text{El coeficiente de determinación } R^2 = \frac{m'^2_{11}}{m'_{20} \cdot m'_{02}} = \frac{m^2 p^2 m^2_{11}}{m^2 m_{20} \cdot p^2 m_{02}} = \frac{m^2_{11}}{m_{20} \cdot m_{02}} = R^2, \text{ es decir, es}$$

**invariante ante cambios de origen y de escala. En consecuencia el coeficiente de correlación es también invariante ante cambios de origen y de escala.**