EJERCICIO MODELO DE ESTADÍSTICA BIDIMENSIONAL

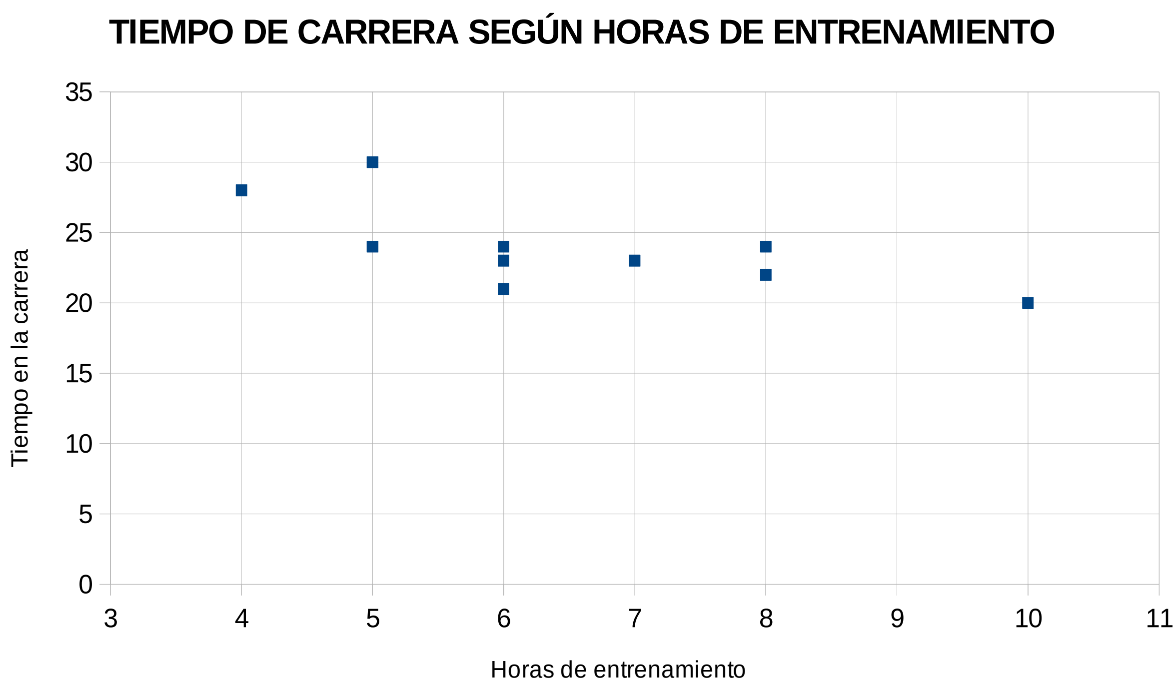
Se ha medido el número medio de horas de entrenamiento a la semana de un grupo de 10 atletas y el tiempo, en minutos, que han hecho en una carrera, obteniendo los siguientes resultados:

Horas de entrenamiento	5	6	6	5	8	6	8	10	7	4
Tiempo carrera	30	23	24	24	22	21	24	20	23	28

a) Estudia la dependencia de estas variables a partir de la nube de puntos.
b) Calcula la covarianza y el coeficiente de correlación.
c) Halla la ecuación de la recta de regresión de Y sobre X y represéntala en el diagrama de dispersión.
d) Halla la ecuación de la recta de regresión de X sobre Y y represéntala en el diagrama de dispersión.
e) Haz una estimación del tiempo que hará en la carrera un atleta que entrene 9 horas de media a la semana.
f) Haz una estimación del número de horas que había dedicado un atleta a entrenar si su tiempo en la carrera fue de 25 minutos.

SOLUCIÓN:

a) La dependencia en variables cuantitativas se estudia comprobando si la nube de puntos se ajusta a una recta (lineal) o a una curva (no lineal) o bien si son independientes (el valor de una no depende del valor de la otra, no se ajustan a un patrón).



En este caso se observa una relación de dependencia lineal entre las horas de entrenamiento y el tiempo en la carrera.
La dependencia es negativa, a más horas de entrenamiento en menos tiempo corrió la carrera.

b)

Horas de entrenamiento xi	**Tiempo de carrera yi**	x_i^2	y_i^2	$x_i\, y_i$
5	30	25	900	150
6	23	36	529	138
6	24	36	576	144
5	24	25	576	120
8	22	64	484	176
6	21	36	441	126
8	24	64	576	192
10	20	100	400	200
7	23	49	529	161
4	28	16	784	112
65	**239**	**451**	**5795**	**1519**

Media de las x: $\bar{x}=\frac{\sum_{i=1}^{N} x_i}{N}=\frac{65}{10}=6{,}5$ Media de las y: $\bar{y}=\frac{\sum_{i=1}^{N} y_i}{N}=\frac{239}{10}=23{,}9$

Varianza de las x: $\sigma_x^2 = \frac{\sum_{i=1}^{N} x_i^2}{N}-\bar{x}^2 = \frac{451}{10}-6{,}5^2 = 2{,}85 \Rightarrow \sigma_x = \sqrt{\sigma_x^2} = 1{,}6882$

Varianza de las y: $\sigma_y^2 = \frac{\sum_{i=1}^{N} y_i^2}{N}-\bar{y}^2 = \frac{5795}{10}-23{,}9^2 = 8{,}29 \Rightarrow \sigma_y = \sqrt{\sigma_y^2} = 2{,}8792$

Covarianza: $\sigma_{xy} = \frac{\sum_{i=1}^{N} x_i y_i}{N}-\bar{x}\,\bar{y} = \frac{1519}{10}-6{,}5\cdot 23{,}9 = -3{,}45$ dependencia negativa

Coeficiente de correlación: $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-3{,}45}{1{,}688\cdot 2{,}879} = -0{,}7099 \approx -0{,}71$

El coeficiente de correlación, es una medida que determina el grado de **dependencia lineal** entre las variables X e Y. No tiene unidades y su valor está entre -1 y 1. Un valor próximo a 0 indica la ausencia de dependencia lineal.

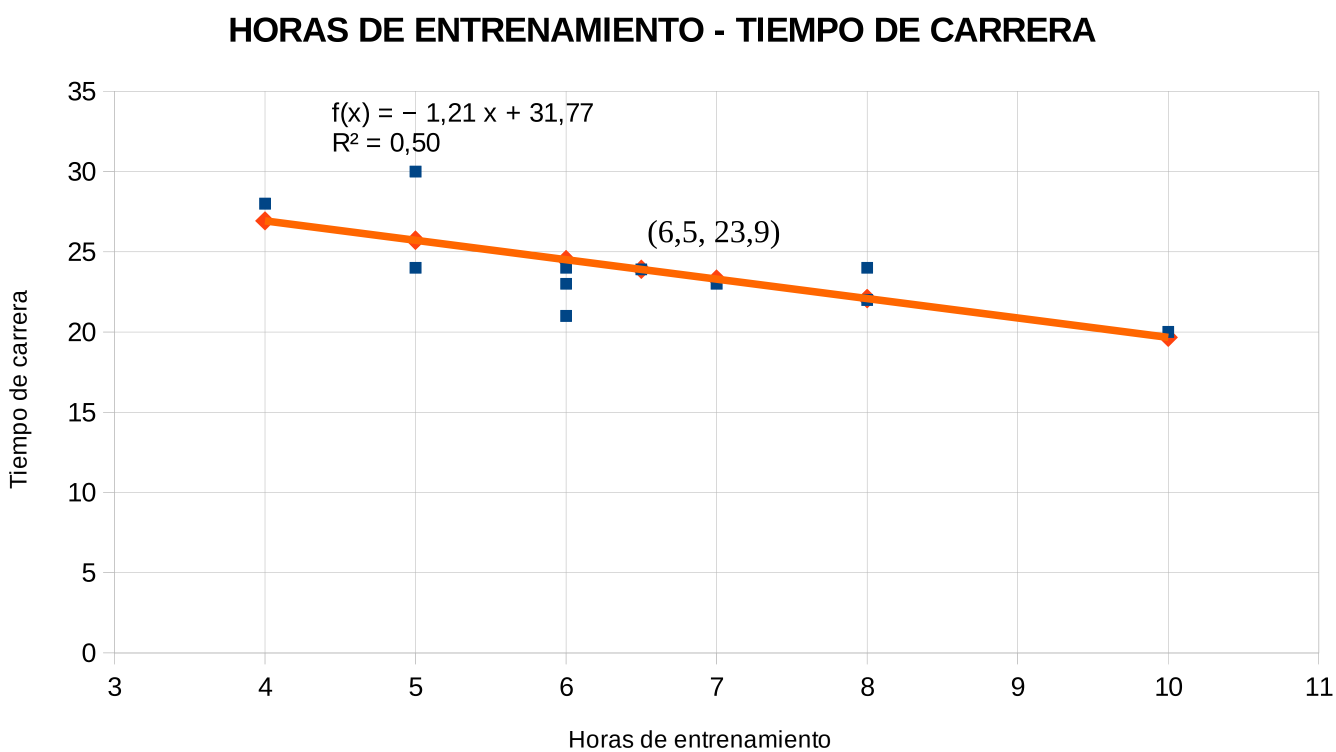
El valor de -0,71 de este ejemplo muestra una dependencia lineal negativa débil. Los puntos de la nube se pueden aproximar por una recta aunque no se ajustarán por completo a ella.

c) De las rectas posibles que podemos ajustar a la nube de puntos elegimos la que hace mínima la suma de las distancias entre las ordenadas de cada punto y las de la recta. A esta recta se le llama recta de regresión de Y sobre X:

$$y-\bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x-\bar{x})$$

En este ejemplo, $y-23{,}9 = \frac{-3{,}45}{2{,}85}(x-6{,}5) \quad \Rightarrow \quad y = -1{,}21(x-6{,}5)+23{,}9$

$$\Rightarrow \quad y = -1{,}21\,x + 31{,}765$$



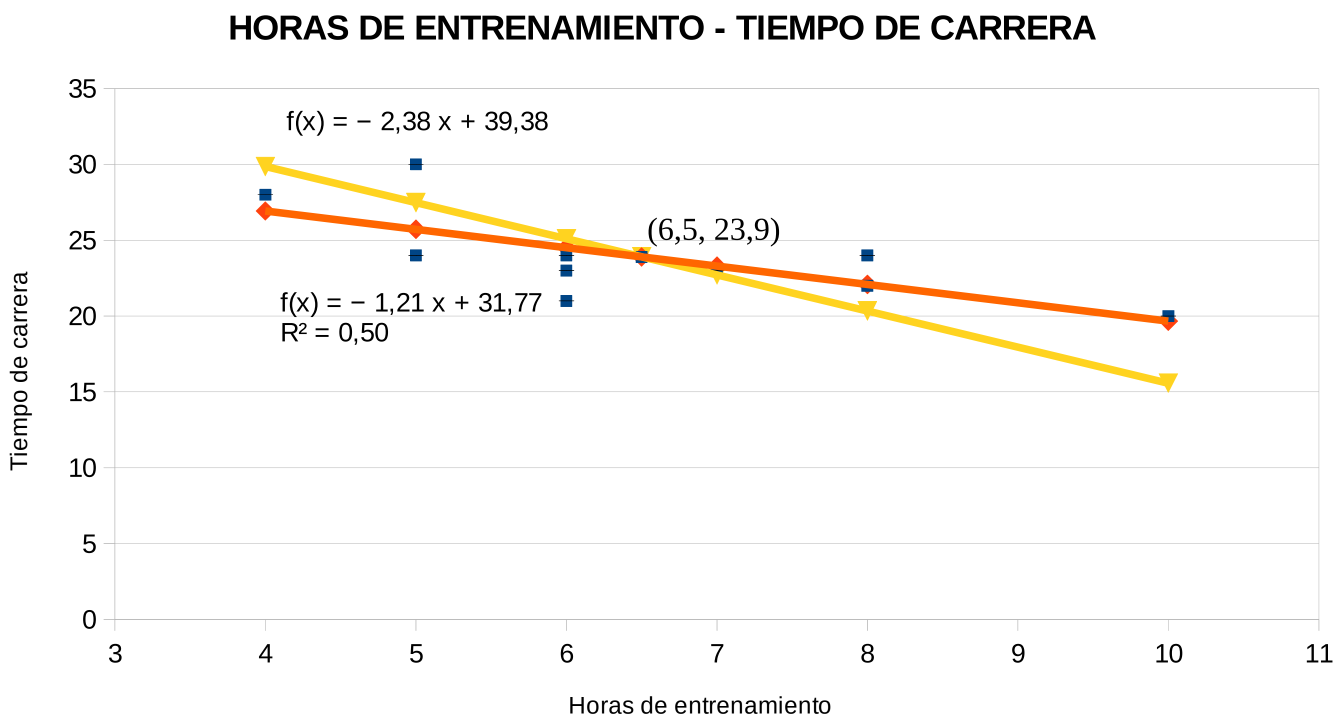
La recta de regresión pasa siempre por el punto $(\bar{x}, \bar{y})=(6{,}5{,}23{,}9)$, centro de gravedad de la nube de puntos.

d) Al minimizar las distancias entre las abscisas de cada punto y las de la recta , obtenemos la recta de regresión de X sobre Y:

$$x-\bar{x} \;=\; \frac{\sigma_{xy}}{\sigma_y^2}(y-\bar{y})$$

En este ejemplo, $x-6{,}5 \;=\; \frac{-3{,}45}{8{,}29}(y-23{,}9)$ $\Rightarrow$ $x \;=\; -0{,}42(y-23{,}9)+6{,}5$

$$\Rightarrow \quad x \;=\; -0{,}42\,y \;+\; 16{,}538$$



Las dos rectas pasan por el centro de gravedad (6,5, 23,9) y el ángulo de ambas rectas no es muy grande debido a la dependencia lineal de las variables.

e) Utilizamos la recta de regresión de Y sobre X:

$\Rightarrow \quad y \;=\; -1{,}21\,x \;+\; 31{,}765 \qquad \Rightarrow y=-1{,}21\cdot 9+31{,}765=20{,}875\approx 21$ minutos

f) Utilizamos la recta de regresión de X sobre Y:

$\Rightarrow \quad x \;=\; -0{,}42\,y \;+\; 16{,}538 \qquad \Rightarrow x=-0{,}42\cdot 25+16{,}538=6{,}038\approx 6$ horas de media a la semana