

Ejercicio 1. Con los datos del ejemplo 2 (esperanza de vida), calcula la covarianza y el coeficiente de correlación lineal.

Country	lifeexpf (y)	birthrate (x)
Somalia	55	46
Tanzania	55	50
Zambia	59	48
Zaire	56	45
Algeria	68	31
Namibia	63	45
Burkina Faso	53	50
Cuba	79	18
Equador	72	28
North Korea	72	24
Mongolia	68	34
Thailand	71	20
Turkey	72	28
France	82	13
Netherlands	81	13

Ejercicio 2. Calcular el coeficiente de correlación para los datos de la tabla siguiente:

x	y
2	5
3	7
4	8
5	13
6	14

Cómo cambiará el coeficiente de correlación si:

- sumamos 5 a la variable x .
- sumamos 5 a ambas variables.
- intercambiamos los valores de x por los de y .

Ejercicio 3. ¿Cuál sería el coeficiente de correlación entre las edades de los cónyuges si las mujeres **siempre** se casaran con hombres 2 años mayores que ellas?

Ejercicio 4. Con los datos del ejemplo 1 (densidad del tráfico) encontrar la recta de regresión que mejor ajusta la velocidad en función de la densidad del tráfico. Por razones que se verán más adelante, tomar la raíz cuadrada de la velocidad.

densidad x	raiz(velocidad) y	densidad x	raiz(velocidad) y
12,7	7,90	18,3	7,16
17	7,12	19,1	7,13
66	4,14	16,5	7,40
50	5,09	22,2	6,82
87,8	3,52	18,6	6,80
81,4	3,66	66	4,11
75,6	3,70	60,3	4,45
66,2	4,23	56	4,60
81,1	3,71	66,3	4,28
62,8	4,23	61,7	4,24
77	3,97	66,6	4,07
89,6	3,55	67,8	4,28

Ejercicio 5 (longitud de la línea de la mano). En un estudio para relacionar la longitud de la línea de la vida en la mano izquierda y la vida de una persona, se han observado los siguientes datos de 50 personas con los siguientes resultados:

$x =$ longitud de línea (en cm)

$y =$ edad al morir (en años)

$$\sum_{i=1}^{50} y_i = 3333, \quad \sum_{i=1}^{50} y_i^2 = 231933, \quad \sum_{i=1}^{50} x_i y_i = 30549,$$

$$\sum_{i=1}^{50} x_i = 459.9, \quad \sum_{i=1}^{50} x_i^2 = 4308.57.$$

- Se pide construir una recta de regresión de y sobre x y encontrar la varianza residual.
- Descomponer la variabilidad y expresarla en una tabla ADEVA (ANOVA).

Ejercicio 6 (salario-escolarización). Un investigador considera que el salario que percibe un individuo es función lineal de sus años de escolarización, esto es, $y_i = \beta_0 + \beta_1 x_i + u_i$, donde y_i representa el salario mensual del individuo i -ésimo, x_i los años de estudio de dicho individuo y u_i es el término de error, que supondremos que verifica las hipótesis habituales del modelo de regresión. El investigador ha obtenido una muestra aleatoria de 100 individuos, de la que conocemos la siguiente información:

$$\sum_{i=1}^{100} y_i = 1180, \quad \sum_{i=1}^{100} y_i^2 = 25543, \quad \sum_{i=1}^{100} x_i y_i = 13469,$$
$$\sum_{i=1}^{100} x_i = 1000, \quad \sum_{i=1}^{100} x_i^2 = 12820.$$

Se pide:

- Obtener estimadores insesgados de los parámetros β_0 , β_1 y σ^2 , así como el coeficiente de determinación. Interpretar los resultados.
- Hallar la descomposición de la variabilidad (tabla ADEVA).

Ejercicio 7.

- a) Hallar un intervalo de confianza para la pendiente de la recta $y=5.9+0.59x$, donde y es el salario mensual y x son los años de escolarización, para un nivel de confianza del 90%.
- b) Lo mismo para el término independiente.

Ejercicio 8. Hallar los intervalos de confianza al 95% para β_0 , β_1 para $y=79.3210-1.3765x$, donde y es la longitud de la línea de la mano y x es la edad al morir.

Ejercicio 9. Hallar un intervalo de confianza al 90% para σ^2 con los datos del ejercicio *longitud de la línea de la mano*.

Ejercicio 10. Hallar un intervalo de confianza al 95% para σ^2 con los datos del ejercicio *salario-escolarización*.

Ejercicio 11. (producción-fertilizantes) Se dispone de los siguientes datos experimentales obtenidos en un campo de cultivo que relacionan la producción con la cantidad de fertilizante aplicado.

Fertilizantes (kg/hect.) x	Producción (kg/hect.) y
100	40
200	45
300	50
400	65
500	70
600	70
700	80

Hallar:

- a) La nube de puntos y dibujar una recta que pase "lo más cerca posible" de todos sus puntos y en especial por el centroide.
- b) La recta de regresión de y sobre x . Interpretar los coeficientes.
- c) La varianza residual.
- d) I.C. al 95% para β_1 .
- e) I.C. al 95% para σ^2 y para σ .
- f) Coeficiente de correlación y coeficiente de determinación. Interpretar los resultados.
- g) Si se aplican 350kg/hect. de fertilizante, ¿qué producción se obtendrá? ¿Y con 1000 kg/hect.?
- h) Realiza un contraste de hipótesis con un nivel de significación del 5% para comprobar si la producción depende del fertilizante.

Ejercicio 12. Dada la ecuación de regresión $y=3+4x$ indicar, mediante el contraste de regresión, si hay evidencias de que y depende linealmente de x con un nivel de significación $\alpha=0.01$ y sabiendo que $R^2=0.4$ y $n=25$.