Soluciones - Estadística bidimensional y regresión lineal

Curso: 4 Eso

Solución del ejercicio 1

a. Totales por deporte: fútbol 14, baloncesto 14, otros 12; total 40. Totales por sexo: chicos 20, chicas 20. Sumamos por columnas para obtener los totales de cada deporte.
b. $\frac{14}{40} = 35\ \%$. Dividimos el total de fútbol entre el total de la clase.
c. $\frac{8}{14} \approx 57{,}1\ \%$ son chicas. Ahora el total de referencia son los 14 alumnos de baloncesto.
d. Chicas: $\frac{4}{20} = 20\ \%$. Chicos: $\frac{10}{20} = 50\ \%$. El fútbol es mucho más popular entre los chicos de esta clase.

Solución del ejercicio 2

a. $r \approx 0{,}995$: correlación positiva y muy fuerte (nube casi sobre una recta creciente).
b. $r \approx -0{,}988$: correlación negativa y muy fuerte (recta decreciente).
c. $r \approx 0{,}096$: correlación casi nula, nube sin forma.
d. No: siempre $-1 \leq r \leq 1$. Un valor de $1{,}3$ indicaría un error de cálculo. Respuesta: $r = 1{,}3$ es imposible.

Solución del ejercicio 3

a. $\bar{x} = \frac{15}{5} = 3$ y $\bar{y} = \frac{21}{5} = 4{,}2$. Sumamos cada variable y dividimos entre 5.
b. $\sum x_i^2 = 1 + 4 + 9 + 16 + 25 = 55$ y $\sum y_i^2 = 4 + 16 + 25 + 16 + 36 = 97$. Elevamos al cuadrado cada dato y sumamos.
c. $\sigma_x^2 = \frac{55}{5} - 3^2 = 2$ y $\sigma_y^2 = \frac{97}{5} - 17{,}64 = 1{,}76$. La varianza es la media de los cuadrados menos el cuadrado de la media.
d. $\sigma_x = \sqrt{2} \approx 1{,}41$ y $\sigma_y \approx 1{,}33$. Hacemos la raíz cuadrada de cada varianza: $\sqrt{1{,}76} \approx 1{,}33$.

Solución del ejercicio 4

a. $\sum x_i y_i = 2 + 8 + 15 + 16 + 30 = 71$. Multiplicamos cada x_i por su y_i y sumamos.
b. $\sigma_{xy} = \dfrac{71}{5} - 3 \cdot 4{,}2 = 14{,}2 - 12{,}6 = 1{,}6$. Usamos las medias del ejercicio anterior, $\bar{x} = 3$ e $\bar{y} = 4{,}2$.
c. Es positiva: cuando x aumenta, y tiende a aumentar (correlación positiva).
d. No: $\sigma_{xy} = \sigma_{yx}$, la covarianza es simétrica. Respuesta: la covarianza no cambia al intercambiar x e y.

Solución del ejercicio 5

a. $r = \dfrac{1{,}6}{1{,}414 \cdot 1{,}327} \approx 0{,}853$.
b. Correlación positiva y fuerte (próxima a 1).
c. $r^2 \approx 0{,}727$: la recta explica aproximadamente el $72{,}7\ \%$ de la variación de y.
d. No: r no cambia al multiplicar una variable por una constante positiva (la covarianza y σ_y se multiplican por 2). Respuesta: r no cambia.

Solución del ejercicio 6

a. $b = \dfrac{1{,}6}{2} = 0{,}8$. Usamos la covarianza 1,6 y la varianza de x, 2.
b. $y - 4{,}2 = 0{,}8(x - 3)$, es decir $y = 0{,}8x + 1{,}8$. Despejamos y y simplificamos: $4{,}2 - 2{,}4 = 1{,}8$.
c. Con $x = 3$: $0{,}8 \cdot 3 + 1{,}8 = 4{,}2 = \bar{y}$. La recta de regresión siempre pasa por el centro de gravedad de la nube.
d. $x = 3{,}5$: $y \approx 4{,}6$ (interpolación, fiable). $x = 8$: $y \approx 8{,}2$ (extrapolación, poco fiable porque está fuera del rango de datos).

Solución del ejercicio 7

a. $y = 70 \cdot 5 + 10 = 360$ km.
b. $y = 70 \cdot 7{,}5 + 10 = 535$ km.
c. No: 15 h está muy fuera del rango $[2, 8]$ de datos y, además, un camionero no puede conducir tantas horas legalmente. La extrapolación no es fiable.
d. La pendiente indica que, en media, se recorren 70 km más por cada hora adicional de conducción (velocidad media).

Solución del ejercicio 8

a. $\bar{x} = 3{,}375$ h y $\bar{y} = 5{,}875$. $\sigma_x \approx 1{,}58$ y $\sigma_y \approx 1{,}9$.
b. $\sigma_{xy} \approx 2{,}922$ y $r \approx 0{,}976$: correlación positiva muy fuerte.

c. $b = \dfrac{\sigma_{xy}}{\sigma_x^2} \approx 1{,}176$ y la recta es $y = 1{,}176x + 1{,}906$.
d. $x = 3{,}5$: nota $\approx 6{,}02$ (interpolación). $x = 10$: nota $\approx 13{,}67$, que supera el 10: la extrapolación no tiene sentido. Respuesta: la recta solo sirve para valores de horas parecidos a los de la muestra.

Solución del ejercicio 9

a. La nube asciende casi en línea recta: correlación positiva fuerte.
b. $r \approx 0{,}995$. Como r está muy cerca de 1, la recta describe bien los datos.
c. $y = 17{,}26x - 204{,}64$ (aproximadamente). Calculamos la pendiente $b = \dfrac{\sigma_{xy}}{\sigma_x^2}$ y la ordenada con las medias.
d. $x = 28$: $y \approx 279$ helados. Pendiente $\approx 17{,}3$: cada grado más vende unos 17 helados más. Respuesta: unos 279 helados.

Solución del ejercicio 10

a. $r \approx -0{,}998$: correlación negativa casi perfecta; a más edad, menos precio.
b. $y = -1{,}289x + 19{,}017$.
c. La pendiente es $\approx -1{,}29$: pierde unos 1289 € de valor por año.
d. A los 4 años: $\approx 13{,}86$ miles de euros. A los 20 años: $-6{,}76$, un precio negativo sin sentido: la extrapolación falla.

Solución del ejercicio 11

a. $\bar{x} = 167{,}5$ cm, $\bar{y} = 62{,}5$ kg y $r \approx 0{,}993$. Las medias se obtienen sumando y dividiendo entre 8.
b. $y = 0{,}886x - 85{,}857$. La pendiente indica unos 0,89 kg más por cada centímetro.
c. $x = 172$: $y \approx 66{,}5$ kg. Sustituimos $x = 172$ en la recta. Respuesta: unos 66,5 kg.
d. $r^2 \approx 0{,}985$: la estatura explica aproximadamente el $98{,}5\ \%$ de la variación del peso; el resto se debe a otros factores.

Solución del ejercicio 12

a. $r \approx -0{,}996$: correlación negativa muy fuerte.
b. $y = -0{,}52x + 9{,}587$.
c. La pendiente es $\approx -0{,}52$: unos 31 minutos menos de sueño por hora adicional de pantalla.
d. No necesariamente. Correlación no implica causalidad: puede haber otros factores, como los horarios o los deberes, o que quien duerme poco use más el móvil.

Solución del ejercicio 13

a. $r \approx 0{,}9999$: casi 1, relación lineal casi perfecta. Los puntos están casi exactamente sobre una recta.
b. $y = 0{,}0615x - 0{,}05$. La ordenada en el origen, $-0{,}05$, es prácticamente cero: con 0 km no se consume nada.
c. La pendiente $\approx 0{,}0615$ L/km equivale a unos $6{,}15$ L cada 100 km. Multiplicamos la pendiente por 100.
d. $x = 350$: $y \approx 21{,}47$ litros. Sustituimos $x = 350$ en la recta.

Solución del ejercicio 14

a. $r \approx -0{,}996$: correlación negativa fuerte. El signo negativo indica que a mayor precio, menos ventas.
b. $y = -12{,}63x + 120{,}16$.
c. La pendiente $\approx -12{,}6$: unos 13 kg menos por cada euro de subida.
d. $x = 4{,}5$: $y \approx 63{,}3$ kg. Ingresos $\approx 4{,}5 \cdot 63{,}3 \approx 285$ €. Los ingresos son el precio por los kilos vendidos.

Solución del ejercicio 15

a. Totales por medio: coche 37, autobús 30, bicicleta 33; total 100.
b. $\dfrac{33}{100} = 33\ \%$.
c. $\dfrac{23}{33} \approx 69{,}7\ \%$.
d. Madrid: $\dfrac{10}{50} = 20\ \%$. Sevilla: $\dfrac{23}{50} = 46\ \%$.
e. Sí: las distribuciones por ciudad son distintas (en Sevilla la bici pesa mucho más). Si no hubiera relación, los porcentajes serían parecidos en las dos ciudades.

Solución del ejercicio 16

a. $r \approx 0{,}995$ y la recta es $y = 1{,}457x + 0{,}4$.
b. $r \approx 0{,}306$ y la recta es $y = 0{,}429x + 3{,}143$.
c. Un solo punto cambia mucho la recta y el coeficiente: r y la pendiente son sensibles a los datos atípicos.
d. Comprobar si es un error de medida o de transcripción, analizarlo por separado y valorar si se justifica excluirlo, pero sin eliminarlo sin razón.

Solución del ejercicio 17

a. $y - 4{,}2 = 0{,}8(x - 3)$, es decir $y = 0{,}8x + 1{,}8$. La pendiente es $b = \dfrac{\sigma_{xy}}{\sigma_x^2} = \dfrac{1{,}6}{2} = 0{,}8$.

b. $b^* = \dfrac{1{,}6}{1{,}76} \approx 0{,}909$, luego $x - 3 = 0{,}909(y - 4{,}2)$. Ahora dividimos la covarianza entre la varianza de y, 1,76.

c. Ambas contienen el punto $(3;\ 4{,}2)$ porque se escriben con $x - \bar{x}$ e $y - \bar{y}$. Al sustituir $x = 3$ e $y = 4{,}2$ los dos miembros valen cero.

d. $b \cdot b^* = 0{,}8 \cdot 0{,}909 \approx 0{,}727 = r^2$. Respuesta: se cumple $r^2 = b \cdot b^*$, porque $0{,}853^2 \approx 0{,}727$.

Solución del ejercicio 18

a. $(-2)^2 = 4$, $(-1)^2 = 1$, $0^2 = 0$, $1^2 = 1$, $2^2 = 4$.

b. $\bar{x} = 0$, $\bar{y} = 2$ y $\sum x_i y_i = -8 - 1 + 0 + 1 + 8 = 0$, luego $\sigma_{xy} = 0 - 0 \cdot 2 = 0$ y $r = 0$.

c. Los puntos forman una parábola en forma de U, una relación fuerte pero no lineal.

d. r solo mide la relación **lineal**: $r = 0$ no significa que no haya relación. Siempre hay que mirar la nube de puntos.

Solución del ejercicio 19

a. Con móvil: 160 (biblioteca 60, no 100). Sin móvil: 40 (biblioteca 30, no 10). Total biblioteca: 90; no: 110.

b. $\dfrac{90}{200} = 45\ \%$.

c. Con móvil: $\dfrac{60}{160} = 37{,}5\ \%$. Sin móvil: $\dfrac{30}{40} = 75\ \%$.

d. No parecen independientes: usar la biblioteca es mucho más frecuente entre quienes no tienen móvil ($75\ \%$ frente al $37{,}5\ \%$). Sin embargo, esto no prueba que el móvil sea la causa.

Solución del ejercicio 20

a. $b = \dfrac{\sigma_{xy}}{\sigma_x^2} = 0$: la recta es horizontal, $y = \bar{y}$; x no ayuda a predecir y.

b. $r = 1$ (todos los puntos alineados con pendiente positiva) y la recta de regresión es exactamente $y = 2x + 1$.

c. $\sigma_{xy} = r\sigma_x\sigma_y = 0{,}9 \cdot 2 \cdot 8 = 14{,}4$ y $b = \dfrac{14{,}4}{4} = 3{,}6$. Recta: $y - 50 = 3{,}6(x - 10)$, es decir $y = 3{,}6x + 14$. Para $x = 12$: $y = 57{,}2$.

d. No. Los incendios grandes causan más daño y requieren más bomberos: la causa común es el tamaño del incendio. Correlación no implica causalidad.