Soluciones - Estadística bidimensional

Curso: 1 Bachillerato Ciencias

Solución del ejercicio 1

a. Totales por filas ($X = 0, 1, 2$): $8,\ 10,\ 7$. Totales por columnas ($Y = 0, 1, 2$): $7,\ 12,\ 6$. Suma: 25 alumnos.

b. $\bar{x} = \dfrac{0 \cdot 8 + 1 \cdot 10 + 2 \cdot 7}{25} = \dfrac{24}{25} = 0{,}96$. $\bar{y} = \dfrac{0 \cdot 7 + 1 \cdot 12 + 2 \cdot 6}{25} = \dfrac{24}{25} = 0{,}96$.

c. La casilla $(1, 1)$ tiene frecuencia 6: $\dfrac{6}{25} = 0{,}24$, es decir, el $24\ \%$.

d. Hay 12 alumnos con una mascota y 6 de ellos tienen un hermano: $\dfrac{6}{12} = 0{,}5$, o sea, el $50\ \%$.

Solución del ejercicio 2

a. La nube sube de izquierda a derecha y los puntos están cerca de una recta imaginaria.

b. La covarianza es positiva porque las dos variables crecen a la vez. Los puntos están bastante alineados, por lo que la correlación es fuerte y positiva (calcularemos $r \approx 0{,}95$).

c. (i) negativa: más años, menos valor. (ii) positiva: más altura, mayor talla. (iii) casi nula: no hay relación entre el día de nacimiento y la nota.

Solución del ejercicio 3

a. $\sum x = 25$, $\sum y = 20$, $\sum x^2 = 4 + 16 + 25 + 49 + 49 = 143$, $\sum y^2 = 1 + 9 + 16 + 25 + 49 = 100$, $\sum xy = 2 + 12 + 20 + 35 + 49 = 118$.

b. $\bar{x} = \dfrac{25}{5} = 5$ y $\bar{y} = \dfrac{20}{5} = 4$.

c. $\sigma_x^2 = \dfrac{143}{5} - 5^2 = 28{,}6 - 25 = 3{,}6$ y $\sigma_y^2 = \dfrac{100}{5} - 4^2 = 20 - 16 = 4$. Entonces $\sigma_x \approx 1{,}90$ y $\sigma_y = 2$.

d. En valor absoluto, $\sigma_y > \sigma_x$. Coeficientes de variación: $\dfrac{1{,}90}{5} = 0{,}38$ y $\dfrac{2}{4} = 0{,}5$. La práctica está relativamente más dispersa.

Solución del ejercicio 4

a. $\sigma_{xy} = \dfrac{118}{5} - 5 \cdot 4 = 23{,}6 - 20 = 3{,}6$.
b. $r = \dfrac{3{,}6}{\sqrt{3{,}6} \cdot 2} = \dfrac{3{,}6}{1{,}897 \cdot 2} \approx 0{,}949$.
c. Es positivo y muy próximo a 1: correlación lineal positiva y fuerte.
d. $r^2 = 0{,}9$. El $90\,\%$ de la variación de la nota práctica queda explicado por la recta que la relaciona con la teórica.

Solución del ejercicio 5

a. $b_{yx} = \dfrac{3{,}6}{3{,}6} = 1$.
b. $y - 4 = 1 \cdot (x - 5) \Rightarrow y = x - 1$.
c. Para $x = 6$: $y = 6 - 1 = 5$.
d. Por cada punto más en la teórica, la nota práctica estimada sube un punto.

Solución del ejercicio 6

a. $b_{xy} = \dfrac{3{,}6}{4} = 0{,}9$ y $x - 5 = 0{,}9(y - 4)$, es decir, $x = 0{,}9y + 1{,}4$.
b. Con $y = 6$: $x = 0{,}9 \cdot 6 + 1{,}4 = 6{,}8$.
c. $b_{yx}\, b_{xy} = 1 \cdot 0{,}9 = 0{,}9 = r^2$. Coincide con el ejercicio 4.
d. Ambas rectas pasan por $(5, 4)$ por construcción: $y - 4 = x - 5$ con $x = 5$ da $y = 4$, y $x - 5 = 0{,}9(y - 4)$ con $y = 4$ da $x = 5$.

Solución del ejercicio 7

a. Al subir la temperatura suben las ventas: correlación positiva y fuerte.
b. $\bar{x} = 28{,}83$, $\bar{y} = 70{,}33$, $\sigma_x^2 = 19{,}81$, $\sigma_y^2 = 418{,}89$, $\sigma_{xy} = 90{,}56$.
c. $r = \dfrac{90{,}56}{4{,}450 \cdot 20{,}467} \approx 0{,}994$. La pendiente es $b_{yx} = \dfrac{90{,}56}{19{,}81} \approx 4{,}572$ y la recta $y - 70{,}33 = 4{,}572(x - 28{,}83)$, es decir, $y \approx 4{,}572x - 61{,}50$.
d. Para $x = 32$: $y \approx 4{,}572 \cdot 32 - 61{,}50 \approx 84{,}8$, unos 85 helados. Es fiable porque $r \approx 0{,}994$ es casi 1 y 32 °C está dentro del rango de datos (22 a 35 °C), por lo que es una interpolación.

Solución del ejercicio 8

a. Con los años el precio baja, así que la correlación es negativa. Con $\sigma_{xy} = -9{,}139$, $\sigma_x = 2{,}794$ y $\sigma_y = 3{,}287$ se obtiene $r \approx -0{,}995$.
b. $b_{yx} = \dfrac{-9{,}139}{7{,}806} \approx -1{,}171$. Además $\bar{x} = 6{,}17$ e $\bar{y} = 8{,}83$, luego $y \approx -1{,}171x + 16{,}05$.

c. Para $x = 6$: $y \approx -1{,}171 \cdot 6 + 16{,}05 \approx 9{,}03$ miles de euros, unos 9 030 €.
d. Cada año de antigüedad resta unos 1 171 € al precio. Para $x = 20$ saldría $y \approx -7{,}4$ miles de euros, un precio negativo: 20 años queda fuera del rango (2 a 10 años) y la extrapolación no vale.

Solución del ejercicio 9

a. $\bar{x} = \dfrac{36}{8} = 4{,}5$, $\bar{y} = \dfrac{47}{8} = 5{,}875$.
$\sum xy = 3 + 8 + 15 + 20 + 30 + 42 + 56 + 72 = 246$, de modo que $\sigma_{xy} = \dfrac{246}{8} - 4{,}5 \cdot 5{,}875 = 30{,}75 - 26{,}4375 = 4{,}3125$.
b. $\sigma_x^2 = \dfrac{204}{8} - 4{,}5^2 = 25{,}5 - 20{,}25 = 5{,}25$ y $b_{yx} = \dfrac{4{,}3125}{5{,}25} = \dfrac{23}{28} \approx 0{,}821$. Recta: $y = \dfrac{23}{28}x + \dfrac{61}{28} \approx 0{,}821x + 2{,}179$.
c. Para $x = 6{,}5$: $y \approx 0{,}821 \cdot 6{,}5 + 2{,}179 \approx 7{,}52$.
d. Como queremos estimar x a partir de y, usamos la recta de X sobre Y: $b_{xy} = \dfrac{4{,}3125}{3{,}609} \approx 1{,}195$ y $x \approx 1{,}195y - 2{,}52$. Con $y = 9$ resulta $x \approx 8{,}2$ horas. Si se despejara x de la recta de Y sobre X saldría $(9 - 2{,}179)/0{,}821 \approx 8{,}3$: es parecido porque r es casi 1, pero el procedimiento correcto es el primero.

Solución del ejercicio 10

a. $\bar{x} = 3$, $\bar{y} = 18{,}6$, $\sigma_x^2 = 2$, $\sigma_y^2 = 21{,}84$ y $\sigma_{xy} = 6{,}6$.
b. $r = \dfrac{6{,}6}{\sqrt{2} \cdot 4{,}673} \approx 0{,}9986$. Pendiente $b_{yx} = \dfrac{6{,}6}{2} = 3{,}3$, recta $y - 18{,}6 = 3{,}3(x - 3)$, es decir, $y = 3{,}3x + 8{,}7$.
c. Con $x = 4{,}5$: $y = 3{,}3 \cdot 4{,}5 + 8{,}7 = 23{,}55$ miles de euros. Con $x = 10$: $y = 3{,}3 \cdot 10 + 8{,}7 = 41{,}7$ miles de euros.
d. La primera: $4{,}5$ está dentro del rango de datos (interpolación). Con 10 se extrapola muy lejos y no sabemos si la relación lineal continúa; las ventas no crecen indefinidamente.

Solución del ejercicio 11

a. $\sum f = 4 + 2 + 5 + 3 + 4 + 2 = 20$. $\sum fx = 40$, luego $\bar{x} = 2$. $\sum fy = 129$, luego $\bar{y} = 6{,}45$.
b. $\sum fx^2 = 92$ da $\sigma_x^2 = 4{,}6 - 4 = 0{,}6$; $\sum fy^2 = 853$ da $\sigma_y^2 = 42{,}65 - 41{,}6025 = 1{,}0475$; $\sum fxy = 246$ da $\sigma_{xy} = 12{,}3 - 2 \cdot 6{,}45 = -0{,}6$.
c. $r = \dfrac{-0{,}6}{\sqrt{0{,}6 \cdot 1{,}0475}} \approx -0{,}757$. Pendiente $b_{yx} = \dfrac{-0{,}6}{0{,}6} = -1$ y recta

$y - 6{,}45 = -(x - 2)$, es decir, $y = -x + 8{,}45$.

d. Con $x = 2{,}5$: $y = -2{,}5 + 8{,}45 = 5{,}95$. El signo negativo indica que, en esta muestra, más horas de pantalla se asocian con notas más bajas (cada hora más, aproximadamente un punto menos). No prueba que las pantallas causen las notas bajas.

Solución del ejercicio 12

a. $\sigma_{xy} = 0{,}9 \cdot 2 \cdot 8 = 14{,}4$.

b. $b_{yx} = \dfrac{14{,}4}{4} = 3{,}6$ y $y - 50 = 3{,}6(x - 10)$, es decir, $y = 3{,}6x + 14$.

c. $b_{xy} = \dfrac{14{,}4}{64} = 0{,}225$ y $x - 10 = 0{,}225(y - 50)$, es decir, $x = 0{,}225y - 1{,}25$.

d. Para $x = 12$: $y = 3{,}6 \cdot 12 + 14 = 57{,}2$. Para $y = 60$: $x = 0{,}225 \cdot 60 - 1{,}25 = 12{,}25$.

Solución del ejercicio 13

a. $r = \dfrac{-5{,}4}{2 \cdot 3} = -0{,}9$: correlación negativa y fuerte.

b. $b_{yx} = \dfrac{-5{,}4}{4} = -1{,}35$, así que $y = -1{,}35x + 12{,}4$. $b_{xy} = \dfrac{-5{,}4}{9} = -0{,}6$, así que $x = -0{,}6y + 8{,}2$.

c. $(-1{,}35)(-0{,}6) = 0{,}81 = (-0{,}9)^2$.

d. Sustituimos $x = -0{,}6y + 8{,}2$ en la primera: $y = -1{,}35(-0{,}6y + 8{,}2) + 12{,}4 = 0{,}81y + 1{,}33$, luego $0{,}19y = 1{,}33$ e $y = 7$ y $x = 4$. Es el punto de medias $(4, 7)$.

Solución del ejercicio 14

a. $r^2 \approx 0{,}990$: el $99\ \%$ de la variación del precio se explica por la edad según el modelo lineal.

b. $x = 3$: $y \approx -3{,}51 + 16{,}05 = 12{,}54$; $x = 8$: $y \approx -9{,}37 + 16{,}05 = 6{,}68$. Ambas están en $[2, 10]$: son interpolaciones.

c. $x = 14$: $y \approx -16{,}39 + 16{,}05 = -0{,}34$, es decir, un precio negativo. No es creíble: es una extrapolación fuera del rango de datos.

d. $-1{,}171x + 16{,}05 > 0 \Rightarrow x$ años. Más allá, el modelo lineal deja de describir la realidad.

Solución del ejercicio 15

a. $\ln 3 = 1{,}099$, $\ln 6 = 1{,}792$, $\ln 12 = 2{,}485$, $\ln 25 = 3{,}219$, $\ln 47 = 3{,}850$. Las diferencias son casi constantes (unos 0,69), así que los puntos están casi

alineados.

b. $a = e^{1{,}103} \approx 3{,}01$ y $b = e^{0{,}693} \approx 2{,}00$. Modelo: $y \approx 3{,}01 \cdot 2^x$.
c. $b \approx 2$: el número de seguidores se duplica cada semana.
d. Para $x = 6$: $y \approx 3{,}01 \cdot 64 \approx 193$ centenas, unos 19 300 seguidores. La extrapolación es poco fiable: ningún canal duplica seguidores indefinidamente.

Solución del ejercicio 16

a. $\bar{x} = \dfrac{80}{6} = 13{,}33$, $\bar{y} = 285$, $\sigma_x^2 = 33{,}22$ y $\sigma_y^2 = 7\,525$.
b. $\sigma_{xy} = -498{,}33$ y $r = \dfrac{-498{,}33}{5{,}764 \cdot 86{,}75} \approx -0{,}997$. Correlación negativa muy fuerte.
c. $b_{yx} = \dfrac{-498{,}33}{33{,}22} = -15$ y $y = -15x + 485$. Cada grado más de temperatura exterior reduce el consumo en unos 15 kWh al mes.
d. A 10 °C: $y = -150 + 485 = 335$ kWh. A 35 °C: $y = -525 + 485 = -40$ kWh, un consumo negativo, imposible. Es una extrapolación que no tiene sentido: la relación lineal solo vale en el rango estudiado (5 a 22 °C).

Solución del ejercicio 17

a. $\bar{x} = \dfrac{1098}{6} = 183$ y $\bar{y} = \dfrac{477}{6} = 79{,}5$.
b. $\sigma_x^2 = 86{,}67$, $\sigma_{xy} = 93{,}17$, $b_{yx} = \dfrac{93{,}17}{86{,}67} \approx 1{,}075$. Recta: $y - 79{,}5 = 1{,}075(x - 183)$, es decir, $y \approx 1{,}075x - 117{,}2$.
c. $\sigma_y^2 = 100{,}92$ y $r = \dfrac{93{,}17}{9{,}309 \cdot 10{,}046} \approx 0{,}996$. Ajuste lineal excelente.
d. Para $x = 195$: $y \approx 1{,}075 \cdot 195 - 117{,}2 \approx 92{,}4$ kg (interpolación, fiable). Para $x = 210$: $y \approx 108{,}5$ kg (extrapolación fuera del rango 170 a 198 cm, poco fiable).

Solución del ejercicio 18

a. $\bar{x} = 3{,}5$, $\bar{y} = 4{,}5$, $\sigma_x^2 = \sigma_y^2 = 2{,}917$, $\sigma_{xy} = 2{,}75$. Entonces $r \approx 0{,}943$ y $b_{yx} = \dfrac{2{,}75}{2{,}917} \approx 0{,}943$. Recta: $y \approx 0{,}943x + 1{,}2$.
b. Con siete datos: $\bar{x} = 4$, $\bar{y} = 6{,}71$, $\sigma_x^2 = 4$, $\sigma_{xy} = 9$. Entonces $b_{yx} = 2{,}25$, $r \approx 0{,}797$ y la recta es $y \approx 2{,}25x - 2{,}29$.
c. Sin el dato atípico: $y(7) \approx 7{,}8$. Con él: $y(7) \approx 13{,}5$. La diferencia es enorme.
d. Un solo dato extremo cambia mucho la pendiente y reduce la correlación. Antes de ajustar conviene dibujar la nube, revisar los datos y decidir si el valor atípico es un error.

Solución del ejercicio 19

a. $\bar{x} = 5{,}5$, $\bar{y} = 3{,}17$, $\sigma_x^2 = 6{,}25$, $\sigma_{xy} = 3{,}75$. Entonces $r = \dfrac{3{,}75}{2{,}5 \cdot 1{,}572} \approx 0{,}954$ y $b_{yx} = \dfrac{3{,}75}{6{,}25} = 0{,}6$, recta $y = 0{,}6x - 0{,}13$.

b. $r \approx 0{,}95$: relación lineal positiva y fuerte. Cuando se venden más helados, hay más rescates.

c. Una correlación fuerte no implica causalidad. Ambas variables dependen de una tercera: la afluencia de gente a la playa en verano (más calor, más bañistas, más helados y más rescates).

d. Por ejemplo, el número de paraguas vendidos y el de resfriados en una ciudad: los dos aumentan en otoño, pero los paraguas no provocan resfriados.

Solución del ejercicio 20

a. $\bar{x}' = 1{,}83$ m y $\bar{y}' = 79\,500$ g; $\sigma_{x'} = \dfrac{9{,}309}{100} = 0{,}0931$ m y $\sigma_{y'} = 1\,000 \cdot 10{,}046 = 10\,046$ g.

b. $\sigma_{x'y'} = \dfrac{1\,000}{100}\sigma_{xy} = 10 \cdot 93{,}17 = 931{,}7$. Como $r' = \dfrac{10\,\sigma_{xy}}{(\sigma_x/100)(1\,000\,\sigma_y)} = \dfrac{10\,\sigma_{xy}}{10\,\sigma_x\sigma_y} = r$, el coeficiente de correlación es el mismo.

c. $b_{y'x'} = \dfrac{\sigma_{x'y'}}{\sigma_{x'}^2} = \dfrac{931{,}7}{0{,}00867} \approx 107\,500$ g por metro, que equivale a 1,075 kg por cm. La pendiente sí depende de las unidades.

d. Con $Y'' = -Y$ la covarianza cambia de signo y $r'' = -r \approx -0{,}996$; la pendiente pasa a $-1{,}075$. La fuerza de la relación es la misma, solo cambia el sentido.