Soluciones - Estadística bidimensional

Curso: 1 Bachillerato Ccss

Solución del ejercicio 1

a. Los puntos $(1,3)$, $(2,4)$, $(3,6)$, $(4,7)$, $(5,10)$ suben de izquierda a derecha y se agrupan cerca de una recta: correlación positiva y fuerte.

b. $\bar{x} = \dfrac{15}{5} = 3$ y $\bar{y} = \dfrac{30}{5} = 6$.

c. $n = 5$; $\sum x = 15$, $\sum y = 30$, $\sum x^2 = 55$, $\sum y^2 = 210$, $\sum xy = 107$. Medias: $\bar{x} = 3$, $\bar{y} = 6$. Varianzas: $\sigma_x^2 = 2$ y $\sigma_y^2 = 6$. Covarianza: $\sigma_{xy} = 3{,}400$. Desviaciones: $\sigma_x = 1{,}414$, $\sigma_y = 2{,}449$.

d. $r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y} = \dfrac{3{,}400}{1{,}414 \cdot 2{,}449} = 0{,}9815$. Es una correlación positiva muy fuerte (muy cerca de 1): a más horas de estudio, mayor nota.

Solución del ejercicio 2

a. Totales: tienen bicicleta $30 + 20 = 50$; no tienen $10 + 40 = 50$. Usan la bicicleta $30 + 10 = 40$; no la usan $20 + 40 = 60$. Total 100.

b. $\dfrac{40}{100} = 0{,}40$: el 40 % usa la bicicleta.

c. $\dfrac{30}{50} = 0{,}60$: el 60 % de los que tienen bicicleta la usa.

d. $\dfrac{10}{50} = 0{,}20$. Los jóvenes que tienen bicicleta la usan el triple de veces (60 % frente a 20 %). Hay relación entre tener y usar, como es lógico.

Solución del ejercicio 3

a. $n = 4$; $\sum x = 20$, $\sum y = 12$, $\sum x^2 = 120$, $\sum y^2 = 46$, $\sum xy = 46$. Medias: $\bar{x} = 5$, $\bar{y} = 3$. Varianzas: $\sigma_x^2 = 5$ y $\sigma_y^2 = 2{,}500$. Covarianza: $\sigma_{xy} = -3{,}500$. Desviaciones: $\sigma_x = 2{,}236$, $\sigma_y = 1{,}581$.

b. La covarianza es negativa ($-3{,}500$): cuando x aumenta, y tiende a disminuir (relación inversa).

c. $r = \dfrac{-3{,}500}{2{,}236 \cdot 1{,}581} = -0{,}9899$. Correlación negativa muy fuerte.

d. La covarianza y r no cambian, porque son simétricas: $\sigma_{xy} = \sigma_{yx}$. Sí cambiaría la recta de regresión (la de x sobre y no es la misma que la de y sobre x).

Solución del ejercicio 4

a. Correlación positiva y muy fuerte: a más renta, más gasto en ocio, y los datos están casi alineados.
b. Correlación negativa y fuerte: cuanto más viejo es el coche, menor es su precio.
c. Correlación casi nula: no hay relación lineal entre las variables (cualquier relación aparente sería casual).
d. Correlación negativa perfecta: todos los puntos están exactamente sobre una recta de pendiente negativa. Es una relación funcional.

Solución del ejercicio 5

a. $n = 6$; $\sum x = 21$, $\sum y = 362$, $\sum x^2 = 91$, $\sum y^2 = 23050$, $\sum xy = 1412$. Medias: $\bar{x} = 3{,}500$, $\bar{y} = 60{,}333$. Varianzas: $\sigma_x^2 = 2{,}917$ y $\sigma_y^2 = 201{,}556$. Covarianza: $\sigma_{xy} = 24{,}167$. Desviaciones: $\sigma_x = 1{,}708$, $\sigma_y = 14{,}197$.
b. $b = \dfrac{\sigma_{xy}}{\sigma_x^2} = \dfrac{24{,}167}{2{,}917} = 8{,}286$. La recta es $y - 60{,}333 = 8{,}286\,(x - 3{,}5)$, es decir, $y = 8{,}286\,x + 31{,}333$.
c. Para $x = 7$: $y = 8{,}286 \cdot 7 + 31{,}333 = 89{,}33$, unos 89 alumnos.
d. $r = 0{,}9967$. Es muy cercano a 1, y $x = 7$ está fuera del rango $(1 \ldots 6)$ pero muy próximo: la predicción es razonable, aunque conviene no extrapolar mucho más.

Solución del ejercicio 6

a. $b = \dfrac{\sigma_{xy}}{\sigma_x^2} = \dfrac{3{,}400}{2} = 1{,}700$. Recta: $y - 6 = 1{,}700(x - 3)$, es decir, $y = 1{,}700\,x + 0{,}900$.
b. $b' = \dfrac{\sigma_{xy}}{\sigma_y^2} = \dfrac{3{,}400}{6} = 0{,}5667$. Recta: $x - 3 = 0{,}5667\,(y - 6)$.
c. La primera recta cumple $y - 6 = b(x - 3)$ y la segunda $x - 3 = b'(y - 6)$. Si $x = 3$ e $y = 6$, ambas se verifican: pasan por $(3,\, 6)$.
d. $b \cdot b' = 1{,}700 \cdot 0{,}5667 = 0{,}9633$ y $r^2 = 0{,}9815^2 = 0{,}9633$. Coinciden (salvo redondeo).

Solución del ejercicio 7

a. $n = 6$; $\sum x = 158$, $\sum y = 1460$, $\sum x^2 = 4334$, $\sum y^2 = 399800$, $\sum xy = 41210$. Medias: $\bar{x} = 26{,}333$, $\bar{y} = 243{,}333$. Varianzas: $\sigma_x^2 = 28{,}889$ y $\sigma_y^2 = 7422{,}222$. Covarianza: $\sigma_{xy} = 460{,}556$. Desviaciones: $\sigma_x = 5{,}375$, $\sigma_y = 86{,}152$. Entonces $r = \dfrac{460{,}56}{5{,}37 \cdot 86{,}15} = 0{,}9946$: correlación positiva muy fuerte.

b. $b = \dfrac{460{,}56}{28{,}89} = 15{,}942$ y $a = \bar{y} - b\bar{x} = 243{,}33 - 15{,}942 \cdot 26{,}33 = -176{,}48$. La recta es $y = 15{,}942\,x - 176{,}48$.
c. Con $x = 30$: $y = 15{,}942 \cdot 30 - 176{,}48 = 301{,}8$, unos 302 helados.
d. La pendiente (15,94) indica que, por cada grado más de temperatura, se venden aproximadamente 16 helados más.

Solución del ejercicio 8

a. $n = 6$; $\sum x = 34$, $\sum y = 208$, $\sum x^2 = 238$, $\sum y^2 = 7954$, $\sum xy = 1362$. Medias: $\bar{x} = 5{,}667$, $\bar{y} = 34{,}667$. Varianzas: $\sigma_x^2 = 7{,}556$ y $\sigma_y^2 = 123{,}889$. Covarianza: $\sigma_{xy} = 30{,}556$. Desviaciones: $\sigma_x = 2{,}749$, $\sigma_y = 11{,}131$. Entonces $r = 0{,}9987$: relación lineal positiva muy fuerte.
b. $b = \dfrac{30{,}556}{7{,}556} = 4{,}044$ y $a = 34{,}667 - 4{,}044 \cdot 5{,}667 = 11{,}750$. Recta: $y = 4{,}044\,x + 11{,}750$.
c. $y(7) = 4{,}044 \cdot 7 + 11{,}750 = 40{,}06$ mil euros.
d. $y(0) = a = 11{,}75$ mil euros: las ventas «base» sin publicidad. Tiene sentido (la tienda vende algo sin anunciarse), aunque 0 está fuera del rango de datos (2 a 10), por lo que es una extrapolación poco fiable.

Solución del ejercicio 9

a. $n = 5$; $\sum x = 21$, $\sum y = 64$, $\sum x^2 = 121$, $\sum y^2 = 898$, $\sum xy = 218$. Medias: $\bar{x} = 4{,}200$, $\bar{y} = 12{,}800$. Varianzas: $\sigma_x^2 = 6{,}560$ y $\sigma_y^2 = 15{,}760$. Covarianza: $\sigma_{xy} = -10{,}160$. Desviaciones: $\sigma_x = 2{,}561$, $\sigma_y = 3{,}970$. Entonces $r = -0{,}9992$: correlación negativa muy fuerte.
b. $b = -1{,}549$ y $a = 19{,}305$. Recta: $y = -1{,}549\,x + 19{,}305$.
c. $y(5) = -1{,}549 \cdot 5 + 19{,}305 = 11{,}56$ miles de euros, unos 11560 €.
d. $y(20) = -1{,}549 \cdot 20 + 19{,}305 = -11{,}67$. Sale un precio negativo, que no es razonable: la recta solo sirve dentro del rango de los datos (1 a 8 años) o cerca de él.

Solución del ejercicio 10

a. $n = 6$; $\sum x = 21$, $\sum y = 38{,}50$, $\sum x^2 = 91$, $\sum y^2 = 255{,}75$, $\sum xy = 122{,}50$. Medias: $\bar{x} = 3{,}500$, $\bar{y} = 6{,}417$. Varianzas: $\sigma_x^2 = 2{,}917$ y $\sigma_y^2 = 1{,}451$. Covarianza: $\sigma_{xy} = -2{,}042$. Desviaciones: $\sigma_x = 1{,}708$, $\sigma_y = 1{,}205$. Entonces $r = -0{,}9923$.
b. $b = \dfrac{-2{,}042}{2{,}917} = -0{,}700$, $a = 8{,}867$. Recta: $y = -0{,}700\,x + 8{,}867$.
c. $y(4{,}5) = -0{,}700 \cdot 4{,}5 + 8{,}867 = 5{,}72$.
d. No. La correlación mide la asociación, no la causa. Pueden existir otras variables (horas de sueño, ambiente de estudio) que influyan en ambas. Correlación no

implica causalidad.

Solución del ejercicio 11

a. $n = 6$; $\sum x = 1035$, $\sum y = 408$, $\sum x^2 = 178975$, $\sum y^2 = 28250$, $\sum xy = 70850$. Medias: $\bar{x} = 172{,}500$, $\bar{y} = 68$. Varianzas: $\sigma_x^2 = 72{,}917$ y $\sigma_y^2 = 84{,}333$. Covarianza: $\sigma_{xy} = 78{,}333$. Desviaciones: $\sigma_x = 8{,}539$, $\sigma_y = 9{,}183$. Entonces $r = 0{,}9989$; $b = 1{,}074$ y $a = -117{,}31$. Recta: $y = 1{,}074\,x - 117{,}31$.
b. $y(178) = 1{,}074 \cdot 178 - 117{,}31 = 73{,}9$ kg.
c. Para estimar x habría que usar la recta de x sobre y: $b' = 0{,}9289$. $x - 172{,}50 = 0{,}9289\,(90 - 68)$ da $x = 192{,}9$ cm. No se puede despejar de la otra recta porque cada recta minimiza un error distinto. Además 90 kg está lejos de los datos: extrapolación.
d. $r^2 = 0{,}9979$: el 99,8 % de la variación del peso se explica por la altura.

Solución del ejercicio 12

a. $n = 5$; $\sum x = 150$, $\sum y = 132$, $\sum x^2 = 5500$, $\sum y^2 = 4004$, $\sum xy = 4680$. Medias: $\bar{x} = 30$, $\bar{y} = 26{,}400$. Varianzas: $\sigma_x^2 = 200$ y $\sigma_y^2 = 103{,}840$. Covarianza: $\sigma_{xy} = 144$. Desviaciones: $\sigma_x = 14{,}142$, $\sigma_y = 10{,}190$. Entonces $b = 0{,}720$, $a = 4{,}800$ y $r = 0{,}9992$.
b. $y(25) = 0{,}720 \cdot 25 + 4{,}800 = 22{,}8$ kg. Está entre 10 y 50 kg: es interpolación y es fiable.
c. $y(100) = 0{,}720 \cdot 100 + 4{,}800 = 76{,}8$ kg. Está muy fuera del rango (10 a 50): extrapolación, poco fiable.
d. El modelo lineal solo describe la tendencia en el intervalo observado. Más allá, la relación puede cambiar: un exceso de abono daña las plantas y baja el rendimiento.

Solución del ejercicio 13

a. $n = 6$; $\sum x = 430$, $\sum y = 5760$, $\sum x^2 = 34050$, $\sum y^2 = 5902600$, $\sum xy = 447500$. Medias: $\bar{x} = 71{,}67$, $\bar{y} = 960$. Varianzas: $\sigma_x^2 = 538{,}89$ y $\sigma_y^2 = 62166{,}67$. Covarianza: $\sigma_{xy} = 5783{,}33$. Desviaciones: $\sigma_x = 23{,}21$, $\sigma_y = 249{,}33$. Entonces $r = 0{,}9992$: correlación positiva muy fuerte.
b. $b = 10{,}73$ y $a = 190{,}88$. Recta: $y = 10{,}73\,x + 190{,}88$. La pendiente indica que cada m² adicional encarece el alquiler en unos 10,7 € al mes.
c. $y(80) = 10{,}73 \cdot 80 + 190{,}88 = 1049{,}4$ €.
d. Residuo: $1300 - 1049{,}4 = 250{,}6$ €. El piso es unos 251 € más caro de lo que predice el modelo (puede deberse a reforma, terraza o ubicación).

Solución del ejercicio 14

a. Con 5 datos: $r = 0{,}9901$; $b = 2$ y $a = -0{,}200$. Recta: $y = 2\,x - 0{,}200$.
b. Con 6 datos: $r = 0{,}2469$; $b = 0{,}457$ y $a = 3{,}400$. Recta: $y = 0{,}457\,x + 3{,}400$.
c. Un solo dato ha reducido r de 0,99 a 0,25 y ha bajado la pendiente de 2 a 0,46. Los coeficientes de la regresión son sensibles a los datos atípicos.
d. Habría que investigar el dato: si se debe a una avería (circunstancia excepcional), lo razonable es retirarlo del análisis y explicarlo; si fuera un dato válido, habría que replantear el modelo.

Solución del ejercicio 15

a. Totales de filas: mujeres $60 + 30 + 30 = 120$ y hombres $30 + 30 + 20 = 80$. Totales de columnas: coche $60 + 30 = 90$, transporte público $30 + 30 = 60$, bicicleta o a pie $30 + 20 = 50$. Total 200.
b. $\dfrac{60}{200} = 0{,}30$.
c. Entre quienes usan el coche (90): mujeres $\dfrac{60}{90} = 0{,}667$ y hombres $\dfrac{30}{90} = 0{,}333$.
d. Mujeres (120): $\dfrac{30}{120} = 0{,}25$. Hombres (80): $\dfrac{30}{80} = 0{,}375$.
e. Si fueran independientes, el porcentaje de usuarios de transporte público sería el mismo (30 %) en ambos grupos. Aquí es el 25 % en mujeres y el 37,5 % en hombres, y también hay diferencias en el coche (50 % de mujeres frente a 37,5 % de hombres). No parece haber independencia: hay cierta relación entre sexo y medio de transporte.

Solución del ejercicio 16

a. $n = 6$; $\sum x = 660$, $\sum y = 40{,}10$, $\sum x^2 = 79600$, $\sum y^2 = 284{,}35$, $\sum xy = 4742$. Medias: $\bar{x} = 110$, $\bar{y} = 6{,}683$. Varianzas: $\sigma_x^2 = 1166{,}667$ y $\sigma_y^2 = 2{,}725$. Covarianza: $\sigma_{xy} = 55{,}167$. Desviaciones: $\sigma_x = 34{,}157$, $\sigma_y = 1{,}651$. Entonces $r = 0{,}9785$.
b. $b = 0{,}0473$ y $a = 1{,}482$. Recta: $y = 0{,}0473\,x + 1{,}482$.
c. $y(110) = 0{,}0473 \cdot 110 + 1{,}482 = 6{,}68$ L/100 km. La tabla entre 100 (5,9) y 120 (6,8) sugiere 6,35: la recta da un valor mayor porque los puntos intermedios quedan por debajo de ella (la nube se curva).
d. Aunque r es alto, el consumo crece cada vez más deprisa (4,8 a 5,1 en el primer tramo y 8,0 a 9,5 en el último): la nube se curva hacia arriba y un modelo cuadrático o exponencial se ajustaría mejor.
e. Estimado: $y(130) = 0{,}0473 \cdot 130 + 1{,}482 = 7{,}63$. Residuo: $7{,}3 - 7{,}63 = -0{,}33$ L/100 km.

Solución del ejercicio 17

a. $n = 8$; $\sum x = 57$, $\sum y = 13080$, $\sum x^2 = 589$, $\sum y^2 = 22109200$, $\sum xy = 104670$. Medias: $\bar{x} = 7{,}12$, $\bar{y} = 1635$. Varianzas: $\sigma_x^2 = 22{,}86$ y $\sigma_y^2 = 90425$. Covarianza: $\sigma_{xy} = 1434{,}38$. Desviaciones: $\sigma_x = 4{,}78$, $\sigma_y = 300{,}71$. Entonces $r = 0{,}9977$.
b. $b = 62{,}75$ y $a = 1187{,}92$. Recta: $y = 62{,}75\,x + 1187{,}92$. Cada año de experiencia supone, en media, unos 62,7 € más al mes.
c. $y(9) = 62{,}75 \cdot 9 + 1187{,}92 = 1752{,}7$ €: interpolación, fiable.
$y(20) = 62{,}75 \cdot 20 + 1187{,}92 = 2442{,}9$ €: está fuera del rango (0 a 15), extrapolación poco fiable.
d. Estimado a 6 años: $1564{,}4$ €. Residuo $= 1650 - 1564{,}4 = 85{,}6$ €: cobra unos 86 € más de lo que predice el modelo.
e. $r^2 = 0{,}9953$: la experiencia explica el 99,5 % de la variabilidad de los salarios.

Solución del ejercicio 18

a. $n = 6$; $\sum x = 9100$, $\sum y = 1270$, $\sum x^2 = 15170000$, $\sum y^2 = 304900$, $\sum xy = 2148000$. Medias: $\bar{x} = 1516{,}7$, $\bar{y} = 211{,}7$. Varianzas: $\sigma_x^2 = 228055{,}6$ y $\sigma_y^2 = 6013{,}9$. Covarianza: $\sigma_{xy} = 36972{,}2$. Desviaciones: $\sigma_x = 477{,}6$, $\sigma_y = 77{,}5$. Entonces $r = 0{,}9983$, muy próximo a 1: la nube está muy alineada.
b. $b = 0{,}1621$ y $a = -34{,}21$. Recta: $y = 0{,}1621\,x - 34{,}21$.
c. $y(1500) = 0{,}1621 \cdot 1500 - 34{,}21 = 209{,}0$ €.
d. En el punto de medias: $\dfrac{\bar{y}}{\bar{x}} = \dfrac{211{,}67}{1516{,}67} = 0{,}1396$, es decir, el 14,0 %.
e. Recta de x sobre y: $x - 1516{,}67 = 6{,}148\,(y - 211{,}67)$ con $b' = 6{,}148$. Para $y = 300$: $x = 2060$ €.

Solución del ejercicio 19

a. $n = 6$; $\sum x = 15$, $\sum y = 336$, $\sum x^2 = 55$, $\sum y^2 = 18922$, $\sum xy = 883$. Medias: $\bar{x} = 2{,}500$, $\bar{y} = 56$. Varianzas: $\sigma_x^2 = 2{,}917$ y $\sigma_y^2 = 17{,}667$. Covarianza: $\sigma_{xy} = 7{,}167$. Desviaciones: $\sigma_x = 1{,}708$, $\sigma_y = 4{,}203$. Entonces $b = 2{,}457$ y $a = 49{,}857$. Recta: $y = 2{,}457\,x + 49{,}857$.
b. $y(7) = 2{,}457 \cdot 7 + 49{,}857 = 67{,}1$ mil habitantes. Está justo fuera del rango (0 a 5), es una extrapolación cercana: razonable.
c. $y(41) = 150{,}6$ mil habitantes. Es una extrapolación a 36 años de distancia del último dato: poco fiable, porque la economía, la natalidad o las migraciones pueden cambiar la tendencia.
d. $r = 0{,}9984$, casi 1. Pero r solo mide el ajuste dentro de los datos observados; no dice nada de lo que pasará fuera de ese rango.
e. La pendiente es $2{,}457$ mil habitantes por año, es decir, unos 2457 por año: es coherente con la afirmación del ayuntamiento (2,4 mil).

Solución del ejercicio 20

a. r no cambia: $r = 0{,}9815$ en ambos casos, porque es invariante frente a cambios lineales de escala con factor positivo.
b. Pendiente inicial: $1{,}700$. Con la nueva escala: 17, es decir, 10 veces mayor (porque y se ha multiplicado por 10).
c. Pendientes: $b = 0{,}8$ y $b' = 0{,}5$. $r^2 = b\,b' = 0{,}4$, luego $r = 0{,}632$ (positivo porque las pendientes lo son). Punto de medias: $y = 0{,}8x + 2$ y $x = 0{,}5y + 1$; sustituyendo, $x = 0{,}5(0{,}8x + 2) + 1 = 0{,}4x + 2$, luego $x = \dfrac{2}{0{,}6} = 3{,}33$ e $y = 0{,}8 \cdot 3{,}33 + 2 = 4{,}67$.
d. $r^2 = b\,b' = (-1{,}2)(-0{,}3) = 0{,}36$, luego $|r| = 0{,}6$. Como las pendientes son negativas, $r = -0{,}6$.