Ejercicios - Estadística bidimensional

Curso: 1 Bachillerato Ciencias · 20

Se debe mostrar el desarrollo de cada ejercicio.

Ejercicio 1

Ejercicio 1 – Hermanos y mascotas

En una clase de 1.º de Bachillerato de 25 alumnos se ha preguntado por el número de hermanos (X) y por el número de mascotas en casa (Y). Los resultados están en la tabla de doble entrada.

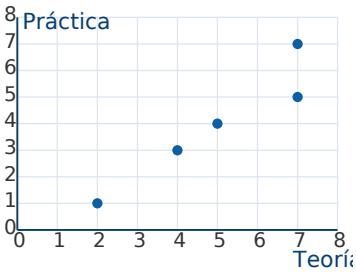
X \ Y	0 mascotas	1 mascota	2 mascotas
0 hermanos	4	3	1
1 hermano	2	6	2
2 hermanos	1	3	3

a. Completa los totales de filas y de columnas y escribe las distribuciones marginales de X y de Y.
b. Calcula las medias $\bar{x}$ y $\bar{y}$ a partir de las distribuciones marginales.
c. ¿Qué porcentaje de alumnos tiene exactamente un hermano y una mascota?
d. Entre los alumnos con una mascota, ¿qué porcentaje tiene un hermano? Esta es una distribución condicionada.

Ejercicio 2

Ejercicio 2 – La nube de puntos

Cinco alumnos han hecho una prueba teórica (X) y una práctica (Y), ambas sobre 10. Las notas son $(2,1)$, $(4,3)$, $(5,4)$, $(7,5)$ y $(7,7)$. La figura muestra la nube de puntos.



a. Describe la forma de la nube: ¿tiende a subir, a bajar o es dispersa?
b. ¿Qué signo tendrá la covarianza? ¿La correlación es fuerte o débil?
c. En cada situación indica si esperas correlación positiva, negativa o casi nula: (i) edad de un coche y su valor; (ii) altura y talla de zapato; (iii) día de nacimiento del mes y nota de matemáticas.

Ejercicio 3

Ejercicio 3 – Medias y varianzas

Con los datos de las cinco notas del ejercicio anterior, $(2,1)$, $(4,3)$, $(5,4)$, $(7,5)$, $(7,7)$, calcula los parámetros de cada variable.

a. Calcula $\sum x$, $\sum y$, $\sum x^2$, $\sum y^2$ y $\sum xy$.
b. Calcula las medias $\bar{x}$ y $\bar{y}$.
c. Calcula las varianzas σ_x^2 y σ_y^2 y las desviaciones típicas.
d. ¿Qué variable está más dispersa? Compara con los coeficientes de variación.

Ejercicio 4

Ejercicio 4 – Covarianza y correlación

Continuamos con los datos $(2,1)$, $(4,3)$, $(5,4)$, $(7,5)$, $(7,7)$, cuyas medias son $\bar{x}=5$ y $\bar{y}=4$ y cuyas varianzas son $\sigma_x^2=3{,}6$ y $\sigma_y^2=4$.

a. Calcula la covarianza $\sigma_{xy}=\dfrac{\sum xy}{n}-\bar{x}\,\bar{y}$.
b. Calcula el coeficiente de correlación lineal $r=\dfrac{\sigma_{xy}}{\sigma_x\sigma_y}$.
c. Interpreta el valor de r.
d. Calcula r^2 e interpreta el resultado como porcentaje.

Ejercicio 5

Ejercicio 5 – Recta de regresión de Y sobre X

Con los mismos datos, $\bar{x}=5$, $\bar{y}=4$, $\sigma_x^2=3{,}6$ y $\sigma_{xy}=3{,}6$, queremos predecir la nota práctica a partir de la teórica.

a. Calcula la pendiente $b_{yx}=\dfrac{\sigma_{xy}}{\sigma_x^2}$.
b. Escribe la recta de regresión de Y sobre X, $y-\bar{y}=b_{yx}(x-\bar{x})$.
c. Estima la nota práctica de un alumno con un 6 en teoría.
d. Interpreta la pendiente.

Ejercicio 6

Ejercicio 6 – Las dos rectas de regresión

Con los datos anteriores, además de la recta de Y sobre X existe la de X sobre Y, que sirve para predecir x a partir de y. Sabemos $\sigma_y^2=4$ y $\sigma_{xy}=3{,}6$.

a. Calcula $b_{xy}=\dfrac{\sigma_{xy}}{\sigma_y^2}$ y escribe la recta de X sobre Y.

b. Estima la nota teórica de un alumno con un 6 en la práctica.
c. Comprueba que $b_{yx} \cdot b_{xy} = r^2$.
d. Comprueba que las dos rectas se cortan en el punto de medias $(\bar{x}, \bar{y})$.

Ejercicio 7

Ejercicio 7 – Helados y temperatura

El dueño de una heladería de Cádiz anota la temperatura máxima del día (X, en °C) y los helados que vende (Y) durante seis días de verano. Quiere saber cuántos helados preparar según el parte meteorológico.

Temperatura (°C)	22	25	28	30	33	35
Helados vendidos	40	55	62	75	90	100

a. Dibuja mentalmente la nube de puntos y di qué tipo de correlación hay.
b. Calcula las medias, las varianzas y la covarianza. Redondea a dos decimales.
c. Calcula r y la recta de regresión de Y sobre X (pendiente con tres decimales).
d. Estima cuántos helados venderá si el parte anuncia 32 °C y explica por qué la predicción es fiable.

Ejercicio 8

Ejercicio 8 – Coche de segunda mano

Una web de coches de ocasión muestra, para el mismo modelo, la edad del vehículo (X, en años) y su precio (Y, en miles de euros). Un estudiante de Murcia quiere comprar uno de 6 años y le interesa saber cuánto es razonable pagar.

Edad (años)	2	4	5	7	9	10
Precio (miles de €)	14	11	10	8	6	4

a. Indica el signo de la correlación y calcula r.
b. Halla la recta de regresión de Y sobre X con tres decimales en la pendiente.
c. Estima el precio de un coche de 6 años.
d. Interpreta la pendiente. ¿Es sensato usar la recta para un coche de 20 años?

Ejercicio 9

Ejercicio 9 – Horas de estudio y nota

Un profesor de un instituto de Zaragoza recoge las horas semanales que dedican a estudiar matemáticas ocho alumnos (X) y su nota en el trimestre (Y).

a. Calcula $\bar{x}$, $\bar{y}$ y la covarianza.
b. Calcula la recta de regresión de Y sobre X en forma de fracción y en decimales.
c. Estima la nota de un alumno que estudia 6,5 horas.
d. ¿Cuántas horas semanales corresponderían a un 9 según la recta? ¿Cuál es el

inconveniente de usar esta recta para ese cálculo?

Ejercicio 10

Ejercicio 10 – Publicidad y ventas

Una tienda de ropa deportiva de Bilbao ha invertido en publicidad (X, en miles de euros al mes) y ha anotado las ventas del mes (Y, en miles de euros) en cinco meses distintos: $(1, 12)$, $(2, 15)$, $(3, 19)$, $(4, 22)$ y $(5, 25)$.

Publicidad (miles de €)	1	2	3	4	5
Ventas (miles de €)	12	15	19	22	25

a. Calcula las medias, las varianzas y la covarianza.
b. Calcula r y la recta de regresión de Y sobre X.
c. Estima las ventas si invierte 4,5 miles de euros y también si invierte 10 miles.
d. ¿Cuál de las dos estimaciones es más fiable? Razónalo.

Ejercicio 11

Ejercicio 11 – Pantallas y notas con frecuencias

En un instituto se ha preguntado a 20 alumnos por las horas diarias que pasan con pantallas de ocio (X) y se ha anotado su nota media (Y). Los datos vienen agrupados con su frecuencia absoluta f.

Horas X	Nota Y	Alumnos f
1	8	4
1	6	2
2	7	5
2	6	3
3	5	4
3	6	2

a. Comprueba que $n = 20$ y calcula $\bar{x}$ e $\bar{y}$ con las frecuencias.
b. Calcula σ_x^2, σ_y^2 y σ_{xy}.
c. Calcula r y la recta de regresión de Y sobre X.
d. Estima la nota media de un alumno que dedica 2,5 horas y comenta qué significa el signo.

Ejercicio 12

Ejercicio 12 – Datos resumidos

De una muestra se conocen los parámetros $\bar{x} = 10$, $\bar{y} = 50$, $\sigma_x = 2$, $\sigma_y = 8$ y el coeficiente de correlación $r = 0{,}9$.

a. Calcula la covarianza $\sigma_{xy} = r\,\sigma_x\sigma_y$.
b. Halla la recta de regresión de Y sobre X.
c. Halla la recta de regresión de X sobre Y.
d. Estima y para $x = 12$ y x para $y = 60$.

Ejercicio 13

Ejercicio 13 – Rectas que se cortan

Una muestra tiene $\bar{x} = 4$, $\bar{y} = 7$, $\sigma_x^2 = 4$, $\sigma_y^2 = 9$ y $\sigma_{xy} = -5{,}4$.

a. Calcula r e interpreta su signo y su valor.
b. Escribe las dos rectas de regresión.
c. Comprueba que $b_{yx}\,b_{xy} = r^2$.
d. Calcula el punto de corte de las dos rectas resolviendo el sistema.

Ejercicio 14

Ejercicio 14 – Interpolar y extrapolar

Con los datos de los coches de segunda mano (ejercicio 8) la recta es $y \approx -1{,}171x + 16{,}05$ (x en años, y en miles de euros) y $r \approx -0{,}995$.

a. Calcula r^2 e interpreta el valor.
b. Estima el precio con 3 años y con 8 años. ¿Son interpolaciones o extrapolaciones?
c. ¿Qué precio da la recta con 14 años? ¿Es creíble?
d. ¿Cuál es el valor máximo de x para que el precio estimado sea positivo? Comenta.

Ejercicio 15

Ejercicio 15 – Seguidores en una red social

Un canal de baloncesto de un instituto de Alicante abre un perfil en una red social. El número de seguidores (en centenas) al cabo de x semanas es el de la tabla. Parece crecer de forma exponencial, $y = a \cdot b^x$, y queremos comprobarlo con un ajuste lineal de los logaritmos.

Semana x	0	1	2	3	4
Seguidores y (centenas)	3	6	12	25	47

a. Calcula $z = \ln y$ para cada semana (tres decimales) y comprueba si los puntos (x, z) están alineados.
b. La recta de regresión de Z sobre X es $z \approx 1{,}103 + 0{,}693x$ ($r \approx 0{,}9998$). Deduce a y b en $y = a \cdot b^x$.
c. ¿Qué significa el valor de b en este contexto?
d. Estima los seguidores en la semana 6 y comenta la fiabilidad.

Ejercicio 16

Ejercicio 16 – Calefacción y temperatura exterior

Una familia de Burgos quiere entender su gasto en calefacción. Ha anotado la temperatura media exterior del mes (X, en °C) y el gas consumido (Y, en kWh) durante seis meses de invierno y primavera.

Temperatura (°C)	5	8	12	15	18	22
Gas (kWh)	420	360	300	250	220	160

a. Calcula las medias y las varianzas de X e Y.
b. Calcula la covarianza y el coeficiente de correlación.
c. Escribe la recta de regresión de Y sobre X e interpreta la pendiente.
d. Estima el consumo de un mes a 10 °C. ¿Qué consumo daría a 35 °C? ¿Tiene sentido?

Ejercicio 17

Ejercicio 17 – Altura y peso de los jugadores

El entrenador de un equipo juvenil de baloncesto de Girona ha medido la altura (X, en cm) y el peso (Y, en kg) de seis jugadores: $(170, 65)$, $(175, 72)$, $(180, 75)$, $(185, 82)$, $(190, 88)$ y $(198, 95)$. Con esos datos quiere estimar el peso de dos fichajes: uno de 195 cm y otro de 210 cm.

a. Calcula las medias y comprueba que $\bar{x} = 183$ e $\bar{y} = 79{,}5$.
b. Calcula σ_x^2, σ_{xy} y la recta de regresión de Y sobre X.
c. Calcula r y valora la calidad del ajuste.
d. Estima el peso de los dos fichajes y razona cuál de las dos estimaciones es más fiable.

Ejercicio 18

Ejercicio 18 – Un dato atípico

Seis alumnos han anotado las horas de ensayo (X) y los puntos obtenidos en un concurso de música (Y): $(1, 2)$, $(2, 3)$, $(3, 5)$, $(4, 4)$, $(5, 6)$, $(6, 7)$. Un séptimo alumno aporta el dato $(7, 20)$, que es claramente atípico (una errata de la hoja de resultados).

a. Con los seis primeros datos calcula r y la recta de regresión de Y sobre X.
b. Añade el dato $(7, 20)$ y calcula de nuevo r y la recta.
c. Compara las predicciones para $x = 7$ con y sin el dato atípico.
d. Explica qué enseña este ejemplo sobre los valores extremos.

Ejercicio 19

Ejercicio 19 – Correlación y causalidad

En seis localidades costeras se han anotado las ventas de helados en la temporada (X, en miles de unidades) y el número de rescates en playas (Y). Los pares son $(2,1)$, $(3,2)$, $(5,2)$, $(6,4)$, $(8,5)$ y $(9,5)$. Un periodista escribe: «comer helados provoca accidentes en el mar».

a. Calcula r y la recta de regresión de Y sobre X.
b. Interpreta el valor de r.
c. Explica por qué la conclusión del periodista no es correcta y propón una variable que explique ambos fenómenos.
d. Da otro ejemplo de dos variables muy correlacionadas sin relación de causa y efecto.

Ejercicio 20

Ejercicio 20 – Desafío: cambio de unidades

Con los datos de los jugadores del ejercicio 17, la altura X está en centímetros y el peso Y en kilogramos, con $b_{yx} \approx 1{,}075$ y $r \approx 0{,}996$. Un compañero propone medir la altura en metros ($X' = X/100$) y el peso en gramos ($Y' = 1\,000\,Y$).

a. Calcula cómo cambian las medias y las desviaciones típicas.
b. Demuestra que la covarianza se multiplica por $\frac{1\,000}{100} = 10$ y que r no cambia.
c. Calcula la nueva pendiente $b_{y'x'}$ e interprétala.
d. Si midiéramos el peso con signo contrario ($Y'' = -Y$), ¿qué ocurre con r y con la pendiente?