Ejercicios - Estadística bidimensional y regresión lineal

Curso: 4 Eso · 20

Se debe mostrar el desarrollo de cada ejercicio.

Ejercicio 1

Ejercicio 1 – Deporte y sexo en la clase

En una clase de 4.º ESO de 40 alumnos de un instituto de Valladolid se pregunta por el deporte favorito. Los chicos (20) eligen: fútbol 10, baloncesto 6, otros 4. Las chicas (20) eligen: fútbol 4, baloncesto 8, otros 8.

	Fútbol	Baloncesto	Otros	Total
Chicos	10	6	4	20
Chicas	4	8	8	20
Total	?	?	?	?

a. Construye la tabla de doble entrada con las frecuencias marginales.
b. ¿Qué porcentaje de la clase prefiere el fútbol?
c. Entre los alumnos que prefieren baloncesto, ¿qué porcentaje son chicas?
d. Entre las chicas, ¿qué porcentaje prefiere fútbol? Compara con los chicos.

Ejercicio 2

Ejercicio 2 – Tipos de correlación

Para cada conjunto de datos, dibuja la nube de puntos, calcula el coeficiente de correlación y clasifica la correlación (fuerte o débil, positiva o negativa).

a. x: $1, 2, 3, 4, 5$ e y: $2, 4, 5, 7, 9$.
b. x: $1, 2, 3, 4, 5$ e y: $9, 7, 6, 3, 2$.
c. x: $1, 2, 3, 4, 5$ e y: $4, 1, 5, 2, 4$.
d. En un mismo conjunto, ¿puede ser $r = 1{,}3$? Razona.

Ejercicio 3

Ejercicio 3 – Medias y desviaciones de dos variables

Se estudian dos variables con los datos $(1, 2)$, $(2, 4)$, $(3, 5)$, $(4, 4)$, $(5, 6)$.

a. Calcula $\bar{x}$ e $\bar{y}$.
b. Calcula $\sum x_i^2$ y $\sum y_i^2$.
c. Calcula las varianzas σ_x^2 y σ_y^2.
d. Calcula las desviaciones típicas.

Ejercicio 4

Ejercicio 4 – Covarianza

Con los mismos datos del ejercicio anterior, $(1,2),\ (2,4),\ (3,5),\ (4,4),\ (5,6)$:

a. Calcula $\sum x_i y_i$.
b. Calcula la covarianza $\sigma_{xy} = \dfrac{\sum x_i y_i}{n} - \bar{x}\bar{y}$.
c. Interpreta el signo.
d. ¿Cambia la covarianza si se intercambian las variables?

Ejercicio 5

Ejercicio 5 – Coeficiente de correlación

Con los datos $(1,2),\ (2,4),\ (3,5),\ (4,4),\ (5,6)$ ya se conocen $\sigma_x = \sqrt{2}$, $\sigma_y \approx 1{,}327$ y $\sigma_{xy} = 1{,}6$.

a. Calcula $r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$.
b. Clasifica la correlación.
c. Calcula r^2 e interprétalo como porcentaje de la variación explicada.
d. Si todos los valores de y se multiplican por 2, ¿cambia r?

Ejercicio 6

Ejercicio 6 – Recta de regresión

Con los datos $(1,2),\ (2,4),\ (3,5),\ (4,4),\ (5,6)$, la recta de regresión de y sobre x es $y - \bar{y} = \dfrac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$.

a. Calcula la pendiente $b = \dfrac{\sigma_{xy}}{\sigma_x^2}$.
b. Escribe la recta en la forma $y = bx + a$.
c. Comprueba que pasa por el punto $(\bar{x}, \bar{y})$.
d. Predice y para $x = 3{,}5$ y para $x = 8$ y comenta la fiabilidad de cada predicción.

Ejercicio 7

Ejercicio 7 – Interpolar y extrapolar

La recta de regresión de los kilómetros recorridos y según las horas de conducción x en una muestra de camioneros que conducen entre 2 y 8 horas es $y = 70x + 10$ y el coeficiente de correlación es $r = 0{,}98$.

a. Predice los kilómetros para $x = 5$ horas.

b. Predice los kilómetros para $x = 7{,}5$ horas.
c. ¿Es razonable predecir para $x = 15$ horas? Razona.
d. Interpreta el significado de la pendiente 70.

Ejercicio 8

Ejercicio 8 – Horas de estudio y nota

Ocho alumnos de 4.º ESO de Santander anotan las horas semanales que estudian matemáticas y la nota de su último examen. Horas: $1, 2, 2, 3, 4, 4, 5, 6$. Notas: $3, 4, 5, 5, 6, 7, 8, 9$.

a. Calcula las medias y las desviaciones típicas de las dos variables.
b. Calcula la covarianza y el coeficiente de correlación.
c. Escribe la recta de regresión de la nota sobre las horas.
d. Predice la nota de un alumno que estudia $3{,}5$ horas y de otro que estudia 10 horas. Comenta.

Ejercicio 9

Ejercicio 9 – Calor y helados

Una heladería de Cádiz anota la temperatura máxima del día (°C) y los helados vendidos. Temperaturas: $18, 21, 24, 27, 30, 33, 36$. Helados: $120, 150, 210, 250, 300, 380, 420$.

a. Dibuja la nube de puntos y estima el tipo de correlación.
b. Calcula r.
c. Obtén la recta de regresión de los helados sobre la temperatura.
d. Predice las ventas con 28 °C y explica qué significa la pendiente.

Ejercicio 10

Ejercicio 10 – La edad del coche y su precio

Un concesionario de coches de segunda mano de Zaragoza registra la edad de siete coches del mismo modelo (en años) y su precio (en miles de euros). Edades: $1, 2, 3, 5, 6, 8, 10$. Precios: $18;\ 16{,}5;\ 15;\ 12{,}5;\ 11;\ 8{,}5;\ 6{,}5$.

a. Calcula el coeficiente de correlación e interprétalo.
b. Halla la recta de regresión del precio sobre la edad.
c. ¿Cuánto pierde de valor, en media, un coche cada año?
d. Predice el precio a los 4 años y a los 20 años. ¿Qué ocurre en el segundo caso?

Ejercicio 11

Ejercicio 11 – Estatura y peso

En un grupo de ocho jóvenes de un equipo de balonmano de Pontevedra se miden la estatura (cm) y el peso (kg). Estaturas: $150, 155, 160, 165, 170, 175, 180, 185$. Pesos: $48, 52, 55, 60, 63, 70, 72, 80$.

a. Calcula las medias y el coeficiente de correlación.
b. Escribe la recta de regresión del peso sobre la estatura.
c. Predice el peso de un joven de 172 cm.
d. Calcula r^2 y explica su significado.

Ejercicio 12

Ejercicio 12 – Pantallas y sueño

Seis estudiantes de un instituto de Alicante anotan las horas diarias que pasan ante la pantalla del móvil y las horas que duermen. Pantalla: $1, 2, 3, 4, 5, 6$. Sueño: $9;\ 8{,}6;\ 8{,}1;\ 7{,}4;\ 7{,}1;\ 6{,}4$.

a. Calcula r y clasifica la correlación.
b. Halla la recta de regresión del sueño sobre la pantalla.
c. ¿Cuántas horas de sueño se pierden, en media, por cada hora adicional de pantalla?
d. ¿Se puede concluir que la pantalla es la causa de dormir menos? Razona.

Ejercicio 13

Ejercicio 13 – Kilómetros y consumo

Un conductor de reparto de Sevilla anota los kilómetros de cinco trayectos y los litros de gasolina consumidos. Kilómetros: $100, 200, 300, 400, 500$. Litros: $6{,}2;\ 12{,}1;\ 18{,}5;\ 24{,}4;\ 30{,}8$.

a. Calcula r.
b. Halla la recta de regresión de los litros sobre los kilómetros.
c. Interpreta la pendiente como consumo cada 100 km.
d. Predice los litros para un trayecto de 350 km.

Ejercicio 14

Ejercicio 14 – Precio y demanda

Un puesto del mercado de un pueblo de Murcia anota, durante seis semanas, el precio del kilo de fresas (euros) y los kilos vendidos. Precios: $2, 3, 4, 5, 6, 7$. Ventas: $95, 82, 70, 55, 48, 30$.

a. Calcula el coeficiente de correlación.
b. Halla la recta de regresión de las ventas sobre el precio.
c. ¿Cuánto disminuye la demanda por cada euro más?
d. Predice la demanda si el precio es $4{,}5$ € y calcula los ingresos esperados.

Ejercicio 15

Ejercicio 15 – El transporte en dos ciudades

Se encuesta a 100 personas de Madrid y Sevilla sobre su medio de transporte habitual. En Madrid (50 personas): coche 22, autobús 18, bicicleta 10. En Sevilla (50 personas): coche 15, autobús 12, bicicleta 23.

	Coche	Autobús	Bicicleta	Total
Madrid	22	18	10	50
Sevilla	15	12	23	50
Total	?	?	?	?

a. Construye la tabla de doble entrada con totales.
b. ¿Qué porcentaje del total usa la bicicleta?
c. Entre quienes usan la bicicleta, ¿qué porcentaje es de Sevilla?
d. Compara el porcentaje de usuarios de bicicleta dentro de cada ciudad.
e. ¿Parece que la ciudad y el medio de transporte están relacionados? Razona con las distribuciones condicionadas.

Ejercicio 16

Ejercicio 16 – El efecto de un dato atípico

Se tienen los datos $(1,2),(2,3),(3,5),(4,6),(5,8),(6,9)$. Después se añade un séptimo dato anómalo, $(7,1)$.

a. Calcula r y la recta de regresión sin el dato anómalo.
b. Calcula r y la recta con los siete datos.
c. Compara y explica qué ha pasado.
d. ¿Qué debe hacer el investigador ante un dato así?

Ejercicio 17

Ejercicio 17 – Dos rectas de regresión

Con los datos $(1,2),(2,4),(3,5),(4,4),(5,6)$ se pueden calcular dos rectas de regresión: la de y sobre x y la de x sobre y.

a. Calcula la recta de y sobre x.
b. Calcula la recta de x sobre y: $x-\bar{x}=\dfrac{\sigma_{xy}}{\sigma_y^2}(y-\bar{y})$.
c. Comprueba que las dos rectas pasan por $(\bar{x},\bar{y})$.
d. Comprueba que $r^2=b\cdot b^*$ siendo b y b^* las dos pendientes (de y sobre x y de x sobre y, respectivamente).

Ejercicio 18

Ejercicio 18 – Una relación perfecta que r no detecta

Se tienen los datos x: $-2, -1, 0, 1, 2$ e $y = x^2$: $4, 1, 0, 1, 4$.

a. Comprueba que $y = x^2$ en todos los pares.
b. Calcula la covarianza y el coeficiente de correlación lineal.
c. Dibuja la nube de puntos y describe su forma.
d. ¿Qué conclusión se saca sobre la utilidad de r?

Ejercicio 19

Ejercicio 19 – Móvil y biblioteca

En un instituto de Gijón se pregunta a los 200 alumnos si tienen móvil propio y si usan la biblioteca. Tienen móvil y usan la biblioteca 60; tienen móvil y no la usan 100; no tienen móvil y usan la biblioteca 30; no tienen móvil y no la usan 10.

	Biblioteca sí	Biblioteca no	Total
Móvil sí	60	100	160
Móvil no	30	10	40
Total	?	?	200

a. Construye la tabla de doble entrada con totales.
b. ¿Qué porcentaje usa la biblioteca?
c. Entre quienes tienen móvil, ¿qué porcentaje usa la biblioteca? ¿Y entre quienes no lo tienen?
d. ¿Parecen independientes las dos variables? Razona.

Ejercicio 20

Ejercicio 20 – Desafío: propiedades de la recta de regresión

Resuelve estos retos.

a. Si $r = 0$, ¿cuál es la pendiente de la recta de regresión de y sobre x y cómo es esa recta?
b. Si todos los puntos están en la recta $y = 2x + 1$, ¿cuánto vale r y cuál es la recta de regresión?
c. Se sabe que $\bar{x} = 10$, $\bar{y} = 50$, $\sigma_x = 2$, $\sigma_y = 8$ y $r = 0{,}9$. Escribe la recta de y sobre x y predice y para $x = 12$.
d. Un estudio da $r = 0{,}95$ entre el número de bomberos en un incendio y el daño causado. ¿Significa que los bomberos causan el daño?