



ACTIVIDADES

1. Variable estadística bidimensional.

1. En un grupo de 24 estudiantes se ha realizado un estudio sobre X = "número de horas diarias de estudio" e Y = "número de suspensos en junio", obteniéndose estos resultados:

(2,1) (0,7) (1,3) (3,1) (3,0) (2,7) (1,0) (2,1) (2,0) (3,1) (2,2) (1,0)
(3,0) (1,2) (1,1) (2,0) (3,1) (1,4) (1,2) (2,1) (1,2) (2,1) (0,6) (2,0)

- Ordenar estos datos en una tabla de frecuencias de doble entrada y deducir las tablas de frecuencias marginales de cada una de las variables.
- ¿Qué número de estudiantes estudiaron tres horas diarias?
- ¿Qué número de estudiantes obtuvieron dos suspensos en junio?
- ¿Qué número de estudiantes estudiaron dos horas diarias y obtuvieron un suspenso?

2. Se toma una muestra de 65 hogares para analizar las variables X = "número de personas por hogar" e Y = "número de vehículos por hogar", obteniéndose estos resultados:

(1,1) (1,1) (1,1) (1,1) (1,1) (1,1) (1,1) (2,1) (2,1) (2,1) (2,1) (2,1) (2,1) (2,1) (2,1) (2,1) (2,1)
(3,1) (3,1) (3,1) (3,1) (3,1) (3,1) (3,1) (3,1) (3,1) (3,1) (3,1) (3,1) (4,1) (4,1) (4,1) (4,1) (4,1) (4,1) (4,1)
(4,1) (4,1) (4,1) (4,1) (4,1) (4,1) (4,1) (4,1) (2,2) (2,2) (3,2) (3,2) (3,2) (3,2) (3,2) (3,2) (3,2)
(4,2) (4,2) (4,2) (4,2) (4,2) (4,2) (4,2) (3,3) (4,3) (4,3) (4,3) (4,3) (4,3) (4,3)

- Ordenar estos datos en una tabla de frecuencias de doble entrada y deducir las tablas de frecuencias marginales de cada una de las variables.
- ¿En cuántos hogares hay tres personas? ¿En cuántos hogares hay dos vehículos?
- ¿En cuántos hogares hay cuatro personas y dos vehículos?

2. Dependencia lineal entre variables.

3. Se toma una muestra de 10 alumnos de 1º Bachillerato para analizar X = "calificación obtenida en Matemáticas" e Y = "calificación obtenida en Física–Química", obteniéndose estos resultados:

(6,7) (7,8) (9,8) (6,6) (4,3) (7,7) (3,4) (8,7) (4,5) (5,5)

Representar el diagrama de dispersión y analizar el tipo de dependencia lineal entre X e Y .

4. La siguiente tabla recoge la información obtenida en una encuesta:

X = número de personas en el hogar	3	3	4	4	5	6	8	9
Y = coste de la factura del gas (en €)	50	55	66	70	83	100	125	140

Representar el diagrama de dispersión y analizar el tipo de dependencia lineal entre X e Y .

5. Se desea analizar la repercusión que pueda tener la lluvia en el número de visitantes a un parque de atracciones. Para ello se anotan, durante los últimos diez años, X = "número de días de lluvia durante la temporada" e Y = "número de visitantes en la temporada":

X = número de días lluviosos	10	14	16	17	19	20	22	23	24	25
Y = número de visitantes (en millones)	2,3	2,1	1,9	1,6	1,5	1,4	1,2	1,1	0,7	0,6

Representar la nube de puntos y analizar el tipo de dependencia lineal entre X e Y.

6. En un grupo de alumnos que han obtenido algún suspenso en junio, se desea analizar si existe alguna relación entre la estatura y el número de suspensos obtenidos. Para ello se toma una muestra de 10 alumnos siendo X = "estatura en cm de la persona" e Y = "número de suspensos":

X = estatura (en cm)	160	162	165	166	168	170	172	174	176	180
Y = número de suspensos en junio	2	5	2	4	1	6	2	3	4	1

Representar la nube de puntos y analizar el tipo de dependencia lineal entre X e Y.

3. Correlación lineal.

7. La siguiente tabla muestra los datos recogidos en una muestra de 9 países sobre las variables X = "renta per cápita expresada en miles de euros" e Y = "esperanza de vida en años".

X = renta per cápita (en miles de €)	12	40	34	6	30	42	2	15	9
Y = esperanza de vida (en años)	65	79	75	63	74	82	60	62	62

Calcular el coeficiente de correlación lineal e interpretar el resultado.

8. La siguiente tabla muestra los datos recogidos en una muestra de 10 estudiantes sobre las variables X = "edad en años" e Y = "tiempo empleado en completar un test de comprensión lectora".

X = edad (en años)	14	16	17	14	15	12	13	13	14	16
Y = tiempo empleado (en minutos)	32	34	36	34	32	34	31	36	38	32

Calcular el coeficiente de correlación lineal e interpretar el resultado.

9. La siguiente tabla muestra los datos recogidos durante siete días diferentes sobre las variables X = "cantidad de sardinas llegadas al puerto en kg" e Y = "precio del kg subastado en lonja en euros".

X = cantidad de sardinas (en kg)	1000	1200	1250	1350	1400	1500	1550
Y = precio del kg subastado (en €)	1,80	1,68	1,65	1,50	1,44	1,32	1,20

Calcular el coeficiente de correlación lineal e interpretar el resultado.

10. La siguiente tabla muestra los datos recogidos a seis bicicletas de competición sobre las variables X = "peso en kilogramos" e Y = "precio en euros".

X = peso de la bicicleta (en kg)	6,45	6,50	6,70	7,15	7,25	7,35
Y = precio de la bicicleta (en €)	6000	4700	4200	2700	2600	3300

Calcular el coeficiente de correlación lineal e interpretar el resultado.

4. Recta de regresión. Estimaciones.

11. La tabla siguiente representa los valores de una distribución bidimensional del gasto en publicidad (en miles de euros) y las ventas conseguidas (en miles de euros), durante los seis primeros meses de promoción de cierto producto:

Gastos (X)	1	2	3	4	5	6
Ventas (Y)	10	17	30	28	39	47

- Calcular el coeficiente de correlación lineal e interpretar el resultado.
- Calcular la recta de regresión de Y sobre X y la recta de regresión de X sobre Y.
- Estimar las ventas que se esperan para un gasto en publicidad de 5 500 €. ¿Es fiable esta estimación? ¿Por qué?
- Para obtener unas ventas de 20 000 €, estimar cuánto habría que gastar en publicidad.

12. La tabla siguiente muestra el número de oficinas bancarias (en cientos) y el número de grandes centros comerciales en ocho ciudades españolas:

Oficinas bancarias (X)	31	19	9	7	6	5	5	5
Grandes centros comerciales (Y)	30	16	12	11	11	9	10	5

- Calcular el coeficiente de correlación lineal e interpretar el resultado.
- Calcular la recta de regresión de Y sobre X y la recta de regresión de X sobre Y.
- Estimar el número de grandes centros comerciales que se esperan para una ciudad con unas 1 500 oficinas bancarias. ¿Es fiable esta estimación? ¿Por qué?

13. Una persona se somete a una dieta de adelgazamiento durante cinco semanas. Su peso en kilogramos al término de cada una de esas semanas viene dado en la siguiente tabla:

Semana (X)	1	2	3	4	5
Peso (Y)	89	87	84	82	79

- Calcular el coeficiente de correlación lineal e interpretar el resultado.
- Calcular la recta de regresión de Y sobre X.
- Estimar el peso que se espera al término de la sexta semana.

14. La tabla siguiente representa los valores de una distribución bidimensional de la talla en centímetros y el peso en kilogramos de diez recién nacidos:

Talla (X)	47	48	48	49	50	50	52	52	54	55
Peso (Y)	2,9	3	3,2	3,1	3,2	3,3	3,5	3,7	3,8	4,1

- Calcular el coeficiente de correlación lineal e interpretar el resultado.
- Calcular la recta de regresión de Y sobre X.
- Estimar el peso que se espera para un recién nacido que midiese 53 cm.

15. La tabla siguiente representa los valores de una distribución bidimensional de la edad en años y el número de horas semanales dedicadas a la práctica deportiva de diez personas:

Edad (X)	10	12	18	22	25	30	31	37	43	50
Horas semanales (Y)	2	3	3	8	1	6	4	9	5	1

- Calcular el coeficiente de correlación lineal e interpretar el resultado.
- Calcular la recta de regresión de Y sobre X.
- Estimar el número de horas de práctica deportiva que se espera para alguien de 24 años. ¿Es fiable esta estimación? ¿Por qué?

16. En un mismo día, se hacen seis mediciones de la temperatura en grados centígrados, que hace según la hora del día, recogidas en la tabla siguiente:

Hora (X)	16	17	18	20	23	24
Temperatura (Y)	23	21	21	19	17	17

- Calcular el coeficiente de correlación lineal e interpretar el resultado.
- Calcular la recta de regresión de Y sobre X.
- Estimar la temperatura que se espera a las 21 horas.

17. La tabla siguiente da la temperatura media (en grados centígrados) de seis inviernos consecutivos y el consumo de energía por habitante (en kilowatios por hora) en una ciudad:

Temperatura media (X)	1	8	3	4	2	6
Consumo (Y)	423	347	401	365	412	354

- Calcular el coeficiente de correlación lineal e interpretar el resultado.
- Calcular la recta de regresión de Y sobre X.
- Averiguar el consumo estimado para una temperatura media de 5°C.

18. La tabla siguiente representa los valores de una distribución bidimensional de la edad en años y el número de horas semanales dedicadas a la lectura de diez personas:

Edad (X)	12	14	18	22	28	30	32	36	42	50
Horas semanales (Y)	5	2	2	6	1	7	3	0	1	9

- Calcula el coeficiente de correlación lineal e interpretar el resultado.
- Calcula la recta de regresión de Y sobre X.
- Estima el número de horas de lectura que se espera para una persona de 29 años. ¿Es fiable esta estimación? ¿Por qué?

SOLUCIONES

1. a) Tabla de frecuencias de doble entrada

$y \backslash x$	0	1	2	3	Total
0		2	3	2	7
1		1	4	3	8
2		3	1		4
3		1			1
4		1			1
6	1				1
7	1		1		2
Total	2	8	9	5	24

Tabla de frecuencias marginales de X

x_i	f_i
0	2
1	8
2	9
3	5
	24

Tabla de frecuencias marginales de Y

y_j	f_j
0	7
1	8
2	4
3	1
4	1
6	1
7	2
	24

b) 5 estudiantes; c) 4 estudiantes; d) 4 estudiantes.

2. a) Tabla de frecuencias de doble entrada

$y \backslash x$	1	2	3	4	Total
1	7	10	11	16	44
2		2	6	7	15
3			1	5	6
Total	7	12	18	28	65

Tabla de frecuencias marginales de X

x_i	f_i
1	7
2	12
3	18
4	28
	65

Tabla de frecuencias marginales de Y

y_j	f_j
1	44
2	15
3	6
	65

b) 18 hogares; en 15 hogares; c) 7 hogares.

3. Dependencia lineal fuerte positiva.

4. Dependencia lineal fuerte positiva.

5. Dependencia lineal fuerte negativa.

6. Dependencia lineal débil o nula.

7. Los parámetros son $\bar{x} = 21,11$; $\bar{y} = 69,11$; $\sigma_x = 14,53$; $\sigma_y = 7,90$; $\sigma_{xy} = 112,2$; el coeficiente de correlación es $r = 0,977$, que indica una dependencia lineal fuerte positiva, es decir, una mayor renta per cápita va ligada a una mayor esperanza de vida.

8. Los parámetros son $\bar{x} = 14,4$; $\bar{y} = 33,9$; $\sigma_x = 1,50$; $\sigma_y = 2,12$; $\sigma_{xy} = 0,14$; el coeficiente de correlación es $r = 0,044$, que indica una dependencia lineal prácticamente nula, es decir, la edad y el tiempo empleado en completar el test no guardan relación de dependencia lineal.

9. El coeficiente de correlación es $r = -0,98$, que indica una dependencia lineal fuerte negativa, es decir, una mayor oferta de sardinas disponible en el mercado va ligada a una bajada del precio por kilogramo.

10. El coeficiente de correlación es $r = -0,90$, que indica una dependencia lineal fuerte negativa, es decir, un menor peso de la bicicleta va ligado a una subida de su precio.

11. a) $\bar{x} = 3,5$; $\bar{y} = 28,5$; $\sigma_x = 1,71$; $\sigma_y = 12,45$; $\sigma_{xy} = 20,75$; $r \cong 0,976$, que indica una dependencia lineal fuerte positiva: un mayor gasto en publicidad va ligado a una mayor venta del producto; b) $\sigma_x^2 = 2,92$; $\sigma_y^2 = 154,92$; las rectas de regresión tienen por ecuación $y = 7,11x + 3,63$; $x = 0,134y - 0,315$, respectivamente; c) Se estiman unas ventas por valor de 42 735 €. Es fiable ya que el coeficiente de correlación r es próximo a 1 y el valor 5 500 € se encuentra dentro del rango [1000, 6000]; d) Habría que gastar unos 2 365 € en publicidad.
12. a) $\bar{x} = 10,875$; $\bar{y} = 13$; $\sigma_x = 8,78$; $\sigma_y = 7,04$; $\sigma_{xy} = 59,25$; $r \cong 0,959$, que indica una dependencia lineal fuerte positiva: un mayor número de oficinas bancarias va ligado a un mayor número de grandes centros comerciales; b) $\sigma_x^2 = 77,11$; $\sigma_y^2 = 49,5$; $y = 0,77x + 4,64$; c) Se espera 16 grandes centros comerciales. Esta estimación es fiable ya que el coeficiente de correlación r es próximo a 1 y el valor 1 500 € se encuentra dentro del rango [500, 3100]
13. a) $\bar{x} = 3$; $\bar{y} = 84,2$; $\sigma_x = 1,414$; $\sigma_y = 3,544$; $\sigma_{xy} = -5$; $r \cong -0,998$, que indica una dependencia lineal fuerte negativa: el aumento del número de semanas de dieta está ligado a una progresiva disminución del peso; b) $\sigma_x^2 = 2$; $\sigma_y^2 = 12,56$; $y = -2,5x + 91,7$; c) Se espera 76,7 kg de peso. Esta estimación es fiable ya que el coeficiente de correlación r es próximo a -1 y el valor 6, aunque está fuera del rango [1, 5], está cerca de uno de los extremos.
14. $\bar{x} = 50,5$; $\bar{y} = 3,38$; $\sigma_x = 2,54$; $\sigma_y = 0,37$; $\sigma_{xy} = 0,90$; $r \cong 0,97$, que indica una dependencia lineal fuerte positiva: una mayor talla va ligada a un mayor peso de los recién nacidos; b) $\sigma_x^2 = 6,45$; $\sigma_y^2 = 0,134$; $y = 0,14x - 3,67$; c) Se espera 3,75 kg de peso. Esta estimación es fiable ya que el coeficiente de correlación r es próximo a 1 y el valor 53 se encuentra dentro del rango [47, 55]
15. $\bar{x} = 27,8$; $\bar{y} = 4,2$; $\sigma_x = 12,36$; $\sigma_y = 2,64$; $\sigma_{xy} = 4,54$; $r \cong 0,139$, que indica una dependencia lineal muy débil positiva: apenas existe correlación entre la edad y el tiempo dedicado a la práctica deportiva; b) $\sigma_x^2 = 152,76$; $\sigma_y^2 = 6,96$; $y = 0,03x + 3,37$; c) Se esperan unas 4 horas. Esta estimación no es fiable ya que el coeficiente de correlación r es próximo a cero.
16. a) $\bar{x} = 19,67$; $\bar{y} = 19,67$; $\sigma_x = 2,98$; $\sigma_y = 2,21$; $\sigma_{xy} = -6,44$; $r \cong -0,978$, que indica una dependencia lineal fuerte negativa: el aumento de la hora desde las 16:00 h hasta las 24:00 h está ligado a una progresiva bajada de las temperaturas; b) $\sigma_x^2 = 8,89$; $\sigma_y^2 = 4,89$; $y = -0,725x + 33,925$; c) Se esperan 18,7°C de temperatura. Esta estimación es fiable ya que el coeficiente de correlación r es próximo a -1 y el valor 21 está dentro del rango [16, 24]
17. a) $\bar{x} = 4$; $\bar{y} = 383,67$; $\sigma_x = 2,38$; $\sigma_y = 29,51$; $\sigma_{xy} = -66,3$; $r \cong -0,944$, que indica una dependencia lineal fuerte negativa: la bajada de las temperaturas en los inviernos está ligada a un progresivo aumento del consumo de energía por habitante; b) $\sigma_x^2 = 5,67$; $\sigma_y^2 = 870,56$; $y = -11,71x + 430,49$; c) Se espera un consumo de casi 372 kw/h por habitante. Esta estimación es fiable ya que el coeficiente de correlación r es próximo a -1 y el valor 5 está dentro del rango [1, 8]
18. $\bar{x} = 28,4$; $\bar{y} = 3,6$; $\sigma_x = 11,62$; $\sigma_y = 2,84$; $\sigma_{xy} = 5,97$; $r \cong 0,181$, que indica una dependencia lineal muy débil positiva: apenas existe correlación entre la edad y el tiempo semanal dedicado a la lectura; b) $\sigma_x^2 = 135,04$; $\sigma_y^2 = 8,04$; $y = 0,044x + 2,35$; c) Se esperan unas 3,6 horas. Esta estimación no es fiable ya que el coeficiente de correlación r es próximo a cero.