

## 6. EJERCICIOS

1. En una investigación biomédica sobre la relación entre el hábito de fumar (X) y la hipertensión (Y) se tomaron estos datos:

	Con hipertensión	Sin hipertensión	
No fumador	20	50	
Fumador moderado	40	25	
Fumador empedernido	60	5	

- Calcula la distribución conjunta de frecuencias relativas en porcentaje.
- Representa estos datos en un diagrama de barras.
- Halla la distribución de  $Y \setminus X$  = “Fumador empedernido”. Halla la moda.
- Halla la distribución de  $X \setminus Y$  = “Con hipertensión”. Halla la moda.
- Halla las distribuciones marginales de X e Y.

2. Se quiere analizar el gasto mensual en telefonía de 30 hogares, para ello se ha estudiado el número de móviles (Y) y el coste mensual en € (X):

Coste (€) \ Nº móviles	1	2	3	4	
[0, 20)	5	3	1	0	
[20, 40)	1	3	3	1	
[40, 60)	0	1	4	5	
[60, 80)	0	0	1	2	

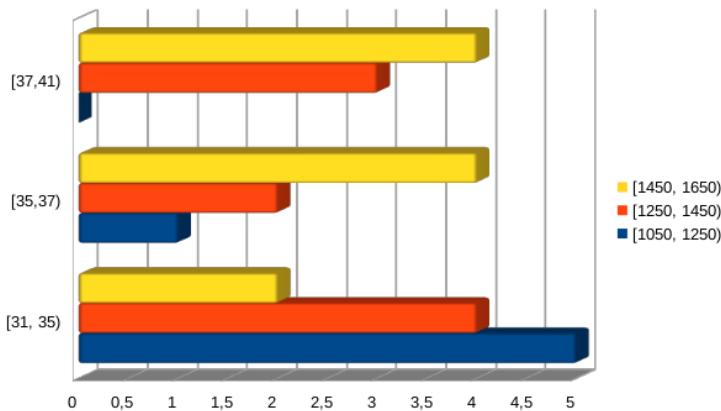
- Halla la distribución del número de móviles condicionado a un gasto entre 40€ y 60€. Halla la moda.
- Halla la distribución del gasto entre los hogares que tienen 2 móviles. ¿Qué porcentaje de estos hogares tienen un gasto entre 20€ y 40€?
- Halla las distribuciones marginales de X e Y.
- Halla la media y la desviación típica de las distribuciones marginales.

3. En una competición de atletismo se ha estudiado la relación entre las horas de entrenamiento diarias (Y) y el puesto conseguido (X) por 20 atletas sabiendo que el 20% obtuvieron un primer puesto. Completa esta tabla de frecuencias con los datos que faltan:

Puesto (X) \ Horas (Y)	6	8	10	12	
1	1	2	1		
2	1	3			11
3	2			1	
		6	8		

- a) Calcula la distribución conjunta de frecuencias relativas en porcentaje. ¿Qué porcentaje de atletas entrenó 10 horas o más?
- b) Halla la distribución del número de horas de entrenamiento condicionado a un primer puesto. Halla la moda.
- c) Halla la distribución de puestos entre los que entrenaron 8 horas. ¿Qué porcentaje de estos atletas tienen obtuvieron un primer o segundo puesto?
- d) Halla las distribuciones marginales de X e Y.
- e) Halla la media y la desviación típica de las distribuciones marginales.

4. El número de horas trabajadas semanalmente (X) y el salario bruto semanal de un grupo de trabajadores de una empresa se recoge en la gráfica:



- a) Pasa los datos a una tabla de doble entrada.
- b) ¿A cuántos trabajadores se estudió?
- c) Halla el número medio de horas trabajadas semanalmente.
- d) ¿Cuántas horas trabaja la mayoría de las personas mejor pagadas?
- e) Calcula el número de horas trabajadas por las personas que tienen un salario de menos de 1450€.

5. Dados los siguientes ejemplos dibuja su diagrama de dispersión y analiza si observas alguna relación lineal entre ellos:

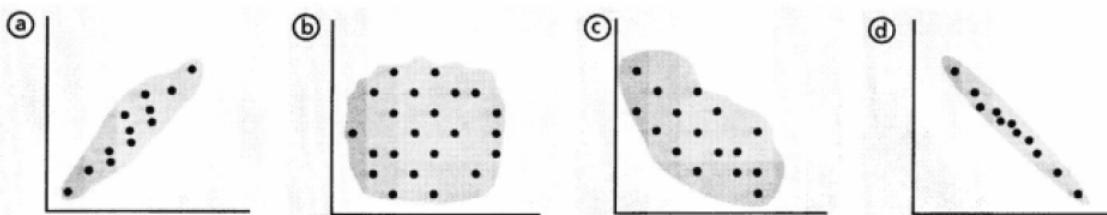
a)

X	2	3	5	5	6	7	8	8
Y	20	18	15	13	10	8	5	6

b)

X	5	9	12	15	17	20	25	30
Y	50	32	4	55	20	2	45	12

6. Los números 0,1; 0,99; 0,6 y 0,89 son los valores absolutos del coeficiente de correlación de las distribuciones bidimensionales cuyas nubes de puntos adjuntamos. Asigna a cada diagrama su coeficiente de correlación cambiando el signo cuando sea necesario.



Indica en cada caso la relación entre las variables.

7. Calcula la covarianza y la correlación lineal de los ejemplos del ejercicio 5. ¿Encuentras relación entre los parámetros calculados y el análisis del diagrama de puntos?

8. Se ha realizado un estudio a ocho empleados de una multinacional sobre el nivel de estrés (escala del 0 al 40) y la satisfacción laboral (escala del 0 al 10):

Nivel de estrés (X)	28	25	16	12	8	31	20	21
Satisfacción laboral (Y)	6	7	8	9	8	4	6	5

¿Se puede afirmar que, cuanto mayor es el nivel de estrés, menor es la satisfacción laboral? Razónalo mediante el diagrama de puntos primero y mediante el coeficiente de correlación después.

**9.** Los salarios brutos mensuales (Y), en euros, y la antigüedad (X), en años, de seis empleados de una empresa fueron:

<b>Antigüedad (X)</b>	0	1	2	3	4	5
<b>Salario bruto (Y)</b>	1200	1375	1600	1780	1900	2200

- a) Dibuja la nube de puntos y analiza si se observa relación lineal entre las variables.
- b) Halla la covarianza y el coeficiente de correlación lineal y compara el resultado con el apartado a)
- c) Halla las rectas de regresión de Y sobre X y de X sobre Y.
- d) Cuando se recogieron los datos había un empleado enfermo con 6 años de antigüedad, ¿podrías estimar cuánto cobra?
- e) El empleado más antiguo de la empresa tiene 20 años de antigüedad, ¿podrías estimar también su salario?
- f) Un empleado dice que cobra 2000€ al mes, ¿qué antigüedad dirías que tiene?

**10.** Una empresa tiene ocho tiendas y cada una realiza su propia campaña publicitaria. Los gastos de cada tienda en publicidad (X) y sus ventas anuales (Y), en miles de euros, vienen dados en la siguiente tabla:

<b>Publicidad (X)</b>	23	34	21	19	18	27
<b>Ventas (Y)</b>	1600	1700	1800	1200	1300	1700

- a) Analiza si la publicidad influye favorablemente en las ventas de forma lineal dibujando un diagrama de dispersión.
- b) La empresa está pensando en abrir otra tienda y se propone invertir 20000 € en publicidad. ¿Qué ingresos por ventas podría esperar?
- c) ¿Qué cantidad debería invertir en publicidad si quiere alcanzar unas ventas anuales de 1,5 millones de euros? ¿Es fiable esta estimación?

**11.** Midiendo la potencia en CV y el consumo en l/100 km en seis modelos de diferentes coches, hemos obtenido los siguientes resultados:

<b>Potencia (X)</b>	95	95	110	115	120	145
<b>Consumo (Y)</b>	4,8	5,1	5,2	6	6,2	7

- a) Halla la recta de regresión de Y sobre X.
- b) Calcula el consumo estimado de un coche de 190 CV. ¿Es fiable esta estimación?

**12.** En una academia para aprender a conducir se han estudiado las semanas de asistencia a clase de sus alumnos y las semanas que tardan en aprobar el examen teórico (desde que se apuntaron a la autoescuela). Los datos correspondientes a seis alumnos son:

<b>Asistencia (X)</b>	6	1	4	3	5	8
<b>Aprobado (Y)</b>	6	5	5	6	5	10

- a) Halla las dos rectas de regresión y represéntalas.
- b) Observando el grado de proximidad entre las dos rectas, ¿cómo crees que será la correlación entre las dos variables?

**13.** La talla media de una muestra de padres es de 1,68 m. con una desviación típica de 5 cm. y la talla media de una muestra de sus hijos es de 1,70 m. con una desviación típica de 7,5 cm. El coeficiente de correlación entre las tallas de hijos y padres es 0,7. Estimar la talla de dos hijos si la talla de sus padres fuera de 1,80 y 1,60 respectivamente.

**14.** En una prueba de natación de 100 m libres un conjunto de 6 nadadores obtienen las siguientes marcas:

55,3 s 54,9 s 58,1 s 52,8 s 56,4 s 57,3 s

a) Calcula la media y desviación típica del conjunto de tiempos.

b) Los mismos nadadores obtienen en la prueba de 100 m mariposa las siguientes marcas:

56,8 s 55,4 s 57,3 s 54,0 s 57,1 s 57,2 s

Calcular el coeficiente de correlación entre ambas pruebas y dar una interpretación. ¿Qué marca obtendría en 100 m mariposa un nadador con una marca de 55 s en 100 m libres?

**15.** Un examen de cierta asignatura consta de dos partes, una teórica (X) y otra práctica (Y). El profesor de la misma quiere ver si existe algún tipo de correlación entre las notas de teoría y práctica. Obtiene que la recta de regresión de Y sobre X es  $4x - 3y = 0$  y la de X sobre Y es  $3x - 2y = 1$ .

a) Calcular el coeficiente de correlación y decir si las variables están o no correlacionadas.

b) Calcular la media de las notas de teoría y práctica.

**16.** Un conjunto de datos bidimensionales ( $x_i, y_i$ ) tiene coeficiente de correlación  $r = -0,9$  siendo las medias marginales 1 y 2, respectivamente. Se sabe que una de las cuatro ecuaciones siguientes corresponde a la recta de regresión de Y sobre X. Selecciona razonadamente dicha recta:

a)  $y = -x + 2$       b)  $y = x + 1$       c)  $3x - y = 1$       d)  $2x + y = 4$

**17.** Sea  $y = 3x - 10$  la recta de regresión de Y sobre X. Sabiendo que  $\bar{x} = 4$ ,  $\sigma_{xy} = 3$  y  $\sigma_y^2 = 16$ , encuentra la media de Y, la varianza de X, el coeficiente de correlación y la recta de X sobre Y.

**18.** Halla las rectas de regresión de esta distribución bidimensional:

X \ Y	8	9	10
3	4	13	3
5	6	7	7

**19.** - Las calificaciones de 40 alumnos en matemáticas II (X) y en métodos estadísticos (Y) han sido las siguientes:

Nota de matemáticas II (X)	3	4	5	6	6	7	7	8	9	10
Nota de métodos estadísticos (Y)	4	5	7	8	6	8	9	10	9	9
Nº alumnos	4	4	2	12	4	5	4	2	1	2

a) Obtener la ecuación de la recta de regresión de calificaciones de estadística respecto de las calificaciones de psicología.

b) ¿Cuál será la nota esperada en estadística para un alumno que obtuvo un 4,5 en psicología?

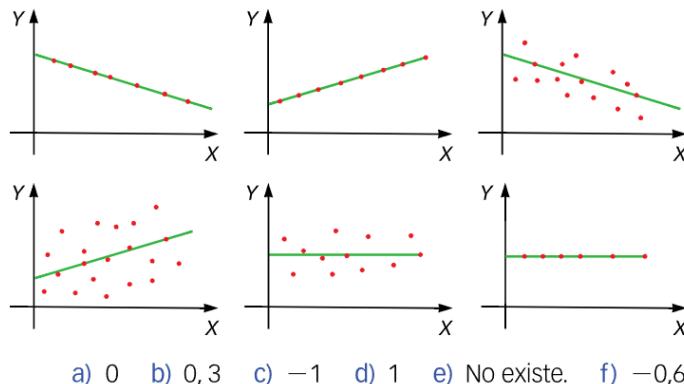
## 6.1. EJERCICIOS REPASO PARA EXAMEN

20. La siguiente tabla muestra los ingresos familiares mensuales de una familia en cientos de euros, X, y los metros cuadrados de la vivienda familiar, Y.

$Y \backslash X$	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)
[0, 50)	20	18	2	1	0
[50, 100)	25	40	30	2	1
[100, 200)	5	10	15	25	3
[200, 250)	0	5	15	20	8
[250, 300)	0	1	2	7	10

- Calcula la distribución conjunta de frecuencias relativas en porcentaje.
- Representa estos datos en un diagrama de barras.
- Halla la distribución de Y condicionado a las familias de ingresos entre 1000€ y 1500€ mensuales. Halla la moda.
- Halla la distribución de X condicionada a las familias que viven en pisos de 250 m<sup>2</sup> o más. Halla la media de ingresos mensuales de estas familias.
- Halla las distribuciones marginales de X e Y. Halla la varianza de ambas variables.

21. Relaciona cada diagrama de dispersión con su coeficiente de correlación:



22. Un modelo muy conocido en las teorías de control de especies es la evolución de las poblaciones de zorros y conejos en función de su interacción. La población de conejos de una zona suele tener oscilaciones relacionadas con la cantidad de zorros que hay en esa misma área. En una zona, en los últimos años, se han realizado ocho censos de animales.

Número de zorros	20	32	16	18	25	30	14	15
Número de conejos	320	500	260	300	400	470	210	240

Si la correlación es fuerte:

- Determina las dos rectas de regresión.
- Estima los conejos que habría si hubiera 10 zorros.
- ¿Cuántos zorros habría, aproximadamente, si hubiéramos contado 350 conejos?
- ¿Cuál de las dos estimaciones es más fiable?

**23.** Se ha medido el peso, X, y la estatura, Y, de los estudiantes de una clase. Su peso medio ha sido de 56 kg, con una desviación típica de 2,5 kg. La ecuación de la recta de regresión que relaciona la estatura y el peso es  $y = 1,8x + 62$ .

- a) ¿Qué estatura puede estimarse en un estudiante que pesa 64 kg?
- b) Y si un estudiante pesara 44 kg, ¿cuál sería su altura?
- c) ¿Cuál es la estatura media de los estudiantes de esa clase?
- d) La pendiente de esa recta es positiva, ¿qué significa esto?

**24.** Encuentra el coeficiente de correlación de la variable bidimensional cuyas rectas de regresión son:

- Recta de regresión de Y sobre X:  $2x - y - 1 = 0$
- Recta de regresión de X sobre Y:  $9x - 4y - 9 = 0$

- a) Halla la media aritmética de cada una de las variables.
- b) ¿Podrías calcular la desviación típica de Y sabiendo que la de la variable X es  $\sqrt{2}$  ?