

Relación de problemas 2

Estadística Descriptiva Bidimensional

1. En una encuesta realizada a 480 familias se han obtenido los siguientes datos en euros sobre los ingresos mensuales (X) y depósitos de Bancos y Cajas de Ahorros (Y)

X / Y	0-200	200-500	500-2000	2000-10000
500-1000	40	12	8	0
1000-1500	16	48	12	4
1500-2500	8	80	92	20
2500-5000	4	40	72	24

- a)* Obtener las distribuciones marginales.
b) Calcular la media y la varianza de las distribuciones marginales.
c) ¿Qué distribución es más homogénea?
d) ¿Qué porcentaje de familias poseen unos ingresos mensuales comprendidos entre 1500 y 2500?
e) Obtener la distribución de los ingresos mensuales para familias con depósitos superiores a 500.
2. A partir de los siguientes datos:

x	1	2	2	3	5	4	1	3	3	4	1	2	5	4	4	4	4	5	5	1	6	5	4	6
y	2	3	1	4	3	2	6	4	1	6	6	5	1	2	5	1	1	2	6	6	2	1	2	5

- a)* Representar los datos en una tabla de doble entrada.
b) Calcular las medias y varianzas marginales.
c) Calcular la covarianza.
3. Dada la siguiente distribución de frecuencias:

X / Y	1	2	3	4	5
100	2	4	6	10	8
200	1	2	3	5	4
300	3	6	9	15	12
400	4	8	12	20	16

Calcular la covarianza, ¿son estadísticamente independientes las variables X e Y ?

4. En una encuesta realizada a 100 familias en la cual se estudian los ingresos (X) y los gastos (Y), se obtuvieron los siguientes datos representados en una tabla de doble entrada:

X / Y	600-1000	1000-1500	1500-2000
1000-1300	11	4	0
1300-1800	8	16	5
1800-2500	10	22	24

- a) En las familias que no superen las 1800 euros de ingresos, calcular qué cantidad se gasta con más frecuencia.
- b) Una entidad bancaria pretende que el 40 % central de las nóminas sea domiciliado. Calcular el valor máximo y el mínimo de este intervalo.
- c) Si una empresa quiere realizar una campaña con aquellas familias que ingresen más de 1600 euros, ¿qué porcentaje de familias serán objetivo de esta campaña?
5. Dada la siguiente tabla, donde X representa el peso (en kg) e Y la altura (en cm) de un grupo de niños,

(X, Y)	[60,90]	(90, 110]	(110,140]
[10,15]	6	3	1
(15,20]	5	10	2
(20,25]	4	1	7
(25,30]	2	2	4

se pide:

- a) Obtener la altura media de los niños que pesan más de 20 kg.
- b) ¿Cuál es el peso más frecuente?
- c) Entre los niños que miden menos de 110 cm, calcular el peso mínimo del 30 % de los niños con más peso.
- d) Entre los niños que pesan entre 15 y 25 kg, calcular el porcentaje que presentan una altura inferior a 117 cm.
6. En el proceso de manufactura de una prenda de vestir se han controlado dos características: el número de días empleado en la fabricación (Y) y el número de errores en el acabado (X), obteniéndose la siguiente distribución conjunta sobre una muestra de 120 unidades:

(X, Y)	3	4	5	6
0	2	5	10	12
1	6	10	28	8
2	15	12	6	6

- a) Independientemente del tiempo empleado, ¿cuál es el número medio de errores cometidos? ¿y el número más frecuente?
- b) ¿Cuánto tiempo se dedica por término medio a la manufactura?
- c) ¿Cuál es el tiempo mínimo que emplea el 30 % de las unidades que más minutos tardan en el proceso de manufactura?

7. En la tabla adjunta se muestran las calificaciones X e Y obtenidas por 200 alumnos en dos asignaturas:

(X, Y)	$[0, 2, 5)$	$[2, 5, 5)$	$[5, 7)$	$[7, 9)$	$[9, 10]$
$[0, 2, 5)$	0	3	15	2	0
$[2, 5, 5)$	4	32	61	33	5
$[5, 7)$	1	5	19	5	0
$[7, 9)$	0	0	12	2	1
$[9, 10]$	0	0	0	0	0

Se pide:

- Representar el histograma para la distribución de las calificaciones de la asignatura correspondientes a Y de los alumnos que han suspendido la asignatura X .
 - ¿Qué asignatura tiene las calificaciones más homogéneas?
 - ¿Qué porcentaje de alumnos aptos en la asignatura Y obtienen una calificación apta en la asignatura X ?
 - ¿Cuál es la nota más frecuente en la asignatura X para los alumnos aptos en la asignatura Y ?
 - ¿Qué nota mínima obtiene el 10% de los alumnos mejores calificados en la asignatura X ?
8. En un estudio sobre la resistencia de un cierto tipo de componente se selecciona una muestra de componentes construidas con diferentes concentraciones de un metal y se anota el número de pruebas de resistencia necesarias hasta que se rompen, obteniéndose los siguientes resultados:

(X, Y)	25-55	55-65	65-95
5	4	2	2
10	1	4	0
15	5	2	0

donde X representa la concentración del metal e Y el número de pruebas antes de la ruptura.

- ¿Qué distribución es más homogénea con respecto a su valor medio?
 - ¿Cuál es el número medio de pruebas necesarias hasta romper las componentes? ¿Y el más frecuente?
 - ¿Qué número mínimo de pruebas requiere el 50% de las componentes más resistentes antes de romperse?
 - De entre las componentes con una concentración superior a 5, calcular el porcentaje de éstas que necesitan 50 pruebas o menos para romperse.
9. El gerente de una empresa afirma que los empleados jóvenes trabajan más horas extras que los adultos. Según la base de datos, se tiene la siguiente información:

(X, Y)	18-35	35-55	55-70
0-5	20	3	12
5-10	8	4	3
10-15	6	2	7
15-20	14	8	8

donde X representa el número de horas extra en un determinado mes e Y representa los grupos de edad.

- a) ¿Cuál es el número total de horas extraordinarias que trabajan los jóvenes de 18 a 35 años en un mes?
- b) ¿Cuál es el número total de horas extraordinarias que trabajan los adultos de 55 a 70 años en un mes?
- c) ¿Cuál es el número medio de horas extraordinarias al mes que trabajan los jóvenes de 18 a 35 años?
- d) ¿Cuál es el número medio de horas extraordinarias al mes que trabajan los adultos de 55 a 70 años?
- e) ¿Qué grupo de edad trabaja más número de horas extraordinarias en esta empresa?
- f) ¿Cuál es el número medio de horas extra que trabaja cualquier trabajador en esta empresa?
- g) ¿Es cierta la afirmación que hace el gerente de la empresa?
10. Se estudia el número de años (X) de experiencia en cierta empresa y el salario mensual (Y) que perciben los trabajadores. Los resultados obtenidos son los siguientes:

$X Y$	5-10	10-15	15-25	25-35
600-1200	21	6	5	3
1200-1500	17	9	8	5
1500-2000	6	12	16	9
2000-3000	2	8	12	18

- a) Calcular el sueldo medio de los trabajadores que llevan menos de 15 años en la empresa.
- b) Calcular el sueldo más frecuente de los trabajadores que llevan más de 15 años en esa empresa.
- c) Calcular el sueldo máximo que perciben el 50 % de los trabajadores que menos cobran de esta empresa.
- d) Calcular el sueldo mínimo que cobra el 20 % de los trabajadores que más cobra en la empresa.
- e) ¿Es más homogénea la distribución de los salarios de los que llevan más de 25 años en la empresa que la distribución de los salarios de los que llevan menos de 10 años?
- f) Calcular el salario medio de los trabajadores de la empresa.
11. Dada la siguiente distribución bidimensional:

$X Y$	1	2	3
-1	0	1	0
0	1	0	1
1	0	1	0

- a) Calcular las medias marginales.
- b) Calcular la covarianza.
- c) Calcular las medias de Y condicionadas a los distintos valores de X .
- d) Calcular las medias de X condicionadas a los distintos valores de Y .
- e) ¿Son X e Y estadísticamente independientes?
- f) ¿Hay alguna contradicción entre los tres apartados anteriores?
12. Una inmobiliaria observa que durante el mes de agosto el número de apartamentos alquilados varía según el precio semanal de los mismos, con los siguientes valores:

Precio	Apartamentos
100-160	9350
160-250	5120
250-370	3640
370-490	1790
490-1000	800

- a) Si se supone una relación lineal entre el precio y los alquileres, ¿cuántos apartamentos se alquilarán a un precio de 480?
- b) ¿En qué medida el número de apartamentos alquilados depende linealmente del precio?

13. Las notas de física (X) y matemáticas (Y) de 10 alumnos fueron:

X	9	7	3	6	7	5	10	8	2	5
Y	8	5	4	2	9	6	10	9	1	5

Suponiendo una relación lineal entre las notas de física y de matemáticas:

- a) Para un alumno que tenga un 4 en física, ¿cuál será su nota en matemáticas?
- b) Para un alumno que tenga un 7 en matemáticas, ¿cuál será su nota en física?

14. A partir de los siguientes datos:

$$\bar{x} = 5; \quad \bar{y} = 8; \quad Cov(x, y) = 15; \quad Var(y) = 20; \quad r^2 = 0,9$$

obtener la varianza de X y las rectas de regresión de Y/X y X/Y .

15. Un empresario sostiene que existe una relación lineal entre las deudas y los recursos de su empresa. Para comprobarlo toma una muestra de 12 elementos, siendo Y las deudas y X los recursos (medidos ambos en millones de euros). Los datos que obtiene son:

$$\sum y_i = 1200; \quad \sum (y_i - \bar{y})^2 = 1800; \quad \sum x_i = 84; \quad \sum (x_i - \bar{x})^2 = 85; \quad r_{xy} = 0,82$$

Estimar cuál sería la deuda de una empresa con unos recursos que asciendan a 8.2 millones.

16. Se eligen 50 matrimonios y se les pregunta la edad al casarse:

(X, Y)	15-20	20-25	25-30	30-35	35-40
15-18	3	2	3	0	0
18-21	0	4	2	2	0
21-24	0	7	10	6	1
24-27	0	0	2	5	3

donde X es la edad del hombre e Y es la edad de la mujer. Calcular:

- a) Rectas de regresión de Y/X y X/Y .
- b) Si una mujer se casó con 22 años, ¿qué edad tendría su pareja cuando se casaron? Y si la mujer se casó a los 24 años, ¿cuál sería la edad de su pareja cuando se casaron?

17. Las calificaciones obtenidas por 10 alumnos en el examen parcial y final de estadística son:

X - Parcial	5	8	6	9	2	3	1	2	4	7
Y - Final	6	6	5	8	1	4	2	1	5	8

- a) Calcular el coeficiente de correlación lineal de Pearson.
 b) La recta de regresión de Y sobre X y la recta de regresión de X sobre Y .

18. Dada la siguiente distribución bidimensional:

X/Y	1	2	3	4
10	1	3	0	0
12	0	1	4	3
14	2	0	0	2
16	4	0	0	0

- a) ¿Son X e Y estadísticamente independientes?
 b) Calcular la recta de regresión de Y sobre X .

19. Dada la siguiente distribución bidimensional:

X/Y	10	15	20
1	1	0	3
2	0	2	1
3	2	1	0

Cuantificar el grado de relación lineal entre las variables.

20. Dada la siguiente distribución bidimensional:

X/Y	2	4	5	7
1	4	5	7	9
2	2	4	8	12
3	1	3	4	6
4	1	2	2	3

- a) Calcular la covarianza y el coeficiente de correlación entre X e Y .
 b) Calcular la recta de regresión de X sobre Y y la de Y sobre X .
 c) Representar gráficamente las dos rectas en el mismo eje. ¿Se cortan en algún punto?
 d) Para un valor de la variable $X = 5$, ¿cuánto valdrá Y según la recta de regresión?
 e) Para un valor de la variable $Y = 6$, ¿cuánto vale X según la recta de regresión?

21. Un psicólogo afirma que a medida que los niños crecen dan menos respuestas inadecuadas en cierta situación experimental. Los datos que ha tomado son:

Edad	2	3	4	4	5	5	6	7	7	9	9	10	11	11	12
Respuestas	11	12	10	13	11	9	10	7	12	8	7	3	6	5	5

Se pide:

- a) Determinar la validez de esta conclusión.

- b) Si un niño de 10 años y medio participa en el experimento, ¿cuál es el número de respuestas inadecuadas que se puede predecir para él?

22. La siguiente tabla muestra los valores de la edad (X) y la presión sanguínea (Y) de 10 mujeres:

X	56	42	72	36	63	47	55	49	38	42
Y	148	126	159	118	149	130	151	142	114	141

Dar una predicción lineal para la presión sanguínea de una mujer de 51 años. Dar una medida de la bondad de dicha predicción.

23. El precio Y en euros y la antigüedad X en años de siete motocicletas de cierto modelo se recogen en la siguiente tabla:

X	2	3	4	5	6	7	8
Y	415	360	312	270	234	204	180

- a) Calcular la recta de regresión de Y sobre X .
 b) Obtener el precio estimado de una motocicleta de ese modelo con 9 años de antigüedad. ¿Es buena esta predicción?

24. Dada la variable bidimensional discreta (X, Y) con la siguiente distribución de frecuencias relativas:

(X, Y)	0.5	1.5
1	0.2	0.1
2	0.1	0.3
3	0.1	0.2

Calcular el coeficiente de correlación de las variables X e Y .

25. Un hipermercado ha decidido ampliar el negocio. Decide estudiar de forma exhaustiva el número de cajas registradoras que va a instalar, para evitar grandes colas. Para ello, se obtienen los siguientes datos procedentes de otros establecimientos similares acerca del número de cajas registradoras (X) y del tiempo medio de espera (Y).

X	10	12	14	16	18	20
Y	59	51	42	32	26	22

Bajo el supuesto de que el tiempo medio de espera depende linealmente del número de cajas registradoras, se pretende saber

- a) ¿Cómo varía el tiempo medio de espera por cada unidad de caja adicional?
 b) Si se instalaran 17 cajas registradoras, ¿cuál sería el tiempo medio de espera? ¿Es fiable esta predicción?

26. Las rectas de regresión entre dos variables son:

$$8x + 2y = 1 \quad \text{y} \quad 16x + 9y - 1 = 0$$

calcular las medias de X e Y y el coeficiente de correlación lineal. ¿Cuál es la recta de Y sobre X y cuál la de X sobre Y ?

27. Dados los siguientes pares de rectas, determinar si se puede asumir que sean rectas de regresión y en su caso calcular r

a) $x + y = 125$
 $16x + 25y = 2675$

b) $2x - 4y = 50$
 $6x + 5y = 75$

c) $4x - 8y = 70$
 $2x - 4y = 50$

28. Dada una distribución bidimensional (X, Y) , donde $\bar{x} = 9$ y la recta de regresión de Y sobre X viene dada por la ecuación $y = -\frac{7}{18}x + \frac{13}{2}$, obtener los datos que faltan en la tabla:

x_i	y_j	n_{ij}
8	y_1	2
12	2	4
x_3	y_3	n_{33}
14	1	4
		16

Determinar la recta de regresión de X sobre Y , así como el coeficiente de determinación y comentar el resultado.

29. Sea la distribución unidimensional

x_i	3	5	8	9
n_i	5	1	2	1

que es una distribución marginal de la bidimensional (X, Y) , de la que se conoce $\sum_j n_j y_j^2 = 3240$, y la ecuación $y = 5x - 20$.

- a) Determinar la recta de regresión de $X|Y$.
 b) Estudiar la bondad del ajuste lineal.
 c) Obtener la varianza de la variable dependiente, así como la varianza residual.

30. Dadas las rectas

$$y = 1 + 2x \quad x = 10 + 5y$$

justificar si es posible que sean las rectas de regresión de $Y|X$ y de $X|Y$ respectivamente, de una misma distribución bidimensional.

31. De una distribución bidimensional (X, Y) se conoce que la media marginal de X es 1, la media marginal de Y es 2 y $r^2 = 0,9$. Si la recta de regresión de $Y|X$ es $y = a - 3x$, obtener el valor de a y el coeficiente de correlación lineal.
32. Un gestor financiero sostiene que existe una relación lineal directa entre las deudas totales de la empresa (Y) y sus recursos propios (X). Para comprobar la validez de aquella creencia considera una muestra piloto de 12 elementos, obteniendo los resultados siguientes:

$$\sum_{j=1}^{12} y_j = 1200; \quad \sum_{i=1}^{12} x_i = 84; \quad \sum_{j=1}^{12} (y_j - \bar{y})^2 = 10800; \quad \sum_{i=1}^{12} (x_i - \bar{x})^2 = 85; \quad r_{XY} = 0,82$$

Se pide:

- a) Obtener la recta de regresión de Y sobre X .
- b) Estimar cuál podría ser el exigible de una empresa de condiciones similares a las de la muestra con un capital de 82 millones de euros.

33. Se supone una distribución bidimensional cuyas rectas de regresión son:

$$x + 4y = 1; \quad x + 5y = 2$$

Calcular el coeficiente de correlación.

34. Se ha estimado la siguiente recta de regresión: $y = 5 + 3x$.

- a) Obtener la recta de X sobre Y sabiendo que $r = 1$.
- b) ¿Sería posible en este modelo que r fuera negativo?

35. En la estimación de un modelo de regresión lineal se obtuvo:

$$\bar{x} = 5; \quad \bar{y} = 8; \quad \sigma_{X,Y} = 15; \quad \sigma_Y^2 = 20; \quad r^2 = 0,9$$

- a) Calcular la varianza de X .
- b) Calcular las rectas de regresión.

36. A partir de un conjunto de datos sobre las variables X e Y se ha calculado la regresión de Y sobre X , obteniéndose los siguientes resultados:

$$y = 10 + 0,45x; \quad r^2 = 0,9; \quad \bar{x} = 20$$

Obtener la recta de regresión de X sobre Y .

37. De una variable estadística bidimensional (X, Y) con frecuencias unitarias se conoce que:

- los valores x_i de la variable (las modalidades) son: 1, 2, 3, 4, 5.
- la recta de regresión de Y sobre X es: $x - 3y + 8 = 0$.
- el coeficiente de correlación es $r = 0,8$.

Se pide:

- a) Hallar \bar{x} y σ_X .
- b) El coeficiente de regresión de Y sobre X .
- c) Calcular \bar{y} y σ_Y .
- d) La ecuación de la recta de regresión de X sobre Y .

38. Si el coeficiente de correlación lineal entre dos variables es -1, decidir si las siguientes afirmaciones son verdaderas o falsas y por qué:

- a) La covarianza es negativa.

- b)* La recta de regresión es decreciente.
- c)* El ajuste de la recta de regresión es perfecto.
- d)* Al calcular la pendiente de la recta de regresión, saldrá un valor negativo.
- e)* El coeficiente de determinación es 1.
- f)* Las variables no son independientes.