2. Estadística Descriptiva Bidimensional

1. Los gastos telefónicos en euros, en invierno (X) y verano (Y), de una serie de familias se indican a continuación.

$X \setminus Y$	$[100, 250]$	$(250, 400]$	$(400, 550]$
$[100, 250]$	75	42	12
$(250, 400]$	41	90	25
$(400, 550]$	11	31	46

Se pide:

a) Calcular las frecuencias, las medias y varianzas marginales de los consumos en invierno y verano.

b) Calcular la distribución condicionada del consumo en los meses de invierno cuando en el verano se consume una cantidad igual o menor de 400 euros, así como la media y varianza de esta distribución.

c) ¿Cuál es el porcentaje de personas que gastan más de 400 euros en invierno y en verano?

d) ¿Cuál es el porcentaje de personas que gastan menos de 300 euros en invierno, si en verano han gastado más de 400?

2. En una clínica se pregunta a cincuenta pacientes por el número de días que llevan ingresados (X) y el número de veces que sus familiares les han visitado (Y), obteniéndose la siguiente tabla:

$X \setminus Y$	0	1	2	3	4	5	6
1	1	2	1	0	0	0	0
2	0	2	3	5	0	0	0
3	6	10	8	1	1	3	1
4	0	0	1	3	0	2	0

Se pide:

a) ¿Cuál es el número medio de días que los pacientes están ingresados?

b) ¿Cuál es el número de veces que van la mayoría de los familiares a ver a los pacientes?

c) ¿Cuál es el número medio de visitas que reciben los pacientes que están ingresados más de 2 días?

d) Calcular la covarianza de estas variables.

3. Se ha medido la capacidad pulmonar (X) y el perímetro toráxico (Y) a 20 atletas, obteniéndose los siguientes resultados:

$X \setminus Y$	$[1, 1.10]$	$(1.10, 1.40]$	$(1.40, 1.70]$	$(1.70, 1.80]$
$[1.80, 1.85]$	3	1	0	1
$(1.85, 1.90]$	0	1	1	2
$(1.90, 2]$	1	0	0	3
$(2, 2.30]$	1	0	1	5

Se pide:

a) Obtener la distribución marginal de la variable X.

b) ¿Cuántos atletas tienen una capacidad pulmonar mayor de 2 y un perímetro toráxico mayor de 1.40?

c) ¿Cuál es el capacidad media pulmonar de los atletas que tienen un perímetro toráxico inferior a 1.40?

d) Calcular la covarianza de estas variables.

4. Dada la siguiente tabla, donde X representa el peso (en kgs) e Y la altura (en cms) de un grupo de niños:

$X \setminus Y$	$[90, 100]$	$(100, 120]$	$(120, 140]$
$[10, 15]$	6	3	1
$(15, 20]$	5	10	2
$(20, 25]$	4	1	7
$(25, 30]$	2	2	4

Se pide:

a) ¿Cuál es el peso más frecuente?

b) Entre los niños que miden menos de 120 cm, calcular el peso mínimo del 30 % de los niños con más peso.

c) Entre los niños que pesan entre 15 y 25 kg, calcular el porcentaje que presentan una altura inferior a 117 cm.

5. Se han obtenido los siguientes datos en una determinada ciudad donde se relaciona la tensión del vapor de agua (Y), con la temperatura (X) en °C.

$X \setminus Y$	$[0.5, 1.5]$	$(1.5, 2.5]$	$(2.5, 5.5]$
$[1, 15]$	4	2	0
$(15, 25]$	1	4	2
$(25, 30]$	0	3	5

Se pide:

a) ¿Cuál es la temperatura que con mayor frecuencia se ha alcanzado cuando la tensión de vapor de agua es superior a 1.5?

b) Decidir que variable es más homogénea, si la de la tensión de vapor de agua cuando la temperatura es inferior o igual a 15°C o la de la tensión de vapor de agua cuando la temperatura es superior a 15°C.

c) ¿Cuál es la mínima tensión de vapor de agua del 40 % de las tensiones más altas?

d) ¿Son independientes estas variables?

6. 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:

$X \setminus Y$	[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?

7. 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:

$Y \setminus X$	[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.

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:

$X \setminus Y$	[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.

a) ¿Qué distribución es más homogénea con respecto a su valor medio?

b) ¿Cuál es el número medio de pruebas necesarias hasta romper las componentes? ¿Y el más frecuente?

c) ¿Qué número mínimo de pruebas requieren el 50 % de las componentes más resistentes antes de romperse?

d) De entre las componentes con una concentración superior a 5, calcular el porcentaje de éstas que necesitan 50 pruebas o menos para romperse.

9. Según un estudio, la distribución de las ventas de cierto artículo según el número de minutos que aparece su publicidad en televisión, es la siguiente:

$X \setminus Y$	[12,35]	(35,60]	(60,120]
[0,2]	2	1	4
(2,5]	6	3	12
(5,10]	4	2	8

donde Y representa el número de minutos al día que aparece la publicidad en televisión y X representa el número de artículos vendidos en una ciudad.

a) Calcular las medias de las ventas para cada uno de los intervalos de tiempo de publicidad.

b) ¿Es independiente el número de ventas de la publicidad en este caso?

10. Dada la siguiente distribución bidimensional:

$X \setminus Y$	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) Si representáramos gráficamente las dos rectas en el mismo eje. ¿Se cortarían en algún punto las rectas?

d) Para un valor de la variable $X = 5$, ¿cuánto vale Y según la recta de regresión?

e) Para un valor de la variable $Y = 6$, ¿cuánto vale X según la recta de regresión?

11. 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 e interpretarlo.

   b) Calcular la recta de regresión de Y sobre X y la recta de regresión de X sobre Y.

12. En una distribución bidimensional de frecuencias unitarias se conoce:

   i) Los valores de Y: 3, 5, 10, 12.

   ii) La recta de regresión de X/Y : $2x + y - 7 = 0$.

   iii) El coeficiente de determinación: R^2 =0.81.

   Calcular la ecuación de la recta de regresión de Y/X.

13. Una variable estadística bidimensional tiene las siguientes rectas de regresión:

$$3x - 4y + 6 = 0 \qquad 3x - y - 3 = 0$$

   Calcular las medias de X e Y y el coeficiente de correlación lineal.

14. Se consideran 50 establecimientos de alimentación del mismo tipo, atendiendo a dos factores, tiempo que llevan funcionando (X, en años) y beneficio anual (Y, en millones de euros).

$Y \setminus X$	[0,5]	(5,10]	(10,15]	(15,20]
[0,1]	0	2	8	10
(1,3]	2	4	4	3
(3,4]	5	4	6	2

   a) Dar una predicción del beneficio anual para un establecimiento con 12 años de antigüedad.

   b) Calcular la varianza residual.

15. El valor promedio de las acciones en un grupo de inmobiliarias y en un grupo de bancos en los años 1980-1989 aparecen en la siguiente tabla

   X = Promedio del precio de las acciones inmobiliarias

   Y = Promedio del precio de las acciones de bancos

Año	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
X	3.5	4	4.2	4.3	4	5.3	5.3	4.9	4	5.5
Y	10.2	10.1	9.7	9.8	9.8	10	9.7	9.1	9.5	9.5

   a) Calcular el coeficiente de correlación lineal.

   b) En el supuesto de que se tengan acciones inmobiliarias, ¿qué le ocurrirá al valor de las acciones si se sabe que las acciones de los bancos aumentarán su valor? ¿En que medida es fiable la anterior estimación?

16. En 100 Ha. de tierra se han recogido datos sobre la producción de trigo (X) y cantidad de abono empleada (Y) por Ha.

$X \setminus Y$	[80,100]	(100,120]	(120,140]
[16,20]	20	0	0
(20,24]	10	30	20
(24,28]	0	5	15

a) ¿Existe alguna relación lineal entre las variables?.

b) Dar una predicción de la producción por Ha. para una cantidad empleada de 110 kg. de abono.

17. Un psicólogo afirma, en base a los datos obtenidos, que a medida que el niño crece menores son las respuestas inadecuadas que da en el transcurso de una situación experimental:

Edad	2	3	4	4	5	5	6	7	7	9	9	10	11	11	12
Nº resp	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) Alberto, de diez años y medio, participa en el experimento, ¿cuál es el número de respuestas inadecuadas que se puede predecir para él?

18. 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 un misma distribución bidimensional.

19. Sea la distribución unidimensional

x_i	3	5	8	9
n_i	5	1	2	1

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

a) Determinar la recta de regresión de X/Y.

b) Estudiar la bondad de ajuste lineal.

c) Obtener la varianza de la variable dependiente, así como la descomposición en varianza explicada y varianza residual.

20. 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 obtuvieron 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	12	18	20
Y	59	51	42	32	26	22

Bajo el supuesto de que el tiempo de espera medio 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 dicho dato?