Tema 2: Estadística descriptiva bidimensional

1. Se lanzan dos dados varias veces, obteniendo los resultados presentados en la siguiente tabla, donde X designa el resultado del primer dado e Y el resultado del segundo:

X	1	2	2	3	5	4	1	3	3	4	1	2	5	4	3	4	4	5	3	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) Construir la tabla de frecuencias.

   b) Calcular las puntuaciones medias de cada dado y ver cuales son más homogéneas.

   c) ¿Qué resultado del segundo dado es más frecuente cuando en el primero sale 3?

   d) Calcular la puntuación máxima del 50 % más bajo de puntuaciones obtenidas con el primer dado si con el segundo se ha obtenido un 2 o un 5.

2. Medidos los pesos, X (en Kg), y las alturas, Y (en cm), a un grupo de individuos, se han obtenido los siguientes resultados:

$X\backslash Y$	160	162	164	166	168	170
48	3	2	2	1	0	0
51	2	3	4	2	2	1
54	1	3	6	8	5	1
57	0	0	1	2	8	3
60	0	0	0	2	4	4

   a) Calcular el peso medio y la altura media y decir cuál es más representativo.

   b) ¿Cuál es el porcentaje de individuos con menos de 55 Kg y más de 165 cm?

   c) Entre los que miden más de 165 cm, ¿qué porcentaje pesa más de 52 Kg?

   d) ¿Cuál es la altura más frecuente de los individuos con peso entre 51 y 57 Kg?

   e) ¿Qué peso medio es más representativo, el de los individuos que miden 164 cm o el de los que miden 168 cm?

3. En una encuesta de familias sobre el número de individuos que la componen (X) y el número de personas activas en ellas (Y) se han obtenido los siguientes resultados:

$X\backslash Y$	1	2	3	4
1	7	0	0	0
2	10	2	0	0
3	11	5	1	0
4	10	6	6	0
5	8	6	4	2
6	1	2	3	1
7	1	0	0	1
8	0	0	1	1

   Calcular la recta de regresión de Y sobre X, y decir si es adecuado suponer esta relación lineal para explicar el comportamiento de Y a partir de X.

4. Se realiza un estudio sobre la tensión de vapor de agua (Y, en ml de Hg) a distintas temperaturas (X, en ºC). Efectuadas 21 medidas, los resultados son:

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

Explicar el comportamiento de la tensión de vapor en términos de la temperatura mediante una función lineal. ¿Es adecuado asumir este tipo de relación?

5. Estudiar la dependencia o independencia de las variables en las siguientes distribuciones. Dar, en cada caso, las curvas de regresión y la covarianza de las dos variables.

$X \backslash Y$	1	2	3	4	5
10	2	4	6	10	8
20	1	2	3	5	4
30	3	6	9	15	12
40	4	8	12	20	16

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

6. Dada la siguiente distribución bidimensional:

$X \backslash Y$	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 estadísticamente independientes X e Y?

b) Calcular y representar las curvas de regresión de X sobre Y y de Y sobre X.

c) Cuantificar el grado en que cada variable es explicada por la otra mediante la correspondiente curva de regresión.

d) ¿Están X e Y correladas linealmente? Dar las rectas de regresión.

7. Para cada una de las distribuciones:

Distribución A

$X \backslash Y$	10	15	20
1	0	2	0
2	1	0	0
3	0	0	3
4	0	1	0

Distribución B

$X \backslash Y$	10	15	20
1	0	2	0
2	1	0	0
3	0	0	3

Distribución C

$X \backslash Y$	10	15	20	25
1	0	3	0	1
2	0	0	1	0
3	2	0	0	0

a) ¿Depende funcionalmente X de Y o Y de X?

b) Calcular las curvas de regresión y comentar los resultados.

8. De una muestra de 24 puestos de venta en un mercado se ha obtenido la siguiente información sobre el número de balanzas (X) y el número de dependientes (Y):

$X\backslash Y$	1	2	3	4
1	1	2	0	0
2	1	2	3	1
3	0	1	2	6
4	0	0	2	3

   a) Determinar las rectas de regresión.

   b) ¿Es apropiado suponer que existe una relación lineal entre las variables?

   c) Predecir, a partir de los resultados, el número de balanzas que puede esperarse en un puesto con seis dependientes. ¿Es fiable esta predicción?

9. Se eligen 50 matrimonios al azar y se les pregunta la edad de ambos al contraer matrimonio. Los resultados se recogen en la siguiente tabla, en la que X denota la edad de la mujer e Y la del hombre:

$X\backslash Y$	$(18,21]$	$(21,25]$	$(25,30]$	$(30,35]$	$(35,40]$
$(18,21]$	3	2	3	0	0
$(21,24]$	0	4	2	2	0
$(24,27]$	0	7	10	6	1
$(27,30]$	0	0	2	5	3

   Estudiar la dependencia lineal entre ambas variables.

10. Calcular el coeficiente de correlación lineal de dos variables con rectas de regresión:

$$x + 4y = 1, \qquad x + 5y = 2.$$

11. En cierta distribución bidimensional, la recta de regresión de Y sobre X es $y = 5x - 20$, $\sum y_j^2 n_{.j} = 3240$, y la distribución marginal de X es:

x_i	3	5	8	9
$n_{i.}$	5	1	2	1

   Determinar la recta de regresión de X sobre Y, y la bondad de los ajustes lineales.

12. De las estadísticas de tiempos de vuelo y consumos de combustible de una compañía aérea, se han obtenido datos relativos a 24 trayectos distintos realizados por el avión DC-9. A partir de estos datos se han obtenido las siguientes medidas:

$$\sum_{i=1}^{24} x_i = 31,470; \quad \sum_{i=1}^{24} x_i^2 = 51,075; \quad \sum_{i=1}^{24} x_i^3 = 93,6; \quad \sum_{i=1}^{24} x_i^4 = 182,977$$

$$\sum_{i=1}^{24} y_i = 219{,}719; \quad \sum_{i=1}^{24} y_i^2 = 2396,504; \quad \sum_{i=1}^{24} x_i y_i = 349,486; \quad \sum_{i=1}^{24} x_i^2 y_i = 633,993,$$

donde la variable Y expresa el consumo de combustible, en miles de libras, correspondiente a un vuelo de duración X (el tiempo se expresa en horas, y se utilizan como unidades de orden inferior fracciones decimales de la hora).

a) Ajustar un modelo del tipo $Y = aX + b$. ¿Qué consumo total se estimaría para un programa de vuelos compuesto de 100 vuelos de media hora, 200 de una hora y 100 de dos horas? ¿Es fiable esta estimación?

b) Ajustar un modelo del tipo $Y = a + bX + cX^2$. ¿Qué consumo total se estimaría para el mismo programa de vuelos del apartado *a*)?

c) ¿Cuál de los dos modelos se ajusta mejor? Razonar la respuesta.

13. La curva de Engel, que expresa el gasto en un determinado bien en función de la renta, adopta en ocasiones la forma de una hipérbola equilátera. Ajustar dicha curva a los siguientes datos, en los que X denota la renta en miles de euros e Y el gasto en euros. Cuantificar la bondad del ajuste:

X	10	12.5	20	25
Y	50	90	160	180

14. Se dispone de la siguiente información referente al gasto en espectáculos (Y, en euros) y la renta disponible mensual (X, en cientos de euros) de 6 familias:

Y	30	50	70	80	120	140
X	9	10	12	15	22	32

Explicar el comportamiento de Y por X mediante:

a) Relación lineal.

b) Hipérbola equilátera.

c) Curva potencial.

d) Curva exponencial.

¿Qué ajuste es más adecuado?