

## Relación de problemas 3

# Variables Estadísticas Bidimensionales

1. Los resultados de la encuesta inicial a los alumnos del curso 2009-2010 reflejan la siguiente información:

$X / Y$	nada	poco	bastante	mucho
nada	0	0	1	2
poco	0	0	18	10
bastante	0	0	7	1
mucho	0	1	0	1

donde la variable llamada  $X$  es la respuesta a la cuestión “*En general, te parece complicada la estadística*” e  $Y$  representa las respuestas a la cuestión “*Confías en aprobar la asignatura*”.

- a) Calcular las frecuencias marginales, tanto las absolutas como las relativas.
  - b) ¿Cuántos alumnos ven la asignatura muy complicada?
  - c) Si dividimos a los alumnos en dos colectivos, los que consideran la asignatura nada o poco complicada frente a los que la consideran bastante o muy complicada, ¿qué porcentaje de alumnos forma cada uno de los grupos?
  - d) De entre los alumnos que ven la asignatura bastante o muy complicada, ¿cuántos de ellos no confían en aprobar (confían poco o nada)? ¿Qué porcentaje supone?
  - e) ¿Cuántos alumnos en total no confían nada en aprobar la asignatura?
  - f) De los alumnos que confían (bastante o mucho) en aprobar la asignatura, ¿qué porcentaje piensa que la asignatura no es complicada (nada o poco)?
2. Los resultados de la encuesta inicial a los alumnos del curso 2009-2010 reflejan la siguiente información:

$Y / Z$	nada	poco	bastante	mucho
nada	0	0	0	0
poco	0	1	0	0
bastante	0	12	13	1
mucho	0	1	11	2

donde la variable llamada  $Y$  representa las respuestas a la cuestión “*Confías en aprobar la asignatura*” y la llamada  $Z$  representa las respuestas a “*Esperas sacar buena nota en esta asignatura*”.

- a) Calcular las frecuencias marginales, tanto las absolutas como las relativas.
  - b) ¿Cuántos alumnos no tienen ninguna esperanza (nada) de aprobar la asignatura?
  - c) ¿Cuántos alumnos no tienen ninguna esperanza (nada) de sacar buena nota en estadística este año?
  - d) ¿Qué porcentaje de alumnos tiene bastante o mucha confianza en aprobar?
  - e) ¿Qué porcentaje de alumnos tiene bastante o mucha confianza en sacar buena nota?
  - f) De los alumnos que tienen bastante o mucha confianza en aprobar, ¿qué porcentaje tiene bastante o mucha esperanza en sacar buena nota?
  - g) De los alumnos que tienen poca confianza en aprobar, ¿qué porcentaje tiene bastante o mucha esperanza de sacar buena nota?
3. 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.
4. A partir de los siguientes datos:

x	1	2	2	3	5	4	1	3	3	4	1	2	5	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

- a) Representar los datos en una tabla de doble entrada.
  - b) Calcular las medias y varianzas marginales.
  - c) Calcular la covarianza.
5. Se eligen 50 matrimonios al azar y se les pregunta la edad de la mujer al casarse ( $X$ ) y la del hombre ( $Y$ ), obteniendo la siguiente tabla:

$X / Y$	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

- a) Edad media de los hombres al casarse.  
 b) Edad media de las mujeres al casarse.  
 c) Porcentaje de mujeres que se casaron entre los 18 y 24 años.  
 d) Edad más frecuente a la que los hombres y mujeres se casan.
6. 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  $Y$ ?

7. 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?
8. Los gastos en telefonía fija, móvil e internet de una serie de familias, medidos en euros, durante el invierno ( $X$ ) y durante el verano ( $Y$ ) se indican a continuación:

(X, Y)	[100, 2550]	(250, 400]	(400, 550]
[100,250]	75	42	12
(250,400]	41	90	25
(400,550]	11	31	46

Se pide

- a) Calcular las medias y varianzas marginales de los consumos en invierno y en verano.  
 b) Calcular la distribución condicionada del consumo en los meses de invierno cuando en verano se consume una cantidad igual o menor a 400 euros, así como la media y varianza de esta distribución.  
 c) ¿Cuál es el porcentaje de personas que gastan menos de 300 euros en invierno si en verano han gastado más de 400?
9. Se han medido la capacidad pulmonar ( $X$ ) y el perímetro torácico ( $Y$ ) de 20 atletas obteniéndose los siguientes resultados:

$(X, Y)$	[1, 1.10]	(1.10, 1.40]	(1.4, 1.70]	(1.70, 1.80]
[1.80,1.85]	3	1	0	1
(1.85,1.90]	0	1	1	2
(1.90,2.00]	1	0	0	3
(2.00,2.30]	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 torácico mayor de 1.40?
  - c) ¿Cuál es la capacidad media pulmonar de los atletas que tienen un perímetro torácico inferior a 1.40?
  - d) Calcular la covarianza entre estas variables.
10. 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.
11. 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 grados centígrados.

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

- 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 qué variable es más homogénea, 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?

12. 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?
13. 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:

- a) Representar el histograma para la distribución de las calificaciones de la asignatura correspondientes a  $Y$  de los alumnos que han suspendido la asignatura  $X$ .
  - b) ¿Qué asignatura tiene las calificaciones más homogéneas?
  - c) ¿Qué porcentaje de alumnos aptos en la asignatura  $Y$  obtienen una calificación apta en la asignatura  $X$ ?
  - d) ¿Cuál es la nota más frecuente en la asignatura  $X$  para los alumnos aptos en la asignatura  $Y$ ?
  - e) ¿Qué nota mínima obtiene el 10 % de los alumnos mejores calificados en la asignatura  $X$ ?
14. 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.

- 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 requiere 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.
15. 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?
16. 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.

17. 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?
18. 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:

$X Y$	12-35	35-60	60-120
1	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 casos de tiempo de publicidad.
  - b) ¿Es independiente el número de ventas de la publicidad en este caso?
19. Dada la siguiente tabla asociada a una variable bidimensional:

$X Y$	2	3	4
12	5	7	8
16	7	3	10
20	43	12	8

- a) Calcular la media de  $X$  condicionada a que  $Y$  vale 4 mediante un cambio de origen y escala.
- b) Calcular la varianza de  $X$  condicionada a que  $Y$  vale 4 mediante un cambio de origen y escala.
- c) Calcular la media marginal de  $Y$  mediante un cambio de origen y escala.
- d) Calcular la moda de  $X$  dado que  $Y$  toma el valor 2.
- e) Calcular la mediana de  $Y$  dado que  $X$  toma el valor 20.
- f) Calcular la covarianza.