EJERCICIOS RESUELTOS TEMA 6:
TABULACIÓN Y REPRESENTACIÓN GRÁFICA
DE DATOS BIVARIANTES

Ejercicio 1.

Se ha realizado una encuesta a 50 estudiantes de la Facultad de Economía y Empresa de Zaragoza con el fin de analizar la posible relación existente entre su "edad" (X) y el "número de créditos matriculados en el presente curso" (Y) y se han obtenido los siguientes resultados:

X Y	**42 a 54**	**54 a 66**	**66 a 78**
18 a 22	7	21	2
22 a 26	4	8	4
26 a 30	1	3	0

a) ¿Son las variables X e Y independientes?
b) Calcular el coeficiente de variación de Pearson del número de créditos en que están matriculados los alumnos cuya edad es superior a 22.
c) Determinar la edad más frecuente de los alumnos matriculados en menos de 66 créditos.

SOLUCIÓN:

a) Con la información dada en el enunciado del ejercicio, hay que incorporar las marcas de cada una de las clases de las dos variables X e Y; además, hay que añadir las frecuencias marginales que se necesitan para el cálculo de las medias de las dos variables.

		42 a 54	54 a 66	66 a 78	
X Y	Marcas de clase	48	60	72	$n_{i.}$
18 a 22	20	7	21	2	30
22 a 26	24	4	8	4	16
26 a 36	31	1	3	0	14
	$n_{.j}$	12	32	6	50

Calculamos los perfiles fila

X Y	42 a 54	54 a 66	66 a 78	Total
18 a 22	23,33%	70,00%	6,67%	100,00%
22 a 26	25,00%	50,00%	25,00%	100,00%
26 a 36	25,00%	75,00%	0,00%	100,00%
Marginal Y	24,00%	64,00%	12,00%	100,00%

Se observa que los perfiles fila son diferentes del marginal por lo que las variables son dependientes.

b) Se pide el coeficiente de variación para la distribución de la variable Y condicionada a que X > 22. En primer lugar, se obtiene la distribución de frecuencias de esta distribución condicionada.

	Y / X > 22	n_i
42 a 54	48	5
54 a 66	60	11
66 a 78	72	4
		20

A continuación, se procede al cálculo de la media y la desviación típica, con las que se calcula el coeficiente de variación.

$$\underline{Y}_{x>22} = \frac{1}{20}(48 \times 5 + 60 \times 11 + 72 \times 4) = 59{,}4$$

$$s^2_{y|x>22} = \frac{1}{20}(48^2 \times 5 + 60^2 \times 11 + 72^2 \times 4) - 59{,}4^2 = 64{,}44$$

$$s_{y|x>22} = \sqrt{64{,}44} = 8{,}03$$

$$CV_{y|x>22} = \frac{8{,}03}{59{,}4} = 0{,}1351$$

c) Se solicita el valor de la moda para la distribución de la variable X condicionada a que Y < 66. Para ello se obtiene en primer lugar la distribución de frecuencias condicionada, y a continuación se aplica la fórmula de la moda, sabiendo que el intervalo en el que se encuentra es el que tiene una mayor densidad.

	X / Y < 66	n_j	a_i	d_i
18 a 22	20	28	4	7
22 a 26	24	12	4	3
26 a 36	28	4	10	0,4
		15		

$$Mo = 18 + \frac{3}{0+3} \times 4 = 22$$

Ejercicio 2.

En la siguiente tabla se recoge para 40 empresas información sobre X: "Volumen de ventas en miles de euros" e Y: "Gasto en publicidad en miles de euros" en los últimos cinco años:

X Y	1-4	4-7	7-10
200-250	0	10	3
250-350	7	0	10
350-500	3	5	2

a) Obtener la desviación típica de la variable "Volumen de ventas", para aquellas empresas que han gastado más de 4000 euros en publicidad.

b) Calcular el menor valor de volumen de ventas que no es superado por el 25% de las empresas con menor gasto en dicho concepto.

SOLUCIÓN:

a) Se pide la desviación típica de la variable X condicionada a que Y sea mayor que 4. En primer lugar, se obtiene la distribución de frecuencias condicionada.

	X / Y > 4	n_i
200-250	225	13
250-350	300	10
350-500	475	7
		30

Se calcula la desviación típica de esta distribución unidimensional.

$$\underline{X}_{y>4} = \frac{1}{30}(225 \times 13 + 300 \times 10 + 475 \times 7) = 308{,}33$$

$$s^2_{x|y>4} = \frac{1}{30}(225^2 \times 13 + 300^2 \times 10 + 475^2 \times 7) - 308{,}33^2 = 9513{,}89$$

$$s_{x|y>4} = \sqrt{9513{,}89} = 97{,}54$$

b) Hay que calcular el primer cuartil de la distribución marginal de X. Se obtiene la distribución de frecuencias de la distribución marginal de X:

	X	n_i	N_i
200-250	225	13	13
250-350	300	17	30
350-500	475	10	40
		40	

Para ver en qué intervalo se encuentra se calculan las frecuencias absolutas acumuladas en primer lugar. A continuación, se averigua en qué posición se encuentra el primer cuartil y para ello se divide el número total de elementos de la distribución entre 4, $\frac{N}{4} = 10$, luego el primer cuartil se encuentra en el intervalo 200-250

$$C_1 = 200 + \frac{250 - 200}{13} \times 10 = 238{,}46$$

Ejercicio 3.

En la siguiente tabla se recoge para 100 empresas del sector TIC información sobre X: "Porcentaje de ventas por comercio electrónico sobre las ventas totales (%)" e Y: "Número de empleados".

X	**Y**	**Número de empresas**
0-10	10-20	30
10-40	20-40	55
40-100	40-50	15

a) Determinar el valor de la desviación típica de la variable "Porcentaje de ventas por comercio electrónico".
b) Calcular entre qué valores de la variable "número de empleados" se encuentra el 50% central de las empresas.

SOLUCIÓN:

a) Necesitamos calcular la desviación típica de la variable X, por lo que consideramos únicamente la distribución marginal de esta variable.

X	**Número de empresas**
5	30
25	55
70	15

$$\underline{X} = \frac{1}{100}(50 \times 30 + 25 \times 55 + 70 \times 15) = 25{,}75$$

$$s_x^2 = \frac{1}{100}(5^2 \times 30 + 25^2 \times 55 + 70^2 \times 15) - 25{,}75^2 = 423{,}19$$

$$s_x = \sqrt{423{,}19} = 20{,}57$$

b) El ejercicio pide calcular los cuartiles primero y tercero de la variable Y. Se obtiene su distribución marginal y se calculan el primer y tercer cuartil.

Y	**Número de empresas (n_i)**	**N_i**
10-20	30	30
20-40	55	85
40-50	15	100

$\frac{N}{4} = 25$, luego el primer cuartil se encuentra en el intervalo 10-20

$$C_1 = 10 + \frac{20 - 10}{30} \times 25 = 18{,}33$$

$\frac{3N}{4} = 75$, luego el tercer cuartil se encuentra en el intervalo 20-40

$$C_3 = 20 + \frac{50 - 40}{55} \times 75 = 33{,}64$$

Por lo tanto, el 50% central de las empresas tienen entre 18,33 y 33,64 empleados.

Ejercicio 4.

Durante un estudio sobre los hábitos alimenticios de los trabajadores en activo de una Comunidad Autónoma, dos de las variables investigadas fueron su nivel de estudios (X) y el número de veces por semana que comen fuera de casa (Y), obteniéndose las siguientes respuestas:

Y: Número de veces por semana que come fuera de casa / ***X: Nivel de estudios***	Menos de dos	Entre dos y cuatro	Más de cuatro
Primaria	200	20	5
Secundaria	1000	40	10
Universidad	700	20	5

Contesta justificadamente a las siguientes cuestiones:

a) ¿Son independientes las variables X (nivel de estudios) e Y (número de veces por semana que come fuera de casa)? ¿Qué interpretación tiene la independencia o no independencia estadística en el contexto de estas variables?

b) Obtener la distribución de frecuencias relativa del número de veces que comen fuera de casa los trabajadores cuyo nivel de estudios es Primaria. ¿Cuál es el número medio de veces que estos trabajadores comen fuera de casa?

c) Obtener y representar de forma conveniente la distribución de frecuencias relativa de la variable Nivel de estudios de un trabajador en activo que come fuera de casa dos o más veces por semana.
d) ¿Por qué en el apartado anterior no se puede calcular la media aritmética? ¿Qué medida de posición central se puede calcular? Calcúlala.

SOLUCIÓN:

a) Si las variables son independientes, se debe cumplir que $f_{ij} = f_{i.} f_{.j}$ $\forall$ i = 1,..,k y j = 1,..,h. Obtenemos las distribuciones marginales:

Y: Número de veces por semana que come fuera de casa / ***X: Nivel de estudios***	Menos de dos	Entre dos y cuatro	Más de cuatro	$n_{i.}$
Primaria	200	20	5	225
Secundaria	1000	40	10	1050
Universidad	700	20	5	725
$n_{.j}$	1900	80	20	2000

$f_{11} = \frac{200}{2000} = 0{,}1$ y $f_{1.}\, f_{.1} = \frac{225}{2000} \times \frac{1900}{2000} = 0{,}1125 \times 0{,}95 = 0{,}1069$, por lo que, al no ser iguales, las variables X e Y no son independientes. Esto indica que hay cierta asociación entre el nivel de estudios de un trabajador en activo y el número de veces por semana que come fuera de casa.

b) Se pide la distribución condicional de Y|X=Primaria, la cual viene dada por:

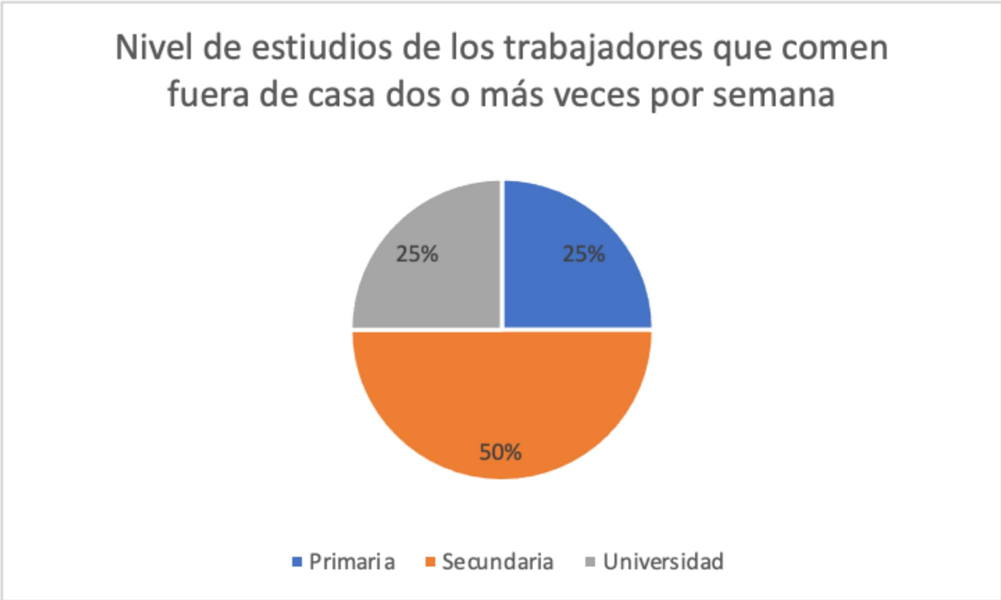
Y: Número de veces por semana que come fuera de casa	$f_{i.\mid X=\text{Primaria}}$
Menos de dos	$100\frac{200}{225} = 88.89\%$
Entre dos y cuatro	$100\frac{20}{225} = 8.89\%$
Más de cuatro	$100\frac{5}{225} = 2.22\%$

$$\underline{Y}_{x=Primaria} = 0 \times 0{,}8889 + 1 \times 0{,}0889 + 2 \times 0{,}0222 = 0{,}1333$$

c) Se pide la distribución de X condicionada a que la variable Y > 2.

X:Nivel de estudios/Y>2	$n_{i\mid Y>2}$	$f_{i\mid Y>2}$
Primaria	25	0,25
Secundaria	50	0,5
Universidad	25	0,25
Total	100	1

Al ser una variable cualitativa, una forma de representación adecuada es mediante un diagrama de sectores.



d) No se puede calcular la media aritmética de la variable X porque esta variable es de tipo nominal. Para este tipo de variables se puede calcular la Moda.

La Moda es el valor más frecuente, en este caso el nivel de Secundaria.