
FUNDAMENTOS DE ESTADÍSTICA

BOLETÍN DE PROBLEMAS 4

A. ROBERTO ESPEJO MOHEDANO.
ARTURO GALLEGO SEGADOR.

-
1. Los siguientes datos representan la altitud en metros sobre el nivel del mar de los Municipios de una provincia:

		n_i
I1	(0 - 100]	26
I2	(100 - 200]	20
I3	(200 - 300]	16
I4	(300 - 400]	10
I5	(400 - 500]	8

Se desea obtener:

- Principales medidas de posición y dispersión para la variable en estudio.
 - ¿Cuántos Municipios tienen una altitud inferior a 210 m?.
2. En un distrito universitario se han escogido 690 alumnos que han solicitado estudiar los Grados de Informática, Electricidad, Electrónica y Mecánica. Las calificaciones obtenidas en Selectividad son las siguientes:

		n_i
I1	(4 - 5]	150
I2	(5 - 6]	200
I3	(6 - 7]	160
I4	(7 - 8]	105
I5	(8 - 9]	65
I6	(9 - 10]	10

La planificación del Centro obliga a que en Informática entren el 12 % de los alumnos con más nota, en Electrónica el 30 % siguiente, en Mecánica el 25 % siguiente y en Electricidad el 20 % siguiente. Determinar:

- Principales medidas de posición y dispersión para las calificaciones obtenidas por los alumnos.
 - Puntuación mínima para entrar en cada Grado.
 - ¿Cuántos alumnos no entrarán en ningún Grado?. ¿Con que calificación se quedarán fuera?.
3. La siguiente tabla presenta los Ingresos (I) y Gastos (G) de un grupo de empresas en millones de unidades monetarias (u.m.) durante un periodo de tiempo determinado:

G : I	(60 - 70]	(70 - 80]	(80 - 90]	(90 - 100]
(60 - 70]	1	4	16	21
(70 - 80]	3	2	22	16
(80 - 90]	2	8	2	4
(90 - 100]	6	2	3	1

-
- ¿Cuál de las dos variables es más homogénea?.
 - Clasificar la variable Gastos por su simetría.
 - Hallar el Gasto medio de las empresas con Ingresos superiores a 80 millones de u.m.
 - ¿Qué proporción de empresas tienen Gastos superiores a la media?.
 - ¿Están relacionadas las dos variables?, en caso afirmativo ¿en qué grado?.
 - ¿Cuánto puede Gastar el citado grupo de empresas, si se prevén para el siguiente ejercicio unos Ingresos de 150.000.000 u.m.?.
 - ¿Qué % de la varianza de G es explicada por el modelo anteriormente calculado?.

Datos auxiliares:

$$\begin{aligned}\sum I &= 9625 & ; & \quad \sum G = 8455 \\ \sum I^2 &= 830425 & ; & \quad \sum G^2 = 643225 \\ \sum IxG &= 715275 & ; & \quad \sum G^3 = 49789375\end{aligned}$$

- En una caja de reclutas se han medido las alturas (en metros) de 130 jóvenes, obteniéndose los resultados siguientes:

		n_i
I1	(1,55 - 1,60]	18
I2	(1,60 - 1,70]	31
I3	(1,70- 1,80]	39
I4	(1,80 - 1,90]	25
I5	(1,90 -2,00]	17

Se desean obtener:

- Principales medidas de posición y dispersión para la variable en estudio.
 - Si se consideran Bajos aquellos jóvenes cuya altura es inferior al percentil 5, ¿cuál es la altura máxima que pueden alcanzar?.
 - Si se consideran Altos aquellos jóvenes cuya altura es superior al percentil 82, ¿cuál es la altura mínima que pueden alcanzar?.
 - ¿En qué percentil se encuentra un joven que mide 1,787 m?.
- Se han estudiado las variaciones en 100 días de cotización en dos empresas A y B del sector eléctrico controladas por un mismo grupo inversor, obteniéndose los siguientes datos:

	Empresa A	Empresa B
Media	110	2,5
Desviación típica	10	0,5
	$R_{AB} = 0,85$	

-
- a) ¿Puede afirmarse que el ajuste lineal entre A y B es bueno?
- b) ¿Qué cotización se puede predecir en el modelo lineal para la empresa B, si la empresa A cotiza a 125 €?
- c) ¿Puede afirmarse que cuanto más alta sea la cotización de A también lo será la de B?
6. Se ha observado que durante el mes de noviembre el Gasto en electricidad y el Ingreso total en 30 familias, expresados en cientos de €, ha sido:

Gasto	2	3	5	9	10	19
Ingresos	70	80	90	130	150	200
N. Familias	2	7	10	5	4	2

- a) Determinar el modelo de regresión lineal que explique el Gasto a partir de los Ingresos totales.
- b) ¿Cuál es el grado de dependencia entre ambas variables?
- c) En el supuesto de que el modelo obtenido sea explicativo, indicar el Gasto estimado en electricidad para una familia con unos Ingresos de 10000 €/mes
7. Entre las vs. as. X e Y existe una intensa correlación lineal. Se sabe que los modelos de regresión lineal entre ambas variables son:

$$Y/X \quad ; \quad \hat{Y} = -3/2X + 3$$

$$X/Y \quad ; \quad \hat{X} = -3/5Y + 2$$

Siendo $S_Y^2 = 4$. A partir de la información disponible, determinar:

- a) El centro de gravedad de la v.a. bivalente (X,Y).
- b) El coeficiente de correlación y el de determinación de cualquiera de los dos modelos anteriores.
8. La máxima autoridad monetaria de un país europeo, dispone que los datos reflejados en la siguiente tabla, ambos en millones de €, donde X = “Cantidad de dinero en circulación” e Y = “Renta Nacional”

Año	X	Y
1980	2	5
1981	2,5	5,5
1982	3,2	6
1983	3,6	7
1984	3,3	7,2
1985	4	7,7
1986	4,2	8,4
1987	4,6	9
1988	4,8	9,7
1989	5	10

a la vista de los datos anteriores, determinar:

- a) El coeficiente de determinación del modelo lineal simple que explique Y a partir de X.
 - b) ¿Cuál es la interpretación del término constante?, ¿cuál el de la pendiente de la recta obtenida?.
9. Los siguientes datos representan las puntuaciones (medidas de 50 a 100), obtenidas por 12 alumnos en las asignaturas de Estadística y Matemáticas que se imparten en el primer curso de Ingeniería:

Estudiante	Sexo	Estadística	Matemáticas
1	Hombre	65	85
2	Hombre	50	74
3	Mujer	55	76
4	Mujer	65	90
5	Hombre	55	85
6	Mujer	70	87
7	Mujer	65	94
8	Hombre	70	93
9	Mujer	55	81
10	Mujer	70	91
11	Hombre	50	76
12	Mujer	55	74

- a) ¿Qué Nota es la más homogénea?.
- b) ¿Están relacionadas ambas notas?.
- c) Determinar la puntuación en Matemáticas que se espera para un alumno que en Estadística ha obtenido 62 puntos. ¿Qué porcentaje de la variación de la Nota de Matemáticas se explica con el modelo calculado?.

Datos auxiliares:

$$\sum_{i=1}^{12} E_i = 725 \quad ; \quad \sum_{i=1}^{12} E_i^2 = 44475$$

$$\sum_{i=1}^{12} M_i = 1006 \quad ; \quad \sum_{i=1}^{12} M_i^2 = 84950$$

$$\sum_{i=1}^{12} E_i M_i = 61335$$

- d) Agrupando las puntuaciones de ambas asignaturas en dos intervalos de igual amplitud, obtener las principales medidas de posición y dispersión para ambas series de datos, interpretando los resultados.
- e) ¿Dependen las puntuaciones de Estadística (según la agrupación anterior) del Sexo del alumno?, en caso afirmativo, ¿en qué grado?. Justificar la respuesta.

-
10. La siguiente tabla reproduce las Exportaciones e Importaciones, (en miles de Euros), de una determinada Empresa durante 13 Años, (el valor de las Importaciones del Año 1988 se ha perdido):

Año	V. auxiliar T	Importación	Exportación
1988	1	***	1493
1989	2	2976	1890
1990	3	3475	2234
1991	4	4177	2847
1992	5	4630	3743
1993	6	5115	4109
1994	7	4955	3816
1995	8	6051	4210
1996	9	6989	4660
1997	10	8396	5134
1998	11	8915	5643
1999	12	8752	5826
2000	13	9012	5984

A la vista de los datos auxiliares, determinar:

- ¿Quién presenta una mayor homogeneidad, las Importaciones o las Exportaciones?.
- ¿Están relacionadas las Importaciones con las Exportaciones?. ¿Influye en esa relación el paso del Tiempo?.
- Obtener un modelo de regresión para tratar de explicar las Importaciones en función de las Exportaciones y del Año.
- Para el modelo anterior, determinar la bondad del mismo y estimar el valor de las Importaciones del año 1988 que se desconoce.

Datos auxiliares:

$$\begin{aligned}
 \sum_{i=2}^{13} T_i &= 90 \quad ; \quad \sum_{i=2}^{13} T_i^2 = 818 \\
 \sum_{i=2}^{13} I_i &= 73443 \quad ; \quad \sum_{i=2}^{13} I_i^2 = 503776091 \\
 \sum_{i=2}^{13} E_i &= 50096 \quad ; \quad \sum_{i=2}^{13} E_i^2 = 229515688 \\
 \sum_{i=2}^{13} T_i I_i &= 637124 \quad ; \quad \sum_{i=2}^{13} T_i E_i^2 = 428688; \quad \sum_{i=2}^{13} I_i E_i^2 = 338908433
 \end{aligned}$$

SOLUCIONES

1. a) Medidas de posición:

$$m_o = (0 - 100] \quad ; \quad m_e = 170 \quad ; \quad \bar{x} = 192,5$$

Medidas de dispersión:

$$S^2 = 17443,75 \quad ; \quad S = 132,074 \quad ; \quad \bar{S}^2 = 17664,5569 \quad ; \quad \bar{S} = 132,908 \quad ; \quad CV = 0,686$$

b) 48

2. a) Medidas de posición:

$$m_o = (5 - 6] \quad ; \quad m_e = 5,975 \quad ; \quad \bar{x} = 6,159$$

Medidas de dispersión:

$$S^2 = 1,708 \quad ; \quad S = 1,3069 \quad ; \quad \bar{S}^2 = 1,71047 \quad ; \quad \bar{S} = 1,30785 \quad ; \quad CV = 0,21219$$

b) Informática: 7,925 ; Electrónica: 6,314 ; Mecánica: 5,388 ; Electricidad: 4,598

c) En ningún grupo: 90 ; Con nota inferior a 4,598, el 13 %

3. a) Homogénea: ingresos ; b) Asimétrica a la derecha ; c) $\bar{G} = 74,823$; d) 44,45 % ;
e) $R_{GI} = -0,462$; $C=0,5503$; $V=0,3805$; f) 44,869 millones de u.m. g) 21,4 %

4. a) Medidas de posición:

$$m_o = (1,70 - 1,80] \quad ; \quad m_e = 1,741 \quad ; \quad \bar{x} = 1,7473$$

Medidas de dispersión:

$$S^2 = 0,0137 \quad ; \quad S = 0,11735 \quad ; \quad \bar{S}^2 = 0,013806 \quad ; \quad \bar{S} = 0,11749 \quad ; \quad CV = 0,06716$$

b) 1,568 c) 1,8744 d) 63,79

5. a) Bueno ($R^2 = 0,7225$) ; b) 3,1375 ; c) Si

6. a) $\hat{G} = -6,111 + 0,1173I$; b) $R_{GI} = 0,9834$; c) $\hat{G}_{(I=100)} = 5,619$

7. a) (2, 0) ; b) $R_{XY} = -0,9486$; $R^2 = 0,8999$

8. a) $R^2 = 0,9625$; b) Ordenada en el origen; modelo: $\hat{Y} = 1,1109 + 1,7263X$

9. a) Matemáticas ; b) Si, $R_{EM} = 0,865$; c) $\hat{M}_{(E=62)} = 85,144$, explicado el 74,8 %

d) $\bar{E} = 60$; $S_E^2 = 25$; $CV_E = 0,0833$

$\bar{M} = 84,83$; $S_E^2 = 24,871$; $CV_E = 0,0587$; $S_{EM} = 21,033$; $R_{EM} = 0,8435$

e) si dependen, $C=0,1666$ y $V=0,169$

10. a) Exportaciones (ligeramente) ; b) $R_{IE} = 0,9668$; $R_{IE,T} = 0,1938$. La dependencia $I \approx E$ se debe parcialmente al paso del tiempo (T)

c) $\hat{I} = 1162,02 + 0,3037E + 491,97T$; d) $R^2 = 0,962$; $\hat{I}_{(E=1493,T=1)} = 2107,41$

2. En un distrito universitario se han escogido 690 alumnos que han solicitado estudiar los Grados de Informática, Electricidad, Electrónica y Mecánica. Las calificaciones obtenidas en Selectividad son las siguientes:

		n_i	x_i	f_i	N_i	F_i
I1	(4 - 5]	150	4,5	5/23	150	5/23
I2	(5 - 6]	200	5,5	20/69	350	20/69
I3	(6 - 7]	160	6,5	16/69	510	13/23
I4	(7 - 8]	105	7,5	7/46	615	41/69
I5	(8 - 9]	65	8,5	11/138	680	68/69
I6	(9 - 10]	10	9,5	1/69	690	1

La planificación del Centro obliga a que en Informática entren el 12 % de los alumnos con más nota, en Electrónica el 30 % siguiente, en Mecánica el 25 % siguiente y en Electricidad el 20 % siguiente. Determinar:

- Principales medidas de posición y dispersión para las calificaciones obtenidas por los alumnos.
- Puntuación mínima para entrar en cada Grado.
- ¿Cuántos alumnos no entrarán en ningún Grado?. ¿Con que calificación se quedarán fuera?.

a) $m_0 = 12 = (5-6]$

$$\bar{x} = \frac{1}{n} \sum x_i n_i = \frac{425}{69} = 6,1594$$

$$s^2 = \left(\frac{1}{n} \sum x_i^2 \right) - \bar{x}^2 = 1,708$$

$$s = \sqrt{s^2} = 1,3069$$

$$s^2 = \frac{n \cdot s}{n-1} = 1,71042$$

$$\bar{s} = \sqrt{s^2} = 1,20785$$

$$CV = \frac{s}{\bar{x}} = 0,21219 \quad \text{el 21,21\% de los alumnos}$$

$$(5-6] \rightarrow \frac{6-5}{0,1072-0,2134} = \frac{m_0-5}{0,5-0,2134} \rightarrow m_0-5 = \frac{0,5-0,2134}{0,1072-0,2134} \rightarrow m_0 = 5,9751$$

$$(5-6] \rightarrow \frac{6-5}{0,5072-0,2134} = \frac{c_1-5}{0,8-0,2134} \rightarrow c_1 = 5,1125$$

$$(7-8] \rightarrow \frac{8-7}{0,8513-0,7391} = \frac{c_3-7}{0,75-0,7391} \rightarrow c_3 = 7,0716$$

b)

Info	Elect	Mec	Elect	I
12%	33%	25%	28%	100%
p_{12}	p_{33}	p_{25}	p_{28}	

$$\frac{8-7}{0,8513-0,7391} = \frac{p_{28}-7}{0,88-0,7391} \rightarrow p_{28} = 7,9258$$

$$p_{25} = 6,214$$

$$p_{33} = 5,39$$

$$p_{12} = 4,598$$

c)

$$1. \quad \frac{5-4}{156} = \frac{4,598-4}{N} \rightarrow N = 0,598 \cdot 150 = 89,7 \approx 90$$

$$2. \quad 690 \cdot 0,12 = 82,8 \approx 90$$

Se ha medido el peso de un determinado artículo, expresado en gramos, los resultados aparecen en la tabla siguiente.

Peso del artículo	n_i
150 - 180	42
180 - 220	46
220 - 230	54
230 - 240	62
240 - 250	86
250 - 260	68
260 - 290	65
290 - 360	37

aproximando con tres cifras decimales, se pide:

1. ¿Existe mucha variabilidad en la medida estudiada?
2. El mercado exige que el porcentaje de artículos fuera del intervalo $(\bar{x} \pm 2s)$ sea inferior al 10%, ¿cumplen los datos anteriores la exigencia del mercado?

Completemos la tabla de frecuencias

Peso del artículo	X_i^*	n_i	$f_i = n_i/N$	N_i	$F_i = N_i/N$
150 - 180	165	42	0.0913	42	0.0913
180 - 220	200	46	0.1	88	0.1913
220 - 230	225	54	0.1174	142	0.3087
230 - 240	235	62	0.1347	204	0.4434
240 - 250	245	86	0.1869	290	0.6303
250 - 260	255	68	0.1478	358	0.7781
260 - 290	275	65	0.1413	423	0.9194
290 - 360	325	37	0.0804	460	1
		$N = 460$	≈ 1		

$$1) \bar{x} = 241,6522$$

$$s^2 = 1491,8345$$

$$s = 38,638$$

$$cv = 0,1598 = 15,98\%$$

$$2) \bar{x} + 2s = 241,6522 + 2(38,638) = 318,9282$$

$$\bar{x} - 2s = 164,3762$$

$$\frac{180 - 150}{0,0913} = \frac{164,3762 - 150}{N} \rightarrow N = 0,0438$$

$$\frac{360 - 290}{1 - 0,9194} = \frac{318,9282 - 290}{N - 0,9194} \rightarrow N = 0,9527$$

$$0,9527 - 0,0438 = 0,9089$$

$$1 - 0,9089 = 0,0916 \text{ cumple}$$

9. Los siguientes datos representan las puntuaciones (medidas de 50 a 100), obtenidas por 12 alumnos en las asignaturas de Estadística y Matemáticas que se imparten en el primer curso de Ingeniería:

Estudiante	Sexo	Estadística	Matemáticas
1	Hombre	65	85
2	Hombre	50	74
3	Mujer	55	76
4	Mujer	65	90
5	Hombre	55	85
6	Mujer	70	87
7	Mujer	65	94
8	Hombre	70	93
9	Mujer	55	81
10	Mujer	70	91
11	Hombre	50	76
12	Mujer	55	74

- a) ¿Qué Nota es la más homogénea?
b) ¿Están relacionadas ambas notas?
c) Determinar la puntuación en Matemáticas que se espera para un alumno que en Estadística ha obtenido 62 puntos. ¿Qué porcentaje de la variación de la Nota de Matemáticas se explica con el modelo calculado?
Datos auxiliares:

$$\sum_{i=1}^{12} E_i = 725 \quad ; \quad \sum_{i=1}^{12} E_i^2 = 44475$$

$$\sum_{i=1}^{12} M_i = 1006 \quad ; \quad \sum_{i=1}^{12} M_i^2 = 84950$$

$$\sum_{i=1}^{12} E_i M_i = 61335$$

- d) Agrupando las puntuaciones de ambas asignaturas en dos intervalos de igual amplitud, obtener las principales medidas de posición y dispersión para ambas series de datos, interpretando los resultados.
e) ¿Dependen las puntuaciones de Estadística (según la agrupación anterior) del Sexo del alumno?, en caso afirmativo, ¿en qué grado?. Justificar la respuesta.

$$\sum_{i=1}^{12} h_i = 1006$$

$$\sum_{i=1}^{12} h_i^2 = 84950$$

$$\bar{E} = \frac{1}{12} \sum_{i=1}^{12} E_i = \frac{1}{12} \cdot 725 = 60.42$$

$$S_E^2 = \frac{1}{12} \sum_{i=1}^{12} E_i^2 - \bar{E}^2 = \frac{1}{12} \cdot 44475 - (60.42)^2 = 55.6196$$

$$S_E = \sqrt{S_E^2} = 7.4615$$

$$CV_E = \frac{S_E}{\bar{E}} = \frac{7.4615}{60.42} = 0.1235 \quad \text{A. 12.35\%}$$

$$\sum_{i=1}^{12} E_i = 805$$

$$\sum_{i=1}^{12} E_i^2 = 50875$$

$$\sum_{i=1}^{12} h_i = 1051$$

$$\sum_{i=1}^{12} h_i^2 = 92145$$

$$CV_h = \frac{8.8877}{61.9171} = 0.1435$$

$$\bar{E} = 61.9171$$

$$S_E^2 = 78.9912$$

$$S_E = 8.8877$$

$$CV_h = \frac{6.8133}{83.9191} = 0.0815$$

$$h = 82.9231$$

$$S_h^2 = 43.2545$$

$$S_h = 6.5773$$

Sexo \ Est	H	M	$n_{i.}$
[50, 60]	3 2,5	3 3,5	6
(60, 70]	2 2,5	4 3,5	6
$n_{.j}$	5	7	$n = 12$

$$f_{i.} \text{ esperada} = \frac{n_{i.} n_{.j}}{n}$$

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(e_{ij} - f_{ij})^2}{f_{ij}} \rightarrow \frac{(3-2,5)^2}{2,5} + \frac{(3-3,5)^2}{3,5} + \dots + \frac{(4-3,5)^2}{3,5} = 0,342$$

$$\chi^2 \in [0, n+1] = [0, 12+1] = [0, 13]$$

\downarrow
 $\text{Est}(2) - 1 = 1$
 $\text{Sexo}(2) - 1 = 1$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{0,342}{0,342 + 12}} = 0,169 \quad C \in [0, 1]$$

$$V = \sqrt{\frac{\chi^2}{n+1}} = \sqrt{\frac{0,342}{12+1}} = 0,169 \quad V \in [0, 1]$$

$t=1$ (de novo pesquisa)