

PROBLEMAS DE ESTADÍSTICA DESCRIPTIVA CON SOLUCIONES

1.- La siguiente tabla recoge el salario bruto anual (en miles de euros) percibido por los trabajadores y trabajadoras de una empresa A:

Hombres	20	16	18	24	30	28	36	24	42	32	26	23	34	28
Mujeres	22	34	26	17	33	25	18	26	34	39	20	25	24	

Por otra parte, se conocen las siguientes características (también en miles de euros) correspondientes a los salarios de una empresa B de la competencia: media = 25; desviación típica = 6; primer cuartil = 22; mediana = 26; percentil 75 = 30.

a) ¿Dónde es más homogénea la distribución de salarios: en la empresa A ó en la B?

Media = 26,8148

Desviación típica = 6,87267

Coefficiente de variación A = 0,25

Coefficiente de variación B = 0,24

$CV(B) < CV(A)$, luego es más homogénea en B.

b) ¿Dónde es mayor la dispersión del 50% central de los datos: entre los hombres de la empresa A o en el conjunto de la empresa B?

Recorrido intercuartílico en A: 9

Recorrido intercuartílico en B: 8

Mayor en A.

En la empresa A, el 10% de los hombres y el 10% de las mujeres ocupan cargos directivos. ¿Dónde es más alto el salario de los directivos: entre los hombres o entre las mujeres?

Percentiles para salarios de hombres: 90,0% = 36,0

Percentiles para salarios de mujeres: 90,0% = 34,0

Para los hombre es más alto.

Incrementamos los salarios anuales de todos los trabajadores de la empresa B en 1000 euros. ¿Cuáles son, tras el incremento la media, la mediana y la desviación típica?

Media = 26

Mediana = 27

Desviación típica = 6

PROBLEMAS DE ESTADÍSTICA DESCRIPTIVA CON SOLUCIONES

2.- a) En cierta localidad existe la sospecha de que el número de nacimientos de bebés registrados durante la última década está estrechamente relacionado con el número de cigüeñas que anidan en la comarca. Estudiar, a partir de los siguientes datos, si esta sospecha está justificada.

Año	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Nacimientos	200	230	180	224	168	304	256	254	160	130
Cigüeñas	505	380	730	160	450	416	200	560	720	310

Coefficiente de Correlación lineal = -0,287633

Los modelos no lineales ofrecen aún una correlación más baja.

b) La siguiente tabla recoge la temperatura media (en grados centígrados) de los últimos 10 años en la comarca citada. Establecer una relación de tipo lineal entre el número de cigüeñas anidadas y la temperatura y estudiar la bondad del ajuste.

Año	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Temperaturas	27	24	30	19	27	25	20	28	30	23

Coefficiente de Correlación = 0,980046

Coefficiente de determinación = 96,04%

Ecuación del modelo: cigüeñas = - 811,5 + 49,58 · temperatura

Predecir el número de cigüeñas que anidarán el próximo año si se prevé una temperatura media de 25° ¿es fiable la predicción?

Para una temperatura de 25 grados se estiman 428 cigüeñas.

Como el coeficiente de correlación está muy próximo a 1, la estimación es muy fiable.

3.- La siguiente tabla recoge los salarios percibidos por los trabajadores de una empresa durante el año 2004:

Salario anual (miles de €)	Número de trabajadores
0-20	12
20-30	10
30-40	14
40-50	14

a) ¿Por encima de qué salario se encuentra el 40% de trabajadores mejor remunerados?

El 40% de los trabajadores se encuentra por encima de 35.714,5 euros (percentil 60).

b) De cara al año 2005, todos los trabajadores experimentaron un aumento de sueldo de 1.000 €. ¿Cuándo fue más homogénea la distribución de salarios: durante el año 2004 ó durante el año 2005?. Justifique la respuesta.

El estudio de la homogeneidad nos remite a la comparación de los coeficientes de variación de los años 2004 y 2005. Se verifica que

$$CV_{2004} = \frac{S_{2004}}{\bar{X}_{2004}} \quad \text{y} \quad CV_{2005} = \frac{S_{2005}}{\bar{X}_{2005}}$$

De cara al 2005, el salario de todos los trabajadores experimenta un aumento de 1.000 euros. Por tanto, usando las propiedades de la desviación típica y de la media, se tiene que

$$S_{2005} = S_{2004} \quad \text{y} \quad \bar{X}_{2005} = \bar{X}_{2004} + 1.000$$

por lo que se deduce que

$$CV_{2005} < CV_{2004}$$

es decir, en el año 2005 la distribución es más homogénea.

PROBLEMAS DE ESTADÍSTICA DESCRIPTIVA CON SOLUCIONES

Una forma alternativa de hacer este apartado consiste en hallar directamente los coeficientes de variación y compararlos: $CV_{2004} = 0,438$ y $CV_{2005} = 0,424$.

4.- Se ha estudiado la relación que existe entre las variables $X = \text{"nº de hermanos"}$ e $Y = \text{"nota (sobre 5 puntos) en un examen de prácticas"}$ en un conjunto de 40 alumnos.

Y X	1	2	3	4	5
1	2	5	4	3	2
2	2	4	2	3	2
3	1	3	1	2	4

a) ¿ Son variables independientes?

Las variables serán independientes si, para todo par i, j , el producto de las frecuencias relativas marginales es igual a la correspondiente frecuencia relativa conjunta. Es decir, si

$$fr_{ij} = fr_i \cdot fr_j, \text{ para todo } i, j.$$

Tomando, por ejemplo, las frecuencias correspondientes al par $(X = 2, Y = 1)$ se comprueba que $\frac{2}{40} \neq \frac{13}{40} \cdot \frac{5}{40}$. Por tanto, las variables no son independientes.

b) ¿Existe relación lineal entre las variables?

El coeficiente de correlación lineal vale 0,16, lo que nos indica que la relación lineal es directa y muy débil.

c) Calcule la varianza marginal de Y.

La varianza marginal de Y es 1,79. Se obtiene a partir de la distribución marginal:

Y	1	2	3	4	5
Frecuencias	5	12	7	8	8

5- Calcule e interprete el coeficiente de correlación lineal entre dos variables X e Y sabiendo que las correspondientes rectas de regresión son: $2x + 3y = 1$ y $x + y = 2$.

La recta de regresión de Y sobre X es: $r: 2x + 3y = 1$, mientras que la de X sobre Y es $s: x + y = 2$ (ya que, de lo contrario, llegaríamos a un absurdo).

Los respectivos coeficientes de regresión son: $b_{X/Y} = -\frac{2}{3}$ y $b_{Y/X} = -1$.

El cuadrado del coeficiente de correlación lineal es igual al producto de los coeficientes de regresión:

$$r^2 = b_{X/Y} \cdot b_{Y/X} = \frac{2}{3}.$$

Dado que $\text{signo}(r) = \text{signo}(b_{X/Y}) = \text{signo}(b_{Y/X})$, se deduce que $r = -\sqrt{\frac{2}{3}} \approx -0,82$; por tanto, existe fuerte dependencia lineal de tipo inverso.