

(En la realización de este examen trabajar con dos decimales y un nivel de significación de 0'05)

Una empresa multinacional con tres centros de producción en el País Vasco, desea realizar un estudio sobre los ingresos mensuales de sus trabajadores en estos tres centros de producción, para lo cual mediante muestreo aleatorio simple obtiene una muestra de los ingresos mensuales de 200 trabajadores obteniéndose los siguientes resultados :

<u>Interv. de ingresos en u.m.</u>	<u>nº de trabajadores</u>
58 - 60	9
61 - 63	20
64 - 66	45
67 - 69	55
70 - 72	43
73 - 75	17
76 - 78	11

a) Para un nivel de significación de 0'05. ¿ Se puede considerar que la distribución de los ingresos de la población de la cual procede la muestra es Normal ? (6 puntos)

b) ¿ Que distribución de probabilidad seguira la varianza muestral en esta muestra ? ¿ porque ? (2 puntos)

La empresa cree que los ingresos medios de sus trabajadores en el País Vasco es de 72 u.m.

c) En base a lo obtenido en la muestra. ¿ Que le podrias decir al respecto para un nivel de significación de 0'05 ? (6 puntos)

La empresa quiere comprobar que en su centro de Sestao los gastos de mantenimiento no tienen nada que ver con el volumen de producción , para lo cual eligen aleatoriamente cinco semanas trabajadas durante el ultimo año obteniendose los siguientes resultados :

<u>Gtos de manten. (u.m.)</u>	<u>Unidades producidas</u>
80	340
83	355
78	320
81	384
78	341

d) Basandote en esta muestra y para un nivel de significación del 5 % ¿ Que le dirias a la empresa respecto a lo que quiere comprobar ? (6 puntos)



Resolución examen de ESTADISTICA II (3-2-1997)

Nombre / Izena

Apellidos / Deiturak

Núm. / Zk.

Asignatura / Irakasgaia

Curso / Maila

Fecha / Data

a) v.a. X : "Ingreso mensual neto de los Trabajadores de una empresa"

Objetivo: v.a. X — Normal?

Tendremos que realizar una prueba de bondad de un ajuste, donde

- 1) $H_0: F(x) = F_0(x)$ La muestra procede de una población Normal
- 2) $H_a: F(x) \neq F_0(x)$ La muestra No procede de una población con Distribución Normal

Como el tamaño de la muestra es grande ($n=200$) y el número de clases $K \geq 5$ aplicaremos la prueba de la CHI-CUADRADA DE PEARSON, comprobando que se cumple el requisito de que todas las frecuencias esperadas son superiores a 5

3) $\alpha = 0.05$

4) $n = 200$

5) $EC = \sum_{i=1}^K \frac{(N_i - E_i)^2}{E_i}$

en nuestro caso $K=7$; N_i nos los dan; y tendremos que obtener E_i .

Para obtener E_i lo primero que nos fijamos es que:

* No tenemos μ ni σ . Los estimaremos en base a la muestra obtenida.

* Existe discontinuidad entre los intervalos. Hay que continuizarlos.

<u>interv.</u>	<u>x_i</u>	<u>N_i</u>
57'5 - 60'5	59	9
60'5 - 63'5	62	20
63'5 - 66'5	65	45
66'5 - 69'5	68	55
69'5 - 72'5	71	43
72'5 - 75'5	74	17
75'5 - 78'5	77	11

$$\hat{\mu}_x = 67'97$$

$$\hat{\sigma}_x = 4'33$$

Para obtener E_i calcularemos primero las probabilidades teóricas P_i teniendo en cuenta que lo que queremos probar es que la v.a. x sigue una distribución Normal.

Probab. Teóricas (P_i)

$$\begin{aligned} P_1 &= P[57'5 \leq x \leq 60'5] = P\left[\frac{57'5 - 67'97}{4'33} \leq \frac{x - 67'97}{4'33} \leq \frac{60'5 - 67'97}{4'33}\right] = \\ &= P[-2'42 \leq z \leq -1'73] = P[z \leq -1'73] - P[z \leq -2'42] = \text{Tablas} \\ &= 0'0418 - 0'0078 = \underline{0'0340} \end{aligned}$$

$$\begin{aligned} P_2 &= P[60'5 \leq x \leq 63'5] = P[-1'73 \leq z \leq -1'03] = P[z \leq -1'03] - P[z \leq -1'73] \\ &= 0'1515 - 0'0418 = \underline{0'1097} \end{aligned}$$

$$\begin{aligned} P_3 &= P[63'5 \leq x \leq 66'5] = P[-1'03 \leq z \leq -0'34] = P[z \leq -0'34] - P[z \leq -1'03] \\ &= 0'3669 - 0'1515 = \underline{0'2154} \end{aligned}$$

$$\begin{aligned} P_4 &= P[66'5 \leq x \leq 69'5] = P[-0'34 \leq z \leq 0'35] = P[z \leq 0'35] - P[z \leq -0'34] \\ &= 0'6368 - 0'3669 = \underline{0'2699} \end{aligned}$$

$$\begin{aligned} P_5 &= P[69'5 \leq x \leq 72'5] = P[0'35 \leq z \leq 1'05] = P[z \leq 1'05] - P[z \leq 0'35] = \\ &= 0'8531 - 0'6368 = \underline{0'2163} \end{aligned}$$

$$\begin{aligned} P_6 &= P[72'5 \leq x \leq 75'5] = P[1'05 \leq z \leq 1'74] = P[z \leq 1'74] - P[z \leq 1'05] \\ &= 0'9591 - 0'8531 = \underline{0'1060} \end{aligned}$$

$$\begin{aligned} P_7 &= P[75'5 \leq x \leq 78'5] = P[1'74 \leq z \leq 2'43] = P[z \leq 2'43] - P[z \leq 1'74] \\ &= 0'9925 - 0'9591 = \underline{0'0334} \end{aligned}$$

Modelo de Freq. Teóricas.

3

<u>Interv.</u>	<u>Pi</u>
57'5 - 60'5	0'0340
60'5 - 63'5	0'1097
63'5 - 66'5	0'2154
66'5 - 69'5	0'2699
69'5 - 72'5	0'2163
72'5 - 75'5	0'1060
75'5 - 78'5	0'0334
	<u>$\Sigma < 1$</u>

Cuando la suma de probabilidades teóricas es menor que 1 hay que replantearse los intervalos y obtenermos el modelo de frecuencias teóricas haciendo
 $E_i = n \cdot P_i$

Modelo de Frecuencias Teóricas

<u>Interv.</u>	<u>Pi</u>	<u>$E_i = n \cdot P_i$</u>	<u>Ni</u>	<u>$(N_i - E_i)^2$</u>	<u>EC</u>
hasta 60'5	0'0418	8'36	9	$(9 - 8'36)^2$	0'05
60'5 - 63'5	0'1097	21'94	20	$(20 - 21'94)^2$	0'17
63'5 - 66'5	0'2154	43'08	45	$(45 - 43'08)^2$	0'09
66'5 - 69'5	0'2699	53'78	55	$(55 - 53'78)^2$	0'02
69'5 - 72'5	0'2163	43'26	43	$(43 - 43'26)^2$	0'01
72'5 - 75'5	0'1060	21'20	17	$(17 - 21'20)^2$	0'83
Más que 75'5	0'0409	8'18	11	$(11 - 8'18)^2$	0'97
	<u>$\Sigma = 1$</u>	<u>200</u>			<u>2'1422</u>

$$P_1 = P[X \leq 60'5] = P[Z \leq -1'43] = \underline{0'0418}$$

$$P_2 = P[X \geq 75'5] = 1 - P[Z \leq 1'74] = 1 - 0'9591 = \underline{0'0409}$$

$$EC = \sum_{i=1}^k \frac{(N_i - E_i)^2}{E_i} = \underline{2'1422}$$

$$6) \text{ Valor crítico: } \chi^2_{k-2-1} = \chi^2_{7-2-1} = \chi^2_{4;0'05} = 9'488$$

$$7) \text{ Regla de decisión: Si } EC > \chi^2_{4;0'05} \text{ Rechazamos } H_0$$

$$8) \text{ Decisión: Como } 2'1422 < 9'488 \text{ No podemos rechazar } H_0$$

- 9) Para un nivel de significación de $\alpha = 0.05$, tenemos evidencia estadística suficiente para aceptar que la muestra proviene de una población cuya distribución de ingresos mensuales es Normal.
- 10) Podemos haber cometido un error de Tipo II al aceptar H_0 cuando esta es falsa, con una probabilidad de β .

b) ¿Distrib. de probabilidad de S^2 ?

Como la naturaleza de la población es Normal, y como
$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$
 se deduce que

$$\frac{(n-1) S^2}{\sigma^2} = \chi_{n-1}^2$$

en nuestro caso $n=200$:
$$\frac{199 S^2}{\sigma^2} = \chi_{199}^2$$

c) La empresa cree que $\mu = 72$ u.m. y para poder decidir algo respecto a esto, se realiza una prueba de hipótesis para $\alpha = 0.05$.

1.) $H_0: \mu = 72$ u.m. Se confirma lo que cree la empresa y los ingresos mensuales por término medio son de 72 u.m.

2.) $H_a: \mu \neq 72$ u.m. No se confirma lo que cree la empresa y por término medio los ingresos mensuales de los trabajadores no son de 72 u.m.

3.) $\alpha = 0.05$

4.) $n = 200$

5.) ¿FC? : v.a $\bar{X}_{200}$: "media de ingresos mensuales en una muestra de 200 trabajadores de la empresa"

v.a $\bar{X}_{200} \rightarrow \text{Normal}$ ($\mu_{\bar{x}} = \mu_x = ?$; $\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} = ?$)

Como no conocemos σ_x^2 , tendremos que estimarlo en base a la muestra calculando s^2 y por lo tanto estaremos ante una v.a. que sigue una distribución t de Student con $(n-1)$ grados de libertad. En este caso como $n=200$ es muy grande, consideramos que la v.a. sigue una distribución Normal.

$$EC = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = t_{n-1} = t_{199} \rightarrow \text{Normal}$$

$$EC = \frac{6797 - 72}{4'33/\sqrt{200}} = \frac{-4'13}{0'31} = -13$$

6.) Valores críticos:

$$Z_{0'025} = +1'96$$

$$Z_{0'975} = -1'96$$

7.) Regla de decisión: $\left. \begin{array}{l} \text{Si el } EC < Z_{0'975} \\ \text{Si el } EC > Z_{0'025} \end{array} \right\} \text{Rechazamos } H_0$

8.) Decisión: Como $-13 < -1'96$ Rechazamos H_0

9.) Tenemos evidencia estadística suficiente para poder afirmar que en el 95% de las ocasiones, la media de los ingresos mensuales de los trabajadores de esa empresa NO es de 72 u.m.

10.) Podemos equivocarnos al rechazar H_0 cuando en realidad esta es verdadera, es decir podemos cometer un error tipo I con una probabilidad de $\alpha = 0'05$

$$P[\text{error tipo I}] = 0'05$$

6.
e) No piden que comprobemos en base a la muestra obtenida si existe alguna relación entre los gts de mantenimiento y las unidades producidas.

Podemos pensar que los datos que tenemos nos indican una cierta relación que podríamos muy presentarse ajustándolos a una recta de regresión, donde X serían las unidades producidas que explicarían los gastos de mantenimiento (Y).

Fu el caso de que no exista relación, la pendiente poblacional tendría que ser $B=0$ por lo que haríamos una prueba de hipótesis para comprobarlo.

1) $H_0: B_0 = 0$ No existe relación entre los gastos de mantenimiento y las unidades producidas

2) $H_a: B \neq 0$ existe algún tipo de relación entre los gastos de mantenimiento y las unidades producidas

3) $\alpha = 0.05$

4) $n = 5$

5) $FC = \frac{b - B_0}{S_b}$

Tendremos que calcular el valor de b mediante la siguiente fórmula:

$$b = \frac{\sum xy - n \bar{x} \cdot \bar{y}}{\sum x^2 - n \bar{x}^2}$$

Para obtener S_b utilizaremos la siguiente fórmula:

$$S_b = \frac{S_{yx}}{\sqrt{\sum x^2 - n \bar{x}^2}} \quad \text{donde} \quad S_{yx} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

<u>X</u>	<u>Y</u>	<u>X.Y</u>	<u>X²</u>	<u>$\hat{Y}_i$</u>	<u>$(Y_i - \hat{Y}_i)^2$</u>
340	80	27200	115600	80'24	0'0576
355	83	29465	126025	82'04	0'9216
320	78	24960	102400	77'84	0'0256
334	81	27054	111556	79'52	2'1904
341	78	26598	116281	80'36	5'5696
<u>Totales</u>	<u>1690</u>	<u>135277</u>	<u>571862</u>		<u>8'7648</u>

media de x : $\bar{x} = \frac{1690}{5} = 338$

media de y : $\bar{y} = \frac{400}{5} = 80$

$$b = \frac{135277 - (5 \cdot 338 \cdot 80)}{571862 - 5 \cdot (338)^2} \approx 0'12$$

$$a = \bar{y} - b \bar{x} = 80 - 0'12 \cdot 338 \approx 39'44$$

luego : $\hat{y} = 39'44 + 0'12x$

$$\hat{y}_1 = 39'44 + 0'12 \cdot 340 = 80'24$$

$$\hat{y}_2 = 39'44 + 0'12 \cdot 355 = 82'04$$

$$\hat{y}_3 = 39'44 + 0'12 \cdot 320 = 77'84$$

$$\hat{y}_4 = 39'44 + 0'12 \cdot 334 = 79'52$$

$$\hat{y}_5 = 39'44 + 0'12 \cdot 341 = 80'36$$

$$s_{y|x} = \sqrt{\frac{8'7648}{3}} \approx 1'71$$

$$s_b = \frac{1'71}{\sqrt{642}} \approx 0'07$$

Entonces: $EC = \frac{0'12 - 0}{0'07} \approx 1'71$ que es una puntuación t de Student con $n-2$ g.l. En nuestro caso t_3 .

6) Valores críticos: $t_{3;0'025} = 3'1824$

$$t_{3;0'975} = -3'1824$$

7) Regla de decisión: $\left. \begin{array}{l} \text{si } EC > t_{3;0'025} \\ \text{si } EC < t_{3;0'975} \end{array} \right\} \text{RECHAZO } H_0$

8) Decisión: Como $-3'1824 < 1'71 < 3'1824$ Aceptamos H_0

9) Para un nivel de confianza del 95% tenemos evidencia estadística suficiente para afirmar que NO existe relación entre los gts de mantenimiento y las unidades producidas.

10) No podemos equivocarnos al haber aceptado H_0 siendo esta falsa y cometer un error de tipo II.

