

### Auxiliar 12 MA34B-03

#### Problema 1

Se quiere evaluar la eficacia de distintas dosis de un fármaco contra la hipertensión arterial, comparándola con la de una dieta sin sal. Para ello se seleccionan al azar 25 hipertensos y se distribuyen aleatoriamente en 5 grupos de tratamiento. Al primero de ellos no se le suministra nada de fármaco ni se le restringe la sal, al segundo se le da una dieta con un contenido pobre en sal, al tercero una dieta sin sal, al cuarto el fármaco a una dosis determinada y al quinto el mismo fármaco a otra dosis. Las presiones arteriales sistólicas de los 25 sujetos al finalizar los tratamientos aparecen en la Tabla 1. Se asuma normalidad de las observaciones.

Tabla 1

	Grupos				
	1	2	3	4	5
	180	172	163	158	147
	173	158	170	146	152
	175	167	158	160	143
	182	160	162	171	155
	181	175	170	155	160
Media	178,2	166,4	164,6	158	151,4
Varianza $s_i^2$	12,56	43,44	22,24	65,2	35,44

$$\frac{1}{n} \sum n_i s_i^2 = 35.776 \quad \frac{1}{n} \sum_i \sum_j x_{ji}^2 - \left( \frac{\sum_i \sum_j x_{ji}}{n} \right)^2 = 116.2016$$

2.1 Formule el test, puntualizando la hipótesis nula y alternativa que permite contestar a la siguiente pregunta “¿Difiere la presión sistólica de un paciente dependiendo del tratamiento al cual es sometido?”.

2.2 Complete la tabla 2 Anova asociada:

Tabla 2

Fuente de variación	Grados de libertad	Suma de los cuadrados	Cuadrados medios	F
Tratamiento (Intergrupo)				
Error (intragrupo)				
Total				

2.3 A su juicio ¿Difiere la presión sistólica de un paciente dependiendo del tratamiento al cual es sometido? Utilice los datos y/o las tablas 1 y 2 anteriores para concluir.

2.4 Mida el grado de relación funcional entre la hipertensión y el tratamiento realizado.

2.5 Realice el test  $H_o : \mu_4 = \mu_5$  contra  $H_1 : \mu_4 > \mu_5$  donde  $\mu_4$  y  $\mu_5$  corresponden a las medias de los grupos 4 y 5 respectivamente. Suponga que las varianzas de los dos grupos  $\sigma_4^2$  y  $\sigma_5^2$  no son iguales pero cumplen la relación  $\frac{\sigma_4^2}{\sigma_5^2} = 2$ . Concluya a un nivel de significación del 5%.

**Sol.:**

Se quiere evaluar la eficacia de distintas dosis de un fármaco contra la hipertensión arterial, comparándola con la de una dieta sin sal. Para ello se seleccionan al azar 25 hipertensos y se distribuyen aleatoriamente en 5 grupos de tratamiento. Al primero de ellos no se le suministra nada de fármaco ni se le restringe la sal, al segundo se le da una dieta con un contenido pobre en sal, al tercero una dieta sin sal, al cuarto el fármaco a una dosis determinada y al quinto el mismo fármaco a otra dosis. Las presiones arteriales sistólicas de los 25 sujetos al finalizar los tratamientos son:

Tabla 1 Grupos					
	1	2	3	4	5
	180	172	163	158	147
	173	158	170	146	152
	175	167	158	160	143
	182	160	162	171	155
	181	175	170	155	160
Media	178,2	166,4	164,6	158	151,4
Varianza $S_i^2$	12,56	43,44	22,24	65,2	35,44

$$\frac{1}{n} \sum n_i S_i^2 = 35.776 \quad \frac{1}{n} \sum_i \sum_j x_{ji}^2 - \left( \frac{\sum_i \sum_j x_{ji}}{n} \right)^2 = 116.2016$$

Asuma normalidad de las observaciones.

2.1 Formule el test, puntualizando la hipótesis nula y alternativa que permite contestar a la siguiente pregunta “¿Difiere la presión sistólica de un paciente dependiendo del tratamiento al cual es sometido?”.

$H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  contra  $H_1 : \text{lo contrario}$

2.2 Construya la tabla Anova asociada completando:

Fuente de variación	GL	SS	MS	F	P valor
tratamiento (Intergrupo)	4	2010,6	20,106	11,2399374	0
Error (intragrupo)	20	894,4	1,7888		
Total	24	2905			

2.3 A su juicio ¿Difiere la presión sistólica de un paciente dependiendo del tratamiento al cual es sometido?. Utilice los datos y/o tabla anteriores (1 y 2) para concluir.

La región de rechazo viene dada por  $R = \{F_{4,20} > C\}$  Luego  $C=2.866$  por tanto rechazamos la igualdad de medias. Difiere la presión sistólica dependiendo del tratamiento.

También podemos calcular el Pvalor cómo  $P(F_{4,20} > 11,23) = 0$

2.4 Mida el grado de relación funcional entre la hipertensión y el tratamiento realizado.

Se usa la razón de correlación:  $\eta^2 = \frac{SCE}{SCT} = 0.69$

2.5 Realice el test  $H_o : \mu_4 = \mu_5$  contra  $H_1 : \mu_4 > \mu_5$  dónde  $\mu_4$  y  $\mu_5$  corresponden a las medias de los grupos 4 y 5 respectivamente. Suponga que las varianzas de los dos grupos  $\sigma_4^2$  y  $\sigma_5^2$  no son iguales pero cumplen la relación  $\frac{\sigma_4^2}{\sigma_5^2} = 2$ .

Concluya a un nivel de significación del 5%.

$$\begin{aligned} H_o : \mu_4 &= \mu_5 \\ H_1 : \mu_4 &> \mu_5 \end{aligned} \Rightarrow \{\bar{X}_4 - \bar{X}_5 > C\}$$

Sabemos que  $\bar{X}_4 - \bar{X}_5 \rightarrow N\left(\mu_4 - \mu_5, \frac{\sigma_4^2}{n_4} + \frac{\sigma_5^2}{n_5}\right)$  y que

$$\frac{n_4 S_4^2}{\sigma_4^2} + \frac{n_5 S_5^2}{\sigma_5^2} \rightarrow \chi_{n_4+n_5-2=8}^2$$

Luego desarrollando obtenemos

$$t_8 = \frac{[(\bar{X}_4 - \bar{X}_5) - (\mu_4 - \mu_5)]\sqrt{n_4 + n_5 - 2}}{\sqrt{\left(\frac{2}{n_4} + \frac{1}{n_5}\right)\left(n_5 S_5^2 + \frac{n_4 S_4^2}{2}\right)}} \quad \text{Recordando que } n_1 = n_2 = 5$$

$$t_8 = \frac{[(\bar{X}_2 - \bar{X}_3) - (\mu_2 - \mu_3)]\sqrt{8}}{\sqrt{\left(\frac{3}{5}\right)\left(5S_5^2 + \frac{5S_4^2}{2}\right)}} \Rightarrow t_8 = \frac{[(\bar{X}_2 - \bar{X}_3) - (\mu_2 - \mu_3)]\sqrt{8}}{\sqrt{\left(3S_5^2 + \frac{3S_4^2}{2}\right)}}$$

Luego el error tipo I viene dado por

$$P\left(t_8 > C \frac{\sqrt{8}}{\sqrt{204.12}}\right) = P(t_8 > 0.1979C) = 0.05 \Rightarrow 0.1979C = 1.860 \Rightarrow C = 9.398$$

La diferencia en la muestra es 6.6 por lo que no rechazo  $H_0$ .

## Test de independencia

Este test se utiliza cuando queremos comprobar si dos variables aleatorias discretas son independientes entre sí. Sean  $X$  e  $Y$  dos v.a. discretas e independientes entre sí.

Para que sean independientes debe cumplirse que

$$\boxed{P(X = i, Y = j) = P(X = i) * P(Y = j)} \quad \forall i, j$$

Sean  $M_{ij}$  = Frecuencia de observaciones obtenidas

$p_{ij} = P(X = i, Y = j)$  bajo hipótesis de independencia

Entonces  $\hat{p}_{ij} = P(X = i) * P(Y = j) = \hat{p}_i * \hat{p}_j$ , con  $\hat{p}_i = \frac{\sum_j M_{ij}}{n}$  y  $\hat{p}_j = \frac{\sum_i M_{ij}}{n}$

$\Rightarrow$  se construye el estadístico

$$\boxed{Q = \sum_{i,j} \frac{(M_{ij} - n\hat{p}_i\hat{p}_j)^2}{n\hat{p}_i\hat{p}_j} \sim \chi^2_{(p-1)(q-1)}}$$

## Problema 2

Se quiere probar si hay una diferencia de ingreso entre hombres y mujeres médicos. Para lo cual se hizo una encuesta a 200 médicos al azar y de manera independiente obteniéndose la siguiente información:

	Ingreso Bajo	Ingreso Alto	Total
Hombres	20	100	120
Mujeres	70	10	80
Total	90	110	200

Estudie la independencia entre las variables Sexo e ingreso

**Sol.:**

De la tabla podemos ver que

- Probabilidad de ser Hombre  $= P_H = \frac{120}{200}$  ,
- Probabilidad de ser Mujer  $= P_M = \frac{80}{200}$
- Probabilidad de Ingreso Bajo  $= P_B = \frac{90}{200}$  ,
- Probabilidad de Ingreso Alto  $= P_A = \frac{110}{200}$

Entonces, si ambas variables son independientes debe cumplirse que:

- $P_{HB} = P_H P_B = \frac{120}{200} * \frac{90}{200} = 0.27$  (probabilidad que sea hombre y de ingreso bajo)
- $P_{HA} = P_H P_A = \frac{120}{200} * \frac{110}{200} = 0.33$  (probabilidad que sea hombre y de ingreso alto)
- $P_{MB} = P_M P_B = \frac{80}{200} * \frac{90}{200} = 0.18$  (probabilidad que sea mujer y de ingreso bajo)
- $P_{MA} = P_M P_A = \frac{80}{200} * \frac{110}{200} = 0.22$  (probabilidad que sea mujer y de ingreso alto)

Tabla de frecuencias teóricas  $n\hat{p}_i\hat{p}_j$

	Ingreso Bajo	Ingreso Alto	Total
Hombres	54	66	120
Mujeres	36	44	80
Total	90	110	200

$$\Rightarrow Q = \sum_{i,j} \frac{(M_{ij} - n\hat{p}_i\hat{p}_j)^2}{n\hat{p}_i\hat{p}_j} = \frac{(20-54)^2}{54} + \frac{(100-66)^2}{66} + \frac{(70-36)^2}{36} + \frac{(10-44)^2}{44} = 97.3$$

$$\Rightarrow Q \sim \chi^2_{(p-1)(q-1)} = \chi^2_{(2-1)(2-1)} = \chi^2_{1*1} = \chi^2_1 \quad y$$

$P(\chi^2_1 > 97.3) < 0.05 \Rightarrow$  se rechaza  $H_0$  y las variables Sexo e Ingreso no son independientes.

## Regresión Lineal

El coeficiente de correlación es un índice que nos muestra el grado de relación lineal que existe entre dos variables X e Y y se escribe como

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(x) * Var(y)}} = \frac{E(xy) - E(x)E(y)}{\sqrt{Var(x) * Var(y)}}$$

Si tenemos dos muestras  $x_1, \dots, x_n$  y  $y_1, \dots, y_n$ , podemos estimar las varianzas y las covarianzas como:

$$Var(x) \rightarrow S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$Var(y) \rightarrow S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$Cov(x, y) \rightarrow \tilde{Cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\Rightarrow \hat{\rho} = \frac{\tilde{Cov}(x, y)}{\sqrt{S_x^2 S_y^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

### Propiedades:

- El coeficiente de correlación es insensible a los cambios de escala
- $\rho = 0 \rightarrow$  ausencia de tendencia lineal
- $\rho \in (-1,0) \cup (0,1) \rightarrow$  tendencia mas o menos lineal
- $\rho = \pm 1 \rightarrow$  relación lineal estricta

### Problema 3

Se considera la variable X : talla del padre e Y: la talla de su hijo en la población chilena. Se supone la distribución conjunta de X e Y normal bivariada de medias  $\mu_x, \mu_y$  iguales a  $\mu$  y varianzas  $\sigma_x^2, \sigma_y^2$  iguales a  $\sigma^2$  y la correlación  $\rho$  positiva. Si ambas variables son obtenidas de una muestra con errores de medición  $X = X' + \varepsilon$  e  $Y = Y' + \eta$ , en donde X e Y son las mediciones X', Y' son los valores verdaderos y  $\varepsilon, \eta$  los errores de medición. Suponiendo que  $\varepsilon, \eta$  son independientes de X' e Y' e independientes entre sí.

- a) Exprese la correlación entre X e Y en función de la correlación  $\rho$  entre X' e Y'. Concluir.  
b) Si se obtiene una muestra de 20 familias con los siguientes resultados

$$\bar{x} = 170,92 \quad \bar{y} = 170,93, \quad \frac{1}{20} \sum x_i^2 = 29.293, \quad \frac{1}{20} \sum y_i^2 = 29.349,$$

$\frac{1}{20} \sum x_i y_i = 29.307$ . Deduzca el coeficiente de correlación entre las tallas observadas del padre y del hijo. Si se supone que la varianza de los errores de medición es igual a 9, dé una estimación de la correlación  $\rho$  entre las tallas verdaderas X' y Y'. Comente.

**Sol.:**

$$a) \rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

Como  $X = X' + \varepsilon$  e  $Y = Y' + \eta$ , si calculamos las varianzas de X e Y tenemos que

$$\begin{aligned} \Rightarrow \sigma_x^2 &= \sigma_{x'}^2 + \text{var}(\varepsilon) \Rightarrow \sigma_x^2 \geq \sigma_{x'}^2 \\ \Rightarrow \sigma_y^2 &= \sigma_{y'}^2 + \text{var}(\eta) \Rightarrow \sigma_y^2 \geq \sigma_{y'}^2 \end{aligned}$$

Por otro lado,

$$Cov(x, y) = Cov(x' + \varepsilon, y' + \eta) = \text{cov}(x', y') + \text{cov}(x', \eta) + \text{cov}(\varepsilon, y') + \text{cov}(\varepsilon, \eta)$$

Por supuesto de independencia  $\text{cov}(x', \eta) = \text{cov}(\varepsilon, y') = \text{cov}(\varepsilon, \eta) = 0$

$$\Rightarrow Cov(x, y) = Cov(x', y')$$

$$\text{luego } \frac{Cov(x, y)}{\sigma_x \sigma_y} \leq \frac{Cov(x', y')}{\sigma_{x'} \sigma_{y'}} \Rightarrow \rho_{xy} \leq \rho_{x'y'}$$

Es decir, con los valores muestrales se subestima la verdadera correlación entre X e Y.

b) Si consideramos los valores obtenidos en las 20 familias, tenemos que

$$\Rightarrow \hat{\rho}_{xy} = \frac{\frac{1}{20} \sum_{i=1}^n x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{20} \sum_{i=1}^n x_i^2 - \bar{x}^2 * \frac{1}{20} \sum_{i=1}^n y_i^2 - \bar{y}^2}} = \frac{91,6444}{102,321} \approx 0,9$$

Por otro lado,  $\hat{\rho}_{x'y'} \approx \frac{\hat{\sigma}_x \hat{\sigma}_y * \hat{\rho}_{xy}}{\hat{\sigma}_{x'} \hat{\sigma}_{y'}}$

Con  $\sigma_x^2 = \frac{1}{20} \sum x_i^2 - \bar{x}^2 = 29.293 - (170,92)^2 = 79,3536 \Rightarrow \sigma_x = 8,91$

$$\sigma_y^2 = \frac{1}{20} \sum y_i^2 - \bar{y}^2 = 29.349 - (170,93)^2 = 131,935 \Rightarrow \sigma_y = 11,486$$

Si las varianzas de los errores de medición es igual a 9, entonces

$$\sigma_{x'}^2 = 79,3536 - 9 = 70,3536 \Rightarrow \sigma_{x'} = 8,39$$

$$\sigma_{y'}^2 = 131,935 - 9 = 122,935 \Rightarrow \sigma_{y'} = 11,09$$

$$\Rightarrow \hat{\rho}_{x'y'} = \frac{8,91 * 11,486 * 0,9}{8,39 * 11,09} = 0,9899$$