Evaluación Continua
Segundo examen parcial

Duración: 90 minutos

- **No está permitido el uso de documentación, salvo el formulario que haya recibido**
- **Use 4 dígitos decimales en todos los cálculos y resultados**

1. **(2.5 Puntos)** Sea $X_1, \ldots, X_n$ (con $n \geq 3$) una m.a.s de una población $N(\mu, \sigma^2)$. Se definen los estimadores de μ:

$$T_1 = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i \qquad T_2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i$$

   a) **(1 Punto)** ¿Son insesgados los estimadores?

   Solution

   T_1 sí lo es, pero no T_2

$$E(T_1) = \frac{(n-2)\mu}{n-2} = \mu \qquad E(T_2) = \frac{n\mu}{n-1} \neq \mu$$

   El sesgo de T_2 es $Sesgo(T_2) = E(T_2) - \mu = \frac{\mu}{n-1}$.

   b) **(0.75 Puntos)** ¿Cuál de ellos tiene menor varianza? Tenga en cuenta que $n \geq 3$.

   Solution

   $V(T_1) = \frac{\sigma^2}{n-2} \qquad V(T_2) = \frac{n\sigma^2}{(n-1)^2}$

   Como $(n-1)^2 \geq n(n-2)$ para cualquier valor de $n \geq 3$, el estimador T_1 tiene mayor varianza que el estimador T_2

   c) **(0.75 Puntos)** Obtenga el error cuadrático medio de cada estimador.

   Solution

   $ECM(T_1) = Sesgo^2(T_1) + V(T_1) = \frac{\sigma^2}{n-2}$

   $ECM(T_2) = Sesgo^2(T_2) + V(T_2) = \frac{\mu^2}{(n-1)^2} + \frac{n}{(n-1)^2}\sigma^2$

2. **(2.5 Puntos)** Se quiere determinar la eficacia de un nuevo tipo de asfalto fonoabsorbente en la reducción del ruido derivado del tráfico. Para ello, se selecciona un vehículo y, en condiciones controladas, se mide el ruido que produce con y sin asfalto fonoabsorbente (101 pasadas para cada condición). Con el asfalto tradicional se obtienen los siguientes resultados:

$$\bar{x}_1 = 60.5\ dBA; \quad \hat{s}_1 = 1.95\ dBA.$$

   Tras instalar el asfalto fonoabsorbente se obtienen los siguientes resultados:

$$\bar{x}_2 = 57.1\ dBA; \quad \hat{s}_2 = 1.75\ dBA.$$

   a) **(1 Punto)** Calcule un intervalo de confianza al 95 % para la diferencia de medias del ruido en ambos escenarios, y justifique si el nuevo material reduce el ruido, tal y como asegura el fabricante.

   Solution

$$\left[(\overline{x}_1 - \overline{x}_2) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}\right] = \left[(60.5 - 57.1) \pm 1.96\sqrt{\frac{1.95^2}{101} + \frac{1.75^2}{101} =}\right] = [2.489; 3.511]$$

   Dado que el intervalo es íntegramente positivo, podemos asegurar que el nuevo asfalto sí reduce el ruido.

   b) **(0.75 Puntos)** Asumiendo el mismo tamaño de muestra para ambas condiciones (es decir, $n_1 = n_2 = n$), determine el tamaño de muestra (n) necesario para que la amplitud del intervalo obtenido en el aparatado anterior sea de 0.5.

   Solution

   $Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{s}_1^2}{n} + \frac{\hat{s}_2^2}{n}} = 0.25$. Luego, $1.96\sqrt{\frac{1.95^2+1.75^2}{n}} = 0.25$, y despejando $n = \frac{1.95^2+1.75^2}{(0.25/1.96)^2} = 421.16 \approx 422$.

   c) **(0.75 Puntos)** Calcule el intervalo de confianza al 95 % de la varianza del ruido tras la instalación del nuevo material. Además, indique qué hipótesis sobre la población aseguran la validez del procedimiento empleado.

   Nota: utilice, si es necesario, la siguiente información $\chi^2_{(100)(0.025)} = 129.564$; $\chi^2_{(100)(0.975)} = 74.216$.

   Solution

   $\sigma_2^2 \epsilon (\frac{100 \cdot 1.75^2}{129.564}; \frac{100 \cdot 1.75^2}{74.216}) = (2.364; 4.126)$

Fábrica	Antes (horas perdidas)	Después (horas perdidas)
1	45	36
2	73	60
3	46	44
4	39	29
5	17	11
6	30	32

3. **(2.5 Puntos)** Los siguientes datos corresponden a un estudio sobre las horas perdidas mensualmente por accidentes laborales en 6 fábricas, antes y después de implantar un programa de seguridad industrial.

   a) (1.25 **Puntos**) Suponiendo que la población es normal, probar que el programa es eficaz para un nivel de significación del 5 %. Para ello, indique las hipótesis nula y alternativa, el estadístico del contraste, la región o área de rechazo y la conclusión a la que llegue.

   Nota: utilice, si es necesario, la siguiente información: $t_{(6)(0.05)} = 1.94$; $t_{(6)(0.975)} = -2.45$; $t_{(5)(0.05)} = 2.01$; $t_{(5)(0.95)} = -2.01$; $t_{(5)(0.025)} = 2.57$.

   Solution

   Sean μ_1 : media de las horas perdidas antes del programa y μ_2 : media de las horas perdidas después del programa.

   Contraste para la comparación de medias (muestras pareadas):

   $H_0 : \mu_1 - \mu_2 = \mu_d = 0$

   $H_1 : \mu_1 - \mu_2 = \mu_d > 0$

   $\bar{d} = 6.335; \quad \hat{S}_d = 5.5377.$

   $t_0 = \frac{\bar{d}-d_0}{\hat{S}_d/\sqrt{n}} = \frac{6.335-0}{5.5377/\sqrt{6}} = 2.8022 > t_{(5)(0.05)} = 2.01$

   Luego, se rechaza H_0 con una significación del 5 %. Por tanto, se puede afirmar que el programa ha sido eficaz.

   b) (0.75 **Puntos**) Calcule el p-valor asociado al contraste anterior. Utilice, si es necesario, la siguiente información:

   Probability Distributions

   Distribution: Student's t

Parameters:	*D. F.*
Dist. 1	5
Dist. 2	6
Dist. 3	
Dist. 4	
Dist. 5	

   Cumulative Distribution

   Distribution: Student's t

   Lower Tail Area (<)

Variable	*Dist. 1*	*Dist. 2*	*Dist. 3*	*Dist. 4*	*Dist. 5*
2,8	0,981003	0,984418			
2,6	0,975875	0,979669			
1,9	0,942068	0,946915			

   Probability Density

Variable	*Dist. 1*	*Dist. 2*	*Dist. 3*	*Dist. 4*	*Dist. 5*
2,8	0,0224158	0,0205328			
2,6	0,0291761	0,0272865			
1,9	0,0743428	0,0736027			

   Upper Tail Area (>)

Variable	*Dist. 1*	*Dist. 2*	*Dist. 3*	*Dist. 4*	*Dist. 5*
2,8	0,018997	0,015582			
2,6	0,024125	0,0203311			
1,9	0,0579318	0,0530849			

   Solution

   $p - valor = P(t_{(5)} > 2.8) = 0.019.$

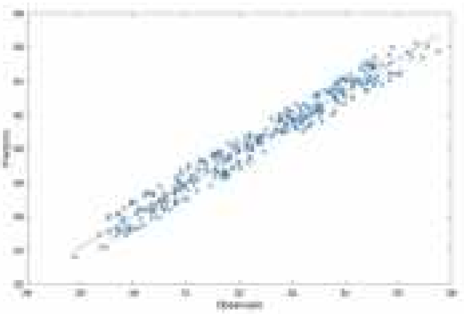
   c) (**0.5 Puntos**) Qué hipótesis sobre la población aseguran la validez del procedimiento empleado en los apartados anteriores.

   Solution

   Al tratarse de muestras pequeñas, es necesario suponer que la distribución de X_1 y X_2 es normal. Por tanto la distribución de $X_1 - X_2$ también será normal, pues es una combinación lineal de dos variables normales. Además, se asumen varianzas desconocidas y distintas.

4. **(2.5 Puntos)** En la siguiente figura se muestran los resultados de una regresión múltiple realizada con Statgraphics. Responda de manera razonada a las siguientes cuestiones.

Multiple Regression - Y
Dependent variable: Y
Independent variables:
X1
X2
X3
X4



		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	2.1747	0.78408	2.7735	0.0058982
X1	1.011	0.011512	87.815	1.6455e-213
X2	0.21182	0.017262	12.271	3.0446e-28
X3	0.10478	0.016577	6.3209	9.5842e-10
X4	1.6559	0.075496	21.934	6.3753e-64

R-squared = 96,5 percent
R-squared (adjusted for d.f.) = 96,3 percent
Residual variance = 0.32

a) **(1 Punto)** Es adecuado el modelo ajustado? Indique sí/no y por qué. Además, escriba la ecuación del modelo propuesto e indique la proporción de variabilidad explicada por dicho modelo.

Solution

El modelo es adecuado, pues todas la variables son significativas (p-valor<0.05). La ecuación del modelo es:

$$\hat{y} = 2.1747 + 1.011 \times X1 + 0.2118 \times X2 + 0.1048 \times X3 + 1.6559 \times X4.$$

Este modelo consigue explicar el 96.5 % de la variabilidad total.

b) **(0.5 Puntos)** Prediga el valor de Y para los valores $X1 = 1$, $X2 = 2$, $X3 = 3$ y $X4 = 4$.

Solution

$\hat{y} = 2.1747 + 1.011 \times 1 + 0.2118 \times 2 + 0.1048 \times 3 + 1.6559 \times 4 = 10.5473.$

c) **(1 Punto)** Teniendo en cuenta que la varianza residual es $S_R^2 = 0.32$, indique cuál sería la distribución de probabilidad de Y cuando $X1 = 1$, $X2 = 2$, $X3 = 3$ y $X4 = 4$. Además, calcule la probabilidad de $Y > 12$.

Solution

$Y \sim N(10.5473; \sqrt{0.32})$ Luego $P(Y > 12) = 1 - P(Y < 12) = 1 - P(Z < \frac{12-10.5473}{\sqrt{0.32}}) = 1 - P(Z < 2.57) = 1 - 0.9949 = 0.0051.$