MÉTODOS ESTADÍSTICOS DE LA INGENIERÍA

SEGUNDA CONVOCATORIA

RESOLUCIÓN DEL EJERCICIO FINAL

Ejercicio 1 (ESTADÍSTICA DESCRIPTIVA)**:**

Un estudio efectuado sobre 240 muestras de aceros especiales, aleados con manganeso, de una calidad determinada ha proporcionado los resultados en % Mn (p/p): $\overline{x} = 1.35\%,\ s = 0.21\%$, supuesta que es una variable aleatoria normal. Para hacer otros análisis, el Departamento de Calidad de una cierta empresa, desea establecer una escala cualitativa donde los posibles valores de manganeso queden clasificados según los siguientes criterios:

CRITERIO	SIGNIFICADO	INTERPRETACIÓN
I	valores excesivamente bajos	5 % inferior
II	valores bajos aceptables	20 % siguiente
III	valores aceptables	50 % central
IV	valores altos aceptables	20 % siguiente
V	valores excesivamente altos	5 % superior

Determine:

(1º) Los límites de concentración de Mn derivados de la anterior clasificación (**6 PUNTOS**).

Se nos ha dado de forma indirecta una tabla de frecuencias. Supondremos que la muestra está distribuida normalmente. Entonces:

$$z = \frac{x - \overline{x}}{s} \Leftrightarrow x = \overline{x} + z \cdot s$$

La forma de trabajar será obtener el límite inferior y superior de cada intervalo de clase para lo que se utilizará la probabilidad de referencia. Por ejemplo, la zona I (valores excesivamente bajos) corresponde al 5% inferior, con lo que la puntuación tipificada asociada será tal que

$$\mathbb{P}(Z \le z_1) = 0.05 \underset{simetria}{\Leftrightarrow} \mathbb{P}(Z \le |z_1|) = 0.95 \Rightarrow z_1 \underset{interpolando}{=} -1.644853627$$

$$\mathbb{P}(z_1 \le Z \le z_2) = 0.20 \Leftrightarrow \mathbb{P}(Z \le z_2) = 0.25 \underset{simetria}{\Leftrightarrow} \mathbb{P}(Z \le |z_2|) = 0.75 \Rightarrow z_2 \underset{interpolando}{=} -0.67448975$$

$$\mathbb{P}(z_2 \le Z \le z_3) = 0.50 \Leftrightarrow \mathbb{P}(Z \le z_3) = 0.75 \underset{simetria}{\Leftrightarrow} z_3 = -z_2 = 0.67448975$$

$$\mathbb{P}(z_3 \le Z \le z_4) = 0.20 \Leftrightarrow \mathbb{P}(Z \le z_4) = 0.95 \underset{simetria}{\Leftrightarrow} z_4 = -z_2 = 1.644853627$$

De esta forma la tabla de frecuencias en términos de concentración de manganeso será

CLASE	DEFINICIÓN	$\ell_i(\%)$	$\mathcal{L}_i(\%)$
I	5%	0	1.004581
II	20%	1.004581	1.208357
III	50%	1.208357	1.491643
IV	20%	1.491643	1.695419
V	5%	1.695419	∞

(2º) Los valores de concentración de Mn que delimitan el 50 % de los valores centrales del estudio realizado ¿Qué estadístico muestral se acaba de determinar? ¿Qué información adicional se puede deducir de este resultado, sin hacer más cálculos adicionales? (**3 PUNTOS**).

De la información proporcionada se deduce que el intervalo central que contiene el 50 % de los valores centrales de la muestra es el intervalo [1.208357, 1.491643], que corresponde al rango intercuartílico, con lo que de forma indirecta se dispone de los cuartiles:

$$Q_1 = 1.208357\ \% \text{ y } Q_3 = 1.491643\ \%$$

(3º) ¿Qué probabilidad hay de que un valor supere el 1.63 %? (**1 PUNTO**).

$$\mathbb{P}(X \geq 1.63) = \mathbb{P}\left(Z \geq z_1 = \frac{1.63 - 1.35}{0.21}\right) = \mathbb{P}(Z \geq z_1 = 1.3333) = 1 - \mathbb{P}(Z \leq z_1 = 1.3333) =$$
$$= 1 - 0.908788726 = 9.1211274\ \%$$

EJERCICIO 2 (PROBABILIDAD**):**

Se conoce que para una resina basada en la etil-celulosa el color, la viscosidad de la solución y el porcentaje de etoxilo son características independientes. Además, al efectuar valoraciones de la resina respecto de cada una de tales características por separado se ha medido que las probabilidades de cometer un error en tal valoración son, respectivamente, 0.03, 0.05 y 0.02.

Sean los siguientes sucesos:

E_1 := "Valoración correcta del color" $\mathbb{P}(\overline{E}_1) = 0.03$

E_2 := "Valoración correcta de la viscosidad" $\mathbb{P}(\overline{E}_2) = 0.05$

E_3 := "Valoración correcta del porcentaje de etoxilo" $\mathbb{P}(\overline{E}_3) = 0.02$

2 PUNTOS

Entonces, la probabilidad de hacer una medición de cualquiera de las tres características por separado sin cometer un error sería, respectivamente:

$$\mathbb{P}(E_1) = 0.97; \mathbb{P}(E_2) = 0.95; \mathbb{P}(E_3) = 0.98;$$

2 PUNTOS

(**1º**) La probabilidad de no cometer errores si se efectuara una valoración simultánea de las tres características es

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = 0.97 \times 0.95 \times 0.98 = 0.903070$$

3 PUNTOS

(**2º**) La probabilidad de cometer algún error al efectuar una valoración en la que intervenga alguna de esas tres características es

$$\mathbb{P}(\overline{E}_1 \cup \overline{E}_2 \cup \overline{E}_3) =$$
$$\mathbb{P}(\overline{E}_1) + \mathbb{P}(\overline{E}_2) + \mathbb{P}(\overline{E}_3) - \mathbb{P}(\overline{E}_1 \cap \overline{E}_2) - \mathbb{P}(\overline{E}_1 \cap \overline{E}_3) - \mathbb{P}(\overline{E}_2 \cap \overline{E}_3) + \mathbb{P}(\overline{E}_1 \cap \overline{E}_2 \cap \overline{E}_3) = 0.0969$$

De otra manera

$$\mathbb{P}(\overline{E}_1 \cup \overline{E}_2 \cup \overline{E}_3) = 1 - \mathbb{P}(E_1 \cap E_2 \cap E_3) = 1 - 0.903070 = 0.096930$$

3 PUNTOS

<u>Ejercicio 3</u> (ESTIMACIÓN y CONTRASTE**):**

Las presiones críticas de dos grupos independientes de recipientes de distintos vidrios, de dos composiciones diferentes, han dado los siguientes valores

GRUPO 1º	100	102	96	106	110	110	120	112	112	90
GRUPO 2º	104	88	100	98	102	92	96	100	96	96

Se supone que las dos poblaciones de composiciones son normales.

Los estadísticos muestrales de ambas m.a.s. son (**2 PUNTOS**):

ESTADÍSTICO	GRUPO 1º	GRUPO 2º
n	10	
$\sum_{i=1}^{10} x_i$	1058	972
$\overline{x}_i$	105.8	97.2
$\sum_{i=1}^{10} x_i^2$	112644	94680
$S \triangleq \hat{s}$	8.8692	4.7329

Y se trata de muestras independientes, dos poblaciones (porque hay dos muestras) y pequeñas muestras (dado que n es inferior a 30 elementos).

(1º- 2 PUNTOS) Aunque no son conocidas, ¿existe evidencia estadística suficiente para determinar que las dos poblaciones poseen varianzas iguales con un nivel de significación α=1%?

Es un problema de contraste bilateral de varianza de poblaciones (que se suponen normalmente distribuidas):

$$\begin{cases} H_0: & \sigma^2_{G1^o} = \sigma^2_{G2^o} \\ H_a: & \sigma^2_{G1^o} \neq \sigma^2_{G2^o} \end{cases} \quad \text{(0.5 PUNTOS)}$$

Por lo tanto, el modelo de probabilidad es el de Fisher-Snedecor dando como valores frontera de la región de admisibilidad de la hipótesis nula (**0.5 PUNTOS**):

$$\begin{cases} F_1 = F_{\alpha=0.5\%,\nu_1=9\,gdl,\nu_2=9\,gdl} = F^{-1}_{\alpha=99.5\%,\nu_1=9,\nu_2=9} = 6.54^{-1} = 0.1529 \\ F_2 = F_{\alpha=99.5\%,\nu_1=9\,gdl,\nu_2=9\,gdl} = 6.54 \end{cases}$$

El estadístico del contraste es (**0.5 PUNTOS**)

$$F=\frac{\hat{s}_{G1^o}^2}{\hat{s}_{G2^o}^2}\cdot\cancel{\frac{\sigma_{G2^o}^2}{\sigma_{G1^o}^2}}=\frac{\hat{s}_{G1^o}^2}{\hat{s}_{G2^o}^2}=\frac{8.8692^2}{4.7329^2}=\frac{78.6627}{22.4003}=3.5117$$

Como $F_1 \le F \le F_2$, a partir de la muestra analizada no existe evidencia estadística para rechazar la hipótesis nula con un nivel de significación α = 1% (lo que implica que los dos conjuntos de datos provienen de poblaciones con varianzas iguales) (**0.5 PUNTOS**).

(2º- 3 PUNTOS) ¿Existe evidencia estadística para decidir si el grupo 2 soporta una mayor presión crítica que el grupo 1 con un nivel de significación α=1%?

En esta ocasión se nos pide analizar un contraste de medias aritméticas: el modelo de probabilidad a utilizar es la t de Student, y el contraste es unilateral (de cola inferior). Las hipótesis son (**0.5 PUNTOS**):

$$\begin{cases} H_0: & \mu_{G1^o}=\mu_{G2^o} \\ H_a: & \mu_{G1^o}<\mu_{G2^o} \end{cases} \Leftrightarrow \begin{cases} H_0: & \mu_{G1^o}-\mu_{G2^o}=0 \\ H_a: & \mu_{G1^o}-\mu_{G2^o}<0 \end{cases}$$

siendo el error probable que hay que aplicar (**0.5 PUNTOS**)

$$\sigma_{\overline{\mu_{G1^o}-\mu_{G2^o}}}=\sqrt{\frac{1}{n}(\hat{s}_1^2+\hat{s}_2^2)}=\sqrt{\frac{1}{10}(8.8692^2+4.7329^2)}=\sqrt{10.1063}=3.1790$$

El estadístico del contraste es (**0.5 PUNTOS**)

$$t=\frac{(\overline{x}_{G1^o}-\overline{x}_{G2^o})-\cancel{(\mu_{G1^o}-\mu_{G2^o})}}{\sigma_{\overline{\mu_{G1^o}-\mu_{G2^o}}}}=\frac{\overline{x}_{G1^o}-\overline{x}_{G2^o}}{\sigma_{\overline{\mu_{G1^o}-\mu_{G2^o}}}}=\frac{8.6}{3.1790}=2.7052$$

El valor crítico que separa la región de aceptación de la de rechazo para un nivel de significación α = 1 % viene dado por

$$t_1=t_{\alpha,\nu}=t_{1\%,\nu=2(n-1)}\equiv -t_{99\%,18\,gdl}=-2.552379618$$

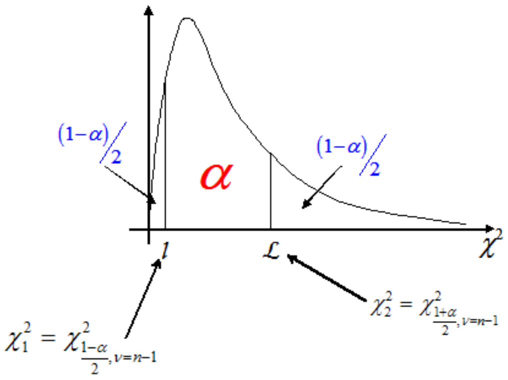
Como $t_1 \le t$ no existe, a partir de la muestra suministrada, evidencia estadística para rechazar la hipótesis nula con un nivel de significación del 1 % (o sea, el grupo 2 no soporta una mayor presión crítica que el grupo 1) (**0.5 PUNTOS**).

(3º- 1 PUNTO) Calcule el p-valor aproximado del contraste, justificando el resultado que se obtenga.

$$t_{p-valor,\nu}=t_{p-valor,\nu=2(n-1)}\equiv t_{p-valor,18\,gdl}=2.7052\Rightarrow p-valor=99.98550929\%$$

(4º- 2 PUNTOS) Calcule el intervalo de estimación para la varianza del grupo 1 con un nivel de confianza 99%. Interprete el resultado de dicha estimación confidencial.

El modelo de probabilidad que interviene es el modelo χ^2, siendo los valores críticos (**1 PUNTO**):



$$\chi_1^2 = \chi^2_{\frac{1-\alpha}{2},\nu=n-1} = \chi^2_{0.5\%,\nu=9gdl} = 1.734932909$$

$$\chi_2^2 = \chi^2_{\frac{1+\alpha}{2},\nu=n-1} = \chi^2_{99.5\%,\nu=9gdl} = 23.58935078$$

El intervalo de estimación pedido satisface (**1 PUNTO**)

$$l = \hat{s}^2_{G1^o}\frac{n-1}{\chi_2^2} \le \sigma^2 \le \mathcal{L} = \hat{s}^2_{G1^o}\frac{n-1}{\chi_1^2} \text{ con } \left[\hat{s}^2_{G1^o}(n-1) = 707.9643\right] \Rightarrow \sigma^2 \in [l,\ \mathcal{L}] = [30.0125,\ 408.0486]$$

Es decir, si se repitieran estimaciones por intervalo en las condiciones del enunciado con otras muestras de tamaño n para esa población, en $[l,\ \mathcal{L}] = [30.01,\ 408.04]$ se esperarían obtener 99 de cada 100 valores de estimación calculados.

NOTA ACLARATORIA: A lo largo de este ejercicio se ha trabajado con unidades de presión crítica, como indica el enunciado.

Ejercicio 4 (VARIABLE ALEATORIA**):**

Sea X la variable aleatoria "las llegadas de los trabajos a una máquina se producen de acuerdo a un proceso de Poisson con una tasa de llegadas de 8 clientes por hora". Sea Y otra variable aleatoria definida como "el tiempo empleado por la máquina en realizar un trabajo según una distribución exponencial de media 5 minutos". Se pide:

(1º) Calcule la probabilidad de que al menos lleguen 4 clientes en la próxima hora (**2 PUNTOS**).

$$\mathbb{P}(X \geq 4 \text{ clientes}) = 1 - \mathbb{P}(X < 4 \text{ clientes}) = 1 - \mathbb{P}(X \leq 3 \text{ clientes}) = 1 - F(3) = 1 - \sum_{i=0}^{3} \frac{e^{-8} 8^i}{i!} = 0.957619888$$

(2º) Calcule la función de distribución de probabilidad de la variable aleatoria X (**2 PUNTOS**).

$$F_X(a) = \sum_{k=0}^{a} \mathbb{P}(X = k) = \sum_{k=0}^{a} \frac{e^{-\lambda} \lambda^i}{i!} = \sum_{k=0}^{a} \frac{e^{-8} 8^i}{i!}$$

Calcule la probabilidad de que lleguen al menos 5 clientes pero no más de ocho.

$$\mathbb{P}(5 \leq X \leq 8 \text{ clientes}) = F(8) - F(4) = 0.592547341\text{-}0.0996324\text{=}0.49291494$$

(3º) La probabilidad de que una máquina emplee 3 o menos minutos para hacer un trabajo (**2 PUNTOS**).

$$\mathbb{P}(Y \leq 3 \text{ minutos}) = \int_0^3 \tfrac{1}{5} e^{-\frac{1}{5}y} dy = -\left[e^{-\frac{1}{5}t}\right]_0^3 = 1 - e^{-\frac{3}{5}} = 0.451188$$

(4º) Si la máquina ha empleado ya 3 minutos en hacer un trabajo dado, ¿cuál es la probabilidad de que use otros tres minutos adicionales para hacer el trabajo? (**2 PUNTOS**).

$$\mathbb{P}(Y \geq 3 \text{ minutos + 3 minutos} \mid Y \geq 3 \text{ minutos}) =$$

$$\begin{cases} \text{(con probabilidad condicionada)} & \dfrac{\mathbb{P}(Y \geq 6 \text{ minutos})}{\mathbb{P}(Y \geq 3 \text{ minutos})} = \dfrac{\int_6^\infty \frac{1}{5} e^{-\frac{1}{5}y} dy}{\int_3^\infty \frac{1}{5} e^{-\frac{1}{5}y} dy} = \dfrac{-\left[e^{-\frac{1}{5}t}\right]_6^\infty}{-\left[e^{-\frac{1}{5}t}\right]_3^\infty} = \dfrac{e^{-\frac{6}{5}}}{e^{-\frac{3}{5}}} = 0.5488 \\ \text{(la v.a. exponencial no tiene memoria)} & \mathbb{P}(Y \geq 3 \text{ minutos}) = \int_3^\infty \frac{1}{5} e^{-\frac{1}{5}y} dy = -\left[e^{-\frac{1}{5}t}\right]_3^\infty = e^{-\frac{3}{5}} = 0.5488 \end{cases}$$

(5º) El 83.64 % de los trabajos son finalizados por la máquina en T minutos, ¿cuánto vale T? (**2 PUNTOS**).

$$\mathbb{P}(Y \leq T \text{ minutos}) = \int_0^T \tfrac{1}{5} e^{-\frac{1}{5}y} dy = -\left[e^{-\frac{1}{5}t}\right]_0^T = 1 - e^{-\frac{T}{5}} = 0.8364 \Leftrightarrow T = 9.05165 \text{ minutos}$$