Universitat d'Alacant
Universidad de Alicante

Examen 17/12/2024

1.- Se ha detectado que los ciberataques que sufre la UA llegan de cuatro continentes (A,B,C y D). Se sospecha que la proporción de ataques es 3:1:1:1 respectivamente. Se han recogido los siguientes intentos de ataque informático en un mes: A: 325: B:110: C: 101: D: 99. ¿Hay motivos para refutar la sospecha? ¿Por qué?

Solución

Se plantea la siguiente hipótesis:

- H_0: Las proporciones de ciberataques son $3:1:1:1$.
- H_1: Las proporciones de ciberataques no son $3:1:1:1$.

Cálculo de frecuencias esperadas

El total de ataques observados es:

$$\text{Total} = 325 + 110 + 101 + 99 = 635$$

Bajo la hipótesis nula, las proporciones esperadas son $\frac{3}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$. Por lo tanto, las frecuencias esperadas se calculan como:

$$E_i = \text{Proporción}_i \times \text{Total}$$

Continente	**Frecuencia Observada** O_i	**Frecuencia Esperada** E_i
A	325	$\frac{3}{6} \times 635 = 317.5$
B	110	$\frac{1}{6} \times 635 = 105.83$
C	101	$\frac{1}{6} \times 635 = 105.83$
D	99	$\frac{1}{6} \times 635 = 105.83$

Estadística de prueba χ^2

La estadística de prueba χ^2 se define como:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Sustituyendo los valores observados y esperados:

$$\chi^2 = \frac{(325 - 317.5)^2}{317.5} + \frac{(110 - 105.83)^2}{105.83} + \frac{(101 - 105.83)^2}{105.83} + \frac{(99 - 105.83)^2}{105.83}$$

Realizando los cálculos, obtenemos:

$$\chi^2 = 1.00$$

Grados de libertad

El número de grados de libertad es:

$$df = k - 1 = 4 - 1 = 3$$

Valor p

El valor p asociado a la estadística $\chi^2 = 1.00$ con 3 grados de libertad se calcula como:

$$p = P(\chi^2 > 1.00) \approx 0.8013$$

Por todo lo expuesto, concluimos que dado que el p-valor ($p \approx 0.8013$) es mucho mayor que el nivel de significación típico ($\alpha = 0.05$), no podemos rechazar la hipótesis nula H_0. Es decir, no hay evidencia suficiente para refutar la sospecha de que las proporciones de ciberataques siguen la distribución $3 : 1 : 1 : 1$.

2.- Con el comando random.normalvariate(5,3) de la librería *random* de *Python*, hemos generado 20 valores de una normal con el siguiente resultado:

$$7.71, 4.23, 7.84, 1.61, 7.42, 6.3, 8.66, 1.81, 11.28, 3.98, 4.32, 5.13, 4.67, 3.68, 6.02, 2.72, 5.82, 5.15, 6.11, 4.61$$

¿Tenemos razones para pensar que el comando funciona mal? ¿Por qué?

Nota: $\bar{X} = 5.4535 \quad S = 2.3$

Solución

Hipótesis

- H_0: Los datos siguen la distribución $N(5, 3^2)$.
- H_1: Los datos no siguen la distribución $N(5, 3^2)$.

Definición de intervalos Al tener una muestra de tamaño 20, sería conveniente definir intervalos donde tuviéramos al menos, 5 valores esperados, por tanto, probabilidad 0.25. Se requiere calcular 4 intervalos disjuntos para una variable aleatoria normal con media $\mu = 5$ y varianza $\sigma^2 = 9$ (desviación estándar $\sigma = 3$), tales que la probabilidad de que la variable esté dentro de cada intervalo sea 0.2.

Método

Primero calculamos los valores de una distribución estándar $N(0, 1)$ y los adaptamos a la escala de $N(5, 9)$ usando:

$$x = \mu + \sigma z$$

Donde z son los valores críticos obtenidos.

Resultados

Los intervalos disjuntos calculados para una N(0,1) son:

$$z_{0.25} = -0.675 \; z_{0.5} = 0 \; z_{0.75} = 0.675$$

y adaptando la escala, tendríamos que cada intervalo contiene aproximadamente el 25 % de la probabilidad de la distribución normal $N(5, 9)$.

Intervalo	Frecuencia Observada (O_i)	Frecuencia Esperada (E_i)
$(-\infty, 2.98)$	3	5
$(2.98, 5.00)$	6	5
$(5.00, 7.02)$	6	5
$(7.02, \infty)$	5	5

El estadístico chi-cuadrado se calcula como:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Sustituyendo los valores:

$$\chi^2 = \frac{(3-5)^2}{5} + \frac{(6-5)^2}{5} + \frac{(6-5)^2}{5} + \frac{(5-5)^2}{5} = 1.2$$

Decisión

- Grados de libertad (gl):

$$\text{gl} = (\text{número de intervalos}) - 1 - (\text{número de parámetros estimados}) = 4 - 1 - 0 = 3$$

- Valor crítico ($\chi^2_{\text{crítico}}$) para $\alpha = 0.05$ y gl = 1:

$$\chi^2_{\text{crítico}} \approx 7.81473$$

- P-Valor $pp = 0.753$

Como $\chi^2 = 1.2 \leq \chi^2_{\text{crítico}} \approx 7.81473$ o equivalentemente $p = 0.753 > 0.05$, no rechazamos la hipótesis nula (H_0).

No hay evidencia suficiente al nivel de significación del 5 % para rechazar que los datos provienen de una distribución normal $N(5, 3^2)$.

3.- Bajo la suposición de que los datos anteriores sí sean de la distribución normal,

a) Da un intervalo de confianza al 95 % de la media de la distribución.

b) ¿Podemos pensar que la media es superior a 5?

c) ¿Podemos pensar que la varianza no es 5.3?

Solución

a) Intervalo de confianza al 95 %.

Dado que asumimos que los datos provienen de una distribución normal, el intervalo de confianza para la media μ al 95 % se calcula con la fórmula:

$$\text{IC} = \bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$$

donde:

- $\bar{x} = 5.45$: Media muestral.
- $s = 2.36$: Desviación estándar muestral.
- $n = 20$: Tamaño de la muestra.
- $t_{0.975} = 2.093$: Cuantil para un nivel de confianza del 95 %.

El margen de error es:

$$\text{ME} = t_{0.975,19} \cdot \frac{s}{\sqrt{n}} = 2.093 \cdot \frac{2.36}{\sqrt{20}} \approx 1.1$$

Por lo tanto, el intervalo de confianza es:

$$I_{95\,\%} = [4.35, 6.56]$$

b) En esta ocasión, planteamos un contraste de hipótesis donde la Hipótesis alternativa sea que la media poblacional sea superior a 5:

- Hipótesis nula (H_0): $\mu \leq 5$ (La media es menor o igual a 5).
- Hipótesis alternativa (H_1): $\mu > 5$ (La media es mayor a 5).

Nivel de significación

Utilizamos un nivel de significación del 5 % ($\alpha = 0.05$).

Estadístico de prueba

El estadístico de prueba se calcula como:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Sustituyendo los valores:

$$t = \frac{5.45 - 5}{2.36/\sqrt{20}} \approx 0.86$$

Región de rechazo

El valor crítico para una cola derecha con $\alpha = 0.05$ y $n - 1 = 19$ grados de libertad es:

$$t_{\text{crítico}} = t_{1-\alpha,\,19} = 1.73$$

Decisión

- Como $t = 0.86 \leq t_{\text{crítico}} = 1.73$, no rechazamos H_0.
- El valor p asociado al estadístico $t = 0.86$ es:

$$p = 0.200$$

  Dado que $p > 0.05$, tampoco rechazamos H_0.

Conclusión

No hay evidencia suficiente al nivel de significación del 5 % para afirmar que la media de los datos es superior a 5.

c) En esta ocasión se plantea el contraste:

- Hipótesis nula (H_0): $\sigma^2 = 5.3$ (La varianza poblacional es igual a 5.3).
- Hipótesis alternativa (H_1): $\sigma^2 \neq 5.3$ (La varianza poblacional es diferente de 5.3).

*Estadístico de prueba

El estadístico de prueba se calcula como:

$$\chi^2 = \frac{(n-1)\cdot s^2}{\sigma_0^2}$$

donde:

- $n = 20$: Tamaño de la muestra.
- $s^2 = 2.36^2 = 5.57$: Varianza muestral.
- $\sigma_0^2 = 5.3$: Varianza bajo la hipótesis nula.

Sustituyendo los valores:

$$\chi^2 = \frac{(20-1)\cdot 5.57}{5.3} \approx 19.93$$

Región de rechazo

El contraste es bilateral con $\alpha = 0.05$. Los valores críticos se obtienen de la distribución chi-cuadrado con $n - 1 = 19$ grados de libertad:

$$\chi^2_{\alpha/2,\,19} = 8.91 \quad \text{y} \quad \chi^2_{1-\alpha/2,\,19} = 32.85$$

La región de rechazo es:

$$\chi^2 < 8.91 \quad \text{o} \quad \chi^2 > 32.85$$

Decisión

Dado que:

- $8.91 \leq \chi^2 = 19.93 \leq 32.85$,
- $p = 0.798 > 0.05$,

no rechazamos la hipótesis nula (H_0).

Conclusión

No hay evidencia suficiente al nivel de significancia del 5 % para rechazar que la varianza poblacional sea igual a 5.3. Por lo tanto, podemos aceptar que:

$$\sigma^2 = 5.3$$

4.- En un experimento sobre dos muestras de betatesters se les hace jugar *hasta que no puedan aguantar más* a dos versiones distintas del juego para medirles el umbral de soporte. Los datos obtenidos (en tiempo de aguante) están en la siguiente tabla:

	Versión 1	Versión 2
n	14	10
$\bar{x}$	16.2	14.9
s^2	12.7	26.4

a) ¿Muestran los datos que hay evidencia de que hay diferencias en el tiempo que se soporta el juego?

b) Si debes realizar uno o varios contrastes, utiliza el *p-valor* en alguno de ellos.

Solución

1. Contrastar si las varianzas son iguales

Para comparar las varianzas de dos poblaciones, se realiza una prueba F, donde la hipótesis nula establece que las varianzas son iguales:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

El estadístico de prueba es:

$$F = \frac{s_1^2}{s_2^2}$$

Donde:

$$s_1^2 = 26.4, \quad s_2^2 = 12.7$$

Sustituyendo los valores:

$$F = \frac{26.4}{12.7} \approx 2.079$$

Los grados de libertad son:

$$df_1 = n_1 - 1 = 10 - 1 = 9, \quad df_2 = n_2 - 1 = 14 - 1 = 13$$

Para un nivel de significación $\alpha = 0.05$, consultamos los valores críticos de $F(9, 13)$ en tablas:

$$F_{\alpha/2,9,13} = 3.68, \quad F_{1-\alpha/2,9,13} = \frac{1}{3.33} \approx 0.30$$

El valor calculado $F = 2.079$ se encuentra entre los valores críticos, lo que indica que el p-valor es mayor a 0.05.

Conclusión: No se rechaza H_0. Las varianzas pueden considerarse iguales.

2. Contrastar si hay diferencias en los tiempos de soporte

Ya que no se rechazó la igualdad de varianzas, realizamos una prueba t de Student para muestras independientes con varianzas iguales. Las hipótesis son:

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2$$

El estadístico de prueba es:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde s_p^2 es la varianza combinada:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Sustituyendo los valores:

$$s_p^2 = \frac{(14 - 1)(12.7) + (10 - 1)(26.4)}{14 + 10 - 2} = \frac{165.1 + 237.6}{22} \approx 18.28$$

$$s_p = \sqrt{18.28} \approx 4.28$$

El estadístico t es:

$$t = \frac{16.2 - 14.9}{4.28\sqrt{\frac{1}{14} + \frac{1}{10}}} = \frac{1.3}{4.28 \cdot 0.446} \approx \frac{1.3}{1.91} \approx 0.681$$

Los grados de libertad son:

$$df = n_1 + n_2 - 2 = 14 + 10 - 2 = 22$$

Para $t = 0.681$ con $df = 22$, el p-valor es mayor a 0.05.

Conclusión: No se rechaza H_0. No hay evidencia significativa para afirmar que existen diferencias en los tiempos de soporte entre las dos versiones del juego.

3. Uso del p-valor

El p-valor se utilizó para ambos contrastes:

- En el contraste de varianzas, el p-valor fue mayor a 0.05, por lo que no se rechazó la igualdad de varianzas.
- En el contraste de medias, el p-valor también fue mayor a 0.05, indicando que no hay evidencia suficiente para rechazar la igualdad de medias.

Resumen:

- Las varianzas pueden considerarse iguales.
- No hay evidencia significativa para afirmar que hay diferencias en los tiempos de soporte entre las dos versiones del juego.

5.- A la hora de plantear la resolución de problemas a un LLM, se han detectado los siguientes porcentajes de errores en función de si se planteaba a GPT4, Gemini, Claude v1, Cohere, Copilot:

	GPT4	*Gemini*	*Claude*	*Cohere*	*Copilot*
	24	33	24	50	32
	37	20	40	20	62
	22	28	63	30	40
	55	12	18	13	15
	23	17	62	42	26
	38	17	30	28	37
	46	57	38	17	52
	25	42	23	73	12
	25	25	37	25	16
	23	63	26	22	25
$\bar{X}_i$	31.27	31.73	35.09	33.73	31.55
$\bar{S}_i$	11,01	16.56	14.86	17.42	15.55

La media global es $\bar{X} = 32,67$ y la varianza global es $S = 14,744$. Hemos comprobado también que las varianzas son homogéneas. ¿Podemos afirmar si hay diferencias entre los modelos planteados? ¿Por qué?

Solución

- Hipótesis nula (H_0): $\mu_{\text{GPT4}} = \mu_{\text{Gemini}} = \mu_{\text{Claude}} = \mu_{\text{Cohere}} = \mu_{\text{Copilot}}$. No hay diferencias en las medias de los modelos.
- Hipótesis alternativa (H_1): Al menos una de las medias es diferente.

Cálculo del estadístico F

- Suma de cuadrados entre grupos (SS_B):

$$SS_B = n_i \sum_{i=1}^{k} (\bar{X}_i - \bar{X})^2 \approx 110.78$$

$$(10 \cdot [(31.27 - 32.67)^2 + (31.73 - 32.67)^2 + (35.09 - 32.67)^2 + (33.73 - 32.67)^2 + (31.55 - 32.67)^2])$$

- Suma de cuadrados dentro de los grupos (SS_W):

$$SS_W = \sum_{i=1}^{k} (n_i - 1) \cdot S_i^2$$

Sustituyendo los valores:

$$SS_W = 9 \cdot [(11.01)^2 + (16.56)^2 + (14.86)^2 + (17.42)^2 + (15.55)^2] = 10453.79$$

- Grados de libertad:

$$df_B = k - 1 = 4, \quad df_W = N - k = 50 - 5 = 45$$

- Media cuadrática:

$$MS_B = \frac{SS_B}{df_B} = \frac{110.78}{4} = 27.70$$

$$MS_W = \frac{SS_W}{df_W} = \frac{10453.79}{45} = 232.31$$

- Estadístico F:

$$F = \frac{MS_B}{MS_W} = \frac{27.70}{232.31} = 0.12$$

Valores críticos y p-valor

- Valor crítico ($F_{\text{crítico}}$):

$$F_{\text{crítico}} = F_{0.95,4,45} = 2.58$$

  (Aproximadamente un valor intermedio entre 2.53 y 2.61 que son los valores en la tabla. No necesitaríamos entrar en mucha precisión en este caso.)

- p-valor:

$$p = 0.975$$

Decisión

Como $F = 0.12 \leq F_{\text{crítico}} = 2.58$ y $p = 0.975 > 0.05$, no rechazamos la hipótesis nula (H_0), por lo que NO hay evidencia suficiente al nivel de significancia del 5 % para afirmar que existen diferencias significativas entre los modelos en términos de porcentaje de error. Esto indicaría que los modelos presentan un comportamiento similar en este aspecto.

2 Puntos por ejercicio.