Prueba Bonus SOL114

Instrucciones: Responde a las preguntas proporcionadas, asegurándote de justificar tus respuestas con cálculos y razonamientos claros.

1. Distribución Normal y Rango de Probabilidad (10%):

   Dada una variable X que sigue una distribución normal estándar (media 0 y desviación estándar 1), identifica dos valores, x_1 y x_2, para los cuales $P(x_1 < X < x_2) = 0.5$.

La CDF de una distribución normal estándar se denota como $\Phi(z)$, donde z es un valor z estándar. Queremos encontrar x_1 y x_2 tal que $P(x_1 < X < x_2) = 0.5$, lo que significa que queremos encontrar los valores de x_1 y x_2 en términos de valores z estándar z_1 y z_2.

La CDF nos da la probabilidad de que X sea menor o igual a un valor dado z:

$$\Phi(z) = P(X \leq z)$$

Entonces, podemos escribir la probabilidad que estamos buscando como:

$$P(x_1 < X < x_2) = P(X < x_2) - P(X < x_1) = \Phi(z_2) - \Phi(z_1)$$

Dado que queremos que esta probabilidad sea igual a 0.5, tenemos:

$$\Phi(z_2) - \Phi(z_1) = 0.5$$

Ahora, podemos buscar en las tablas de valores z estándar (o utilizar una calculadora o software estadístico) para encontrar z_1 y z_2 que satisfagan esta ecuación. Estos valores corresponderán a los percentiles en la distribución normal estándar.

En una tabla de valores z estándar, encontrarás que $z_1 \approx -0.6745$ y $z_2 \approx 0.6745$ para que $\Phi(z_2) - \Phi(z_1) = 0.5$.

Finalmente, para encontrar los valores correspondientes en la escala original, puedes usar la media y la desviación estándar de la distribución normal estándar (media 0 y desviación estándar 1):

$$x_1 = \mu + z_1 \cdot \sigma = 0 + (-0.6745) \cdot 1 = -0.6745$$

$$x_2 = \mu + z_2 \cdot \sigma = 0 + 0.6745 \cdot 1 = 0.6745$$

Entonces, los valores x_1 y x_2 que satisfacen $P(x_1 < X < x_2) = 0.5$ para una distribución normal estándar son $x_1 \approx -0.6745$ y $x_2 \approx 0.6745$.

2. Evaluando la Probabilidad de Aprobación (20%):

   Considera una variable Y que representa aprobar (1) o reprobar (0) un test, con la condición de aprobación establecida en un puntaje de 2 o más ($X \geq 2$).

- (5%) Calcula la probabilidad de aprobar el test, $P(Y = 1)$.
- (10%) Describe la distribución de Y, incluyendo su función de masa de probabilidad.
- (5%) Determina y explica el valor esperado $E(Y)$ y la varianza $Var(Y)$ de Y.

Probabilidad de aprobar el test, $P(Y = 1)$ (5%):

- Dado que Y es una variable que representa aprobar (1) o reprobar (0) un test, y la condición de aprobación es que $X \geq 2$, Y sigue una distribución de Bernoulli con parámetro de éxito p.
- Entonces, $P(Y = 1) = p$, donde p es la probabilidad de aprobar el test.
- Para encontrar p, podemos usar la probabilidad acumulada de la distribución normal estándar, ya que Y está relacionado con X de la siguiente manera:
- $P(Y = 1) = P(X \geq 2) = 1 - P(X < 2) = 1 - \Phi(2)$.
- Usando tablas de valores z o una calculadora, encontramos que $\Phi(2) \approx 0.9772$. Luego, calculamos:
- $P(Y = 1) = 1 - 0.9772 = 0.0228$.
- Por lo tanto, la probabilidad de aprobar el test $P(Y = 1)$ es aproximadamente 0.0228 o 2.28%.

Descripción de la distribución de Y y su función de masa de probabilidad (10%):

- La variable Y sigue una distribución de Bernoulli, que es una distribución discreta con dos posibles resultados: 0 (reprobar) o 1 (aprobar).
- Su función de masa de probabilidad (PMF) está dada por:

- $P(Y = y) = p^y(1-p)^{1-y}$, donde y puede ser 0 o 1, y p es la probabilidad de éxito (en este caso, la probabilidad de aprobar el test).
- Entonces, $P(Y = 0) = (1 - p)$ y $P(Y = 1) = p$.

Valor esperado $E(Y)$ y varianza $Var(Y)$ de Y (5%):

- El valor esperado (media) de Y se calcula como:
- $E(Y) = \sum_y y \cdot P(Y = y)$.
- Para la distribución de Bernoulli, $E(Y) = p$.
- La varianza de Y se calcula como:
- $Var(Y) = E(Y) \cdot (1 - E(Y))$.
- Para la distribución de Bernoulli, $Var(Y) = p \cdot (1 - p)$.
- Usando el valor calculado de p, tenemos $E(Y) = 0.0228$ y $Var(Y) = 0.0223$ (aproximadamente).

3. Análisis de Puntajes de Estudiantes (15%):

Cuentas con los puntajes de 25 estudiantes distribuidos en dos grupos, A y B. En el grupo A hay 16 estudiantes y en el grupo B hay 9 estudiantes. Los puntajes se muestran en la siguiente tabla.

Estudiante	Grupo	Puntaje (X)
A1	A	1.76
A2	A	0.40
A3	A	0.98
A4	A	2.24
A5	A	1.87
A6	A	-0.98
A7	A	0.95
A8	A	-0.15
A9	A	-0.10
A10	A	0.41
A11	A	0.14
A12	A	1.45
A13	A	0.76
A14	A	0.12
A15	A	0.44
A16	A	0.33
B1	B	1.49
B2	B	-0.21
B3	B	0.31
B4	B	-0.85
B5	B	-2.55
B6	B	0.65
B7	B	0.86
B8	B	-0.74
B9	B	2.27

Sin conocimiento previo sobre la distribución poblacional, se te pide:

- (5%) Calcular la proporción de estudiantes que aprueban el test en cada grupo.

- (10%) Determinar la distribución muestral de las proporciones de aprobación.

La proporción de estudiantes que aprueban el test en cada grupo es la siguiente:

- Grupo A: 0.0625 o 6.25%, con 1 estudiante aprobado de 16.
- Grupo B: 0.1111 o 11.11%, con 1 estudiante aprobado de 9.

Dado que las muestras son pequeñas (especialmente en el Grupo B), no deberiamos aplicar directamente el Teorema del Límite Central para asumir que la distribución de estas proporciones es normal.

Sin emargo, si aplicamos el Teorema del Límite Central (TLC) a las muestras pequeñas, obtendriamos lo siguiente: asumimos que las distribuciones de las proporciones son aproximadamente normales, con los siguientes parámetros:

1. Media de la Distribución Muestral de la Proporción ($\hat{p}$):
   - La media $\mu_{\hat{p}}$ es igual a la proporción de la muestra.
2. Desviación Estándar de la Distribución Muestral de la Proporción ($\sigma_{\hat{p}}$):
   - La fórmula es $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, donde p es la proporción de la muestra y n es el tamaño de la muestra.

Vamos a realizar estos cálculos para ambos grupos, recordando que las medias de las distribuciones muestrales son las proporciones que ya calculamos: 0.0625 para el Grupo A y 0.1111 para el Grupo B. Ahora, calcularemos las desviaciones estándar.

Bajo la hipótesis de aplicar el Teorema del Límite Central a estas muestras pequeñas, obtenemos los siguientes resultados:

- Grupo A: La desviación estándar de la distribución muestral de la proporción es aproximadamente 0.0605.
- Grupo B: La desviación estándar de la distribución muestral de la proporción es aproximadamente 0.1048.

Estos valores representan la variabilidad de las proporciones de aprobados que podríamos esperar si tomáramos muchas muestras del mismo tamaño de las respectivas poblaciones.Es importante recordar que su validez es limitada debido al pequeño tamaño de las muestras, especialmente en el Grupo B. En la práctica, estos resultados deberían interpretarse con cautela.

4. Comparativa de Desempeño entre Grupos (45%):

   Utiliza los datos para examinar si hay diferencias en el rendimiento de los grupos A y B en el test:

   - (15%) Calcula y compara los intervalos de confianza al 95% para las proporciones de aprobación de cada grupo.
   - (15%) Establece un intervalo de confianza al 95% para la diferencia entre estas proporciones.
   - (15%) Realiza un test de hipótesis para la diferencia de proporciones con un nivel de significancia del 5%.

Intervalos de Confianza al 95% para las Proporciones de Aprobación de Cada Grupo

Grupo A

La proporción de aprobación en el Grupo A es $\hat{p}_A = 0.0625$ con un tamaño de muestra $n_A = 16$. Calculamos el intervalo de confianza al 95% usando la fórmula:

$$\text{IC}_A = 0.0625 \pm 1.96 \times \sqrt{\frac{0.0625 \times (1 - 0.0625)}{16}}$$

Este cálculo nos da un intervalo de confianza al 95% para el Grupo A de $(-0.0561, 0.1811)$.

Grupo B

Similarmente, para el Grupo B con $\hat{p}_B = 0.1111$ y $n_B = 9$, el intervalo de confianza se calcula como:

$$\text{IC}_B = 0.1111 \pm 1.96 \times \sqrt{\frac{0.1111 \times (1 - 0.1111)}{9}}$$

Lo que resulta en un intervalo de $(-0.0942, 0.3164)$ para el Grupo B.

Intervalo de Confianza al 95% para la Diferencia entre Estas Proporciones

Para calcular la diferencia entre las proporciones de aprobación, utilizamos la siguiente fórmula:

$$\mathtt{IC}_{\mathtt{dif}} = (\hat{p}_A - \hat{p}_B) \pm 1.96 \times \sqrt{\frac{\hat{p}_A \times (1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B \times (1-\hat{p}_B)}{n_B}}$$

Sustituyendo los valores, obtenemos un intervalo de confianza para la diferencia de $(-0.2857, 0.1885)$, que incluye el cero.

Test de Hipótesis para la Diferencia de Proporciones

Finalmente, para el test de hipótesis, el estadístico Z se calcula como:

$$Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\hat{p}_A \times (1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B \times (1-\hat{p}_B)}{n_B}}}$$

Este cálculo nos da un valor Z de -0.4018. Dado que este valor está dentro del rango de -1.96 y 1.96, no rechazamos la hipótesis nula y concluimos que no hay evidencia suficiente para afirmar una diferencia significativa en las proporciones de aprobación entre ambos grupos con un nivel de significancia del 5%.

En resumen, estos análisis sugieren que no hay diferencias significativas en el rendimiento de los grupos A y B en el test, según los datos proporcionados.

5. Conclusiones Basadas en los Datos (10%):

   En base a tus hallazgos, redacta conclusiones sustantivas sobre las similitudes o diferencias en el rendimiento de los grupos A y B en el test.

Teniendo en cuenta las limitaciones significativas de los datos disponibles, nuestras conclusiones sobre las similitudes o diferencias en el rendimiento de los grupos A y B en el test deben ser abordadas con cautela:

- Limitaciones Debido al Tamaño de la Muestra: Dado que tanto el grupo A como el B tienen muestras pequeñas (16 y 9 estudiantes respectivamente), nuestras estimaciones estadísticas pueden no ser muy confiables. El pequeño tamaño de la muestra reduce la capacidad de detectar una diferencia real en el rendimiento entre los grupos, si es que existe. Esto significa que incluso si hubiera una diferencia significativa en el rendimiento, es posible que nuestros análisis no tengan el poder estadístico suficiente para identificarla.

- Proporciones Muy Cercanas a Cero: Las proporciones de aprobación en ambos grupos son muy bajas (6.25% en el grupo A y 11.11% en el grupo B), lo que plantea un desafío adicional para el análisis estadístico. Las proporciones cercanas a cero pueden conducir a estimaciones inestables y a intervalos de confianza que no son muy informativos, como se observó en los rangos negativos incluidos en los intervalos de confianza calculados.

- Aplicación del Teorema del Límite Central (TLC): Normalmente, el TLC permite asumir una distribución aproximadamente normal para grandes muestras, pero en este caso, el pequeño tamaño de las muestras y las proporciones cercanas a cero hacen que esta aproximación sea menos válida. Esto limita la capacidad de usar métodos estadísticos que dependen de esta normalidad, como los intervalos de confianza y las pruebas de hipótesis que hemos realizado.

- Resultados de las Pruebas de Hipótesis: Aunque la prueba de hipótesis para la diferencia de proporciones sugiere que no hay diferencia significativa entre los grupos, esta conclusión debe ser considerada con precaución. La falta de poder estadístico debido al pequeño tamaño de la muestra y las bajas proporciones implica que podríamos estar simplemente incapaces de detectar diferencias que sí existen.

En resumen, debido a las limitaciones significativas en términos del tamaño de la muestra y las bajas proporciones de aprobación, debemos ser extremadamente cautelosos al sacar conclusiones definitivas sobre las similitudes o diferencias en el rendimiento de los grupos A y B en el test. Es posible que se requiera una investigación adicional con muestras más grandes para obtener conclusiones más robustas.