Análisis e Interpretación de Datos

Dra. Mariana-Edith Miranda-Varela

21-julio-2025

Estadística Inferencial: del análisis muestral a la predicción poblacional

UNIR LA UNIVERSIDAD EN INTERNET

Problema cerrado

Es un tipo de problemas bien definidos y específicos, es decir, no tienen ambigüedad o incertidumbre. Además tiene una solución única correcta, la cual es verificable.

Ejemplos:

- Eficiencia de un nuevo medicamento
- Diferencia en el rendimiento académico
- Preferencia de un producto

Contraste de hipótesis

Procedimiento formal estadístico para decidir si una afirmación acerca de una población es verdadera o no a partir de los datos.

- **Hipótesis nula** (H_0) es una afirmación que se va a probar si es aceptada o rechazada. Esta hipótesis representa lo conocido o lo que ya está establecido.

- **Hipótesis alternativa** (H_1) expresa lo novedoso o lo que contradice a lo establecido o un punto de vista conservador.

Ejemplos de hipótesis

- Eficiencia de un nuevo medicamento (XX)
  Efectos secundarios

  H_0: El nuevo medicamento XX *no presenta más efectos* secundarios en los pacientes que un tratamiento estándar.
  H_1: El nuevo medicamento XX presenta más efectos secundarios en los pacientes que un tratamiento estándar.

  La prueba para las hipótesis anteriores es de una cola o unilateral, ya que se quiere validar si el nuevo medicamento produce más efectos secundarios que los conocidos.

$$H_1\colon\ \theta < \theta_0 \text{ o } H_1\colon\ \theta > \theta_0$$

Ejemplos de hipótesis

- Eficiencia de un nuevo medicamento (XX)

  Efectos secundarios

  H_0: El nuevo medicamento XX *tiene la misma incidencia* de efectos secundarios en los pacientes, que un tratamiento estándar.

  H_1: El nuevo medicamento XX presenta una incidencia diferente en los efectos secundarios en los pacientes, que un tratamiento estándar.

  La prueba para las hipótesis anteriores es de dos colas o bilateral, ya que se quiere validar si el nuevo medicamento produce una taza de efectos secundarios diferente a los conocidos, esto es menos o más efectos.

$$H_1:\ \theta \neq \theta_0$$

Prueba de una hipótesis estadística

1. Recolección de datos muestrales
2. Calcular el valor de una estadística de prueba
   - Región de no rechazo
   - Región de rechazo

- Pruebas unilaterales

  La hipótesis alternativa indica un cambio en una dirección, > o <, con respecto a la hipótesis nula.

- Pruebas bilaterales

  La hipótesis alternativa no indica una dirección para el cambio.

Nivel de significación

- La significancia es fundamental al momento de contrastar hipótesis.
- Se presenta cuando los estadísticos que se emplean toman valores a partir de los cuales se rechaza H_0.
- Se representa por α e indica la máxima probabilidad de cometer un error tipo I, esto es, rechazar H_0 siendo verdadera.
- Generalmente son 0.01, 0.05 o 0.1

Región de aceptación y de rechazo

FUENTE: https://repositorio-uapa.cuaieed.unam.mx/repositorio/moodle/pluginfile.php/1419/mod_resource/content/1/contenido/index.html

Errores de tipo I y de tipo II

Decisión	Estado real	
	H_0 es verdadera	H_0 es falsa
Rechazar H_0	**Error de tipo I** α	Decisión correcta $1-\beta$
No rechazar H_0	Decisión correcta $1-\alpha$	**Error de tipo II** β

- Reducir al máximo las probabilidades de cometer errores de tipo I y II
- La probabilidad de cometer estos errores es inversamente proporcional.

$$P[\text{rechazar } H_0 | H_0 \text{ es verdadera}] = \alpha$$
$$P[\text{no rechazar } H_0 | H_0 \text{ es falsa}] = \beta$$

Calcular la estadística de prueba

- Muestras aleatorias de poblaciones con distribución de probabilidad normal $N(\mu, \sigma)$.

- Se conoce la media poblacional, μ, mientras que la varianza poblacional σ se puede o no conocer.

- A partir de la naturaleza de los datos e hipótesis, determinar el tipo de prueba

  - Datos categóricos
    - Prueba de chi-cuadrado
    - Prueba binomial

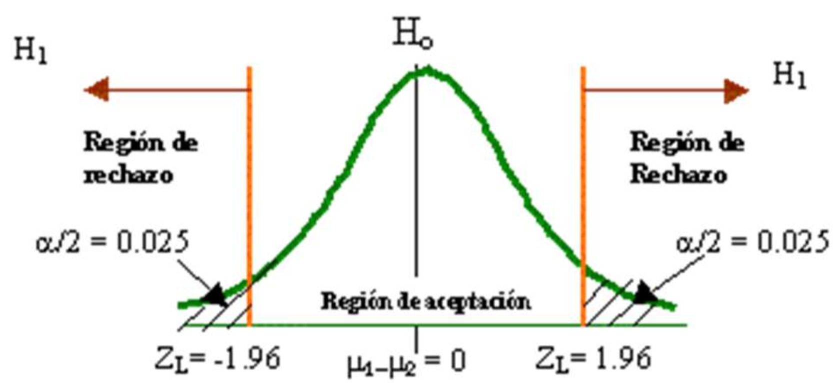
Calcular la estadística de prueba

- Datos numéricos
  - Prueba t de Student
    - Una muestra
    - Dos muestras independientes
    - Dos muestras pareadas
  - Prueba z
  - ANOVA (Análisis de Varianza)
  - ANOVA de medidas repetidas
- Datos de correlación y regresión
  - Correlación de Pearson
  - Regresión lineal simple
  - Regresión lineal múltiple
  - Regresión logística

Prueba z

- Muestra de tamaño n, con parámetros μ y σ conocida
- Se emplea el estadístico

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$



https://virtual.uptc.edu.co/ova/estadistica/docs/libros/estadistica1/cap02c.html

Prueba de chi-cuadrado ($\boldsymbol{X^2}$)

- Se conoce la varianza de una población con distribución N(μ, σ) y se desconoce μ, en un nivel de confianza 1- α.

$$X_{n-1^2} = \frac{(n-1)s^2}{\sigma^2}$$

- Compara la proporción de éxitos observada en una muestra con una proporción esperada bajo una hipótesis nula.

- Ejemplo:
  Verificar si una proporción de pacientes que mejoran con un tratamiento es igual a la proporción esperada del 50%

Prueba binomial

- Compara proporciones entre 2 o más grupos para probar la independencia entre dos variables categóricas

- Ejemplo:
  Determinar si existe una relación entre el género y la preferencia por un producto

Prueba t de Student

- Una muestra
  - Compara su media con un valor conocido

- Dos muestras independientes
  - Compara las medias de dos grupos independientes
  - Estudio de dos tratamientos diferentes

- Muestras pareadas
  - Compara las medias de dos conjuntos emparejados.
  - Evaluar la eficacia de un tratamiento (antes y después)

Valor crítico o *p-value*

- Región crítica
  - Identificar el valor crítico asociado al nivel de significancia
    - Tablas estadísticas
    - Software
  - Definir la región de rechazo en la distribución de la estadística de prueba.

- *P-value*
  - Probabilidad de obtener un valor del estadístico de prueba que sea al menos tan extremo como el obtenido a partir de los datos muestrales.

Ejemplo

Un jefe de seguridad afirma que el estacionamiento es usado por personas que no son clientes, en promedio por más de 80 minutos con varianza de 20 minutos^2.

Si se toma una muestra de 25 vehículos que se encontraban en el estacionamiento y que pertenecen a personas que no son clientes, y se encuentra un promedio de 78 minutos, esta muestra sustenta lo que el jefe de seguridad afirma con $\alpha = 0.05$.

Ejemplo

Solución

1. Formulación de las hipótesis:
   $H_o: \mu \leq 80$ - el tiempo promedio de los no clientes no es mayor a 80 min
   $H_1: \mu > 80$ - el tiempo promedio de los no clientes es mayor a 80 min

2. α = 0.05

3. La estadística de prueba es

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

y se supone que los datos tienen una distribución Normal.

Ejemplo

4. Región crítica

$$RC=(Z_{(1\text{-}0.05)}, \propto)=(1.645, \propto)\text{, rechazar } H_0 \text{ si } Z \in RC$$

5. Valor de la estadística de prueba

$$Z = \frac{78-80}{4.4721/5} = -2.236$$

6. Decisión y conclusión

Como $Z \notin RC$; no se rechaza la hipótesis nula; entonces, las personas que utilizan el estacionamiento y no son clientes en promedio presentan un tiempo de uso no mayor de 80 minutos

Avisos

Manual de uso de IA

Actividad 1

Fecha de entrega

4 de agosto 23:59

muchas gracias.

unir LA UNIVERSIDAD EN INTERNET