Pruebas de Hipótesis y su Aplicación Práctica - Día 2

Change Status

pending 40 min

Learning Objectives

1. Comprender el marco formal para validar hipótesis con datos estadísticos
2. Distinguir entre errores tipo I y tipo II en pruebas estadísticas
3. Aplicar pruebas t para comparación de medias en contextos empresariales

Theory | Practice | Quiz | Evidence

Activities and Learning

Task 1: Marco de Pruebas de Hipótesis (10 minutos)

Las pruebas de hipótesis proporcionan un método sistemático y riguroso para tomar decisiones basadas en datos, minimizando el riesgo de conclusiones erróneas.

Hipótesis Nula vs Hipótesis Alternativa

Hipótesis nula (H_0): La hipótesis de "no hay diferencia" o "no hay efecto".

Ejemplos:
"La tasa de conversión del sitio web nuevo es igual a la del sitio antiguo"
"No hay diferencia en satisfacción entre clientes nuevos y existentes"
"El precio promocional no afecta las ventas"

Hipótesis alternativa (H_1 o H_a): La hipótesis que queremos probar.

Ejemplos:
"La tasa de conversión del sitio web nuevo es mayor"
"Los clientes nuevos están menos satisfechos"

"El precio promocional aumenta las ventas"

Importante: Siempre se formula H_0 como la hipótesis conservadora (status quo).

Nivel de Significancia (α) y Valor p

Nivel de significancia (α): Probabilidad máxima de cometer error tipo I que estamos dispuestos a aceptar.

Valores comunes: 0.05 (5%), 0.01 (1%), 0.10 (10%)
Interpretación: Si $\alpha = 0.05$, aceptamos 5% de riesgo de concluir falsamente que hay diferencia

Valor p: Probabilidad de observar los datos (o más extremos) si H_0 es verdadera.

Interpretación:
$p < \alpha$: Rechazamos H_0 (evidencia suficiente contra la hipótesis nula)
$p \geq \alpha$: No rechazamos H_0 (no hay evidencia suficiente)

Error común: "$p = 0.06$ significa que hay 6% de probabilidad de que H_0 sea falsa" (INCORRECTO)

Task 2: Errores Tipo I y Tipo II (8 minutos)

Las pruebas estadísticas no son perfectas; siempre existe riesgo de error. Entender estos errores es crucial para interpretación correcta de resultados.

Error Tipo I (Falso Positivo)

Definición: Rechazar H_0 cuando H_0 es realmente verdadera.

Probabilidad: α (nivel de significancia)
Ejemplo: Concluir que un nuevo medicamento funciona cuando realmente no funciona
Consecuencias: Implementar cambios innecesarios, desperdiciar recursos

Factores que afectan α:

Tamaño de muestra: Muestras grandes reducen α para mismo efecto
Variabilidad: Más variabilidad aumenta α
Elección de α: Más conservador = menor α
Error Tipo II (Falso Negativo)

Definición: No rechazar H_0 cuando H_0 es realmente falsa.

Probabilidad: β
Ejemplo: No detectar que un nuevo medicamento funciona cuando realmente funciona

Consecuencias: Perder oportunidades, no implementar mejoras efectivas

Poder estadístico: $1 - \beta$ = Probabilidad de detectar efecto real cuando existe.

Matriz de Confusión Estadística

Realidad → Decisión ↓	**H_0 Verdadera**	**H_1 Verdadera**
No rechazar H_0	Decisión correcta (1-α)	Error Tipo II (β)
Rechazar H_0	Error Tipo I (α)	Decisión correcta (1-β)

Equilibrio α vs β:

Estudios exploratorios: α más laxo (0.10) para detectar señales débiles
Decisiones críticas: α muy conservador (0.01) para evitar falsos positivos
Poder: Aumentar tamaño de muestra reduce tanto α como β

Task 3: Pruebas t para Comparación de Medias (12 minutos)

Las pruebas t son las herramientas más comunes para comparar medias entre grupos, especialmente útiles cuando no conocemos la varianza poblacional.

Prueba t de Una Muestra

Compara la media de una muestra con un valor conocido o hipotético.

Fórmula: $t = (\bar{x} - \mu_0) / (s/\sqrt{n})$

$\bar{x}$: Media muestral
μ_0: Valor hipotético bajo H_0
s: Desviación estándar muestral
n: Tamaño de muestra

Aplicación práctica:

```python
from scipy import stats
import numpy as np

# Datos: Satisfacción de clientes después de cambio de servicio
satisfaccion = np.array([4.2, 3.8, 4.5, 4.1, 3.9, 4.3, 4.0, 4.4, 3.7, 4.6])

# Prueba: ¿Es la satisfacción significativamente mayor que 4.0?
t_stat, p_value = stats.ttest_1samp(satisfaccion, 4.0, alternative='greater')

print(f"Estadístico t: {t_stat:.3f}")
```

```python
print(f"Valor p: {p_value:.4f}")
print(f"Conclusión: {'Rechazamos H₀' if p_value < 0.05 else 'No rechazamos H₀'}")
```

Prueba t de Dos Muestras Independientes

Compara medias de dos grupos independientes.

Fórmula: $t = (\bar{x}_1 - \bar{x}_2) / \sqrt{s_1^2/n_1 + s_2^2/n_2}$

Asunción: Varianzas iguales (o usar corrección Welch si no)

Ejemplo empresarial: Comparar efectividad de dos estrategias de marketing.

```python
# Datos: Conversiones de dos campañas de marketing
campana_A = np.random.normal(0.12, 0.03, 100)  # 12% conversión, n=100
campana_B = np.random.normal(0.15, 0.04, 120)  # 15% conversión, n=120

# Prueba: ¿Es la campaña B significativamente mejor?
t_stat, p_value = stats.ttest_ind(campana_B, campana_A, alternative='greater')

print(f"Media campaña A: {campana_A.mean():.3f}")
print(f"Media campaña B: {campana_B.mean():.3f}")
print(f"Diferencia: {campana_B.mean() - campana_A.mean():.3f}")
print(f"Estadístico t: {t_stat:.3f}")
print(f"Valor p: {p_value:.4f}")
```

Prueba t Pareada (Before-After)

Compara medidas antes y después en las mismas unidades.

Ventajas: Controla variabilidad individual, más poder estadístico.

Ejemplo: Efecto de entrenamiento en productividad.

```python
# Datos pareados: Productividad antes y después de entrenamiento
antes = np.random.normal(85, 10, 50)
despues = antes + np.random.normal(5, 8, 50)  # Mejora promedio de 5 puntos

# Prueba pareada
t_stat, p_value = stats.ttest_rel(despues, antes, alternative='greater')

print(f"Productividad antes: {antes.mean():.1f}")
```

TalentOps
Dashboard
Career Path
Forms
Profile

```python
print(f"Productividad después: {despues.mean():.1f}")
print(f"Mejora promedio: {despues.mean() - antes.mean():.1f}")
print(f"Estadístico t: {t_stat:.3f}")
print(f"Valor p: {p_value:.4f}")
```

Consideraciones Prácticas

Tamaño del efecto: Más importante que significancia estadística.

Cohen's d: $d = (\mu_1 - \mu_2) / \sigma$ (pooled)
Interpretación: $d = 0.2$ (pequeño), 0.5 (medio), 0.8 (grande)

Tamaño de muestra: Afecta poder estadístico.

Fórmula aproximada: $n \approx 16 / d^2$ para detectar efecto de tamaño d

Asunciones de pruebas t:

Normalidad: Datos aproximadamente normales ($n \geq 30$ ayuda)
Independencia: Observaciones independientes
Homocedasticidad: Varianzas similares (para pruebas independientes)

Sign Out
Toggle theme: Light Theme