Diferencias Estadísticamente Significativas en Comparaciones A/B

Fundamentos teóricos, formulación matemática e interpretación gráfica

---

1. Introducción

En investigación de mercados, marketing digital y experimentación A/B, una de las preguntas más frecuentes es determinar si la diferencia observada entre dos grupos es real o simplemente producto del azar.

Una diferencia se considera **estadísticamente significativa** cuando, bajo el supuesto de que no existe diferencia real entre los grupos (hipótesis nula), la probabilidad de observar una diferencia igual o mayor a la obtenida es suficientemente baja.

Formalmente, esto se evalúa mediante pruebas de hipótesis.

---

2. Marco Conceptual

2.1 Hipótesis

Sea A y B dos grupos comparables.

- **Hipótesis nula (H_0):** No existe diferencia real entre los grupos.
- **Hipótesis alternativa (H_1):** Existe diferencia entre los grupos.

Se define un nivel de significancia α (comúnmente 0.05).

Regla de decisión:

$$\text{Si } p\text{-value} < \alpha \Rightarrow \text{Rechazar } H_0$$
$$\text{Si } p\text{-value} \geq \alpha \Rightarrow \text{No rechazar } H_0$$

---

3. Comparación de Proporciones (z-test)

3.1 Contexto

Se utiliza cuando se comparan porcentajes o tasas de éxito.

Ejemplo:
Grupo A: 40 de 100 personas compran.
Grupo B: 55 de 120 personas compran.

3.2 Formulación Matemática

Estimadores:

$$\hat{p}_A = \frac{x_A}{n_A}$$
$$\hat{p}_B = \frac{x_B}{n_B}$$

Proporción combinada (pooled):

$$\hat{p} = \frac{x_A + x_B}{n_A + n_B}$$

Error estándar:

$$SE = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

Estadístico de prueba:

$$z = \frac{\hat{p}_A - \hat{p}_B}{SE}$$

Cálculo del p-value bilateral:

$$p\text{-value} = 2(1 - \Phi(|\, z \,|))$$

donde Φ es la función de distribución acumulada de la normal estándar.

4. Representación Gráfica del z-test

El gráfico representa una **distribución normal estándar**:

$$Z \sim N(0,1)$$

Sus componentes son:

1. Curva azul:
   Distribución normal estándar centrada en 0.

2. Líneas rojas punteadas (±1.96):
   Valores críticos para un contraste bilateral con:

$$\alpha = 0.05$$
$$z_{\alpha/2} = \pm 1.96$$

Estas delimitan la **zona crítica**.

3. Línea negra vertical:
   Estadístico observado:

$$z_{obs} = -0.8698$$

4. Área sombreada naranja:
   Representa el **p-value bilateral**, es decir:

$$p\text{-value} = 2(1 - \Phi(|\, z_{obs}\,|))$$

En el gráfico:

$$p\text{-value} = 0.38443$$

4.2 Interpretación matemática del caso mostrado

Se compara el estadístico observado con el valor crítico:

$$|z_{obs}| = 0.8698$$
$$z_{crit} = 1.96$$

Como:

$$|z_{obs}| < z_{crit}$$

El estadístico cae dentro de la región central (zona de no rechazo).

Adicionalmente:

$$p\text{-value} = 0.38443 > 0.05$$

Por lo tanto:

$$\text{No se rechaza } H_0$$

4.3 Interpretación conceptual

El gráfico muestra que:

- La diferencia observada está relativamente cerca del centro.
- No se encuentra en las colas extremas.
- La probabilidad de observar una diferencia igual o mayor bajo la hipótesis nula es del 38.4%.

Esto implica que la diferencia podría explicarse razonablemente por variabilidad muestral.

4.4 Significado estadístico

Bajo la hipótesis nula:

$$H_0: p_A = p_B$$

La diferencia estandarizada fue:

$$z = \frac{\hat{p}_A - \hat{p}_B}{SE}$$

Dado que el valor no cae en la región crítica, no existe evidencia estadística suficiente para afirmar que las proporciones son diferentes al nivel del 5%.

Importante:

No rechazar H_0 no implica que las proporciones sean iguales, sino que con esta muestra no existe evidencia suficiente para demostrar diferencia.

4.5 Interpretación aplicada en marketing

En un contexto de prueba A/B:

- Si A tiene 40% y B 45%,
- Y obtenemos z = -0.8698 con p = 0.38,

La conclusión sería:

No hay evidencia estadística para afirmar que B convierte mejor que A.

Desde una perspectiva de negocio:

- No se justifica un cambio estratégico basado únicamente en esta diferencia.
- Puede requerirse mayor tamaño de muestra.
- Puede analizarse tamaño de efecto y potencia estadística.

4.6 Conclusión técnica del gráfico

El gráfico ilustra visualmente tres conceptos fundamentales:

1. Región crítica definida por α.
2. Posición del estadístico observado.
3. Área asociada al p-value.

Este tipo de visualización permite comprender que la significancia estadística es una comparación entre la magnitud de la diferencia observada y la variabilidad esperada bajo la hipótesis nula.

---

5. Comparación de Medias (t-test de Welch)

5.1 Contexto

Se utiliza cuando se comparan promedios de variables métricas.

Ejemplo:
Grupo A: media = 10.5, desviación estándar = 2.1
Grupo B: media = 9.7, desviación estándar = 2.4

5.2 Formulación Matemática

Estadístico t:

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

Grados de libertad aproximados (Welch):

$$\nu = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{\left(\frac{s_A^2}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B - 1}}$$

Cálculo del p-value:

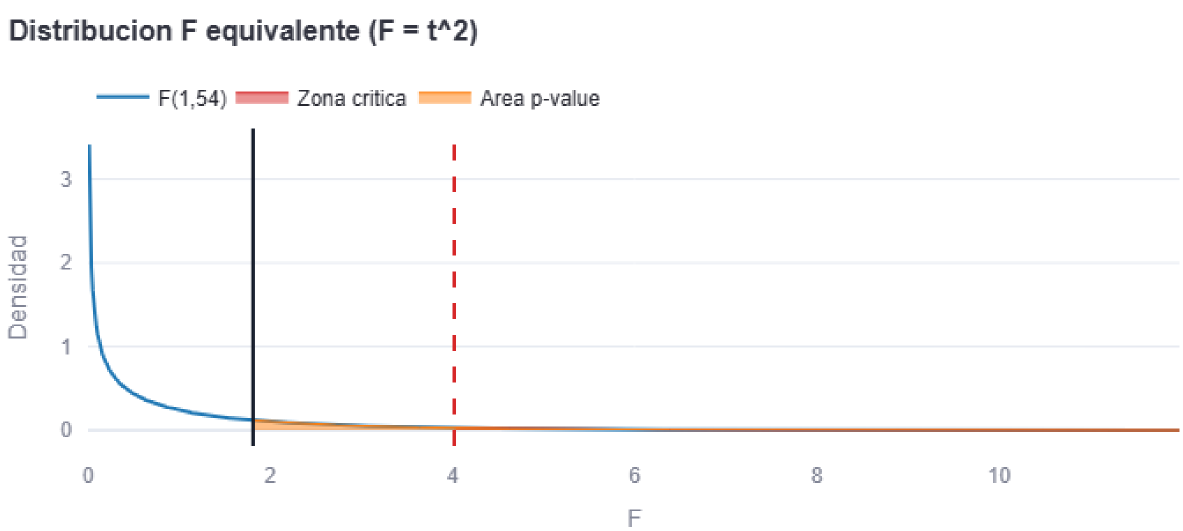
$$p\text{-value} = 2\left(1 - F_{t,\nu}(|\ t\ |)\right)$$

donde $F_{t,\nu}$es la distribución acumulada t con v grados de libertad.

---

6. Representación Gráfica del t-test y su Equivalencia con la Distribución F

A continuación se presenta la interpretación académica de los gráficos correspondientes al contraste de medias mediante t-test de Welch y su equivalencia con la distribución F.



6.1 Fundamento Teórico del t-test

El contraste de medias con varianzas posiblemente distintas se formula como:

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

donde:

- $\bar{x}_A, \bar{x}_B$= medias muestrales
- s_A, s_B= desviaciones estándar
- n_A, n_B= tamaños de muestra

Los grados de libertad aproximados (Welch) son:

$$\nu = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{\left(\frac{s_A^2}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B - 1}}$$

El p-value bilateral se calcula como:

$$p\text{-value} = 2\left(1 - F_{t,\nu}(\mid t \mid)\right)$$

6.2 Interpretación del Primer Gráfico: Distribución t

En el gráfico de la distribución t se observan:

- Curva azul: distribución t con $\nu = 54$grados de libertad.
- Líneas rojas punteadas: valores críticos $\pm t_{\alpha/2}$.
- Línea negra vertical: estadístico observado t_{obs}.
- Área naranja: región correspondiente al p-value.

Regla visual de decisión

Si:

$$\mid t_{obs} \mid > t_{crit}$$

entonces el estadístico cae en la zona crítica y se rechaza H_0.

En el gráfico mostrado:

- t_{obs}se encuentra dentro de la región central.
- El área p-value es relativamente amplia.

Por lo tanto:

$$p\text{-value} > 0.05$$

No existe evidencia suficiente para afirmar diferencia significativa entre las medias.

6.3 Interpretación del Segundo Gráfico: Distribución F Equivalente

Una propiedad fundamental es:

$$F = t^2$$

Cuando el contraste t tiene 1 grado de libertad en el numerador, el contraste es equivalente a una prueba F:

$$F(1, \nu)$$

En el gráfico F:

- La curva es asimétrica (sesgada a la derecha).
- Solo existe región crítica en la cola derecha.
- La línea negra representa $F_{obs} = t_{obs}^2$.

Si:

$$F_{obs} > F_{crit}$$

se rechaza H_0.

En el caso mostrado:

$$F_{obs} < F_{crit}$$

Por lo tanto, el resultado es consistente con el obtenido mediante la prueba t: no se rechaza la hipótesis nula.

6.4 Relación Conceptual entre t y F

Para contrastes de dos medias:

$$t^2 = F$$

Esto implica:

- El t-test bilateral es matemáticamente equivalente a un F-test con 1 grado de libertad en el numerador.
- Ambos conducen al mismo p-value.
- La diferencia radica únicamente en la forma de la distribución.

6.5 Interpretación Académica Integral

Los gráficos permiten visualizar tres elementos esenciales:

1. La región crítica determinada por α.
2. La posición del estadístico observado.
3. El área asociada al p-value.

Desde una perspectiva estadística:

- El contraste estandariza la diferencia observada respecto a su variabilidad.
- La decisión depende de la probabilidad de observar dicho valor bajo H_0.

Desde una perspectiva aplicada:

- Si el estadístico no cae en zona crítica, no existe evidencia suficiente para afirmar diferencia.
- Esto no implica igualdad, sino insuficiencia de evidencia con los datos disponibles.

6.6 Conclusión Técnica

La representación gráfica del t-test y su equivalencia con la distribución F permiten comprender que:

- La significancia estadística es una comparación probabilística.
- El rechazo de la hipótesis nula ocurre únicamente cuando el estadístico observado se encuentra en regiones de baja probabilidad.
- La decisión es coherente independientemente de si se utiliza el marco t o F.

7. Intervalos de Confianza

Otra forma equivalente de evaluar significancia es mediante intervalos de confianza.

Para medias:

$$IC = \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Para proporciones:

$$IC = \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Si los intervalos de confianza de A y B no se superponen sustancialmente, es probable que exista diferencia significativa.

---

8. Interpretación Correcta

Es fundamental evitar errores comunes:

1. **Significativo ≠ importante.**
   Puede existir significancia estadística con efecto pequeño.
2. **No significativo ≠ igualdad.**
   Puede haber falta de potencia estadística.
3. El p-value no mide magnitud de efecto, sino evidencia contra H_0.

---

9. Aplicación en Marketing y Negocios

En experimentación A/B:

- Determina si una versión genera mayor conversión.
- Reduce riesgo en decisiones comerciales.
- Justifica decisiones frente a stakeholders.

La significancia estadística es un mecanismo de control de riesgo, no una garantía absoluta de éxito.

---

10. Conclusión

Las diferencias estadísticamente significativas se fundamentan en:

1. Comparar una diferencia observada contra su variabilidad esperada.
2. Estandarizar esa diferencia mediante un estadístico (z o t).
3. Evaluar la probabilidad asociada bajo la hipótesis nula.

La decisión depende del nivel de significancia α y del p-value calculado.

Desde una perspectiva académica, este marco permite separar variabilidad aleatoria de efectos reales. Desde una perspectiva comercial, permite tomar decisiones con respaldo cuantitativo y control de incertidumbre.