

# 2.3. Pruebas de Proporciones

2. Estadística y Probabilidad Básica

Christian E. Galarza

**Programa New Dimensions**

# Introducción a las Pruebas de Proporciones

- Las pruebas de proporciones se utilizan para comparar proporciones entre grupos.
- Pueden ser de una muestra, dos muestras o múltiples muestras.
- Las pruebas chi-cuadrado son comunes en este contexto.

# Prueba de una Proporción

- Permite evaluar si es razonable pensar que la proporción de real es igual, mayor, o menor a alguna cantidad propuesta.
- Pueden ser pruebas de una cola o dos.
- La idea, como siempre, es a través de una muestra mostrar si existe evidencia significativa que apunte a que la hipótesis nula es falsa.

# Contraste de Hipótesis para una Proporción

## Ejemplo 1: Una cola a la izquierda

Se ha afirmado que la proporción de estudiantes en una universidad que poseen su propio vehículo es 0.5. Una encuesta reciente, realizada a una muestra de 200 estudiantes, encontró que 90 de ellos poseen un vehículo. ¿Hay evidencia suficiente para afirmar que la proporción ha disminuido?

$$H_0 : p = 0,5$$

$$H_1 : p < 0,5$$

# Contraste de Hipótesis para una Proporción

## Ejemplo 2: Una cola a la derecha

Una compañía farmacéutica afirma que solo el 8 % de los usuarios experimentan efectos secundarios con su nuevo medicamento. Sin embargo, en una muestra de 300 pacientes, 35 reportaron haber experimentado efectos secundarios. ¿Hay evidencia suficiente para afirmar que la proporción es mayor que el 8 %?

$$H_0 : p = 0,08$$

$$H_1 : p > 0,08$$

# Contraste de Hipótesis para una Proporción: Dos Colas

Se afirma que la proporción de ciudadanos que están a favor de una nueva ley es del 60%. Una organización ha encuestado a 250 ciudadanos, y 135 expresaron su apoyo. ¿Hay evidencia suficiente para afirmar que la proporción ha cambiado?

$$H_0 : p = 0,6$$

$$H_1 : p \neq 0,6$$

# Cálculo de intervalos de confianza para proporciones

El intervalo de confianza para proporciones nos da un rango en el que es probable que se encuentre una proporción poblacional. Al igual que con las medias, nos proporciona una idea de la incertidumbre asociada con nuestras estimaciones de proporciones.

Es exactamente lo que se levanta en casos electorales cuando queremos estimar la proporción de intención de voto para un determinado candidato. **¿Han escuchado antes de que se dice “con un 95 % de confianza”?**

# Fórmula del intervalo de confianza para proporciones

Un intervalo de confianza del  $(1 - \alpha) \%$  de confianza para una proporción  $p$  está dado por:

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

donde  $\hat{p}$  es la proporción muestral y  $Z_{\alpha/2}$  es el valor crítico de una distribución normal (el cual dependerá del nivel de confianza).

**El intervalo de confianza arriba funciona para tamaños de muestra grandes ( $n > 30$ ). Esto es porque el mismo sigue del teorema de límite central.**



# Prueba de Dos Proporciones

- Compara las proporciones de dos grupos.
- La hipótesis nula  $H_0$ : las proporciones son iguales.
- La hipótesis alterna  $H_a$ : las proporciones son diferentes.

Puede especificarse que una es mayor o menor que la otra, teniéndose en ese caso una prueba de una sola cola.

# Notación Matemática para la Prueba de Dos Proporciones

- Supongamos que las proporciones de la población son  $p_1$  y  $p_2$ .
- Las proporciones de la muestra son  $\hat{p}_1$  y  $\hat{p}_2$ .
- $n_1$  y  $n_2$  son los tamaños de las dos muestras.

**Hipótesis nula y alterna:**

$$H_0 : p_1 = p_2$$

$$H_a : p_1 \neq p_2$$

# Notación Matemática para la Prueba de Dos Proporciones

**Estadístico de prueba (Z):**

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p_c(1 - p_c) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

donde  $p$  es la proporción combinada:

$$p_c = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{n_1}{n_1 + n_2} \cdot p_1 + \frac{n_2}{n_1 + n_2} \cdot p_2$$

# Ejemplo: Comparación de Proporciones

Una compañía quiere saber si la proporción de hombres que compra su producto es diferente a la proporción de mujeres que lo hace. Se toman dos muestras: de 300 hombres, 90 compraron el producto; de 250 mujeres, 65 compraron el producto. ¿Las proporciones de compradores son diferentes entre hombres y mujeres?

# Ejemplo: Comparación de Proporciones

- **Hipótesis:**

$$H_0 : p_m = p_h$$

$$H_a : p_m \neq p_h$$

$$H_0 : p_m - p_h = 0$$

$$H_a : p_m - p_h \neq 0$$

# Ejemplo: Comparación de Proporciones

- **Datos:**  $n_m = 250$ ,  $\hat{p}_m = 0,26$ ,  $n_h = 300$ ,  $\hat{p}_h = 0,3$
- **Estadístico de prueba (Z):**  $Z \approx -1,42$
- **Valor p:**  $p \approx 0,31$

# Ejemplo: Comparación de Proporciones

- **Decisión:** Como  $p = 0,31$ , no hay suficiente evidencia para rechazar  $H_0$ .

**No hay evidencia para afirmar que las proporciones de compradores son diferentes entre hombres y mujeres.**

# Ejemplo Práctico en R: Comparación de Proporciones de Votantes

Supongamos que queremos comparar la proporción de votantes en dos regiones diferentes.

- Región A: 500 votantes de 1000 elegibles (50 %).
- Región B: 300 votantes de 800 elegibles (37.5 %).

Utilizaremos una prueba de dos proporciones para determinar si hay una diferencia significativa en las proporciones de votantes entre las dos regiones.



# Código en R: Prueba de Dos Proporciones

```
# Datos
votantes <- c(500, 300)
elegibles <- c(1000, 800)

# Realizar la prueba de proporciones
resultado <- prop.test(votantes, elegibles)
print(resultado)
```

# Prueba de Independencia Chi cuadrada: Introducción

- Utilizada para probar la independencia entre dos variables categóricas.
- Se basa en una tabla de contingencia, la cual muestra la distribución conjunta de las categorías de las variables.
- Compara las frecuencias observadas con las frecuencias esperadas.

# Formulación Matemática

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Donde:

- $\chi^2$  es el estadístico Chi cuadrada.
- $O_{ij}$  es la frecuencia observada.
- $E_{ij}$  es la frecuencia esperada:

$$E_{ij} = \frac{\text{Total de fila } i \times \text{Total de columna } j}{\text{Total general}}$$

# Ejemplo 1

Un investigador quiere saber si la elección de una carrera (Medicina, Ingeniería, Arte) está relacionada con el género del estudiante. Se realiza una encuesta a 200 estudiantes y se obtiene la siguiente tabla de contingencia:

|         | Medicina | Ingeniería | Arte |
|---------|----------|------------|------|
| Hombres | 40       | 30         | 10   |
| Mujeres | 20       | 20         | 80   |

# Ejemplo 2

Una escuela desea conocer si el rendimiento académico (Alto, Medio, Bajo) está relacionado con la participación en actividades extraescolares (Sí, No).

| Rendimiento / Extraescolares | Sí | No |
|------------------------------|----|----|
| Alto                         | 35 | 20 |
| Medio                        | 25 | 30 |
| Bajo                         | 10 | 40 |

# Prueba Chi cuadrada para Bondad de Ajuste: Introducción

- Evalúa la diferencia entre una distribución observada y una teórica o esperada.
- Útil para verificar si una muestra proviene de una población con una distribución conocida.
- Se basa en calcular las diferencias entre frecuencias observadas y esperadas.

# Ejemplo: Dado Balanceado

Un jugador sospecha que un dado no está balanceado. Lanza el dado 600 veces y registra los resultados:

| Número del dado | Observados | Esperados |
|-----------------|------------|-----------|
| 1               | 110        | 100       |
| 2               | 95         | 100       |
| 3               | 108        | 100       |
| 4               | 103        | 100       |
| 5               | 92         | 100       |
| 6               | 92         | 100       |

¿El dado está balanceado?

# Formulación Matemática

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

Donde:

- $\chi^2$  es el estadístico Chi cuadrada.
- $O_i$  son las frecuencias observadas.
- $E_i$  son las frecuencias esperadas.



# Ejemplo 1

Un restaurante afirma que las preferencias de postres de sus clientes se distribuyen de manera uniforme entre tres opciones: Pastel, Helado y Frutas. Se observan las elecciones de 120 clientes y se obtiene:

| Postre | Observados | Esperados |
|--------|------------|-----------|
| Pastel | 50         | 40        |
| Helado | 45         | 40        |
| Frutas | 25         | 40        |

# Ejemplo 2

Una tienda de mascotas afirma que las ventas de tres tipos de animales (Perros, Gatos, Pájaros) se distribuyen en proporciones de 40

| Animal  | Observados | Esperados |
|---------|------------|-----------|
| Perros  | 70         | 60        |
| Gatos   | 50         | 60        |
| Pájaros | 30         | 30        |

# Importancia y Aplicaciones

- La prueba de bondad de ajuste es fundamental para validar suposiciones sobre distribuciones en muchos campos.
- Se aplica en marketing, biología, ciencias sociales y más.
- Es un primer paso esencial antes de aplicar técnicas que requieran suposiciones sobre la distribución de datos.