Contrastes sobre proporciones Tema 9

1. Una proporción

2. Dos proporciones

   2.1 Dos proporciones independientes
   2.2 Dos proporciones relacionadas (Prueba de McNemar)

3. Más de dos proporciones relacionadas (Prueba de Cochran)

1. Contraste sobre una proporción

Objetivo: Realizar inferencias acerca de una única proporción poblacional.

1. Hipótesis

a) Bilateral: H_0: $\pi = \pi_0$; H_1: $\pi \neq \pi_0$

b) U. derecho: H_0: $\pi \leq \pi_0$; H_1: $\pi > \pi_0$

c) U. izquierdo: H_0: $\pi \geq \pi_0$; H_1: $\pi < \pi_0$

2. Supuestos

Muestra aleatoria

π constante en cada extracción

3. Estadístico de contraste

3.1. Muestra pequeña ($n \leq 25$):

$$X = \sum_{i=1}^{n} X_i \quad \text{(número de éxitos)}$$

X ~ Binomial (n, π_0)

3.2 Muestra grande (n> 25)

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{P - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

Z se distribuye según la normal (0, 1)

4. Zona crítica

a) Bilateral: $Z \leq z_{\alpha/2}$ y $Z \geq z_{1-\alpha/2}$
b) Unilateral derecho: $Z \geq z_{1-\alpha}$
c) Unilateral izquierdo: $Z \leq z_{\alpha}$

5. Decisión:

5.1 Muestra pequeña

Rechazar H_0 si *p* (**nivel crítico**) es menor o igual que α (unilateral) o $\alpha/2$ (bilateral)

5.2 Muestra grande:

Rechazar H_0 si el estadístico *Z* cae en la zona crítica

<u>Ejemplo</u>: Se está realizando una investigación sobre tabaquismo juvenil. Se ha tomado una muestra de 15 alumnos de instituto y se ha encontrado que 5 de ellos fuman habitualmente. Contrastar la hipótesis de que el porcentaje de jóvenes fumadores es del 40% utilizando α=0,01.

1. Hipótesis

H_0: $\pi = 0{,}4$

H_1: $\pi \neq 0{,}4$

2. Supuestos

Muestra aleatoria

$\pi = 0{,}4$ constante en cada extracción

3. Estadístico de contraste

Dado que la muestra es pequeña:

nº de éxitos : $X = 5$

X ~ Binomial (n=15, π_0=0,4)

4. Zona crítica

$X \leq 1$ $P(X \leq 1) = 0{,}005$

$X \geq 12$ $P(X \geq 12) = 1 - 0{,}998 = 0{,}002$

nivel crítico:

$p = (2)P(X \leq 5) = (2)0{,}403 = 0{,}806$

5. Decisión:

X cae en la zona de aceptación.

$p = 0{,}806 > 0{,}01 = \alpha$

Mantener H_0

Utilizando la aproximación normal (no es necesario):

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{5-(15)0{,}4}{\sqrt{(15)0{,}4(1-0{,}4)}} = \frac{5-6}{\sqrt{3{,}6}} = -0{,}53$$

${}_{0,005}Z = -2{,}575$; ${}_{0,995}Z = 2{,}575$

Cálculo del nivel crítico

Unilateral izquierdo	
H_0: $\pi \geq 0,4$ H_1: $\pi < 0,4$	$p = P(X \leq 5) = 0,403$
Unilateral derecho	
H_0: $\pi \leq 0,4$ H_1: $\pi > 0,4$	$p = P(X \geq 5) = 1\text{-}P(X \leq 4)$ $= 1\text{-} 0,217$ $= 0,783$
Bilateral	
H_0: $\pi = 0,4$ H_1: $\pi \neq 0,4$	$(2)p = P(X \leq 5) = (2)0,403 = 0,806$ (2 por el menor de ambos)

2.1. Dos proporciones independientes

Objetivo: Contrastar si son iguales dos proporciones procedentes de dos poblaciones diferentes.

1. Hipótesis

Bilateral: H_0: $\pi_1 = \pi_2$; H_1: $\pi_1 \neq \pi_2$

U. derecho: H_0: $\pi_1 \leq \pi_2$.; H_1: $\pi_1 > \pi_2$

U. izquierdo: H_0: $\pi_1 \geq \pi_2$; H_1: $\pi_1 < \pi_2$

2. Supuestos

Muestra aleatoria

π_1 y π_2 constantes en cada extracción

3. Estadístico de contraste

Muestra 1: n_1, P_1

Muestra 2: n_2, P_2

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Z se distribuye normal (0, 1)

4. Zona crítica

Bilateral:
$Z \leq {}_{\alpha/2}z$
$Z \geq {}_{1-\alpha/2}z$

Unilateral derecho:
$Z \geq {}_{1-\alpha}z$

Unilateral izquierdo:
$Z \leq {}_{\alpha}z$

5. Decisión

Rechazar H_0 si el estadístico de contraste cae en la zona crítica (o si $p \leq \alpha$)

Ejemplo: Continuando con la investigación sobre tabaquismo, desea contrastarse si fuman más las chicas que los chicos. Se toma una muestra de 20 chicas y se encuentra que fuman 12. En una muestra de 18 chicos fuman 8. Realizar el contraste con α=0,01.

1. Hipótesis

H_0: $\pi_1 \leq \pi_2$

H_1: $\pi_1 > \pi_2$

2. Supuestos

Muestra aleatoria

π_1 y π_2 constantes en cada extracción

3. Estadístico de contraste

$n_1 = 20$; $P_1 = 12/20 = 0{,}60$

$n_2 = 18$; $P_2 = 8/18 = 0{,}44$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{12+8}{20+18} = 0{,}53$$

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0{,}6 - 0{,}44}{\sqrt{0{,}53(1-0{,}53)\left(\frac{1}{20} + \frac{1}{18}\right)}} = 0{,}99$$

4. Zona crítica: $Z \geq {}_{0,99}z = 2{,}33$

5. Decisión: Mantener H_0. No puede concluirse que las chicas fuman más.

2.2. Dos proporciones relacionadas (Prueba de McNemar)

Objetivo: Contrastar si son iguales dos proporciones poblacionales cuando se utiliza un diseño de medidas repetidas (antes-después).

3. Estadístico de contraste

		Después		
		1	2	
Antes	1	n_{11}	n_{12}	n_{1+}
	2	n_{21}	n_{22}	n_{2+}
		n_{+1}	n_{+2}	m

Calcular $n = n_{12} + n_{21}$

3.1 Muestra pequeña ($n \leq 25$)

$T = n_{12}$

T ~ Binomial (n, π=0,5)

3.2 Muestra grande (n > 25)

$$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

$X^2 \sim \chi^2$ con 1 grado de libertad

4. Regla de decisión

4.1 Estadístico T :
Utilizar el nivel crítico

4.2 Estadístico X^2 :

a) Bilateral : $X^2 \geq {}_{1-\alpha}\chi_1^2$

b) U. Derecho : $X^2 \geq {}_{1-2\alpha}\chi_1^2$

c) U. Izquierdo : $X^2 \geq {}_{1-2\alpha}\chi_1^2$

<u>Ejemplo</u>: Se ha tomado un grupo de 40 jóvenes y se ha encontrado que fumaban 25. Se les proporciona información sobre los perjuicios del tabaco y se convence a 5 fumadores para que dejen el tabaco. Después de recibir la información son 19 los jóvenes que no fuman ¿Puede concluirse con α=0,01 que la información es eficaz?

Los datos pueden organizarse:

		Después		
		Si	No	
Antes	Si		5	25
	No			
			19	40

1. Hipótesis

U. derecho: H_0: $\pi_1 \leq \pi_2$.; H_1: $\pi_1 > \pi_2$

2. Supuestos

Muestra aleatoria de *m* pares

π_1 y π_2 constantes en cada extracción

3. Estadístico de contraste

Completar la tabla:

		Después		
		Si	No	
Antes	Si	20	5	25
	No	1	14	15
		21	19	40

Calcular $n = n_{12} + n_{21} = 6$

$T = 5$; $T \sim$ Binomial $(6, \pi = 0{,}5)$

4. Regla de decisión

$$
\begin{aligned}
P(T \geq 5) &= 1 - P(T \leq 4) \\
&= 1 - 0{,}891 \\
&= 0{,}109 > \alpha = 0{,}01
\end{aligned}
$$

5. Decisión

Mantener H_0

Utilizando X^2 (no es necesario):

3. Estadístico de contraste

$$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} = \frac{(5-1)^2}{5+1} = 2{,}67$$

$X^2 \sim \chi^2$ con 1 grado de libertad

4. Regla de decisión ($2\alpha = 0{,}02$)

Zona crítica: $X^2 \geq {}_{1-2\alpha}\chi^2 = {}_{0,98}\chi^2 = 5{,}41$

5. Decisión

Mantener H_0

3. Más de dos proporciones relacionadas (Prueba de Cochran)

Objetivo: Contrastar si son iguales más de dos proporciones poblacionales cuando se utiliza un diseño de medidas repetidas.

1. Hipótesis

H_0: $\pi_{+1} = \pi_{+2} = ... = \pi_{+J}$

H_1: $\pi_{+j} \neq \pi_{+j'}$

2. Supuestos

Muestra aleatoria

π_{+j} constante en cada extracción

3. Estadístico de contraste

$$Q = \frac{J(J-1)\sum_{j=1}^{J} T_{+j}^2 - (J-1)T^2}{JT - \sum_{i=1}^{n} T_{i+}^2}$$

J: Número de proporciones (grupos)
T: Total de la muestra
T_{+j}: Total de cada tratamiento *j*
T_{i+}: Total de cada sujeto *i*
$Q \sim \chi^2$ con *J*-1 grados de libertad

4. Zona crítica: $Q \geq {}_{1-\alpha}\chi^2_{J-1}$

5. Decisión. Rechazar H_0 si *Q* cae en la zona crítica

<u>Ejemplo</u>: 10 sujetos deciden participar en un tratamiento antitabaco. Para realizar un seguimiento de su efectividad, una vez finalizado el tratamiento se toman datos al cabo de uno, dos y tres meses. La siguiente tabla indica qué sujetos han vuelto a fumar. ¿Puede concluirse que la proporción de fumadores se mantiene estable con α=0,05?

Un mes	1	0	0	1	0	1	0	0	0	0
Dos meses	1	1	0	0	0	1	0	1	0	0
Tres meses	1	1	0	0	1	0	0	1	1	0

1. Hipótesis

H_0: $\pi_{+1} = \pi_{+2} = \pi_{+3}$

H_1: $\pi_{+j} \neq \pi_{j'}$

2. Supuestos

Muestra aleatoria

π_{+j} constante en cada extracción

2. Estadístico de contraste

Un mes	1	0	0	1	0	1	0	0	0	0	3
Dos meses	1	1	0	0	0	1	0	1	0	0	4
Tres meses	1	1	0	0	1	0	0	1	1	0	5
	3	2	0	1	1	2	0	2	1	0	12

T_{+j}

T_{i+} T

$$Q = \frac{J(J-1)\sum_{j=1}^{J} T_{+j}^{2} - (J-1)T^{2}}{JT - \sum_{i=1}^{n} T_{i+}^{2}}$$

$$= \frac{3(2)(3^{2} + 4^{2} + 5^{2}) - 2(12^{2})}{3(12) - (3^{2} + 2^{2} + \cdots + 0^{2})} = 1$$

$$Q \sim \chi^{2}_{J-1=2}$$

3. Zona crítica:

$$Q \geq {}_{1-\alpha}\chi^{2}_{J-1} = {}_{0,95}\chi^{2}_{2} = 5,99$$

4. Decisión.

Mantener H_0. La proporción de fumadores no cambia durante los tres meses.

Formulario del tema 9

Contraste sobre una proporción

X ~ Binomial (n, π_0)

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{P - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

Z ~ normal (0, 1)

Dos proporciones independientes

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Z ~ normal (0, 1)

Dos proporciones relacionadas (McNemar)

$$T = n_{12}$$

$$T \sim \text{Binomial}\ (n,\ \pi{=}0{,}5)$$

$$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

$$X^2 \sim \chi^2_1$$

Más de dos proporciones relacionadas (Cochran)

$$Q = \frac{J(J-1)\sum_{j=1}^{J} T_{+j}^2 - (J-1)T^2}{JT - \sum_{i=1}^{n} T_{i+}^2}$$

$$Q \sim \chi^2_{J-1}$$

Ejercicios recomendados del libro:

11.1
11.2
11.3
11.14