

# Bioestadística y uso de software científico



## TEMA 4 DATOS CATEGÓRICOS COMPARACIÓN DE PROPORCIONES

# Índice



- Tablas de contingencia y grados de libertad
- Ji-cuadrado
- Comparación de proporciones
  - Tabla 2x2
- Diferencia de proporciones
  - Intervalo de confianza
- Test de Fisher
- Test de McNemar

# ¿Están asociados el cáncer y el grupo sanguíneo?



<b>Grupo sanguíneo Cáncer</b>	<b>0</b>	<b>A</b>	<b>B</b>	<b>AB</b>	<b>Total</b>
<b>Sí</b>	<b>20</b>	<b>40</b>	<b>20</b>	<b>20</b>	<b>100</b>
<b>No</b>	<b>40</b>	<b>40</b>	<b>10</b>	<b>10</b>	<b>100</b>
<b>Total</b>	<b>60</b>	<b>80</b>	<b>30</b>	<b>30</b>	<b>200</b>

Tabla observada (O)

# ¿Están asociados el cáncer y el grupo sanguíneo?



## 1. Calcular la tabla esperada al azar

Grupo sanguíneo Cáncer	0	A	B	AB	Total
Sí	20	40	20	20	100
No	40	40	10	10	100
Total	60	80	30	30	200

Tabla observada (O)

Tabla esperada al azar  
(E)

Grupo sanguíneo Cáncer	0	A	B	AB	Total
Sí					100
No					100
Total	60	80	30	30	200

# ¿Están asociados el cáncer y el grupo sanguíneo?



Tabla esperada al azar (E)

<b>Grupo sanguíneo Cáncer</b>	<b>0</b>	<b>A</b>	<b>B</b>	<b>AB</b>	<b>Total</b>
<b>Sí</b>	$\frac{60 \times 100}{200}$				<b>100</b>
<b>No</b>					<b>100</b>
<b>Total</b>	<b>60</b>	<b>80</b>	<b>30</b>	<b>30</b>	<b>200</b>

# ¿Están asociados el cáncer y el grupo sanguíneo?



Tabla esperada al azar (E)

Grupo sanguíneo Cáncer	O	A	B	AB	Total
Sí	$\frac{60 \times 100}{200}$				100
No					100
Total	60	80	30	30	200

# ¿Están asociados el cáncer y el grupo sanguíneo?



<b>Grupo sanguíneo Cáncer</b>	<b>0</b>	<b>A</b>	<b>B</b>	<b>AB</b>	<b>Total</b>
<b>Sí</b>	<b>20</b>	<b>40</b>	<b>20</b>	<b>20</b>	<b>100</b>
<b>No</b>	<b>40</b>	<b>40</b>	<b>10</b>	<b>10</b>	<b>100</b>
<b>Total</b>	<b>60</b>	<b>80</b>	<b>30</b>	<b>30</b>	<b>200</b>

Tabla observada (O)

Tabla esperada (E)

<b>Grupo sanguíneo Cáncer</b>	<b>0</b>	<b>A</b>	<b>B</b>	<b>AB</b>	<b>Total</b>
<b>Sí</b>	<b>30</b>	<b>35</b>	<b>15</b>	<b>15</b>	<b>100</b>
<b>No</b>	<b>30</b>	<b>35</b>	<b>15</b>	<b>15</b>	<b>100</b>
<b>Total</b>	<b>60</b>	<b>80</b>	<b>30</b>	<b>30</b>	<b>200</b>

## ¿Están asociados el cáncer y el grupo sanguíneo?



2. Calcular:

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

3. Calcular los grados de libertad (g.l.)

$$g.l. = (m - 1) \times (n - 1)$$

m = Número de categorías en una variable.

n = Número de categorías en la otra variable.



# Tablas de contingencia



Grupo sanguíneo Cáncer	0	A	B	AB	Total
Sí					100
No					100
Total	60	80	30	30	200

Grados de libertad =  
 $(4-1)(2-1) = 3$

Grupo sanguíneo Cáncer	0	A	B	AB	Total
Sí	20	40	20		100
No					100
Total	60	80	30	30	200

# Tablas de contingencia



<b>Grupo sanguíneo Cáncer</b>	<b>0</b>	<b>A</b>	<b>B</b>	<b>AB</b>	<b>Total</b>
<b>Sí</b>			<b>10</b>		<b>100</b>
<b>No</b>	<b>40</b>	<b>40</b>			<b>100</b>
<b>Total</b>	<b>60</b>	<b>80</b>	<b>30</b>	<b>30</b>	<b>200</b>

<b>Grupo sanguíneo Cáncer</b>	<b>0</b>	<b>A</b>	<b>B</b>	<b>AB</b>	<b>Total</b>
<b>Sí</b>	<b>20</b>			<b>20</b>	<b>100</b>
<b>No</b>		<b>40</b>			<b>100</b>
<b>Total</b>	<b>60</b>	<b>80</b>	<b>30</b>	<b>30</b>	<b>200</b>

## ¿Están asociados el cáncer y el grupo sanguíneo?



4. Buscar el valor p en la tabla  $\chi^2$  con  $(m-1) \times (n-1)$  grados de libertad

[http://es.wikibooks.org/wiki/Tablas estad%C3%ADsticas/Distribuci%C3%B3n\\_chi-cuadrado](http://es.wikibooks.org/wiki/Tablas_estad%C3%ADsticas/Distribuci%C3%B3n_chi-cuadrado)

# Índice



- Tablas de contingencia y grados de libertad
- **Distribución Ji-cuadrado**
- Comparación de proporciones
  - Tabla 2x2
- Diferencia de proporciones
  - Intervalo de confianza
- Test de Fisher
- Test de McNemar

# Ji-cuadrado ( $\chi^2$ )



- Se puede utilizar en cualquier tabla de contingencia.
- Requiere que en ninguna casilla el valor esperado sea  $<5$

# Índice



- Tablas de contingencia y grados de libertad
- Distribución Ji-cuadrado
- Comparación de proporciones
  - Tabla 2x2
- Diferencia de proporciones
  - Intervalo de confianza
- Test de Fisher
- Test de McNemar

# Comparar dos proporciones



- La tabla de contingencia más habitual es la tabla tetracórica (2x2)  $\rightarrow$  1 grado de libertad

	<b>Diabetes</b>	<b>No diabetes</b>	<b>Total</b>
<b>Glucosuria</b>	<b>60</b>	<b>8</b>	<b>68</b>
<b>No glucosuria</b>	<b>140</b>	<b>792</b>	<b>932</b>
<b>Total</b>	<b>200</b>	<b>800</b>	<b>1000</b>

# Comparar dos proporciones



- $H_0$ : La proporción de diabéticos es igual en los que tienen glucosuria ( $\pi_1$ ) y en los que no la tienen ( $\pi_2$ )

	Diabetes	No diabetes	Total
Glucosuria	60	8	68
No glucosuria	140	792	932
Total	200	800	1000



# Comparar dos proporciones



- $H_0$ : La proporción de diabéticos es igual en los que tienen glucosuria ( $\pi_1$ ) y en los que no la tienen ( $\pi_2$ )

$$H_0 : \pi_1 = \pi_2$$

$$p_1 = \frac{60}{68} = 0,88$$

$$p_2 = \frac{140}{932} = 0,15$$

	Diabetes	No diabetes	Total
Glucosuria	60	8	68
No glucosuria	140	792	932
Total	200	800	1000

# Comparar dos proporciones: cálculo de p



	Diabetes	No diabetes	Total
Glucosuria	60	8	68
No glucosuria	140	792	932
Total	200	800	1000

Tabla O

	Diabetes	No diabetes	Total
Glucosuria			68
No glucosuria			932
Total	200	800	1000

Tabla E

# Comparar dos proporciones: cálculo de p



	Diabetes	No diabetes	Total
Glucosuria	60	8	68
No glucosuria	140	792	932
Total	200	800	1000

Tabla O

	Diabetes	No diabetes	Total
Glucosuria	13,6	54,4	68
No glucosuria	186,4	745,6	932
Total	200	800	1000

Tabla E

## Comparar dos proporciones: cálculo de p



$$\frac{(60 - 13,6)^2}{13,6} + \frac{(8 - 54,4)^2}{54,4} + \frac{(140 - 186,4)^2}{186,4} + \frac{(792 - 745,6)^2}{745,6} = 212,32$$

Ver la tabla de ji-cuadrado con un grado de libertad:

[http://es.wikibooks.org/wiki/Tablas estad%C3%ADsticas/Distribuci%C3%B3n\\_chi-cuadrado](http://es.wikibooks.org/wiki/Tablas_estad%C3%ADsticas/Distribuci%C3%B3n_chi-cuadrado)

# Índice



- Tablas de contingencia y grados de libertad
- Distribución Ji-cuadrado
- Comparación de proporciones
  - Tabla 2x2
- Diferencia de proporciones
  - Intervalo de confianza
- Test de Fisher
- Test de McNemar

## Comparar dos proporciones: intervalo de confianza



- $H_0$ : La proporción de diabéticos es igual en los que tienen glucosuria ( $\pi_1$ ) y en los que no la tienen ( $\pi_2$ )

$$H_0 : \pi_1 = \pi_2 \Leftrightarrow \pi_1 - \pi_2 = 0$$

$$p_1 = \frac{60}{68} = 0,88$$

$$p_2 = \frac{140}{932} = 0,15$$

$$p = \frac{200}{1000} = 0,2$$

## Comparar dos proporciones: intervalo de confianza



1. Calcular la diferencia entre las proporciones observadas

$$p_1 - p_2 = 0,88 - 0,15 = 0,73$$

## Comparar dos proporciones: intervalo de confianza



1. Calcular la diferencia entre las proporciones observadas

$$p_1 - p_2 = 0,88 - 0,15 = 0,73$$

2. Calcular el error estándar de la diferencia de proporciones

$$EEDP = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} = \sqrt{\frac{0,2 \times 0,8}{68} + \frac{0,2 \times 0,8}{932}} = 0,05$$



## Comparar dos proporciones: intervalo de confianza



3. El intervalo de confianza al  $1-\alpha$  es:

$$dp \pm z_{\alpha/2} \times EEDP$$

4. El intervalo de confianza al 95% es:

$$\begin{aligned} dp \pm 1,96 \times EEDP &= \\ = 0,73 \pm 1,96 \times 0,05 &= (0,632 \text{ a } 0,828) \end{aligned}$$

# Prueba Z

## Otra forma de calcular el valor p



1. Calcular:

$$z = \frac{dp}{EEDP}$$

2. Buscar el resultado en la tabla de la distribución normal

# Índice



- Tablas de contingencia y grados de libertad
- Distribución Ji-cuadrado
- Comparación de proporciones
  - Tabla 2x2
- Diferencia de proporciones
  - Intervalo de confianza
- Test de Fisher
- Test de McNemar

# Test exacto de Fisher



- Se puede utilizar en cualquier tabla de contingencia
- Es imprescindible usarlo cuando en la tabla esperada aparece alguna casilla  $<5$
- Calcula cuál es la probabilidad de que ocurra al azar la tabla observada o un resultado más desfavorable para la hipótesis nula
- **Calcular sólo con ordenador**

# Índice



- Tablas de contingencia y grados de libertad
- Distribución Ji-cuadrado
- Comparación de proporciones
  - Tabla 2x2
- Diferencia de proporciones
  - Intervalo de confianza
- Test de Fisher
- Test de McNemar

# Test de McNemar para datos emparejados



- Los métodos anteriores asumen que los individuos son independientes entre sí
- Emparejamiento:
  - Queremos saber si la diabetes se presenta de forma familiar.
  - Seleccionamos 100 parejas padre-hijo
  - Los individuos no son independientes →
  - → Hace falta otra forma de analizar los datos

# Test de McNemar para datos emparejados



		Padre		Total
		Diabético	No diabético	
Hijo	Diabético	20	10	30
	No diabético	15	55	70
Total		35	65	100

La tabla está formada por parejas

# Test de McNemar para datos emparejados



		Padre		Total
		Diabético	No diabético	
Hijo	Diabético	20	10	30
	No diabético	15	55	70
Total		35	65	100

Para calcular el test de McNemar sólo se utilizan las parejas discordantes (marcadas en amarillo)



# Test de McNemar para datos emparejados



		Padre		Total
		Diabético	No diabético	
Hijo	Diabético	20	10	30
	No diabético	15	55	70
Total		35	65	100

Calcular:  $\chi^2_{McNemar} = \frac{(10-15)^2}{10+15} = 1$

Buscar el valor p en la tabla  $\chi^2$  con 1 g.l.

# Test de McNemar para datos emparejados



		Sujeto 1		
		Enfermo	No enfermo	Total
Sujeto 2	Enfermo	A	B	
	No enfermo	C	D	
Total				

Calcular:

$$\chi_{McNemar}^2 = \frac{(B - C)^2}{B + C}$$

Buscar el valor p en la tabla  $\chi^2$  con 1 g.l.