

OCE 313
TÉCNICAS DE ANÁLISIS NO PARAMÉTRICO

CLASE 6 - TABLAS DE CONTINGENCIA

Dr. José Gallardo

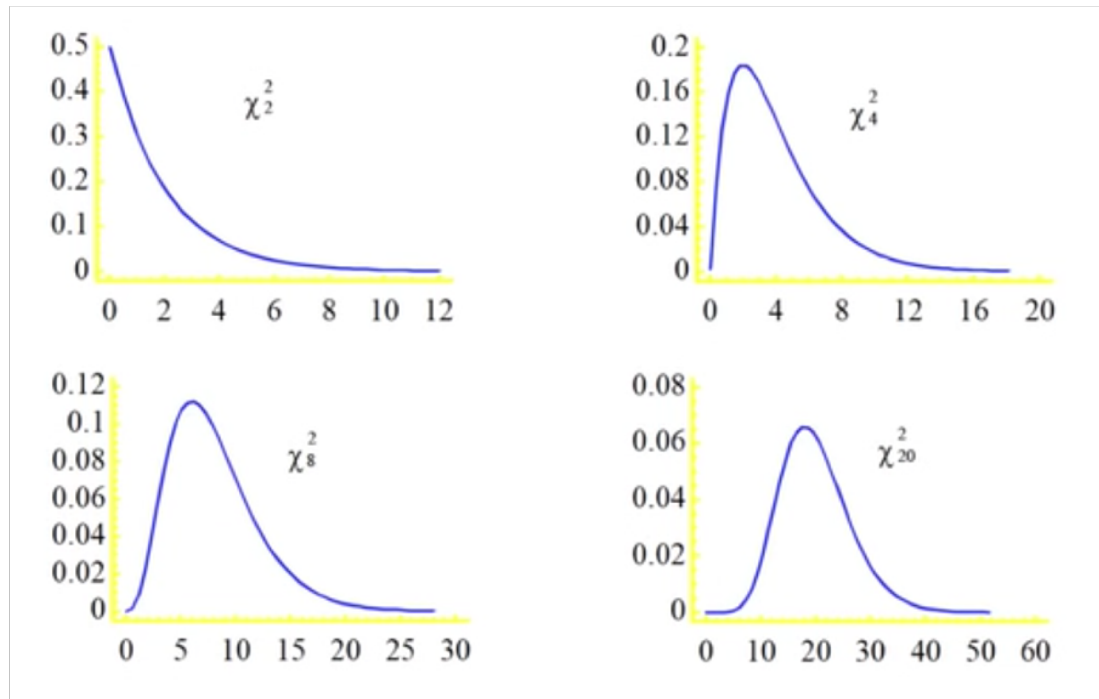
Abril 2021

PLAN DE LA CLASE

- Distribución chi cuadrado.
- **Tablas de contingencia**
- Test chi cuadrado.
- **Test exacto de Fisher.**
- Interpretación test no paramétricos con R.
- **Prácticas R test de correlación, chi cuadrado y fisher.**

Distribución Chi-cuadrado (χ^2)

- Corresponde a la **distribución muestral de varianzas (s^2)** de una población normal.
- El único parámetro que se necesita para modelar la distribución es n , denominado grados de libertad.
- La esperanza matemática es $= n$ y la varianza $= 2n$.



Tablas de contingencia o tabla de clasificación cruzada.

Tratamiento	Respuesta +	Respuesta -
SI	a	b
NO	c	d

- Se usan comúnmente para **resumir datos de variables categóricas** (ej. Cualitativas, dicotómicas).
- Se utilizan para **investigar la asociación de dos o más variables categóricas** una de las cuales es una variable respuesta y otra es una variable predictora.

Prueba de Chi-cuadrado (χ^2)

Hipótesis

H_0 : La variable predictora y la variable respuesta son independientes.

H_1 : La variable predictora y la variable respuesta NO son independientes.

Supuestos:

1) Los datos provienen de una muestra aleatoria de la población de interés.

2) El tamaño de muestra es lo suficientemente grande para que el número esperado en las categorías sea mayor 5 y que ninguna frecuencia sea menor que 1.

Ejemplo prueba Chi-cuadrado

Frecuencia observada

Tratamiento	Respuesta +	Respuesta -	Totales
1	52	17	69
2	54	29	83
3	26	11	37
Totales	132	57	189

Hipótesis

$$H_0: P_{+t_1} = P_{+t_2} = P_{+t_3}$$

$$H_1: P_{+t_1} \neq P_{+t_2}$$

$$H_1: P_{+t_1} \neq P_{+t_3}$$

$$H_1: P_{+t_2} \neq P_{+t_3}$$

Frecuencia observada

$$f_{o11} = 52$$

$$f_{o12} = 17$$

Etc.

Ejemplo prueba Chi-cuadrado cont.

Frecuencia esperada

Tratamiento	Respuesta +	Respuesta -	Totales
1	48,19	20,81	69
2	57,97	25,03	83
3	25,84	11,16	37
Totales	132	57	189

Valor Chi²

$$\chi^2 = \sum (f_o - f_e)^2 / f_e$$

$$\chi^2 = (52 - 48,19)^2 / 48,19 + \dots$$

$$\chi^2 = 0,30 + 0,27 + 0,00 + 0,70 + 0,63 + 0,00$$

$$\chi^2 = 1,9$$

Frecuencia esperada 11

$$f_{e11} = N_+ * N_{T1} / N$$

$$f_{e11} = 132 * 69 / 189 = 48,19$$

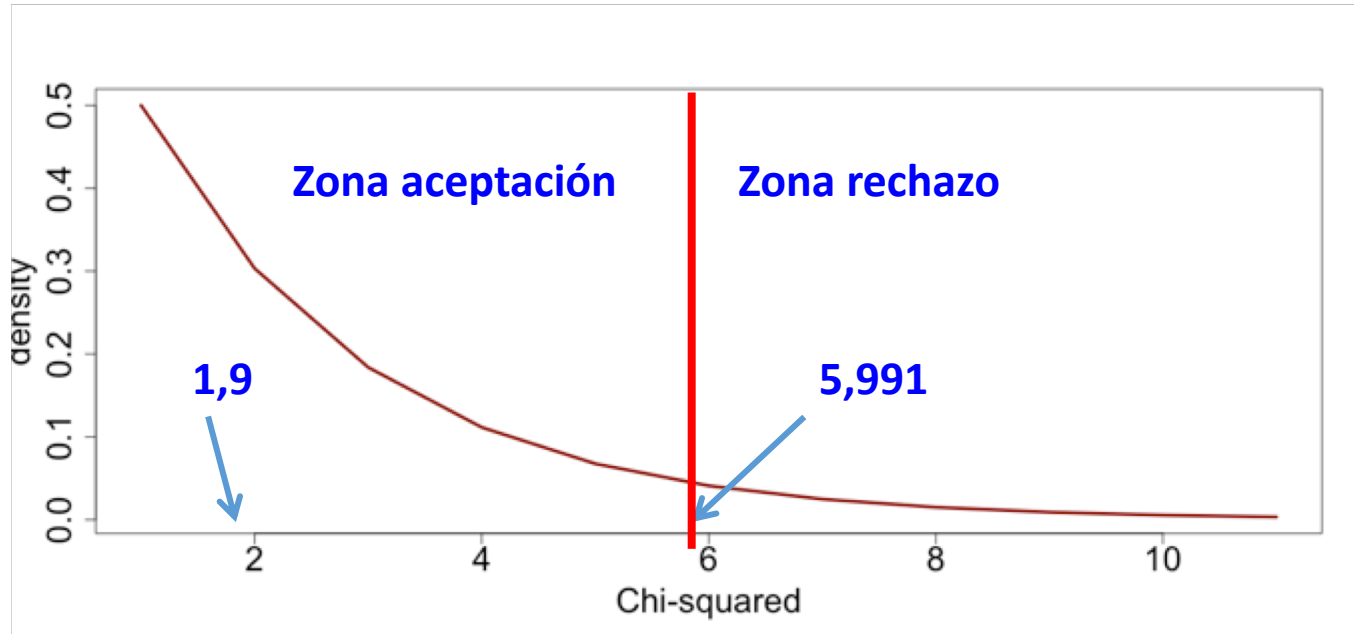
Ejemplo prueba Chi-cuadrado cont.

χ^2 obtenido = 1,9

Alfa = 0,05

g.l. = 3 tratamientos – 1 = 2

χ^2 valor crítico de aceptación o rechazo = 5,991



Conclusión: No se rechaza la hipótesis nula y por lo tanto concluimos que no hay diferencia en los tratamientos.

Test de probabilidad exacta de Fisher

- Se utiliza en tablas de contingencia de 2x2.
- Se utiliza cuando **NO** se cumple el supuesto del test paramétrico Chi-2 respecto de que el número esperado en las categorías sea mayor 5.
- Requiere **calcular las probabilidades individuales para las distintas maneras** en que pueden aparecer las frecuencias dentro de las 4 celdas, manteniendo constantes las frecuencias marginales, **sumando las probabilidades correspondiente a las pregunta de interes.**

Test de probabilidad exacta de Fisher.

Tratamiento	Respuesta +	Respuesta -	Totales
1	a	b	$g = a + b$
2	c	d	$h = c + d$
Totales	$e = a + c$	$f = b + d$	n

Probabilidad exacta de Fisher

$P(a) = \text{Respuesta + del trat. 1}$

$$P(a) = \frac{e! * f! * g! * h!}{a! * b! * c! * d! * n!}$$

Ejemplo Test de probabilidad exacta de Fisher.

Tratamiento	Respuesta +	Respuesta -	Totales
1	7	3	10
2	1	9	10
Totales	8	12	20

Preguntas:

¿ El tratamiento 1 fue más efectivo en generar una respuesta + ?

¿ Es posible que este resultado positivo sea solo por azar ?

Tablas de contingencia alternativas.

Trat.	R +	R -	Tot.
1	<u>0</u>	10	10
2	8	2	10
Tot.	8	12	20

Trat.	R +	R -	Tot.
1	<u>1</u>	9	10
2	7	3	10
Tot.	8	12	20

Trat.	R +	R -	Tot.
1	<u>2</u>	8	10
2	6	4	10
Tot.	8	12	20

Trat.	R +	R -	Tot.
1	<u>3</u>	7	10
2	5	5	10
Tot.	8	12	20

Trat.	R +	R -	Tot.
1	<u>4</u>	6	10
2	4	6	10
Tot.	8	12	20

Trat.	R +	R -	Tot.
1	<u>5</u>	5	10
2	3	7	10
Tot.	8	12	20

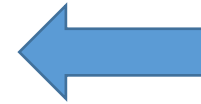
Trat.	R +	R -	Tot.
1	<u>6</u>	4	10
2	2	8	10
Tot.	8	12	20

Trat.	R +	R -	Tot.
1	<u>7</u>	3	10
2	1	9	10
Tot.	8	12	20

Trat.	R +	R -	Tot.
1	<u>8</u>	2	10
2	0	10	10
Tot.	8	12	20

Cálculo de la probabilidad exacta de Fisher.

a	$P(a)$	$P(a)$
0	$\frac{8! 12! 10! 10!}{0! 10! 8! 2! 20!}$	0,00036
1		0,00953
2		0,07502
3		0,24006
4		0,35008
5		0,24006
6		0,07502
7	$\frac{8! 12! 10! 10!}{7! 3! 1! 9! 20!}$	0,00953
8		0,00036



Trat.	R +	R -	Tot.
SI	<u>0</u>	10	10
NO	8	2	10
Tot.	8	12	20



Trat.	R +	R -	Tot.
SI	<u>7</u>	3	10
NO	1	9	10
Tot.	8	12	20

Tarea: complete la información faltante en $P(a)$.

Prueba de hipótesis

Hipótesis

$$H_0: P_{+t1} = P_{+t2}$$

$$H_1: P_{+t1} > P_{+t2}$$

La probabilidad de obtener un valor de **a (respuesta positiva del trat 1)** mayor o igual que 7 = $P(7) + P(8) = 0,00953 + 0,00036 = \mathbf{0,00989}$.

Conclusión: Se rechaza la hipótesis nula con un alfa de 0,01 y por lo tanto concluimos que el tratamiento 1 es más efectivo que el tratamiento 2 en generar una respuesta positiva.

Coeficiente de correlación (cor.test)

Test for Association/Correlation Between Paired Samples

`cor.test {stats}` R Documentation

Description

Test for association between paired samples, using one of Pearson's product moment correlation coefficient, Kendall's tau or Spearman's rho.

Default S3 method:

```
cor.test(x, y,  
         alternative = c("two.sided", "less", "greater"),  
         method = c("pearson", "kendall", "spearman"),  
         exact = NULL, conf.level = 0.95, ...)
```

Coeficiente de correlación de Spearman

```
cor.test(X,Y, method = "spearman", alternative = "two.sided")
```

Spearman's rank correlation rho

data: X and Y

S = 16, p-value = 0.4167

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

-0.6

Test de Chi-squared en R (chisq.test)

chisq.test {stats} R Documentation Pearson's Chi-squared Test for Count Data

Description

chisq.test performs chi-squared contingency table tests and goodness-of-fit tests.

Usage

```
chisq.test(x, y = NULL, correct = TRUE,  
           p = rep(1/length(x), length(x)), rescale.p = FALSE,  
           simulate.p.value = FALSE, B = 2000)
```

Test de Chi-squared en R (chisq.test)

```
M <- c(52, 54, 26, 17, 29, 11)  
dim(3,2)  
chisq.test(M)
```

Pearson's Chi-squared test

data: M

X-squared = 1.9025, df = 2, p-value = 0.3863

Test de fisher en R (fisher.test)

fisher.test {stats} R Documentation
Fisher's Exact Test for Count Data

Description

Performs Fisher's exact test for testing the null of independence of rows and columns in a contingency table with fixed marginals.

Usage

```
fisher.test(x, y = NULL, workspace = 200000,  
hybrid = FALSE, control = list(), or = 1,  
alternative = "two.sided", conf.int = TRUE,  
conf.level = 0.95, simulate.p.value = FALSE, B  
= 2000)
```

Test de fisher en R (fisher.test)

```
Prueba_fisher <- matrix(c(7, 3, 1, 9), nrow= 2,  
  dimnames = list(response = c("positive", "negative"), treatment = c("T1","T2")))
```

```
Fisht<- fisher.test(Prueba_fisher, alternative = "greater")
```

Fisht

Fisher's Exact Test for Count Data

```
data:  Prueba_fisher  
p-value = 0.009883  
alternative hypothesis: true odds ratio is greater than 1  
95 percent confidence interval:  
 1.836563      Inf  
sample estimates:  
odds ratio  
 17.27587
```

RESUMEN DE LO APRENDIDO

1. **Revisión de conceptos de permutación.**
2. Recordatorio correlación Pearson.
3. **Funciones monótonas.**
4. Aplicación de concepto de permutación a prueba de correlación de Spearman.