

CLASE 06 - TABLAS DE CONTINGENCIA

OCE 313 - Técnicas de análisis no paramétricos.

Dr. José Gallardo Matus

Pontificia Universidad Católica de Valparaíso

18 April 2022

PLAN DE LA CLASE

1.- Introducción

- ▶ Tablas de contingencia
- ▶ Prueba de chi cuadrado.
- ▶ Distribución chi cuadrado.
- ▶ Test exacto de Fisher.
- ▶ Interpretación test no paramétricos con R.

2). Práctica con R y Rstudio cloud.

- ▶ Prácticas R Prueba chi cuadrado y fisher.
- ▶ Realizar gráficas avanzadas con ggplot2.

TABLAS DE CONTINGENCIA

- ▶ 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.

Tratamiento	Respuesta +	Respuesta -
Si	a	c
No	b	d

PRUEBA DE CHI CUADRADO

Esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula.

Hipótesis

H₀: Variable predictora y respuesta son independientes.

H₁: Variable predictora y respuesta NO son independientes.

Supuestos:

- Los datos provienen de una muestra aleatoria.
- 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.

ESTUDIO DE CASO: ABUNDANCIA DE TIBURONES

En la siguiente tabla se muestra la abundancia de tiburones en 2 islas Chilenas.

Especie	Tiburón mako	Tiburón azul
Isla 1	30	80
Isla 2	15	95

Hipótesis

H_0 : La abundancia de tiburones mako y azul es independiente de las islas.

H_1 : La abundancia de tiburones mako y azul es dependiente de las islas.

CÁLCULO DE ESTADÍSTICO CHI CUADRADO

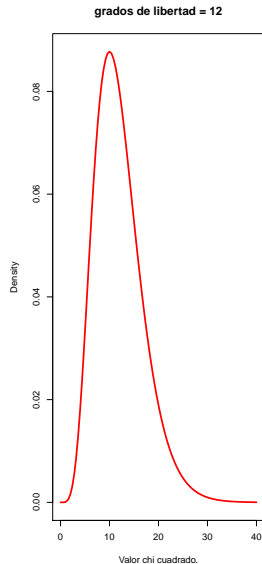
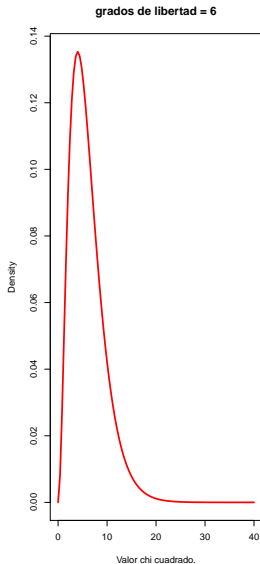
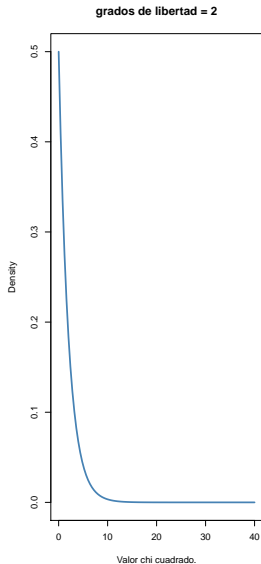
¿Cómo se calcula el estadístico Chi cuadrado?

$$\chi^2 = \sum \frac{(\text{freq. obs.} - \text{freq. esp.})^2}{(\text{freq. esperada})} = \sum \frac{(O - E)^2}{(E)}$$

Frecuencia esperada

```
##          [,1] [,2]  
## [1,] 22.5 87.5  
## [2,] 22.5 87.5  
  
## X-squared  
##  6.285714
```

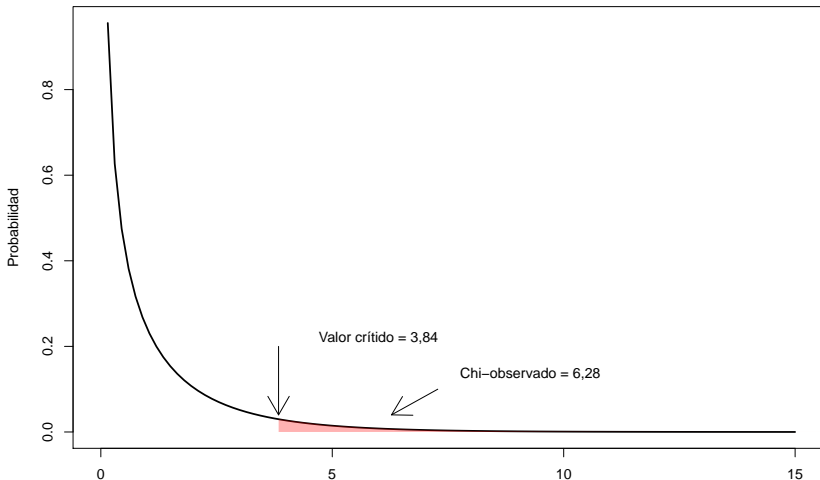
DISTRIBUCIÓN CHI CUADRADO



VALOR CRÍTICO Y DISTRIBUCIÓN CHI-2

Se rechaza hipótesis nula, y por lo tanto concluimos que no hay diferencia en los tratamientos.

Valor crítico para $p=0,05$ y 1 gl = 3,84



PRUEBA DE CHI CUADRADO CON R

```
# Crea matriz de datos
datos <- c(30, 15, 80, 95)
dim(datos) <- c(2,2)
# Test de Chi-squared en R (chisq.test)
chisq.test(datos, correct = FALSE)

##
##  Pearson's Chi-squared test
##
## data:  datos
## X-squared = 6.2857, df = 1, p-value = 0.01217
```

PRUEBA DE FISHER

- ▶ Se utiliza en tablas de contingencia.
- ▶ Se prefiere cuando el número de observaciones es pequeño.
- ▶ Pero es válido para cualquier tamaño de observaciones.
- ▶ 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.

ESTUDIO DE CASO: PREFERENCIA DE ALIMENTACION EN MOLUSCOS

En la siguiente tabla se muestra la preferencia de alimentación de una especie de molusco.

Sexo	Food A	Food B
Macho	7	3
Hembra	1	9

Hipótesis

H₀: La preferencia de alimentos A y B es independiente del sexo.

H₁: La preferencia de alimentos A es mayor en Machos

Fuente: Modificado de Clifford and Taylor, 2008

PROBABILIDAD EXACTA DE FISHER

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

Probabilidad exacta de Fisher.

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

PROBABILIDAD EXACTA DE FISHER

Preferencia de alimentación con totales marginales.

Sexo	Food A	Food B	Total
Macho	7	3	10
Hembra	1	9	10
Total	8	12	20

Probabilidad exacta de Fisher.

$$P(a) = \frac{8! * 12! * 10! * 10!}{7! * 3! * 1! * 9! * 20!}$$

TABLAS DE CONTINGENCIA ALTERNATIVAS

Generadas de todas las permutaciones posibles.

Sexo	A	B	Tot.
M	<u>0</u>	10	10
H	8	2	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>1</u>	9	10
H	7	3	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>2</u>	8	10
H	6	4	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>3</u>	7	10
H	5	5	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>4</u>	6	10
H	4	6	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>5</u>	5	10
H	3	7	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>6</u>	4	10
H	2	8	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>7</u>	3	10
H	1	9	10
<u>Tot.</u>	8	12	20

Sexo	A	B	Tot.
M	<u>8</u>	2	10
H	0	10	10
<u>Tot.</u>	8	12	20

DISTRIBUCIÓN DE PROBABILIDAD

H_0 : La preferencia de alimentos A y B es independiente del sexo.

H_1 : La preferencia de alimentos A es mayor en Machos.

a	$P(a)$
0	0,00036
1	0,00953
2	0,07502
3	0,24006
4	0,35008
5	0,24006
6	0,07502
7	0,00953
8	0,00036

La probabilidad de obtener un valor de a mayor o igual que 7 = $P(7) + P(8) = 0,00953 + 0,00036 = 0,00989$.

Conclusión: Se rechaza H_0 .

PRUEBA EXACTA DE FISHER CON R

```
Prueba_fisher <- matrix(c(7, 3, 1, 9), nrow= 2)

fisher.test(Prueba_fisher, alternative = "greater")

##
##  Fisher's Exact Test for Count Data
##
## data:  Prueba_fisher
## p-value = 0.009883
## alternative hypothesis: true odds ratio is greater than
## 95 percent confidence interval:
##  1.836563      Inf
## sample estimates:
## odds ratio
##  17.27587
```


RESUMEN DE LA CLASE

- ▶ Tablas de contingencia.
- ▶ Prueba de chi cuadrado.
- ▶ Distribución chi cuadrado.
- ▶ Test exacto de Fisher.
- ▶ Interpretación test no paramétricos con R.