

Proves d'Hipòtesis amb Variables Qualitatives

Curs d'Estadística Bàsica per a la Recerca Biomèdica
UEB – VHIR

Miriam Mota-Foix i Alex Sánchez-Pla

Miriam.mota@vhir.org alex.sanchez@vhir.org

QUÈ VEUREM?

1. INTRODUCCIO
2. Contrastos per a una proporció
3. Contrastos per a dues proporcions
4. Anàlisi de taules de contingència
 1. Prova khi-quadrat
 2. Test de Fisher
 3. Test de McNemar

Introducció

- Si la variable d'interès és *categòrica* la forma de resumir-la és mitjançant una *proporció*.
- Podem plantejar diversos tests
 - Sobre una variable (*proves amb proporcions*)
 - La proporció (% d'afectats) coincideix amb un valor donat?
 - La proporció (% d'afectats) és la mateixa en dues poblacions?
 - Sobre dues variables (*khi-quadrat i relacionades*)
 - Hi ha associació entre dues variables categòriques?
 - Hi ha relació entre els valors d'una variable categòrica abans i després d'un tractament?

• Study relating smoking and Cancer

Load data: "dadescancer.csv"

	Smoking X=1	Non smoking X=0	TOTAL
CANCER Y=1	190	87	277
NO CANCER Y=0	60	163	223
TOTAL	250	250	500

0 0	00000000 00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000000 00000000 00000000 00000000	0 0

00000000 00000000 00000000 00000000 00000000 00000000 00000000	0 0
0 0	00000000 00000000 00000000 00000000 00000000 00000000 00000000

00000000 00000000 00000000 00000000 00000000 00000000 00000000	00000000 00000000 00000000 00000000 00000000 00000000 00000000
00000000 00000000 00000000 00000000 00000000 00000000 00000000	00000000 00000000 00000000 00000000 00000000 00000000 00000000

Dades Estadístics Gràfics Models Distributions Eines Ajuda

Resums

Taula de contingències

Mitjanes

Proporcions

Variàncies

Tests no paramètrics

Anàlisi dimensional

Ajustament de models

Taula de doble entrada...

Taula d'entrades múltiples...

Introduir i analitzar una taula de doble entrada...

Taula de doble entrada

Dades Estadístics

Variable de fila (triï'n una)

cancer fumar sexe

Variable de columna (triï'n una)

cancer fumar sexe

Expressió de selecció

<tots els casos vàlids>

Ajuda

Taula de doble entrada

Dades Estadístics

Calcula percentatges

Percentatges de files

Percentatges de columnes

Percentatges del total

Sense percentatges

Test d'hipòtesi

Test Khi-quadrat d'independència

Components de l'estadístic Khi quadrat

Imprimeix les freqüències esperades

Test exacte de Fisher

Ajuda

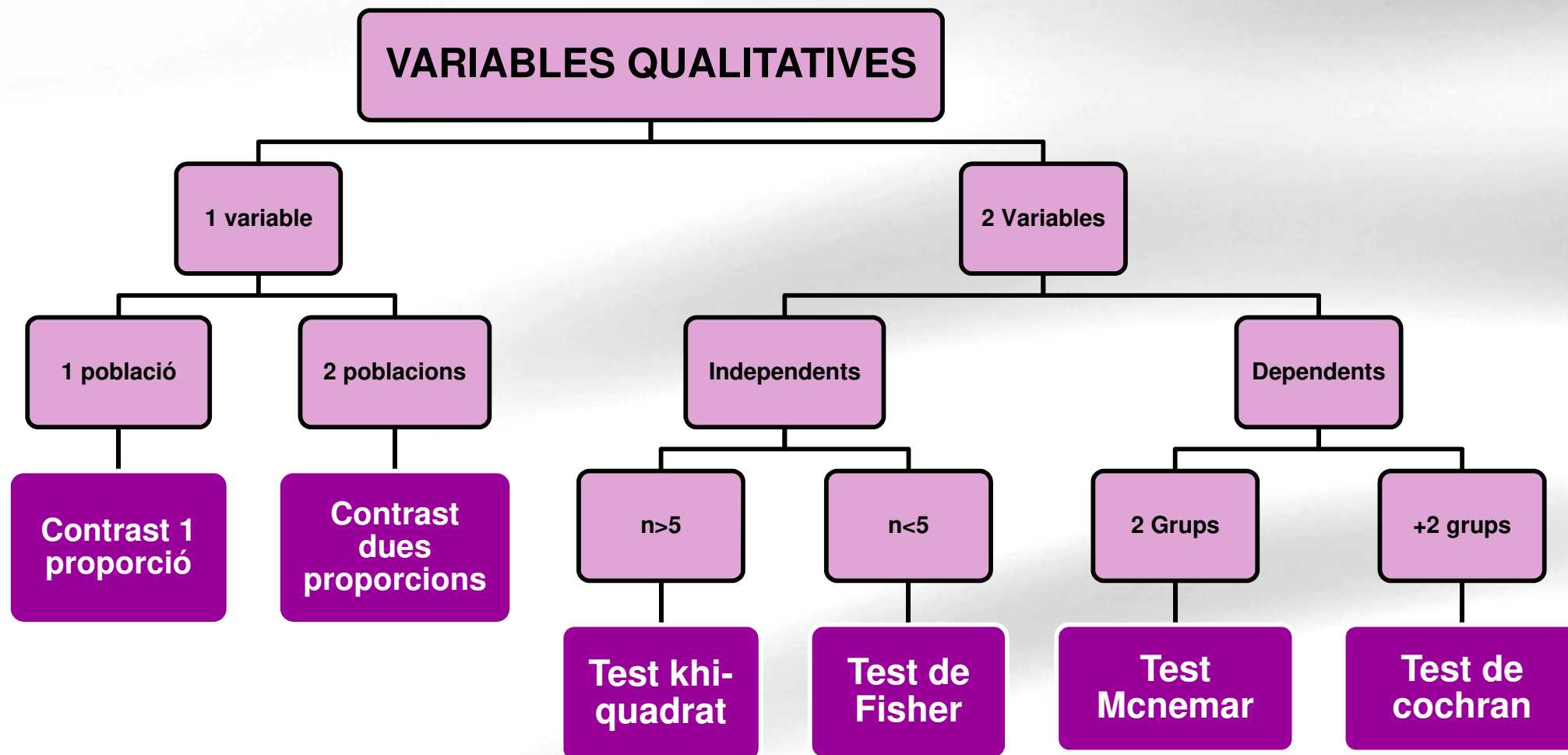
Reset

D'acord

Anulla

Apply

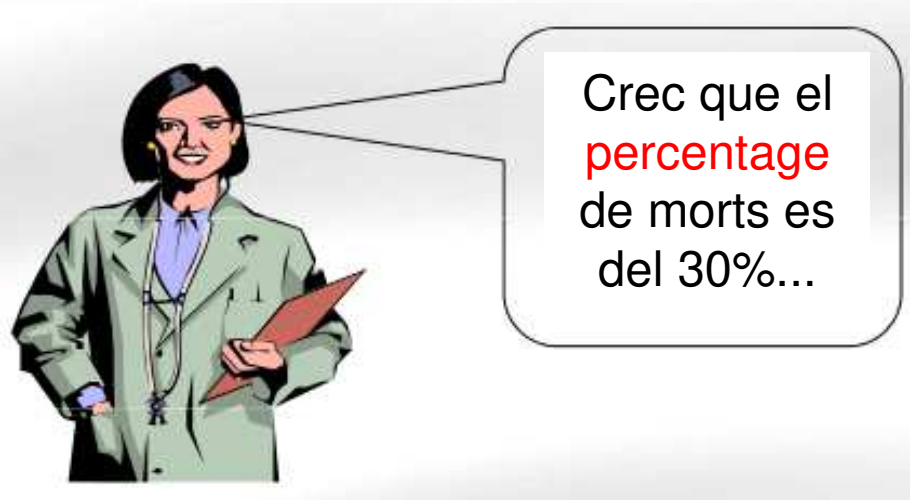
Resum



Contrast per a una proporció

- Objectiu: comprovar si la proporció de successos en una variable categòrica observada en una població s'ajusta a una proporció de referència p_0 .
- Suposarem que es disposa d'una mostra de mida n .
- El contrast es planteja com:
 - $H_0: p = p_0$ vs $H_1: p \neq p_0$
ó
 - $H_0: p = p_0$ vs $H_1: p < p_0$
 - $H_0: p = p_0$ vs $H_1: p > p_0$

Contrast de proporcions amb R



Exercici 1) diabetes.sav:

- Explora les dades diabetes per tenir una idea del percentatge de morts en la nostra mostra.
- Fes el mateix per a la segona hipòtesis.

Contrast de proporcions amb R

- Els tests de proporcions amb R commander necessiten que proporcionem tot el vector de dades.
- La instrucció d'R `prop.test` permet fer-ho amb els comptatges

```
> prop.test(x=25, n=149, p=0.3)
```

```
1-sample proportions test with  
continuity correction
```

```
data: 25 out of 149, null  
probability 0.3
```

```
X-squared = 11.781, df = 1,
```

```
p-value = 0.0005983
```

```
alternative hypothesis: true p is  
not equal to 0.3
```

```
95 percent confidence interval:
```

```
0.1134978 0.2396854
```

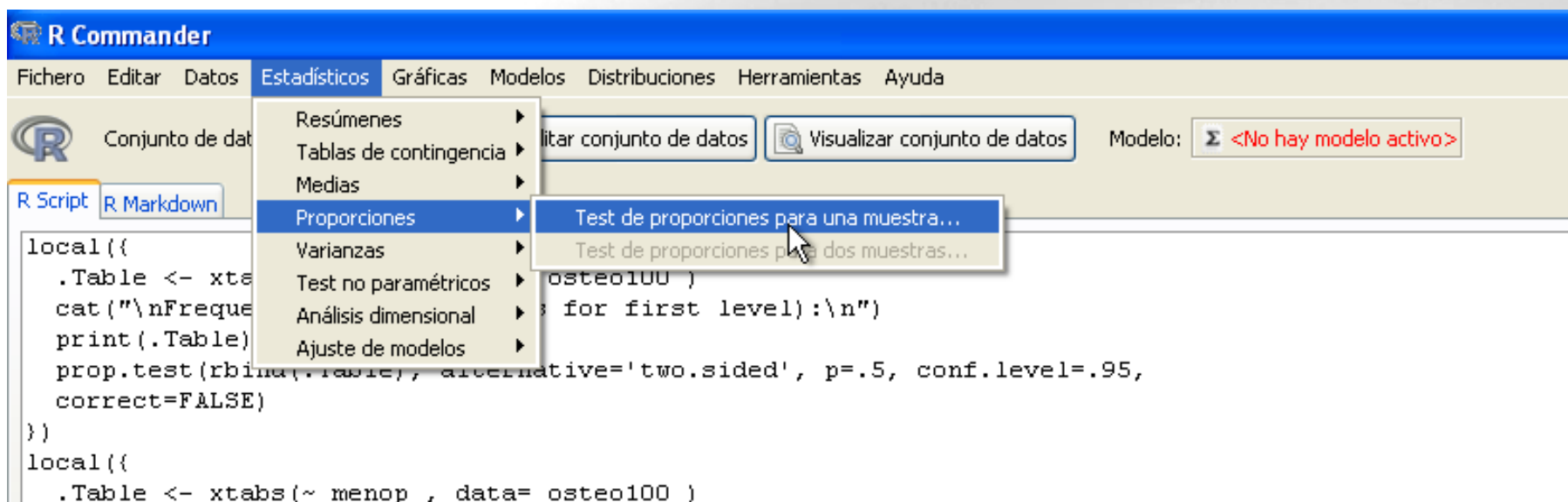
```
sample estimates:
```

```
p
```

```
0.1677852
```

Contrasts de proporcions amb Rcmdr

- Els tests de proporcions amb R commander necessiten que proporcionem tot el vector de dades.
- Proporció morts a la “població” = ?



ALERTA! Com estan endreçats els nivells de les variables?

Datos/Modificar variables del conjunto de datos activo / Reordenar niveles del factor

Contrastos per 2 proporcions

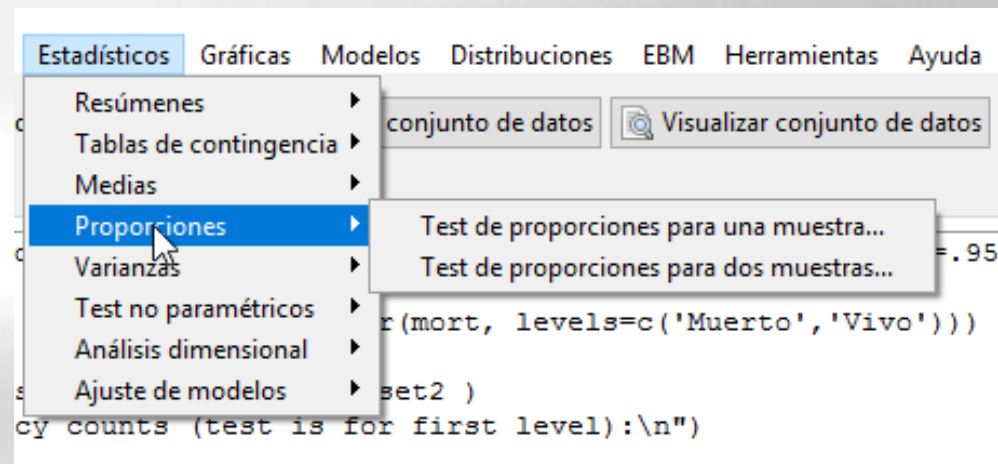
Comparació entre (dues) proporcions

- Objectiu: comprovar si una variable binomial mesurada en dues poblacions diferents presenta la mateixa proporció de successos en totes dues.
- es disposa d'una mostra aleatòria simple per a cadascuna de les poblacions, de grandària respectiva n_1 i n_2 , \tilde{p}_i és la freqüència relativa de successos a la població i .
- El contrast es planteja com:
 - $H_0: p_1 = p_2$ vs $H_1: p_1 \neq p_2$
ó
 - $H_0: p_1 = p_2$ vs $H_1: p_1 < p_2$
 - $H_0: p_1 = p_2$ vs $H_1: p_1 > p_2$

Comparació entre (dues) proporcions



Crec que les
persones amb chd
tenen un
percentatge de
morts different als
que no tenen chd



- Compare % by group $H_0: p_1 = p_2$
- Test z for comparing proportions

Percentage table:

		mort		Total	Count
chd		Muerto	Vivo		
No		13.1	86.9	100	99
Si		24.0	76.0	100	50

Frequency table:

		mort	
chd		Muerto	Vivo
No		13	86
Si		12	38

2-sample test for equality of proportions without continuity correction

data: .Table

X-squared = 2.8105, df = 1, p-value = 0.09365

alternative hypothesis: two.sided

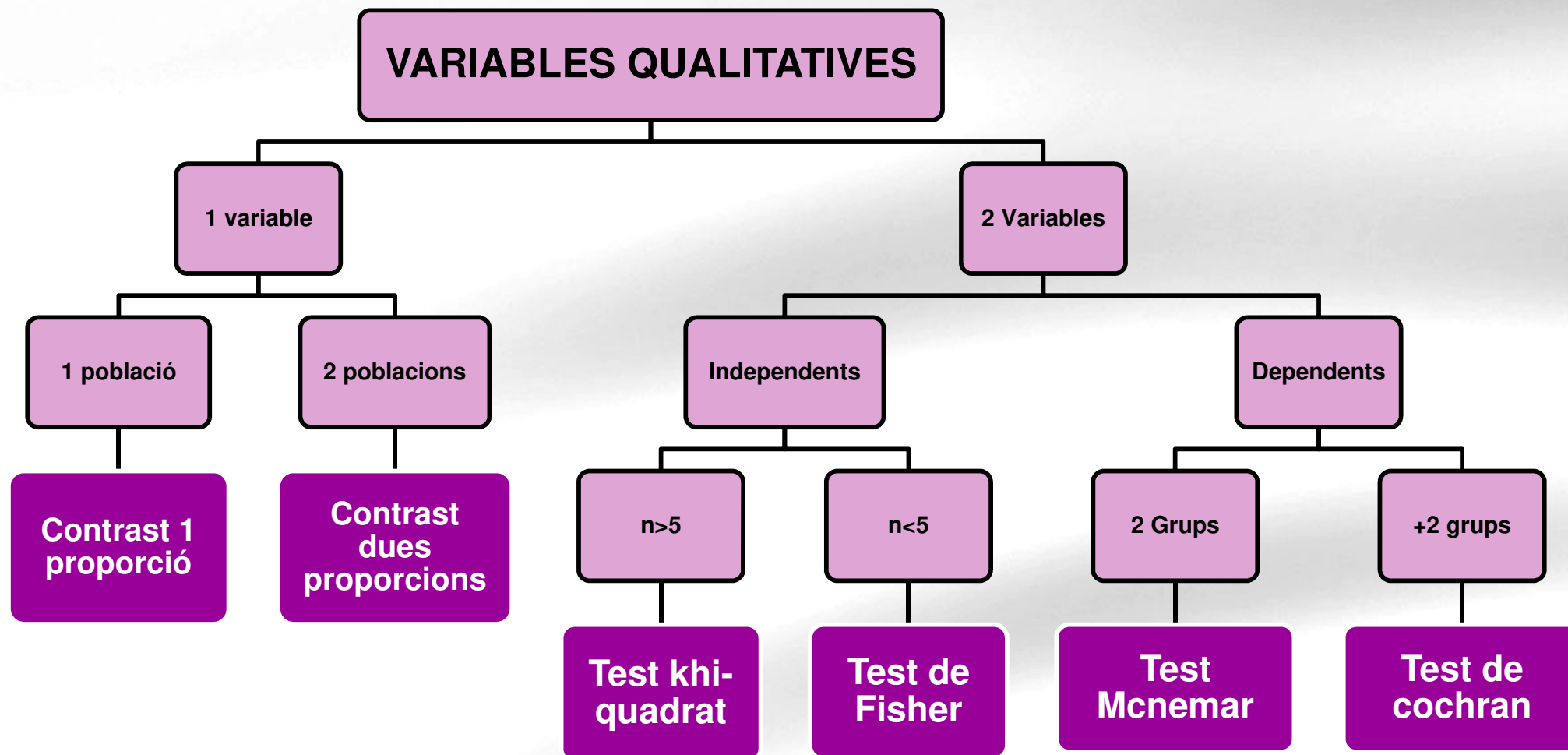
95 percent confidence interval:

-0.24448030 0.02710656

sample estimates:

prop 1	prop 2
0.1313131	0.2400000

Resum



Relació entre variables categòriques

- **taula de contingència** : classificació d'observacions d'acord a 2 característiques qualitatives, una d'elles determina les files i l'altre les columnes.

Clasif	B_1	B_2	...	B_s	Total
A_1	n_{11}	n_{12}	...	n_{1s}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	...	n_{2s}	$n_{2\bullet}$
...	
A_r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$		$n_{\bullet s}$	N

- quan la mostra correspon a **una única població** amb individus classificats d'acord a 2 variables qualitatives, l'objectiu és determinar si existeix relació entre les variables: és una **prova d'independència**.
- quan **cada fila** correspon a una **mostra d'una població diferent** l'objectiu és determinar si les diferents poblacions tenen diferències significatives en la variable estudiada: és una **prova d'homogeneïtat** de poblacions.

Prova d'independència Khi-quadrat

- Objectiu: comprovar en una única població la **possible dependència** de dues variables categòriques A , amb categories (A_1, \dots, A_r) , i B , amb categories (B_1, \dots, B_s) :
- Per tant
 - H_0 : La gravetat del tumor i la mutació **son independents**.
 - H_1 : La gravetat del tumor i la mutació **estan relacionats**
- Aquest test ***requereix que un 80% de les categories tinguin 5 o més observacions***

El test de Fisher

- El test khi quadrat pot ser inexacte amb mostres petites.
- El test de Fisher calcula la probabilitat d'observar el valor de test igual o més gran al observat construint totes les possibles taules
- S'anomena test *exacte* perquè no fa servir una distribució sino que es calcula de nou per cada taula de dades → **Millor opció que la khi-quadrat**

2 variables independents

- Compare H0: Two variables are independent
- Chi squared test

R Two-Sample Proportions Test

Statistics Graphs Models Distributions Tools Help

Summaries
Contingency tables
Means
Proportions
Variances
Nonparametric tests
Dimensional analysis

Two-way table...
Multi-way table...
Enter and analyze two-way table...

Groups (pick one)
grupo
sexo

Response Variable (pick one)
grupo
sexo

Compute Percentages
☐ Row percentages
☐ Column percentages
☒ Percentages of total
☐ No percentages

Hypothesis Tests
☒ Chi-square test of independence
☐ Components of chi-square statistic
☒ Print expected frequencies
☒ Fisher's exact test

Frequency table:
cancer

fumar	Cancer	No cancer
Fuma	190	60
No fuma	87	163

Pearson's Chi-squared test

data: .Table
X-squared = 85.874, df = 1, p-value < 2.2e-16

Expected counts:
cancer

fumar	Cancer	No cancer
Fuma	138.5	111.5
No fuma	138.5	111.5

Fisher's Exact Test for Count Data

data: .Table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
3.945907 8.936465
sample estimates:
odds ratio
5.909114

Calcul 2 variables R

Introducir una tabla de doble entrada

Tabla Estadísticos

Número de filas: 2

Número de columnas: 2

Introducir las frecuencias:

	Canc	No c
Fuma	15	11
No F	20	3

Tabla Estadísticos

Calcular porcentajes

☐ Porcentajes por filas

☐ Porcentajes por columnas

☒ Porcentajes totales

☐ Sin porcentajes

Test de hipótesis

☒ Test de independencia Chi-cuadrado

☒ Componentes del estadístico Chi-cuadrado

☒ Imprimir las frecuencias esperadas

☒ Test exacto de Fisher

Ayuda Reiniciar Aceptar

Pearson's Chi-squared test

data: .Table

X-squared = 5.1212, df = 1, p-value = 0.02363

Fisher's Exact Test for Count Data

data: .Table

p-value = 0.03015

alternative hypothesis: true
odds ratio is not equal to 1

El test de Mc Nemar

- Test de dades aparellades amb dades categòriques.
 - Per exemple si interessa determinar si un individu presenta certa característica abans o després d'una intervenció.
- En un aquest estudi trobem 4 possibles resultats
 - Característica present abans i després de la intervenció
 - Característica absent abans i després de la intervenció
 - Característica present abans i absent després de la intervenció.
 - Característica absent abans i present després de la intervenció.

Taula per al test de McNemar

Condició 1\2

Present

Absent

Present

n_{11}

n_{12}

Absent

n_{21}

n_{22}

Test de Mc Nemar

- H_0 : La probabilitat que la característica estigui present no canvia al fer la intervenció
- H_A : La probabilitat es veu afectada per la intervenció.
- *No es tracta com una taula de contingència sino com un test de proporcions aparellades.*

Exemple Mcnemar

- Subjectes de l'estudi: 165 dones a les que es va demanar d'informar, després d'haver rebut un implant de silicona, si aquest s'havia deteriorat.
 - Aquesta possibilitat es va confirmar quirúrgicament
- Condicions (Cadascuna en totes les dones)
 - Auto informe de Presencia/Absencia de problemes
 - Confirmació quirúrgica de Presencia/Absencia

SELF * SURGICAL Crosstabulation

Count		SURGICAL		Total
		Rupture	No Rupture	
SELF	Rupture	69	28	97
	No Rupture	5	63	68
Total		74	91	165

```
.Table <- matrix(c(69,28,5,63), 2, 2, byrow=TRUE)
mcnemar.test(.Table)
```