RAPPELS PRINCIPES DES TESTS STATISTIQUES ET TESTS PARAMETRIQUES

Comparaison de plusieurs pourcentages observés

Exemple: Effet des Antécédents sur la pathologie digestive, Echantillon de 32 sujets

16 non malades

16 malades

table(DIG,ATCD)

$P_0 = ?\ \%$

$P_1 = ?\ \%$

$P_2 = ?\ \%$

Proportion de pathol. digestive selon les antécédents

prop.table(table(DIG,ATCD),2)

	ATCD0	ATCD1	ATCD2
DIG0	10	5	1
DIG1	0	5	11

	ATCD0	ATCD1	ATCD2
DIG0	1	0,5	0,08
DIG1	0	0,5	0,92

1. Hypothèses:

H0: $P_1=P_2=P_3$ le pourcentage de malades est identique quelque soit les antécédents

H1: 1 ≠ au moins 1 pourcentage est différent

2. Prédictions:

Sous H0 on doit observer P=16/32=50% de pathol. digest.

	ATCD 0	ATCD 1	ATCD 2	
DIG 0	10 5	5 5	1 6	16
DIG 1	0 5	5 5	11 6	16
	10	10	12	32

1. Hypothèses

2. Prédictions

Sous H0 et si les conditions d'application sont respectées

	ATCD 0	ATCD 1	ATCD 2	
DIG 0	10 **5**	5 **5**	1 **6**	16
DIG 1	0 **5**	5 **5**	11 **6**	16
	10	10	12	32

Conditions

- **$C_{ij}>5$**
- Indépendance des individus

$$\chi^2 = \sum \frac{\left(O_{ij} - C_{ij}\right)^2}{C_{ij}}$$

1. Hypothèses

2. Prédictions

3. Confrontation: observation ↔ théorie sous H0

chisq.test(DIG, ATCD, correct=FALSE)

Données

χ^2_0 sous H

ddl

Petit « p »

Conditions d'application:

Où trouver les Cij?

C<-chisq.test(DIG, ATCD, correct=FALSE)

attributes(C)

```
$names
[1] "statistic" "parameter" "p.value"   "method"    "data.name" "observed"
[7] "expected"  "residuals"

$class
[1] "htest"
```

C$expected

```
        ATCD
DIG      0  1  2
    0    5  5  6
    1    5  5  6
```

Conditions d'application:

remarque

chisq.test(DIG, ATCD, ~~correct=TRUE~~)

SI **$3<C_{ij}<5$** !

Pas de correction de continuité de Yates pour plus de 2 pourcentages
il faut...

$\Rightarrow$ regrouper des classes

1. Hypothèses
2. Prédictions
3. Confrontation
4. Interprétation

- $p<0,05$
- Test significatif
- Rejet de H0 au risque $\alpha=5\%$
- Il y a, au moins, une différence entre les 3 pourcentages
- Dans le sens « Les patients ayant plus d'antécédents ont plus de pathologie digestive »

ATTENTION

1. On rejette l'hypothèse nulle $P_1 = P_2 = P_3$

$\Rightarrow$ une égalité au moins est fausse

$\Rightarrow$ il y a au moins une différence

$\Rightarrow$ mais on n'a pas testé laquelle:

$P_1 \neq P_2$ ou $P_1 \neq P_3$ ou $P_2 \neq P_3$ **???**

!

2. On ne peut pas tester **ENSUITE** les moyennes 2 à 2 sinon **α** ↑↑

inégalité de Bonferroni

$$\alpha_{total} \leq \sum_{i=1}^{k\ tests} \alpha_i$$

Exercice

- fichier TABAC.csv
- Y a-t-il une différence pourcentage d'hommes en fonction des antécédents ?

table(SEXE,ATCD)

	ATCD0	ATCD1	ATCD2
S0	6	6	4
S1	4	4	8

prop.table(table(SEXE,ATCD),2)

	ATCD0	ATCD1	ATCD2
S0	0,6	0,6	0,33
S1	0,4	0,4	0,67

Exercice

1. Hypothèses

H0: $P_0=P_1=P_2$, il n'y a pas de différence entre les pourcentages

H1: il y a, au moins, une différence

2. Prédictions, conditions d'application

Sous H0 et si les conditions d'application sont respectées

$$\chi^2 = \sum \frac{\left(O_{ij} - C_{ij}\right)^2}{C_{ij}}$$

Conditions

- **$C_{ij}>5$**
- Indépendance des individus

1. Hypothèses

2. Prédictions

3. Confrontation: observation ↔ théorie sous H0

chisq.test(SEXE, ATCD)

```
 Pearson's Chi-squared test

data:  SEXE and ATCD
X-squared = 2.1333, df = 2, p-value = 0.3442
```

C$expected

```
        ATCD
SEXE  0  1  2
   0  5  5  6
   1  5  5  6
```

1. Hypothèses
2. Prédictions
3. Confrontation
4. Interprétation

- $p>0,05$
- Test non significatif
- Non rejet de H0 au risque β
- On ne met pas en évidence de différence entre les 3 pourcentages d'hommmes

Références

Jean Bouyer: *Méthodes statistiques, Médecine-Biologie*, éditions INSERM

STA UNIV, Pr Jean Gaudart, Faculté de Médecine de Marseille