

# Tests Non Parametriques en R (4)

October 17, 2018

Greta Laage, Luc Adjengue

## Contents

<b>1</b>	<b>Test d'ajustement du <math>\chi^2</math></b>	<b>2</b>
1.1	Étape 1: importer les données et créer un tableau de contingence . . . . .	2
1.2	Étape 2: effectuer le test avec la fonction <code>chisq.test()</code> . . . . .	2
1.3	Deuxième exemple . . . . .	2
<b>2</b>	<b>Tests d'indépendance (tableaux de contingence)</b>	<b>3</b>
2.1	Étape 1 : Définir le tableau de contingence . . . . .	3
2.2	Étape 2 : Effectuer le test sur le tableau . . . . .	3

# 1 Test d'ajustement du $\chi^2$

Le test vise à vérifier si les données dont on dispose proviennent d'une population avec une distribution donnée. Il peut s'agir d'une loi connue (normale, exponentielle, Poisson, géométrique, uniforme, etc.), ou d'une distribution quelconque. La procédure passe par la comparaison des effectifs observés de l'échantillon par rapport aux effectifs attendus si la variable aléatoire suivait une certaine loi.

L'hypothèse testée est donc  $H_0 : X \sim F$  où  $F(\theta)$  est une distribution de paramètre  $\theta$ .

## 1.1 Étape 1: importer les données et créer un tableau de contingence

Supposons par exemple qu'on a des tulipes rouges, jaunes et blanches et on souhaite savoir si la couleur des tulipes suit une distribution uniforme. On a 81 tulipes rouges, 50 jaunes et 27 blanches.

Si la couleur était uniformément distribuée, alors la proportion attendue pour chaque couleur serait  $1/3$ ,  $1/3$  et  $1/3$ .

```
In [1]: d7 <- c(81,50,27)
```

## 1.2 Étape 2: effectuer le test avec la fonction `chisq.test()`

On précise le vecteur de probabilités attendus suivant la loi uniforme.

```
In [2]: res7 <- chisq.test(d7,p = c(1/3,1/3,1/3))
      res7
```

Chi-squared test for given probabilities

```
data: d7
X-squared = 27.886, df = 2, p-value = 8.803e-07
```

La valeur-p est très proche de 0, donc au seuil de 5%, on peut dire que la couleur des tulipes ne suit pas une distribution uniforme.

## 1.3 Deuxième exemple

```
In [3]: # Exemple 10.26 du livre: on teste si la loi uniforme est un bon modèle.
      O = c(94,93,112,101,104,95,100,99,108,94)
      E = c(100,100,100,100,100,100,100,100,100,100)
```

```
In [4]: chisq.test(O)
```

Chi-squared test for given probabilities

```
data: O
X-squared = 3.72, df = 9, p-value = 0.9289
```

## 2 Tests d'indépendance (tableaux de contingence)

On teste si deux variables aléatoires sont indépendantes. On utilise la fonction `chisq.test()` de R. Le test se fait en plusieurs étapes:

- Importer les données
- Créer un tableau de contingence
- Utiliser la fonction `chisq.test()`

Les données peuvent être importées depuis un fichier csv ou autre. La fonction **`chisq.test()`** prend en paramètres le tableau de contingence pour les tests d'indépendance de deux variables aléatoires.

### 2.1 Etape 1 : Définir le tableau de contingence

```
In [2]: # Données de l'exemple 10.29 du livre.  
d = data.frame(regime1 = c(160,40),  
               regime2 = c(140,60),  
               regime3 = c(40,60),  
               row.names = c("Cadres", "Employés"))
```

### 2.2 Etape 2 : Effectuer le test sur le tableau

La fonction `chisq` donne la statistique  $u_0$  du test et la p-value à partir desquelles on peut conclure si on rejette ou non l'hypothèse  $H_0$ .

```
In [5]: test = chisq.test(d)  
test
```

Pearson's Chi-squared test

```
data: d  
X-squared = 49.632, df = 2, p-value = 1.669e-11
```

Observations: On retrouve bien la valeur de la statistique  $u_0$  du livre.

La valeur-p est très petite (inférieure au seuil) donc les variables en ligne et en colonne sont significativement associées (ie non indépendantes), au sens statistique au seuil de 5%.

```
In [7]: # On peut accéder aux comptes observés et attendus:  
test$observed
```

	regime1	regime2	regime3
Cadres	160	140	40
Employés	40	60	60

```
In [8]: test$expected
```

	regime1	regime2	regime3
Cadres	136	136	68
Employés	64	64	32