

Analyses de fréquences

Ce guide présente la marche à suivre pour réaliser les analyses de fréquences prévues à la Leçon 5.

Les points traités sont les suivants :

- [Test de khi-carré](#) ;
- [Tableau de contingence](#) ;
- [Résidus de Pearson](#) ;
- [Détermination des fréquences observées et théoriques](#) ;
- [Correction de continuité de Yates](#).

R Commander offre des approches différentes selon que l'on souhaite réaliser le test de khi-carré ou le tableau de contingence. Une approche unique sera préconisée dans ce guide afin de simplifier la compréhension de la matière.

Test de khi-carré

La présentation du test de khi-carré se base sur le deuxième cas de l'Exemple 1 sur [Wikipedia](#) sur la détermination de l'équilibrage d'un dé.

1) On crée tout d'abord le fichier de données à l'aide de l'approche présentée dans le guide [Création et édition d'un jeu de données](#). Le fichier de données est `des` ;

2) À l'onglet [Script R](#), veuillez saisir ou copier-coller la commande suivante :

```
out <- chisq.test(x=des, p = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)); out
```

où :

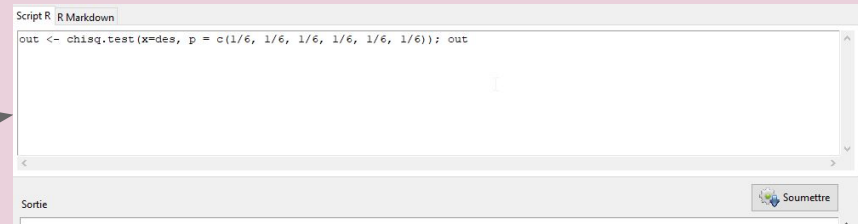
- `out` : objet où sont sauvegardés les résultats ;
- `chisq.test` : fonction qui permet de réaliser le test ;
- `des` : nom du jeu de données ;
- `p = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)` : hypothèses à tester ;

3) Une fois cette commande saisie, veuillez la surligner. Cliquez ensuite sur le bouton [Soumettre](#) afin de soumettre la commande ;


Les résultats du test apparaîtront dans la zone [Sortie](#). On peut conclure que le dé est truqué.

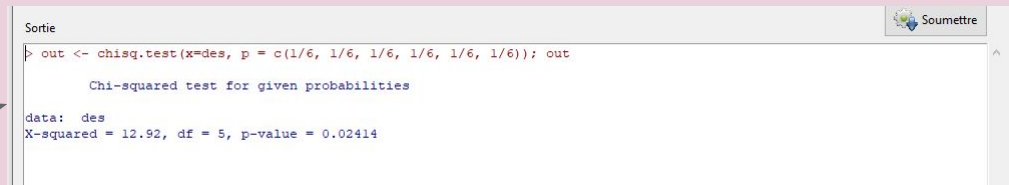


	Face.1	Face.2	Face.3	Face.4	Face.5	Face.6
Effectif	89	131	93	92	104	91



```
Script R R Markdown
out <- chisq.test(x=des, p = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)); out
```

Sortie 



```
Sortie
> out <- chisq.test(x=des, p = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)); out

Chi-squared test for given probabilities

data:  des
X-squared = 12.92, df = 5, p-value = 0.02414
```

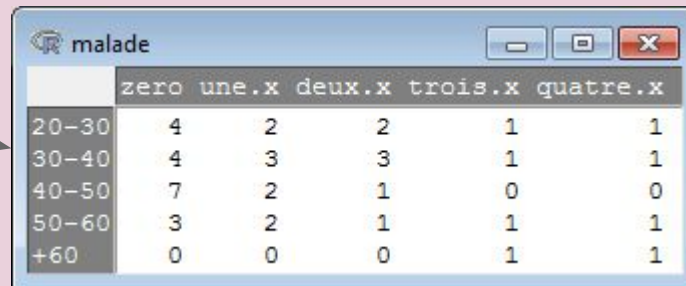
Tableau de contingence

Le test du tableau de contingence est le même que celui du test de khi-carré.
Un exemple tiré de [Wikipédia](https://fr.wikipedia.org/wiki/Chi-carr%C3%A9) sera utilisé pour présenter la marche à suivre.

Le jeu de données présente le nombre de fois que l'on tombe malade en fonction de l'âge :

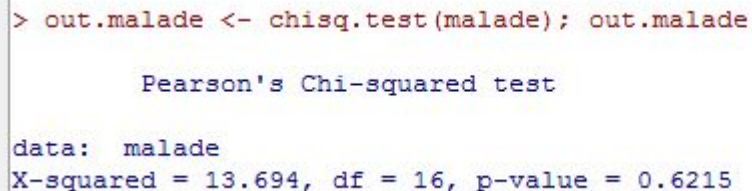
On obtient les résultats de ce test en soumettant les commandes suivantes :

```
out.malade <- chisq.test(malade)  
out.malade
```



A screenshot of a R console window titled "malade". It displays a contingency table with age groups as rows and the number of times one gets sick as columns. The columns are labeled "zero", "une.x", "deux.x", "trois.x", and "quatre.x". The rows are labeled "20-30", "30-40", "40-50", "50-60", and "+60".

	zero	une.x	deux.x	trois.x	quatre.x
20-30	4	2	2	1	1
30-40	4	3	3	1	1
40-50	7	2	1	0	0
50-60	3	2	1	1	1
+60	0	0	0	1	1



A screenshot of the R console output for the command `out.malade <- chisq.test(malade); out.malade`. The output shows the results of a Pearson's Chi-squared test, including the data name, the chi-squared value, degrees of freedom, and the p-value.

```
> out.malade <- chisq.test(malade); out.malade  
  
Pearson's Chi-squared test  
  
data:  malade  
X-squared = 13.694, df = 16, p-value = 0.6215
```

On peut observer que les différences avec le test de khi-carré réalisé plus tôt consistent en un changement du jeu de données (`malade`) et l'absence d'hypothèse. Il n'était pas nécessaire d'inscrire d'hypothèse étant donné qu'il y en a pas de spécifiée.

Résidus de Pearson

Les résidus de Pearson sont calculés par défaut par la fonction `chisq.test()`. On accède à ces valeurs en soumettant une commande qui comprend :

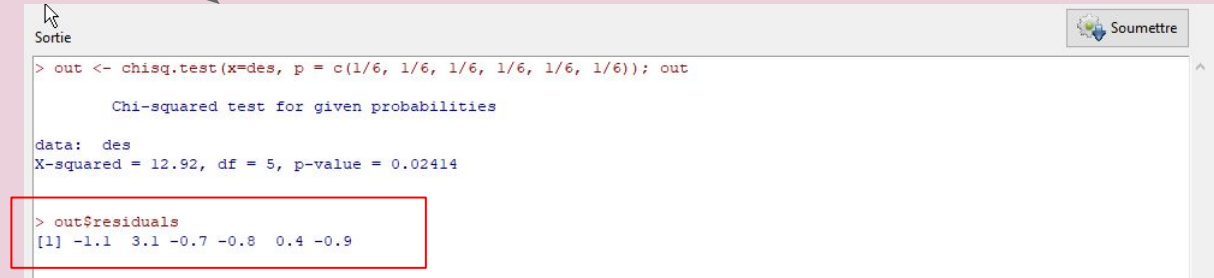
- L'objet où sont sauvegardés les résultats du test de khi carré ;
- L'opérateur \$;
- La composante `residuals`.

À partir de l'exemple sur détermination de l'équilibrage d'un dé présenté plus tôt, la commande à soumettre est la suivante :

```
out$residuals
```

Les résidus apparaissent dans la zone Sortie.

On remarque à la deuxième position que le résidu est supérieur à 2. Cette position correspond à la Face.2 du dé. On peut considérer que cette face survient plus souvent que les autres.



```
> out <- chisq.test(x=des, p = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)); out

      Chi-squared test for given probabilities

data:  des
X-squared = 12.92, df = 5, p-value = 0.02414

> out$residuals
[1] -1.1  3.1 -0.7 -0.8  0.4 -0.9
```

Détermination des fréquences observées et théoriques

Il est parfois nécessaire de connaître les fréquences observées et théoriques afin de pousser plus loin l'interprétation des résultats ou pour vérifier l'une des conditions d'utilisation. La présentation de la marche à suivre pour obtenir ces fréquences se base sur l'Exemple 5.3 des notes de cours (Leçon 5, p. 8).

1) Il faut tout d'abord réaliser le test. Dans le cas de l'Exemple 5.3, la commande est la suivante :


```
out.select <- chisq.test(select, p = c(0.340, 0.101, 0.104, 0.455))  
out.select
```

2) On obtient les fréquences observées en soumettant la commande qui suit :

```
out.select$observed
```

où :


`out.select` : nom de l'objet contenant les résultats du test;
`observed` : nom de la variable contenant les valeurs des fréquences observées.



```
> out.select$observed  
[1] 25 22 30 40
```

3) On obtient les fréquences théoriques en soumettant la commande suivante :

```
out.select$expected
```



```
> out.select$expected  
[1] 39.780 11.817 12.168 53.235
```

Correction de continuité de Yates


On applique la correction de continuité de Yates lorsque le test a un degré de liberté. L'Exemple 5.7 des notes de cours (Leçon 5, p. 16) sera utilisé pour présenter la marche à suivre.

On obtient les résultats du test corrigé en soumettant les commandes suivantes :

```
out.tab <- chisq.test(x = freq.obs,  
correct = TRUE)
```

```
out.tab
```

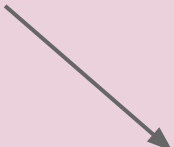
On a ajouté l'option `correct = TRUE` afin d'indiquer que la correction doit être appliquée.



```
> out.tab <- chisq.test(x = freq.obs, correct = TRUE); out.tab  
  
Pearson's Chi-squared test with Yates' continuity correction  
  
data: freq.obs  
X-squared = 33.112, df = 1, p-value = 8.7e-09
```

Comme mentionné dans les notes de cours (p. 18 de la Leçon 5), la commande `chisq.test()` n'applique la correction de continuité de Yates que sur les tableaux de contingence de 2 x 2. Il existe quand même des situations où une telle correction est nécessaire. Il est alors possible de la réaliser à l'aide de la commande suivante :

```
khi.corr <- sum(((abs(out.tab$observed - out.tab$expected) -  
0.5)^2)/out.tab$expected)  
khi.corr
```



```
> khi.corr <- sum(((abs(out.tab$observed - out.tab$expected) - 0.5)^2)/out.tab$expected)  
  
> khi.corr  
[1] 33.11197
```

Cette commande est semblable à celle présentée à la page 13 des notes de cours. On a simplement remplacé le nom de l'objet où se trouvent les résultats du test (`khi2` → `out.tab`)