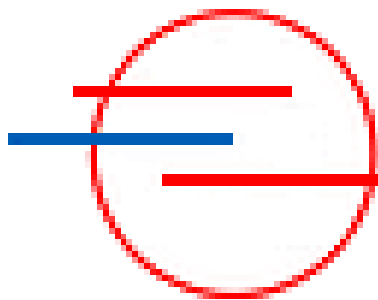


<https://adjectif.net.shs.parisdescartes.fr/spip.php?article176>



# Construire une table de contingence avec R

- Outils et méthodologies - Outils et techniques de recherches -



Outils  
&  
Méthodo

Date de mise en ligne : lundi 27 août 2012

---

Copyright © Adjectif - Tous droits réservés

---

[Dans cet article](#) paru sur Adjectif, plusieurs types d'instruments logiciels permettant de traiter des données quantitatives ont été présentés. Cette contribution explique comment créer des tableaux de contingence avec le logiciel R.

Comment décrire le comportement de deux variables qualitatives l'une par rapport à l'autre ?

Une façon de procéder est de produire des tableaux à double entrée. On verra ainsi comment deux variables qualitatives se répartissent l'une par rapport à l'autre : on aura une vue des relations entre les modalités de chaque variable. Sous certaines conditions nous pourrions appliquer un test de khi-deux de contingence qui nous aidera à nous prononcer sur l'existence d'un lien statistique entre ces deux variables qualitatives. Nous allons donc voir dans ce document comment construire simplement avec R des tableaux croisés appelés aussi tableaux à double entrée ou tables de contingence. *Un article ultérieur explicitera le type de résultats qu'on peut en obtenir.*

### Exemple 1 : Comparer les effets de deux médicaments

Pour comparer l'efficacité de deux médicaments agissant sur la même maladie, mais aux prix très différents, une enquête a été effectuée sur les guérisons obtenues en suivant chacun des traitements. Les résultats sont consignés dans le tableau ci-dessous.

[<https://adjectif.net.shs.parisdescartes.fr/IMG/png/100000000000024000000061A3B545DF.png>]

En examinant la première colonne du tableau, on voit que sur 50 patients ayant pris le médicament cher, 44 ont guéri. Dans la deuxième colonne, 156 sur 200 patients ont guéri. Mêmes choses pour les lignes : sur les 200 patients guéris, 156 ont pris le médicament bon marché, etc.

### Exemple 2 : Dans quelles disciplines des enseignants font-ils faire des recherches sur internet aux élèves ?

On a posé les deux questions suivantes à 179 enseignants « Faites-vous faire des recherches sur internet à vos élèves ? » et « Quelle discipline enseignez-vous ? ». On a ensuite regroupé les réponses à la deuxième question en deux groupes : les disciplines pour lesquels les élèves reçoivent le plus d'heures de cours par semaine, que l'on appelle « disciplines importantes » et les « disciplines secondaires » pour lesquels les élèves reçoivent le moins d'heures de cours.

On obtient deux variables qualitatives qui se répartissent comme présenté sur le tableau ci-dessous.

[<https://adjectif.net.shs.parisdescartes.fr/local/cache-vignettes/L264xH64/10000000000012C00000049E515EA1A-7926e.png>]

En colonnes figure le nombre de réponses à la question « Faites-vous faire des recherches sur internet à vos élèves ? ».

On voit que ce type de tableau présente des informations précises et nous renseigne sur la structure des données.

## Comment produire des tableaux croisés avec R ?

Lorsque l'on traite des données quantitatives, la base dans laquelle figurent les variables se présente le plus souvent avec les individus en ligne et les variables en colonne :

utilordiclasse	utilordielev	discipline	age	anaiss						
une_fois_par_sem_ou_moins	oui	lettres	[30_40[	1970	une_fois_par_sem_ou_moins	non	math	[50_et_+[	1952	
plus_d_une_fois_par_sem	oui	hg	[30_40[	1977						
plus_d_une_fois_par_sem	oui	autres	[30_40[	1980						
plus_d_une_fois_par_sem	oui	langues	[20_30[	1982						
...	...	...	...	...						

Dans ce cas précis, chaque ligne correspond aux réponses d'un individu et chaque colonne aux réponses de tous les individus à une question. La première ligne est réservée aux noms de variables comme « utilordiclasse », « utilordielev », « discipline », etc.

Dans un tableur, ce même fichier se présente ainsi :

[<https://adjectif.net.shs.parisdescartes.fr/IMG/png/10000000000001EB000000BA4771D755.png>]

Ce qui nous intéresse ici, c'est de pouvoir lire en un coup d'oeil combien d'individus sont concernés par deux modalités. On veut par exemple connaître le nombre d'individus appartenant à la classe d'âge 30-40 ans (4ème colonne) et qui a répondu « oui » à la question « Faites-vous faire des recherches sur internet à vos élèves ? » (2ème colonne). Voyons comment procéder avec R.

## Identifier le répertoire de travail de R

Pour importer des données dans R, il est très commode d'identifier le répertoire de travail de R, c'est à dire le dossier par défaut où R lira et écrira des fichiers. Pour le connaître, il suffit, après avoir lancé R, de taper la commande `getwd()`, R vous répondra le nom du dossier. Je vous invite maintenant à déposer à cet endroit une petite base au [format.csv](#), dans laquelle figure des réponses à 4 questions d'un questionnaire rempli par 173 enseignants de collège que vous pouvez télécharger [en cliquant sur ce lien](#).

## Importer les données

Si vous ouvrez le fichier `bdd.csv` que vous venez de télécharger avec un éditeur de texte (*bloc note*, *notepad*, *notepad++*, *gedit*, *smultron*, etc.) vous remarquerez que la première ligne correspond aux noms de variables et que les colonnes sont séparées par des « ; ».

Pour importer les données dans R, nous allons utiliser la commande `read.csv2()` qui permet de lire les documents de type « .csv » ou « .txt ». Pour que cette commande fonctionne bien il est nécessaire de préciser certains éléments comme le séparateur de colonnes et la présence des noms de variables sur la première ligne du tableau.

La commande sera donc :

```
tableaubrut <- read.csv2('bdd.csv', sep=';', header=TRUE)
```

Où :

- 1) 'bdd.csv' est le nom du fichier qui se trouve dans le répertoire de travail de R,
- 2) sep=';' » précise le séparateur de champ,
- 3) header=TRUE indique que la première ligne du tableau est constituée par des noms de variables,
- 4) tableaubrut est l'objet dans lequel on met le tableau.

## Construire une table de contingence avec R

Après avoir exécuté cette commande on tape dans R « `tableaubrut` », R affiche le tableau d'origine, ça y est les données sont importées nous allons pouvoir les manipuler.

### Manipuler les lignes et les colonnes d'un tableau

R est adapté pour manipuler des fragments de tableaux grâce à l'utilisation de crochets [ ].

Si l'on veut isoler uniquement la valeur de la première case de la première colonne du tableau contenu dans l'objet « `tableaubrut` » on tape la commande suivante : `tableaubrut[1,1]`

R permet d'isoler des éléments de cet objet en utilisant le crochet [ ] cette opération s'appelle l'indexation. On peut l'appliquer très souvent et elle permet de manipuler des données importantes et complexes avec économie et fiabilité. On peut mettre cette valeur dans un objet comme ceci par exemple :

```
tableaubrut[1,1]»valeurpremierelignepremierecolonne
```

ou comme cela :

```
tableaubrut[1,1]»valeur.premiere.ligne.premiere.colonne
```

Si on voulait isoler toutes les valeurs de la première ligne du tableau contenu dans l'objet `tableaubrut`, on utiliserait la commande suivante :

```
tableaubrut[1,]
```

Pour isoler toutes les valeurs de la première colonne :

```
tableaubrut[,1]
```

Même chose pour la deuxième colonne :

```
tableaubrut[,2]
```

ou pour la deuxième ligne :

```
tableaubrut[2,]
```

### Créer le tableau

Pour créer le tableau on va isoler la première colonne dans un objet et la deuxième dans un second objet :

```
tableaubrut[,1]->premierecolonne
```

```
tableaubrut[,2]->deuxiemecolonne
```

On va ensuite utiliser la commande `table()` qui permet de construire un tableau croisé :

```
table(premierecolonne , deuxiemecolonne)
```

Mais rien n'oblige à créer des objets intermédiaires on aurait aussi bien pu créer le tableau directement ainsi :

```
table(tableaubrut[,1] , tableaubrut[,2])->tableaucontingence
```

En résumé deux lignes suffisent pour créer un tableau croisé avec R :

```
read.csv2('bdd.csv', sep=';', header=TRUE)->tableaubrut
```

```
table(tableaubrut[,1] , tableaubrut[,2])->tableaucontingence
```