

Quelques fonctions R

Robin J. Ryder

Pour toutes ces fonctions, n'hésitez pas à utiliser l'aide de R : `help(foo)` ou `?foo`. Cela vous donnera d'une part la syntaxe de la fonction, d'autre part toutes les options.

1 Lecture de données

- `file.choose()` : Permet de naviguer dans l'arborescence de fichiers pour obtenir le chemin vers votre fichier de données.
- `read.table()`, `read.csv()` : Lire un fichier tableur ou CSV. Parmi les options utiles : `header`, `sep`, `dec`, `skip`
- `scan()` : pour lire les fichiers texte sans structure particulière. Demande plus de travail.

2 Gestion de données

- `as.character()`, `as.numeric()`... : pour convertir d'un format à un autre
- `grepl()` : pour chercher un motif à l'intérieur d'une chaîne de caractères
- `merge()` : pour fusionner deux bases de données qui ont des noms de lignes identiques
- `rbind()`, `cbind()` : pour superposer ou accoler deux matrices ayant respectivement le même nombre de colonnes ou de lignes

3 Étude simples de données

- `attach(BDD)` : permet d'accéder plus facilement aux colonnes de la base BDD. Par exemple, plutôt que `BDD$NomDeColonne`, on pourra alors utiliser simplement `NomDeColonne`
- `head()` : donne les premières entrées d'un vecteur ou d'une matrice
- `summary()` : donne quelques statistiques résumées d'un vecteur : moyenne, quantiles...
- `table(x,y)` : donne un tableau croisé des valeurs prises par les vecteurs `x` et `y`
- `hist()` : pour faire un histogramme

- `lm(y ~ x1 + x2)`, `glm()` : pour faire une régression linéaire (y expliqué par x_1 et x_2), ou un modèle linéaire généralisé. Il est souvent utile d'appliquer la fonction `summary()` au résultat pour une présentation facile à interpréter.

4 Tests statistiques usuels

- `t.test()` : test de Student
- `chisq.test()` : test du χ^2
- `binom.test()`, `prop.test()` : test binomial
- `wilcox.test()` : tests de Wilcoxon et de Mann-Whitney
- `cor.test()` : test de corrélation
- `ks.test()` : test de Kolmogorov-Smirnov

5 Sauvegarde de données

- `write.csv()` : pour créer un fichier `.csv` contenant une base de données
- `save()` : pour enregistrer toutes les variables créées dans R.
- `load()` : pour charger des variables sauvées avec `save()`