

MSS1

Codage en R

Pour tout problème : la commande `help(nom de la fonction qu'on recherche)`

Affecter une valeur : `x← valeur` Quand on appelle la valeur le numéro entre crochet est le numéro de l'indice en première ligne

Faire une séquence :

- commande `seq()`, s'utilise pour afficher une séquence de nombres avec un début et une fin précisée, `seq(from= , to= , by=)` (début, fin, pas). Exemple `seq(from=1, to=10, by=2)` affiche 1 3 5 7 9
- commande `c()` qui affiche la séquence de nombre entre parenthèses. Exemple `c(5, 6, 3)` affiche 5 6 3
- commande `rep()`, s'utilise pour faire une répétition de nombre ou de séquence. Exemple `rep(c(1,5,3),3)` affiche 1 5 3 1 5 3 1 5 3

Loi normale :

- `rnorm()`, r pour random. Affiche au hasard une distribution qui suit une loi normale, `rnorm(n, mean= ,sd=)` (nombre de données, moyenne, variance). Exemple `rnorm(150, mean=180, variance=5)` affiche une séquence de 150 valeur de moyenne 180 et de variance 5
- `dnorm()`, d pour density. Affiche la densité de probabilité des données, `dnorm(n, mean= ,sd=)`
- `qnorm()`
- `pnorm()`

Affichage :

- Histogramme : commande `hist()`
`hist(x)` affiche un nombre de valeurs calculées par défaut avec le log de l'effectif
Pour changer le nombre de colonnes affichées dans l'histogramme on utilise `breaks=nombre de colonnes qu'on veut`, R en affiche un nombre approximatif autour du nombre choisi. Exemple : `hist(x, breaks=20)` affiche une vingtaine de colonnes
Titre, sous-titre et nom des axes : `hist([...], main= « ... », sub= « ... », xlab= « ... », ylab= « ... »)`
Mettre une couleur : `hist([...], col= « ... »)`
- Afficher nuage de points ou courbes : commande `plot()`,
Pour tracer en points : `plot(x)`
Pour relier les points : `plot(x, type= « b »)` ou `plot(x, type= « o »)`
Pour tracer que des lignes : `plot(x, type= « l »)`
Titre, sous-titre et nom des axes : `plot([...], main= « ... », sub= « ... », xlab= « ... », ylab= « ... »)`
Mettre une couleur : `plot([...], col= « ... »)`
- Boîte à moustache : commande `boxplot()`

Prendre des données d'un fichier :

- Changer le répertoire courant : Fichier → « changer le répertoire courant » → se placer dans le répertoire du fichier (attention qu'il soit bien visible non zippé etc)
- Prendre les données d'un fichier `read.csv(" filename ", header=T, sep='\"t\"')`, pour des fichiers de type csv qu'on utilisera principalement. **Header** sert à signaler qu'il y a une ligne de titres qui ne sera donc pas prise en compte en tant que données telles quelles, **sep** sert à signaler par quoi les données sont séparées ici une tabulation.

- Afficher seulement quelques lignes et/ou colonnes → `doc[lignes, colonnes]`
- Afficher le haut ou le bas d'un document : commande `head()` ou
- Enlever les NA : commande `na.omit(doc)` /\ il faut le faire après avoir sélectionné les données qu'on voulait traiter sinon ce sont toutes les lignes contenant des NA qui sont supprimées même si cela ne concerne pas ce qu'on veut traiter
- Pour afficher min, max, quantiles, moyenne, médiane et effectif : commande `summary(doc)`

■

Pour renommer des classes qui sont pareilles mais n'ont pas le même nom par faute de frappe ou autre on fait, on applique à la commande `level(tbl)←` une concaténation `c()` en donnant le nom à chaque élément par ex : s'il y a pour des couleurs 'Marron', 'Marron ' et 'Vert' on fera `levels(tbl)← c(« M », «M », « V »)` on renomme ainsi la catégorie Marron en M en sommant les deux catégories (puisque c'est les mêmes)

`table(tbl)` donne le nombre d'individus par classe.

`chisq(deux variables)` permet d'avoir le chi des deux variables (connaitre l'indépendance des variables par rapport à H_0)