

TP-1 : Introduction à R

La première séance est dédiée à la prise en main du logiciel libre R de traitement de données et d'analyse statistiques.

1 Création et enregistrement d'un fichier R

- Connectez-vous sous **Linux**.
- Créez dans votre espace personnel (répertoire `/S`) un dossier `AnalyseDonnees` dans lequel vous mettrez tous vos fichiers R et les fichiers de données.
- Ouvrir **RStudio**, allez dans "File/New file" puis cliquez sur **R Script**.
- Allez immédiatement dans "File/Save as" et sauvegardez le fichier vide sous `TP1.r` dans le dossier `AnalyseDonnees`.

2 Quelques règles de base...

Pour effacer toutes les variables de la console, tapez

```
rm(list=ls())
```

Pour introduire une ligne de commentaire dans votre fichier, commencez la ligne par un `#` :

```
# Votre commentaire
```

Pour exécuter votre fichier, tapez la commande "**Ctrl + Alt + R**".

Pour affecter à la variable `a` la valeur `2`, tapez

```
a=2
```

On notera que, lors de l'exécution de cette instruction, le contenu de la variable n'est pas affiché. Pour afficher la valeur de `a`, tapez

```
a
```

ou

```
print(a)
```

Pour affecter à `x` le vecteur ligne `(1, 2, 3, 4, 5, 6)`, tapez

```
x=c(1,2,3,4,5,6)
```

Pour afficher le deuxième élément du vecteur `x`, tapez

```
x[2]
```

Pour afficher les éléments entre les positions `3` et `5`, tapez

```
x[3:5]
```

Pour accéder à l'aide d'une commande, il suffit d'utiliser la commande `help`. Par exemple pour obtenir l'aide relative à la commande `combine`, tapez :

```
help(c)
```

Pour effectuer une recherche dans l'aide, il suffit de mettre `??` devant le mot recherché :

```
??combine
```

3 Un peu de statistique

1. Affectez la valeur 2 à la variable a , 3 à la variable b puis la somme de ces deux variables à la variable c . Affichez c .
2. Créez le vecteur $E = (0.1, 0.12, 5.1, 5.7, 6.2, 11.9)$.
3. Calculez la moyenne de E sur votre feuille ou à l'aide d'une calculatrice.
4. Calculez la moyenne $Ebar$ de E en utilisant les fonctions `sum` et `length`.
5. Calculez la moyenne moy de E en utilisant la fonction `mean`.
6. Que donne la commande `E*E`?
7. Calculez la somme $S2$ des carrés des observations de E .
8. On suppose ici que E est une population, calculez la variance $sigma2$ et l'écart-type $sigma$.
9. On suppose ici que E est un échantillon. Calculez la variance empirique $hatsigma2$ et l'écart-type empirique $hatsigma$.
10. Utilisez maintenant les commandes `var` pour calculer la variance et `sd` pour calculer l'écart-type. Obtient-on le résultat pour une population ou pour un échantillon ?
En fait, les fonctions `var` et `sd` de R nous fournissent les estimateurs sans biais de la variance et de l'écart-type, à savoir

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

11. À partir de cette formule, faites coïncider les résultats des questions 8 et 9.
12. Définir les échantillons $E = (12, 15, 20, 25, 28)$ et $F = (19.3, 19.8, 20, 20.2, 20.7)$. Calculez la moyenne et l'écart type empiriques de E et F . Que remarquez-vous ?

4 Inclure une bibliothèque et traitement de données

Dans cette partie nous souhaitons inclure une bibliothèque de données géostatistiques. Pour cela, tapez la commande

```
library(geoR)
```

Si vous souhaitez installer cette librairie sur votre ordinateur personnel : tapez la commande

```
install.packages("geoR")
```

Il faut ensuite charger la librairie. Pour cela, tapez la commande

```
library(geoR)
```

Pour afficher la liste des fichiers disponibles, tapez

```
data(package='geoR')
```

1. Afficher `camg`.
2. Que donne `camg[2,8]` ?
3. Que donne `camg[1:10,]` ?
4. Affichez la deuxième ligne de `camg`.
5. Que donne `camg[camg[,4]==3,]` ?
6. Affichez toutes les lignes de `camg` telles que la valeur en `mg2040` vaut 35.
7. Que donne `camg$region` ?
8. Que donne `summary(camg)` ?
9. Affecter à `d2` les 20 premières lignes de la colonne `region` de `camg`.
10. En utilisant la commande `sort`, affectez à `d2ord`, le vecteur des valeurs de `d2` triées par ordre croissant.