

Mémento R sur les Statistiques descriptives et inférentielles

Table des matières

I	Gestion de données	2
I.1	Effacer la mémoire	2
I.2	Importation des données	2
I.3	Création d'un vecteur ou d'une matrice	2
I.4	Création d'un vecteur associé à une séquence	2
I.5	Classement des éléments d'un vecteur par ordre croissant (resp. décroissant)	2
I.6	Nature d'un objet (Structure)	2
I.7	Définition de facteurs et variables	2
I.8	Sélection partielle de données	2
II	Statistiques descriptives	3
II.1	Calcul de moyenne, quantiles, variance et écart-type (estimés)	3
II.2	Représentations graphiques	3
II.3	Régression linéaire	3

Florent ARNAL
florent.arnal@u-bordeaux.fr

I Gestion de données

I.1 Effacer la mémoire

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

I.2 Importation des données

Il faut, dans un premier temps, créer un fichier sur Excel et l'enregistrer au format csv.

En outre, le répertoire peut être changé en passant par :

Session / Set Working Directory / Choose Directory ...

On peut également utiliser **Import Dataset**. Les commandes R pour importer et visualiser les données sont alors :

```
donnees = read.csv2("nom.csv", sep=";", dec="...", header=TRUE, stringsAsFactors = T)
```

I.3 Création d'un vecteur ou d'une matrice

```
x = c( ... )  
M = matrix(data=... , nrow=... , ncol=... , byrow=... )
```

I.4 Création d'un vecteur associé à une séquence

```
y=seq(from= ... , to=... , by=... )
```

I.5 Classement des éléments d'un vecteur par ordre croissant (resp. décroissant)

```
sort(vecteur)  
sort(vecteur, decreasing = TRUE)
```

I.6 Nature d'un objet (Structure)

```
str( ... )
```

I.7 Définition de facteurs et variables

```
Facteur = as.factor(Facteur)  
Variable = as.numeric(Variable)  
ou  
donnees = transform(donnees, Facteur = as.factor(Facteur))
```

I.8 Sélection partielle de données

```
subset(Objet étudié, Modalité == " Modalité à considérer ")  
subset(Objet étudié, Modalité1 == ... & Modalité2 == ... )
```

II Statistiques descriptives

II.1 Calcul de moyenne, quantiles, variance et écart-type (estimés)

```
mean(...)  
quantile(x=..., p=0.75) # renvoie le quantile d'ordre 0,75 (3ème quartile)  
var(...)  
sd(...)
```

II.2 Représentations graphiques

1. Partage de la fenêtre graphique

```
par( mfrow=c(..., ...) )
```

Pour revenir à un affichage "classique", on peut utiliser

```
dev.off()
```

2. Choix "automatique" de couleurs :

```
terrain.colors(n=...)  
rainbow(n=...)  
colorRampPalette(...)
```

3. Création d'un boxplot :

```
boxplot(x, y, main = "...", names=c("...", "..."), col=c("...", "..."))
```

4. Création d'un histogramme :

```
hist(x, col="...", main="...")
```

5. Création d'un diagramme en barres :

```
barplot(x, col="...", main="...")
```

II.3 Régression linéaire

1. Recherche des coefficients de corrélation et de l'ajustement affine du type $y = ax + b$

```
# Coefficient de corrélation  
cor(x, y) # Modele linéaire (lm = linear model)  
modele = lm(y ~ x)  
modele  
# Récupération des coefficients a et b  
b = modele$coefficients[1] # intercept = ordonnée à l'origine  
a = modele$coefficients[2]
```

2. Nuage de points et droite d'ajustement d'une série (x, y)

```
plot(y ~ x)  
abline(modele)
```

3. Résidus et valeurs prédites (pour $x = 7$)

```
resid( modele )  
predict( modele ), newdata=data.frame(x=c(7)))
```