**Chargement d’un fichier**

mesdata = read.table("data/monfichier.txt", header = TRUE)

Tu peux nommer mesdata comme tu veux

**Affichage**

mesdata

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| age | poids | genre | titi | tata |
| 20 | 60 | H | truc | bidule |
| 21 | 61 | F | machin | machine |

**Accès à une colonne (série)**

mesdata$nom\_colonne, par exemple mesdata$age pour l’age donne 20, 21

**Stats**

Moyenne = mean par exemple mean(mesdata$poids)

Médiane = median par exemple median(mesdata$titi)

Ecart type = sd par exemple sd(mesdata$tata)

Variance = var

Résumé (tout sauf sd) = summary par exemple summary(mesdata$age)

1er quartile = q1 = quantile(mesdata$age, 0.25)

3ème quartile = q3 = quantile(mesdata$age , 0.75)

Nombre d’élément = length par exemple length(mesdata$poids)

Min et max

**Dessins**

Hist(mesdata$poids)

Boxplot(mesdata$poids)

**Variables qualitatives**

Répartition = table(mesdata$genre) on obtient H=1 et F=1

**Filtre (Sous tableau, opérateur [])**

mesdata\_H = mesdata[mesdata$genre == “H”] => on obtient que les lignes où genre == H

mesdata\_titi = mesdata[mesdata$titi == “truc”] => on obtient que les lignes où titi == truc

mesdata\_age = mesdata[mesdata$age < 50] => on obtient que les lignes où age inférieur à 50

**Filtre sur une colonne (série)**

mesdata\_age\_H = mesdata$age[mesdata$genre == “H”] => on obtient uniquement les ages où genre == H, dans le crochet on filtre sur les lignes et avant le crochet on filtre sur les colonnes

mesdata\_tata\_titi = mesdata$tata[mesdata$titi == “truc”] => on obtient uniquement les tata où titi == “truc”

mesdata\_age\_H\_poids = mesdata$age[mesdata$genre == “H” & mesdata$poids > 50] => on obtient uniquement les ages où genre = “H” ET poids > 50

**Vérification de la normalité (obligatoire si length < 30)**

Méthode graphique : qqnorm et qqline par exemple qqnorm(mesdata\_H) puis qqline(mesdata\_H)

Méthode shapiro : shapiro\_test(mesdata\_H), la pvalue doit être > 5%

**Intervalle de confiance**

CI(mesdata\_H, 0.95)

**Comparer une moyenne observée à une moyenne théorique**

t.test(femme\_g1, mu=29, alternative="greater", conf.level = 0.98): va tester si la moyenne de femme\_g1 > 29 avec un intervalle de confiance de 0.98

**Comparer 2 echantillons (2 series)**

Test de fisher = var.test(femme\_g1, femme\_g2)

Test de student = t.test(femme\_g1, femme\_g2, alternative="greater", var.equal=TRUE, conf.level = 0.98)

Test de student pour valeurs appariés (ssi dosage$a a le même nombre de données que dosage$b) = t.test(dosage$a, dosage$b, paired=TRUE, conf.level = 0.98)