Atelier 3

IOOS Sterenn

17/10/2019

## Tableaux de données

Il y a un type d’objets que nous avons déjà abordé dans le chapitre Premier travail avec les données, il s’agit du tableau de données ou data frame en anglais.

###Propriétés et création

Dans R, les tableaux de données sont tout simplement des listes avec quelques propriétés spéficiques :

les tableaux de données ne peuvent contenir que des vecteurs ;  
tous les vecteurs d’un tableau de données ont la même longueur ;  
tous les éléments d’un tableau de données sont nommés et ont chacun un nom unique.

Dès lors, un tableau de données correspond aux fichiers de données que l’on a l’habitude de manipuler dans d’autres logiciels de statistiques comme SPSS ou Stata. Les variables sont organisées en colonnes et les observations en lignes.

On peut créer un tableau de données avec la fonction data.frame :

df <- data.frame(sexe = c("f", "f", "h", "h"), age = c(52, 31, 29, 35), blond = c(FALSE, TRUE, TRUE, FALSE))  
df

## sexe age blond  
## 1 f 52 FALSE  
## 2 f 31 TRUE  
## 3 h 29 TRUE  
## 4 h 35 FALSE

class(df$sexe)

## [1] "factor"

class(df$age)

## [1] "numeric"

class(df$blond)

## [1] "logical"

str(df)

## 'data.frame': 4 obs. of 3 variables:  
## $ sexe : Factor w/ 2 levels "f","h": 1 1 2 2  
## $ age : num 52 31 29 35  
## $ blond: logi FALSE TRUE TRUE FALSE

La fonction data.frame a un gros défaut : si on ne désactive pas l’option stringsAsFactors elle transforme les chaînes de caractères, ici la variable sexe en facteurs (un type de vecteur que nous aborderons plus en détail dans un prochain chapitre)

df <- data.frame(sexe = c("f", "f", "h", "h"), age = c(52, 31, 29, 35), blond = c(FALSE, TRUE, TRUE, FALSE), stringsAsFactors = FALSE)  
df

## sexe age blond  
## 1 f 52 FALSE  
## 2 f 31 TRUE  
## 3 h 29 TRUE  
## 4 h 35 FALSE

str(df)

## 'data.frame': 4 obs. of 3 variables:  
## $ sexe : chr "f" "f" "h" "h"  
## $ age : num 52 31 29 35  
## $ blond: logi FALSE TRUE TRUE FALSE

Un tableau de données étant une liste, la fonction length renverra le nombre d’éléments de la liste, donc dans le cas présent le nombre de variables et names leurs noms :

length(df)

## [1] 3

names(df)

## [1] "sexe" "age" "blond"

## Importer des données par internet

nutriage <- read.csv("https://afalco.000webhostapp.com/cursos/nutriage.csv",header=TRUE)  
length(nutriage)

## [1] 13

names(nutriage)

## [1] "sexe" "situation" "the"   
## [4] "cafe" "taille" "poids"   
## [7] "age" "viande" "poisson"   
## [10] "fruit\_crus" "fruit\_legume\_cuits" "chocol"   
## [13] "matgras"

Taille de l’échantillon ?

length(nutriage$age)

## [1] 226

Consommation de viande chez les hommes et chez les femmes

class(nutriage$sexe)

## [1] "integer"

viande.h <- nutriage$viande[nutriage$sexe==1]  
viande.f <- nutriage$viande[nutriage$sexe==2]  
  
length(viande.h)

## [1] 85

length(viande.f)

## [1] 141

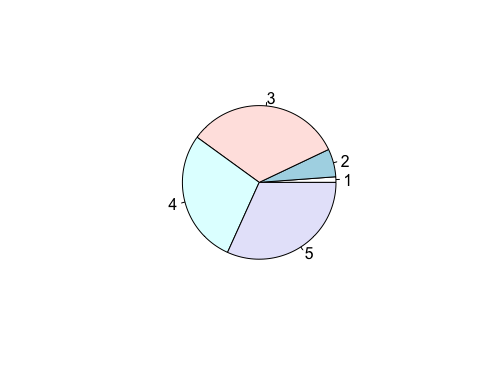
100\*(table(viande.h))/(length(viande.h))

## viande.h  
## 1 2 3 4 5   
## 1.176471 5.882353 32.941176 28.235294 31.764706

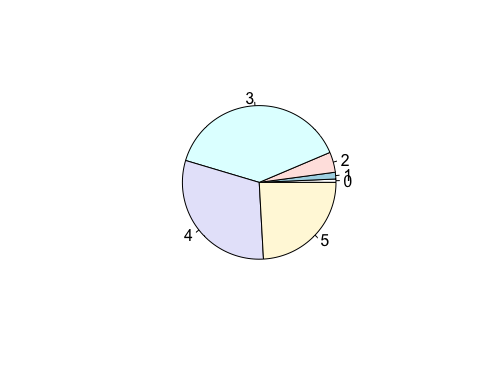
100\*(table(viande.f))/(length(viande.f))

## viande.f  
## 0 1 2 3 4 5   
## 0.7092199 1.4184397 4.2553191 39.0070922 30.4964539 24.1134752

pie(table(viande.h))



pie(table(viande.f))



Consommation de chocolat en fonctionet du sexe

class(nutriage$sexe)

## [1] "integer"

choco.h <- nutriage$chocol[nutriage$sexe==1]  
choco.f <- nutriage$chocol[nutriage$sexe==2]  
  
length(choco.h)

## [1] 85

length(choco.f)

## [1] 141

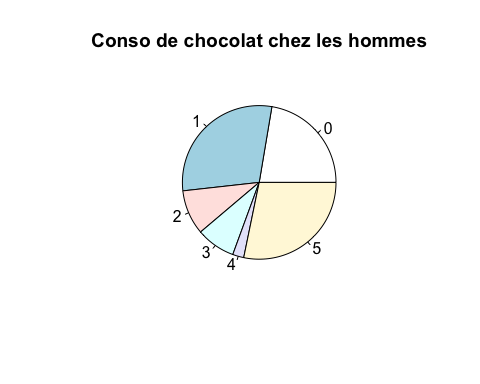
100\*(table(choco.h))/(length(choco.h))

## choco.h  
## 0 1 2 3 4 5   
## 22.352941 29.411765 9.411765 8.235294 2.352941 28.235294

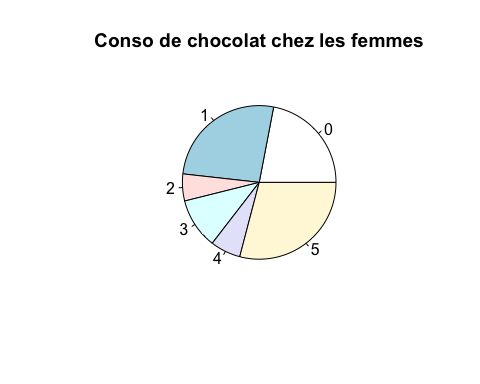
100\*(table(choco.f))/(length(choco.f))

## choco.f  
## 0 1 2 3 4 5   
## 21.985816 26.241135 5.673759 10.638298 6.382979 29.078014

pie(table(choco.h),main = "Conso de chocolat chez les hommes")



pie(table(choco.f), main = "Conso de chocolat chez les femmes")



La distribution de l’age de l’échantillon

table(nutriage$age)

##   
## 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89   
## 6 10 11 8 17 13 13 14 16 17 12 15 12 18 7 3 4 4 2 5 5 2 3 2 4   
## 90 91   
## 1 2

viande.groupe1 <- nutriage$viande[nutriage$age <79]  
viande.groupe2 <- nutriage$viande[nutriage$age >78]  
  
100\*(table(viande.groupe1)/length(viande.groupe1))

## viande.groupe1  
## 1 2 3 4 5   
## 1.098901 5.494505 34.615385 31.318681 27.472527

100\*(table(viande.groupe2)/length(viande.groupe2))

## viande.groupe2  
## 0 1 2 3 4 5   
## 2.272727 2.272727 2.272727 45.454545 22.727273 25.000000

Distribution de la variable poids

table(nutriage$poids)

##   
## 38 40 42 44 45 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67   
## 1 1 2 1 4 4 1 6 3 2 9 6 7 5 5 8 1 13 3 7 8 8 6 5 7   
## 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 85 86 87 90 91 92 94 95 96   
## 6 5 8 3 2 5 9 18 4 3 3 2 9 2 4 2 4 1 1 7 1 1 1 1 1

summary(nutriage$poids)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 38.00 57.25 66.00 66.48 75.00 96.00

Consomation de viande en fonction du poids

viande.poids1 <- nutriage$viande[nutriage$poids <66]  
viande.poids2 <- nutriage$viande[nutriage$poids >66]  
  
100\*(table(viande.poids1)/length(viande.poids1))

## viande.poids1  
## 0 1 2 3 4 5   
## 0.9009009 1.8018018 4.5045045 37.8378378 27.9279279 27.0270270

100\*(table(viande.poids2)/length(viande.poids2))

## viande.poids2  
## 1 2 3 4 5   
## 0.9090909 5.4545455 35.4545455 30.0000000 28.1818182

Consommation de poisson en focntion du poids

class(nutriage$poids)

## [1] "integer"

poisson.L <- nutriage$poisson[nutriage$poids>66]  
poisson.l <- nutriage$poisson[nutriage$poids<66]  
length(poisson.L)

## [1] 110

length(poisson.l)

## [1] 111

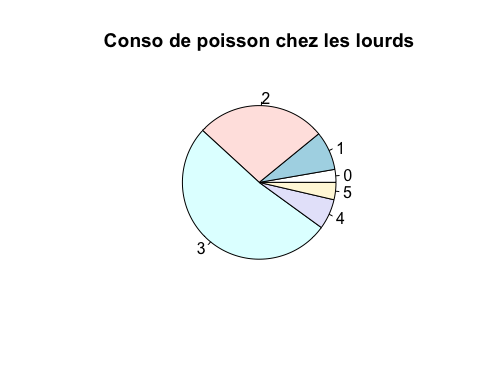
100\*(table(poisson.L))/(length(poisson.l))

## poisson.L  
## 0 1 2 3 4 5   
## 2.702703 8.108108 27.027027 51.351351 6.306306 3.603604

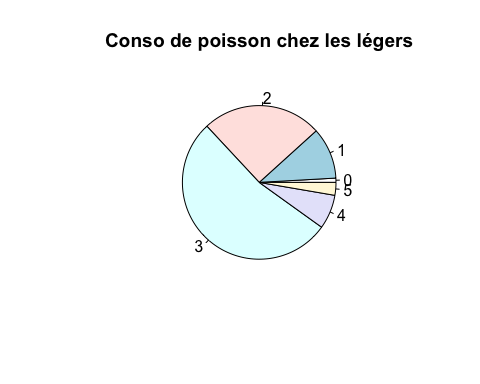
100\*(table(poisson.l))/(length(poisson.l))

## poisson.l  
## 0 1 2 3 4 5   
## 0.9009009 10.8108108 25.2252252 53.1531532 7.2072072 2.7027027

pie(table(poisson.L),main = "Conso de poisson chez les lourds")



pie(table(poisson.l), main = "Conso de poisson chez les légers")



On va utiliser l’IMC

imc <- nutriage$poids/(nutriage$taille/100)^2  
nutriage <- data.frame(nutriage,imc)  
names(nutriage)

## [1] "sexe" "situation" "the"   
## [4] "cafe" "taille" "poids"   
## [7] "age" "viande" "poisson"   
## [10] "fruit\_crus" "fruit\_legume\_cuits" "chocol"   
## [13] "matgras" "imc"

summary(imc)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 16.33 22.21 24.62 24.64 27.01 36.98

viande.groupe1 <- nutriage$viande[nutriage$icm < median(nutriage$imc)]  
viande.groupe1

## integer(0)

viande.groupe2 <- nutriage$viande[nutriage$icm >= median(nutriage$imc)]  
100\*(table(viande.groupe1)/length(viande.groupe1))

## numeric(0)

100\*(table(viande.groupe2)/length(viande.groupe2))

## numeric(0)