Excercice

Thibaut Goetsch et Bineta FAYE

2023-09-08

## Exercice

**Réalisez une fonction qui calcule la moyenne (avec vérification si bien numérique). Comparerz avec la fonction mean()**

moyenne<-function(x){  
 if (is.numeric(x))  
 return(sum(x)/length(x[!is.na(x)]))  
}  
moyenne(1:10)

## [1] 5.5

mean(1:10)

## [1] 5.5

**Réalisez une fonction qui calcule la variance (avec vérification si bien numérique). Comparez avec var()**

variance<-function(x){  
 if (is.numeric(x))  
 return(sum((x-mean(x))^2)/length(x))  
}  
variance(1:10)

## [1] 8.25

var(1:10)

## [1] 9.166667

**Chargez le jeu de données iris et mettez le dans une variable df**

df<-iris  
head(df)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

mean(df[df$Species=="setosa" ,1])

## [1] 5.006

**Pour chaque espèce d’iris présente dans la base de données, calculez la moyenne et la variance pour chaque variable**

#levels(iris$Species)  
print("Sepal.Length")

## [1] "Sepal.Length"

cat("moyenne setosa" ,moyenne(df$Sepal.Length[df$Species=="setosa" ]),"\n")

## moyenne setosa 5.006

cat("moyenne versicolor" ,moyenne(df$Sepal.Length[df$Species=="versicolor" ]),"\n")

## moyenne versicolor 5.936

cat("moyenne virginica" ,moyenne(df$Sepal.Length[df$Species=="virginica" ]),"\n")

## moyenne virginica 6.588

print("Petal.Width")

## [1] "Petal.Width"

cat("moyenne setosa" ,moyenne(df$Petal.Width[df$Species=="setosa" ]),"\n")

## moyenne setosa 0.246

cat("moyenne versicolor" ,moyenne(df$Petal.Width[df$Species=="versicolor" ]),"\n")

## moyenne versicolor 1.326

cat("moyenne virginica" ,moyenne(df$Petal.Width[df$Species=="virginica" ]),"\n")

## moyenne virginica 2.026

## Var

print("Petal.Width")

## [1] "Petal.Width"

cat("variance setosa" ,variance(df$Petal.Width[df$Species=="setosa" ]),"\n")

## variance setosa 0.010884

cat("variance versicolor" ,variance(df$Petal.Width[df$Species=="versicolor" ]),"\n")

## variance versicolor 0.038324

cat("variance virginica" ,variance(df$Petal.Width[df$Species=="virginica" ]),"\n")

## variance virginica 0.073924

**Créez maintenant une boucle pour les calculer**

colonne<-colnames(df)  
espece<-levels(iris$Species)  
for(i in 1:4){  
 cat(colonne[i],"\n")  
 for(e in espece){  
 cat("moyenne",e,moyenne(df[df$Species==e,i]),"\n")  
 cat("variance",e,variance(df[df$Species==e,i]),"\n")  
 }}

## Sepal.Length   
## moyenne setosa 5.006   
## variance setosa 0.121764   
## moyenne versicolor 5.936   
## variance versicolor 0.261104   
## moyenne virginica 6.588   
## variance virginica 0.396256   
## Sepal.Width   
## moyenne setosa 3.428   
## variance setosa 0.140816   
## moyenne versicolor 2.77   
## variance versicolor 0.0965   
## moyenne virginica 2.974   
## variance virginica 0.101924   
## Petal.Length   
## moyenne setosa 1.462   
## variance setosa 0.029556   
## moyenne versicolor 4.26   
## variance versicolor 0.2164   
## moyenne virginica 5.552   
## variance virginica 0.298496   
## Petal.Width   
## moyenne setosa 0.246   
## variance setosa 0.010884   
## moyenne versicolor 1.326   
## variance versicolor 0.038324   
## moyenne virginica 2.026   
## variance virginica 0.073924

**Mettez votre boucle dans une fonction qui affiche le nom de l’espèce, le nom de la variable et les différents résultats à chaque fois**

colonne<-colnames(df)  
espece<-levels(iris$Species)  
fct<-function(espece,df){  
for(e in espece){   
for(i in 1:4){  
 cat(e,colonne[i],"moyenne",moyenne(df[df$Species==e,i]),"variance",variance(df[df$Species==e,i]),"\n")  
}}  
}  
fct(espece,df)

## setosa Sepal.Length moyenne 5.006 variance 0.121764   
## setosa Sepal.Width moyenne 3.428 variance 0.140816   
## setosa Petal.Length moyenne 1.462 variance 0.029556   
## setosa Petal.Width moyenne 0.246 variance 0.010884   
## versicolor Sepal.Length moyenne 5.936 variance 0.261104   
## versicolor Sepal.Width moyenne 2.77 variance 0.0965   
## versicolor Petal.Length moyenne 4.26 variance 0.2164   
## versicolor Petal.Width moyenne 1.326 variance 0.038324   
## virginica Sepal.Length moyenne 6.588 variance 0.396256   
## virginica Sepal.Width moyenne 2.974 variance 0.101924   
## virginica Petal.Length moyenne 5.552 variance 0.298496   
## virginica Petal.Width moyenne 2.026 variance 0.073924

**Améliorez votre fonction pour extraire ces résultats dans une nouvelle base de données plus petite que vous aurez créé**

colonne<-colnames(df)  
espece<-levels(iris$Species)  
fct\_simple<-function(espece,df){  
esp=c()  
vrbl=c()  
moy=c()  
varia=c()  
for(e in espece){   
for(i in 1:4){  
 esp<-c(esp,e)  
 vrbl<-c(vrbl,colonne[i])  
 moy<-c(moy,moyenne(df[df$Species==e,i]))  
 varia<-c(varia,variance(df[df$Species==e,i]))  
}}  
return(data.frame(esp,vrbl,moy,varia))  
}

fct\_simple(espece,df)

## esp vrbl moy varia  
## 1 setosa Sepal.Length 5.006 0.121764  
## 2 setosa Sepal.Width 3.428 0.140816  
## 3 setosa Petal.Length 1.462 0.029556  
## 4 setosa Petal.Width 0.246 0.010884  
## 5 versicolor Sepal.Length 5.936 0.261104  
## 6 versicolor Sepal.Width 2.770 0.096500  
## 7 versicolor Petal.Length 4.260 0.216400  
## 8 versicolor Petal.Width 1.326 0.038324  
## 9 virginica Sepal.Length 6.588 0.396256  
## 10 virginica Sepal.Width 2.974 0.101924  
## 11 virginica Petal.Length 5.552 0.298496  
## 12 virginica Petal.Width 2.026 0.073924

**(Avancé) Utilisez les fonctions dplyr pour obtenir le même résultat de la manière la plus courte possible**

library(dplyr)

## Warning: le package 'dplyr' a été compilé avec la version R 4.1.3

##   
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':  
##   
## filter, lag

## Les objets suivants sont masqués depuis 'package:base':  
##   
## intersect, setdiff, setequal, union

df|> group\_by(Species)|>summarise(moyenne=mean(Sepal.Length))

## # A tibble: 3 x 2  
## Species moyenne  
## <fct> <dbl>  
## 1 setosa 5.01  
## 2 versicolor 5.94  
## 3 virginica 6.59

###  
df|> group\_by(Species)|>summarise(moyenne=mean(Sepal.Width))

## # A tibble: 3 x 2  
## Species moyenne  
## <fct> <dbl>  
## 1 setosa 3.43  
## 2 versicolor 2.77  
## 3 virginica 2.97

###  
df|> group\_by(Species)|>summarise(moyenne=mean(Petal.Length))

## # A tibble: 3 x 2  
## Species moyenne  
## <fct> <dbl>  
## 1 setosa 1.46  
## 2 versicolor 4.26  
## 3 virginica 5.55

###  
df|> group\_by(Species)|>summarise(moyenne=mean(Petal.Width))

## # A tibble: 3 x 2  
## Species moyenne  
## <fct> <dbl>  
## 1 setosa 0.246  
## 2 versicolor 1.33   
## 3 virginica 2.03

df|> group\_by(Species)|>summarise(variance=var(Sepal.Width))

## # A tibble: 3 x 2  
## Species variance  
## <fct> <dbl>  
## 1 setosa 0.144   
## 2 versicolor 0.0985  
## 3 virginica 0.104

email contact: [bineta.faye@chru-strasbourg.fr](mailto:bineta.faye@chru-strasbourg.fr)