

TD de la 3^o séance

Travailler sous R

1. ouvrir R puis un nouveau script pour les commandes et vérifier votre répertoire de travail

```
getwd()
```

si votre répertoire de travail n'est pas le répertoire documents/AtelierR alors changer le répertoire de travail avec la fonction setwd()

```
setwd(dirname(file.choose()))
```

2. vérifier le contenu et le format du fichier CSV à importer, il s'agit du fichier sauvegardé lors de la précédente séance: TDR3.csv - utiliser la fonction file.show()

```
file.show(file.choose())
```

3. Créer un objet R appelé Mydata qui contiendra les données du fichier CSV. Utiliser la fonction read.table()

Informer R que la première ligne contient le nom des colonnes. Demander d'afficher les noms des variables avec le paramètre header=T ou header=TRUE

Nous avons vu que le séparateur décimal était une décimale alors que R utilise l'écriture anglo-saxonne avec un point, Utiliser le paramètre dec="," pour lui demander de transformer les virgules en point.

Enfin rappeler que le séparateur des colonnes est un point virgule avec le paramètres sep = ";" "

```
Mydata<-read.table(file.choose(), header=T ,dec="," , sep=";")
```

4. L'objet créé contient-il bien l'information ?

```
str(Mydata)
```

5. Décrire une variable quantitative

```
mean(Mydata$DUREEMOIS)
```

```
median(Mydata$DUREEMOIS)
```

```
var(Mydata$DUREEMOIS)
```

```
sd(Mydata$DUREEMOIS)
```

```
sqrt(var(Mydata$DUREEMOIS))
```

```
summary(Mydata$DUREEMOIS)
```

6. Intervalle de confiance d'une moyenne

```
Result<-t.test(Mydata$DUREEMOIS, conf.level=0.95)
```

```
Result
```

```
attributes(Result)
```

```
lc95<-t.test(Mydata$DUREEMOIS, conf.level=0.95)$conf.int
```

```
lc95
```

7. Décrire une variable qualitative par rapport à une ou plusieurs autres (Tableau croisé)

```
aggregate(list(Age=Mydata$AGEAN), list(Fever=Mydata$FEVER2),mean)
```

```
aggregate(list(Age=Mydata$AGEAN), list(Sexe=Mydata$SEXE), mean)
aggregate(list(Age=Mydata$AGEAN), list(Fever=Mydata$FEVER2,Sexe=Mydata$SEXE),mean)
```

```
by(Mydata$AGEAN,Mydata$SEXE,mean)
by(Mydata$AGEAN,Mydata$SEXE,var)
by(Mydata$AGEAN,Mydata$SEXE,summary)
tapply(Mydata$DUREEMOIS,list(Mydata$SEXE,Mydata$FEVER2),mean)
```

Travail en autonomie:

Quelle est la valeur moyenne de la fièvre en fonction de la présence ou non de toux?

Quel est le nombre de cas de fièvre par année?

8. Décrire une variable qualitative

```
table(Mydata$SEXE)#pour connaître les effectifs de chacun des groupes
```

#mais pour avoir les pourcentages aussi

```
Tableau1<-
rbind(table(Mydata$SEXE),round(table(Mydata$SEXE)*100/sum(table(Mydata$SEXE)),digits =2))
Tableau1
Tableau2<-prop.table(table(Mydata$SEXE))
Tableau2
Tableau3<-round(Tableau2*100,digits=2)
Tableau3
```

Effectifs cumulés

```
cumsum(table(Mydata$SEXE))
round(cumsum(table(Mydata$SEXE)/sum(table(Mydata$SEXE)))*100, digits=2)
```

Tables de contingence

```
Table1<-table(Mydata$TOUX)
margin.table(Table1)#Effectif total de la table
```

```
Table2<-table(Mydata$TOUX,Mydata$SEXE)
Total<-margin.table(Table2)#Effectif total de la table
Efflign<-margin.table(Table2, 1) #total des effectifs par ligne
Effcol<-margin.table(Table2, 2) #total des effectifs par colonne
Pcol<-prop.table(Table2,margin=2) #Pourcentage par colonne
Plign<-prop.table(Table2, margin=1)#Pourcentage par ligne
prop.table(Efflign)#Pourcentage du total des effectifs par ligne
prop.table(Effcol) #Pourcentage du total des effectifs par colonne
Tablecont<-cbind(Table2,Efflign) #ajouterunecolonnecontenantletotal
Tablecont
Totalcol<-c(Effcol>Total) #créer un vecteur contenant les effectifs des colonnes et le total global
Tablecont<-rbind(Tablecont>Totalcol) #l'ajouter sur une ligne au tableau de contingence
Tablecont
```

Enregistrer votre table Mydata dans un format.CSV pour pouvoir garder les variables créées au cours de ce TD. Vous nommerez le fichier de sortie DataTD2Rvers20180406

Exercice en autonomie

Dans cet exercice, nous utilisons les données du TD1.

- A1. Créer une variable Debsympt qui contient des dates entre le 1 janvier 2008 et le 31 décembre 2011. De même créer la variable Finsympt avec des dates supérieures de 1 à 31 jours. (utiliser les fonctions paste() et sample()).
- A2. Ajouter les variables Debsympt et Finsympt dans Mydata et modifier les noms
- A3. Calculer la durée du symptôme en créant la variable Duree .
Calculer la durée moyenne de maladie.
- A4. Décrire la variable AGE
- A5. Créer la classe d'âge (Classage) en 3 classes :]18,22]]22,25],]25 et plus]
- A6. Ajouter les variables Duree et Classage dans Mydata et modifier les noms pour respecter la casse.
- A7. Calculer l'IMC (Poids en kg/Taille en mètre élevée au carré) et l'inclure dans Mydata
- A8. Ajouter la variable Surpoids qui prend la valeur 1 si IMC>24.9 et 0 dans les autres cas
- A9. Ajouter la variable Imclass avec 4 classes en fonction de la valeur de l'IMC
<18.5=Insuff pondérale ; [18,5-24,9] normal, [25-29.9] surpoids, >29.9 obésité
- A10. Calculer la moyenne d'âge par sexe.
- A11. Calculer l'effectif dans chaque classe d'âge.
- A12. Créer un tableau Tab1 contenant l'effectif des classes d'âge par sexe.
- A13. Créer un tableau Tab2 contenant la proportion des classes d'âge par sexe.
- A14. Créer un tableau Tab3 contenant les effectifs et les proportions des classes par sexe.
- A15. Créer un tableau Tab4 contenant en plus les totaux des lignes et ensuite un tableau Tab5 contenant les totaux des colonnes ainsi que les titres des colonnes (homme, %, Femme, %, Total) et des lignes

Enregistrer votre table Mydata dans un format.CSV pour pouvoir garder les variables créées au cours de ce TD. Vous nommerez le fichier de sortie DataTD3.csv