

```
#####
# TD 3 travail en Autonomie #
#####

#Définir le répertoire de travail

setwd(dirname(file.choose()))

#Importer les données du fichier créé en première semaine

Mydata<-read.table(file.choose(),header=T,sep=";", dec =',')

# 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()).

set.seed(10)
Debsympt<-
paste(sample(1:31,100,replace=T),sample(1:12,100,replace=T),sampl
e(2008:2011,100,replace=T),sep="/")

Debsympt<-as.Date(Debsympt,"%d/%m/%Y")

# ou créer un vecteur comprenant toutes les dates
# An1<-as.Date("01/01/2008", "%d/%m/%Y")
# An2<-as.Date("31/12/2011", "%d/%m/%Y")
# Periode<-seq(from=An1,to=An2,by="days")
#puis en extraire 100 valeurs avec sample()
# set.seed(10) #permet d'avoir le même tirage aléatoire
# Debsympt<-sample(Periode,100,replace=T)

set.seed(10)
Finsympt<-Debsympt+sample(1:31,100,replace=T)

# A2. Ajouter les variables Debsympt et Finsympt dans Mydata et
modifier les noms

Mydata<-data.frame(Mydata,Debsympt,Finsympt)

# A3. Calculer la durée du symptôme en créant la variable Duree

Duree<-Finsympt-Debsympt
Duree<-as.numeric(Duree)

Calculer la durée moyenne de maladie.

mean(Duree,na.rm=T)

# A4. Décrire la variable AGE

summary(Mydata$AGE)

# A5. Créer la classe d'âge (Classage) en 3
classes : ]18,22] ]22,25],]25 et plus]

Clage<-
cut(Mydata$AGE,breaks=c(18,22,25,max(Mydata$AGE)))
summary(Clage)

# A6. Ajouter les variables Duree et Classage dans Mydata et modifier
les noms pour respecter la casse.

Mydata<-data.frame(Mydata,Duree,Clage)
names(Mydata)[6:9]<-
c("DTEDEBUT","DTEFIN","DUREE","CLASSAGE")

# A7. Calculer l'IMC (Poids en kg/Taille en mètre élevée au carré) et
l'inclure dans Mydata. Attention la variable Taille est exprimée dans la
base en centimètres.

Imc<-round(Mydata$POIDS/((Mydata$TAILLE/100)^2),1)
```

```

Mydata<-data.frame(Mydata,IMC=Imc)

#ou
# Mydata<-data.frame(Mydata,Imc)
# names(Mydata)[11]<-"IMC"

# A8. Ajouter la variable Surpoids qui prend la valeur 1 si IMC>24.9 et
0 dans les autres cas

Surpoids<-ifelse(Mydata$IMC>25,1,0)
Mydata<-data.frame(Mydata, Surpoids)
names(Mydata)[11]<-"SURPOIDS"

#ou Mydata<-data.frame(Mydata,SURPOIDS=Surpoids)

# 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

Mydata$IMCLASS<-
cut(Mydata$IMC,breaks=c(min(Mydata$IMC)-0.1,18.4,24.9,29.9,40))

#Vérifier le contenu de la variable créée
summary(Mydata$IMCLASS)

#A10. Calculer la moyenne d'âge par sexe.

by(Mydata$AGE,Mydata$SEXE,mean,na.rm=T)

# A11. Calculer l'effectif dans chaque classe d'âge.

summary(Mydata$CLASSAGE)

# A12. Créer un tableau Tab1 contenant l'effectif par classe d'âge et
par sexe.

Tab1<-table(Mydata$CLASSAGE,Mydata$SEXE)
Tab1

#A13. Créer un tableau Tab2 contenant la proportion des classes d'âge
par sexe.

Tab2<-prop.table(Tab1,2)
Tab2

#A14. Créer un tableau Tab3 contenant les effectifs et les proportions
des classes par sexe.

Tab3<-cbind(Tab1[,1],round(Tab2[,1]*100,digits=1),Tab1[,
2],round(Tab2[,2]*100,digits=1))
Tab3

#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

Tab4<-cbind(Tab3,margin.table(Tab1,1))
Tab5<-rbind(Tab4,margin.table(Tab4,2))
colnames(Tab5)<-c("Male", "%", "Female", "%", "Total")
rownames(Tab5)[4]<-"Total"
Tab5

# 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_2.csv

write.csv2(Mydata,"DataTD3_2.csv",row.names=FALSE)

```