Christopher PHILET

Statistique descriptive : Etude descriptives de données
Fiche TP B : Analyse descriptive du fichier Nutriage

- **9.1**

```
#nutriage <-
read.table("http://www.biostatisticien.eu/springeR/nutriage.txt",header=TRUE)
require(gdata)
nutriage <- read.xls("nutriage.xls",header=T)
attach(nutriage)
```

- **9.2**

```
names(which.max(table(situation)))
names(which.max(table(chocol)))
names(which.max(table(taille)))
```

- **9.3**

```
res <- hist(taille,breaks=seq(140,190,by=5),right=T,plot=F)
ind <- which.max(res$count)
classe.modale <- paste(res$breaks[ind],res$breaks[ind+1],sep="-")
# La classe modale est la classe ]155;160]
```

- **9.4**

```
ma.mediane <- function(x) {
if (is.numeric(x)) return(median(x))
if (is.ordered(x)) {
N <- length(x)
if (N%%2) return(sort(x)[(N+1)/2]) else {
inf <- sort(x)[N/2]
sup <- sort(x)[N/2+1]
if (inf==sup) return(inf) else return(list(inf,sup))
}}
stop("Calcul de médiane impossible pour ce type")
}
ma.mediane(as.ordered(chocol))
ma.mediane(as.ordered(fruit_crus))
```

- **9.5**

```
table(chocol)
table(fruit_crus)
```

- **9.6**

- **9.7**

```
quartile.sur.freq <- function(x,quart) {
# x est le tableau des fréquences
tab.freq.cum <- cumsum(x/sum(x))
index <- order(tab.freq.cum < quart)[1]
f1 <- tab.freq.cum[index]
f2 <- tab.freq.cum[index-1]
x1 <- as.numeric(names(f1))
x2 <- as.numeric(names(f2))
quartile <- as.numeric(x1 + (x2-x1)*(quart-f1)/(f2-f1))
return(quartile)
}

tab <- res$counts
names(tab) <- res$breaks[-1]
quartile.sur.freq(tab,0.25)
quartile.sur.freq(tab,0.5)
quartile.sur.freq(tab,0.75)
```

- **9.8**

```
bornes <- res$breaks
plot(bornes,ecdf(taille)(bornes),type="l",main=paste("Polygone des
fréquences cumulées",
"de la variable
taille",sep="\n"),ylab="Fréquences",col="darkolivegreen",lwd=3)
abline(h=c(0.25,0.5,0.75))
locator(1)$x
```

- **9.9**

```
mean(chocol)
mean(taille)
```

- **9.10**

```
table(chocol)
sum((0:5)*as.numeric(table(chocol)))/sum(table(chocol))
```

- **9.11**

```
table(the)
sum(c(0:6,9,10)*as.numeric(table(the)))/sum(table(the))
```

- **9.12**

```
sum(res$mids*res$counts)/sum(res$counts)
```

- **9.13**

```
diff(range(chocol))
```

- **9.14**

```
diff(range(poids))
```

- **9.15**

```
boxplot(poids)
```

- **9.16**

- **9.17**

```
var.pop <- function(x) var(x)*(length(x)-1)/length(x)
sd.pop <- function(x) sqrt(var.pop(x))
sd.pop(taille)
```

- **9.18**

- **9.19**

```
eta2 <- function(x, gpe) {
moyennes <- tapply(x, gpe, mean)
effectifs <- tapply(x, gpe, length)
varinter <- (sum(effectifs * (moyennes - mean(x))^2))
vartot <- (var(x) * (length(x) - 1))
res <- varinter/vartot
list(var.tot=vartot,var.inter=varinter,var.intra=vartot-
varinter,eta2=res)
}

res <- eta2(the,sexe)
```