

Exercice : l'indice de Quételet

A.B. Dufour, D. Chessel & J.R. Lobry

La masse corporelle d'un individu est mesurée à partir du poids et de la taille. L'indice le plus utilisé est celui proposé par Adolphe Quételet (1796-1874), astronome et mathématicien belge (http://statbel.fgov.be/info/quetelet_fr.asp)

1 Questions

1. Construire la variable " indice de Quételet " à partir des données de t3var (fiche 1.2).

$$imc = \frac{poids}{taille^2}$$

La taille est exprimée en mètre et le poids en kilogrammes.

Cet indice s'appelle aujourd'hui " indice de masse corporelle " (imc) ou " body mass index " (bmi). Il permet de mesurer la corpulence de l'homme adulte. L'Organisation Mondiale de la Santé a défini les critères suivants : maigreur (inférieur à 18.5), normal (de 18.5 à 25), risque de surpoids (de 25 à 30), obésité (supérieur à 30). Mais l'indice de Quételet n'a qu'une valeur indicative. Pour déterminer l'existence d'une obésité réelle, il faut faire d'autres mesures destinées à établir exactement la proportion de masse grasse, car c'est l'excès de masse grasse qui représente un facteur de risque.

2. Calculer les paramètres statistiques élémentaires de cette nouvelle variable sur l'ensemble des individus et en fonction de chaque sexe.
3. Construire l'histogramme de cette nouvelle variable sur l'ensemble des individus.
4. Comparer graphiquement les indices de masse corporelle chez les hommes et chez les femmes à l'aide d'histogrammes et de fonctions de répartitions empiriques (utiliser la fonction `ecdf` de la librairie `stepfun`).
5. En utilisant la fonction `boxplot`, représenter l'indice de Quételet en fonction du sexe.
6. Construire sur un seul graphique les nuages de points de l'indice de Quételet en fonction de la taille, en fonction du poids, pour chaque sexe. Commenter.

7. Représenter, sur un seul graphique, le nuage de points de l'indice de Quételet en spécifiant hommes et femmes. Placer les bornes des critères de l'O.M.S. Commenter.

2 Réponses

1. Les données

```
options(digits = 4)
t3var <- read.table("http://pbil.univ-lyon1.fr/R/donnees/t3var.txt",
  h = T)
poi <- t3var$poi
tai <- t3var$tai * 0.01
sexe <- factor(t3var$sexe)
imc <- poi/(tai^2)
imc
```

```
[1] 20.76 19.96 17.24 18.17 17.72 19.29 18.29 20.16 20.31 20.90 19.47 26.77 19.92
[14] 23.04 21.46 20.43 19.71 24.66 22.34 20.48 19.63 18.82 25.01 23.37 18.34 23.04
[27] 21.60 20.16 23.66 20.48 19.25 22.09 23.99 18.07 17.99 22.27 19.95 19.49 22.31
[40] 21.22 23.15 19.05 16.92 19.26 18.38 21.74 22.40 21.56 21.67 20.37 22.41 24.93
[53] 21.97 18.13 24.96 22.22 19.82 21.88 25.66 22.28 23.20 25.96 23.80 20.52 20.89
[66] 21.39
```

2. Paramètres statistiques élémentaires

```
summary(imc)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 16.9   19.5   20.9   21.2   22.4   26.8
```

```
mean(imc)
```

```
[1] 21.16
```

```
var(imc)
```

```
[1] 5.137
```

```
sd(imc)
```

```
[1] 2.267
```

```
imcs <- split(imc, sexe)
imcs
```

```
$f
 [1] 19.96 17.24 18.17 17.72 19.29 18.29 20.31 19.47 20.43 19.71 19.63 18.82 18.34
[14] 23.66 20.48 18.07 17.99 22.27 19.95 19.49 22.31 19.05 16.92 18.13 20.89
```

```
$h
 [1] 20.76 20.16 20.90 26.77 19.92 23.04 21.46 24.66 22.34 20.48 25.01 23.37 23.04
[14] 21.60 20.16 19.25 22.09 23.99 21.22 23.15 19.26 18.38 21.74 22.40 21.56 21.67
[27] 20.37 22.41 24.93 21.97 24.96 22.22 19.82 21.88 25.66 22.28 23.20 25.96 23.80
[40] 20.52 21.39
```

```
lapply(imcs, mean)
```

```
$f
 [1] 19.46
```

```
$h
 [1] 22.19
```

```
lapply(imcs, var)
```

```

$f
[1] 2.673

$h
[1] 3.858

lapply(imcs, sd)

$f
[1] 1.635

$h
[1] 1.964

lapply(imcs, summary)

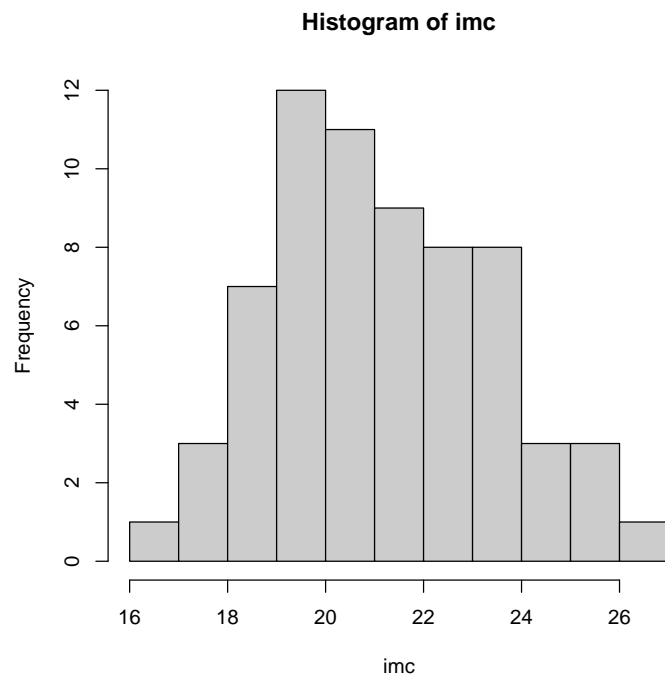
$f
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 16.9   18.2   19.5   19.5   20.3   23.7

$h
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 18.4   20.8   22.0   22.2   23.2   26.8

```

3. Histogramme sur l'ensemble des individus

```
hist(imc, col = grey(0.8))
```



4. Représentations graphiques par sexe

```
lapply(imcs, range)
```

```

$f
[1] 16.92 23.66

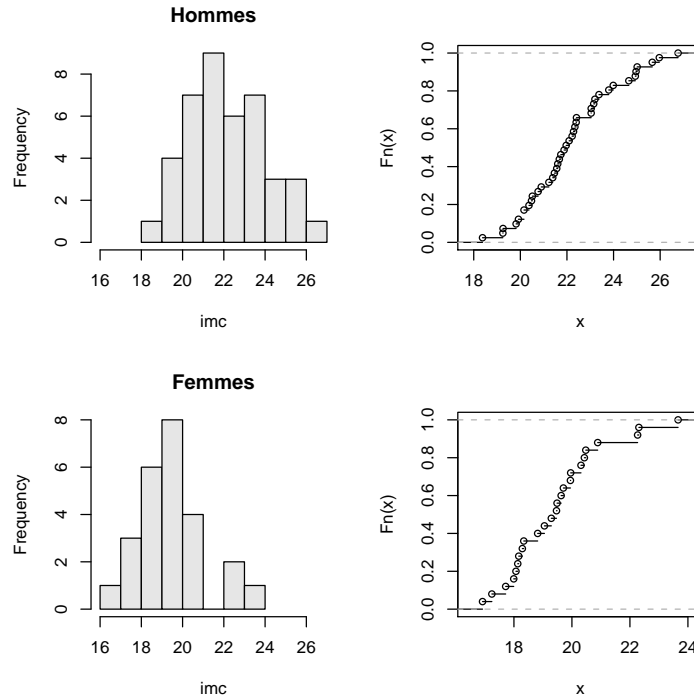
$h
[1] 18.38 26.77

```

```

par(mfcol = c(2, 2))
par(mar = c(5, 4, 3, 2))
hist(imc[sexe == "h"], main = "Hommes", xlab = "imc", xlim = c(16,
27), col = grey(0.9))
hist(imc[sexe == "f"], main = "Femmes", xlab = "imc", xlim = c(16,
27), col = grey(0.9))
plot(ecdf(imc[sexe == "h"]), main = "")
plot(ecdf(imc[sexe == "f"]), main = "")

```

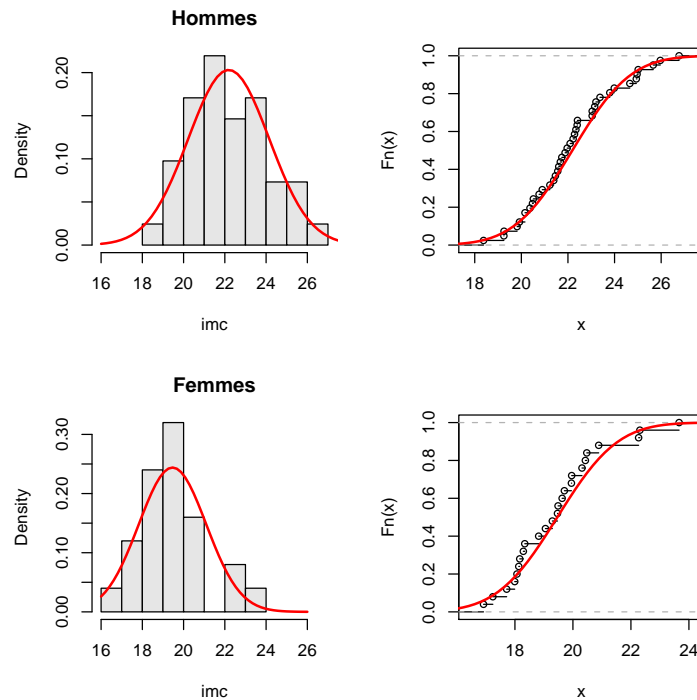


Passage à la modélisation

```

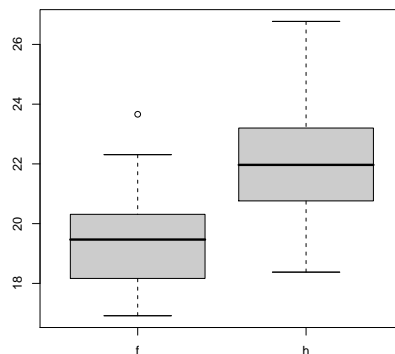
par(mfcol = c(2, 2))
par(mar = c(5, 4, 3, 2))
hist(imch <- imc[sexe == "h"], main = "Hommes", xlab = "imc", xlim = c(16,
27), col = grey(0.9), prob = T)
lines(x0 <- seq(16, 28, le = 50), dnorm(x0, mean(imch), sd(imch)),
lwd = 2, col = "red")
hist(imcf <- imc[sexe == "f"], main = "Femmes", xlab = "imc", xlim = c(16,
27), col = grey(0.9), prob = T)
lines(x1 <- seq(16, 26, le = 50), dnorm(x1, mean(imcf), sd(imcf)),
lwd = 2, col = "red")
plot(ecdf(imch), main = "")
lines(x0, pnorm(x0, mean(imch), sd(imch)), lwd = 2, col = "red")
plot(ecdf(imc[sexe == "f"]), main = "")
lines(x1, pnorm(x1, mean(imcf), sd(imcf)), lwd = 2, col = "red")

```



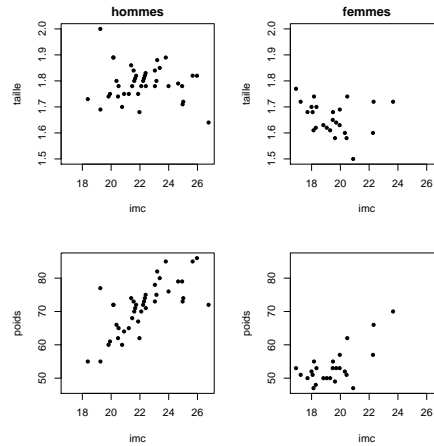
5. Boîte à moustaches

```
boxplot(imc ~ sexe, col = grey(0.8))
```



6. Relation entre l'imc, la taille et le poids

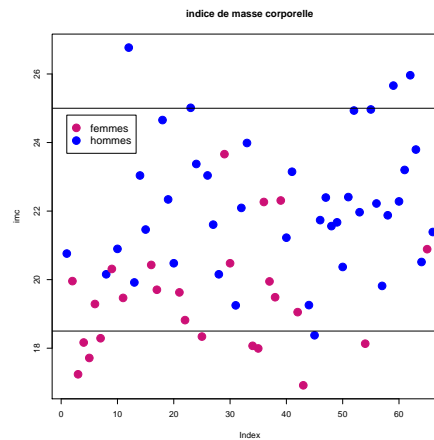
```
par(mfcol = c(2, 2))
par(mar = c(5, 4, 2, 2))
plot(imc[sexe == "h"], tai[sexe == "h"], xlim = c(min(imc), max(imc)),
      ylim = c(min(tai), max(tai)), pch = 20, xlab = "imc", ylab = "taille",
      main = "hommes")
plot(imc[sexe == "h"], poi[sexe == "h"], xlim = c(min(imc), max(imc)),
      ylim = c(min(poi), max(poi)), pch = 20, xlab = "imc", ylab = "poids")
plot(imc[sexe == "f"], tai[sexe == "f"], xlim = c(min(imc), max(imc)),
      ylim = c(min(tai), max(tai)), pch = 20, xlab = "imc", ylab = "taille",
      main = "femmes")
plot(imc[sexe == "f"], poi[sexe == "f"], xlim = c(min(imc), max(imc)),
      ylim = c(min(poi), max(poi)), pch = 20, xlab = "imc", ylab = "poids")
```



7. Un seul graphique

Version de base

```
plot(imc, pch = 19, cex = 2, col = c("deppink3", "blue")[unclass(sexe)])
abline(h = 18.5)
abline(h = 25)
title("indice de masse corporelle")
l10 = c("femmes", "hommes")
legend(1, 24.8, l10, pch = 21, pt.cex = 2, cex = 1.25, col = c("deppink3",
"blue"), pt.bg = c("deppink3", "blue"))
```



Version Graphe de Cleveland

```
neoimc = c(sort(imcf), sort(imch))
neosex = as.factor(rep(c("f", "h"), c(25, 41)))
moyennes = c(mean(imcf), mean(imch))
dotchart(neoimc, main = "indice de masse corporelle", pch = 21,
bg = grey(0.8), cex = 1.25, groups = neosex, gdata = moyennes,
gpch = 19, gcol = c("deppink3", "blue"))
abline(v = 18.5, lwd = 2, col = grey(0.6))
abline(v = 25, lwd = 2, col = grey(0.6))
```

indice de masse corporelle

