Atelier 5 Repondu

Antonio Falcó

16/11/2019

Indice de masse corporelle (IMC) chez des enfants

Un échantillon de dossiers d'enfants a ete saisi. Ce sont des enfants vus lors d'une visite en 1er section de maternelle en 1996-1997 dans des ecoles de Bordeaux (Gironde, France). L'échantillon est constitue de 152 enfants ages de 3 ou 4 ans.

Variables et codage

Description	Unite ou Codage	Variable
Sexe	`F` pour fille ; `G` pour garcon	`SEXE`
Ecole située en zone d'education prioritaire	`O` pour oui; `N` pour non	`zep`
Poids	`Kg` (arrondi à 100g près)	`poids`
Âge à la date de la visite	Années	`an`
Âge à la date de la visite	Mois	`mois`
Taille	Cm (arrondi à 0.5cm près)	`taille`

Vous avez de telecharcher le fichier `imcenfant.csv` de le site:

https://github.com/afalco/Atelier5/

```
imcenfant <- read.csv2("~/Dropbox/Public/Biostatistique/Donnees/imcenfant.csv")
names(imcenfant)
```

```
## [1] "SEXE"    "zep"     "poids"   "an"      "mois"    "taille"
```

1. Calculer l'IMC de tous les enfants et l'âge et rajoutez une colonne pour l'âge et une autre pour l' IMC.

```
# Utiliser pour repondre la question 1
IMC <- imcenfant$poids/(imcenfant$taille/100)^2
age <- imcenfant$an+imcenfant$mois/12
imcenfant <- data.frame(imcenfant, age,IMC)
```

2. Extrayez les enfants ayant un IMC < 15 et un age <= 3.5 ans.

```
# Utiliser pour repondre la question 2
enfants.sexe <-imcenfant$SEXE[(imcenfant$IMC < 15 & imcenfant$age <= 3.5)]
enfants.sexe
```

```
## [1] G F G F F G G F
## Levels: F G
```

3. Donnez le nombre d'enfants vérifiant les conditions ci-dessus.

```
# Utiliser pour repondre la question 3
length(enfants.sexe)
```

```
## [1] 8
```

4. Est-ce que la plus part des enfants ayant un IMC < 15 et un age <= 3.5 ans sont dans une école située en zone d'éducation prioritaire?

```r
# Utiliser pour repondre la question 4
enfants.zep <-imcenfant$zep[(imcenfant$IMC < 15 & imcenfant$age <= 3.5)]
enfants.zep
```

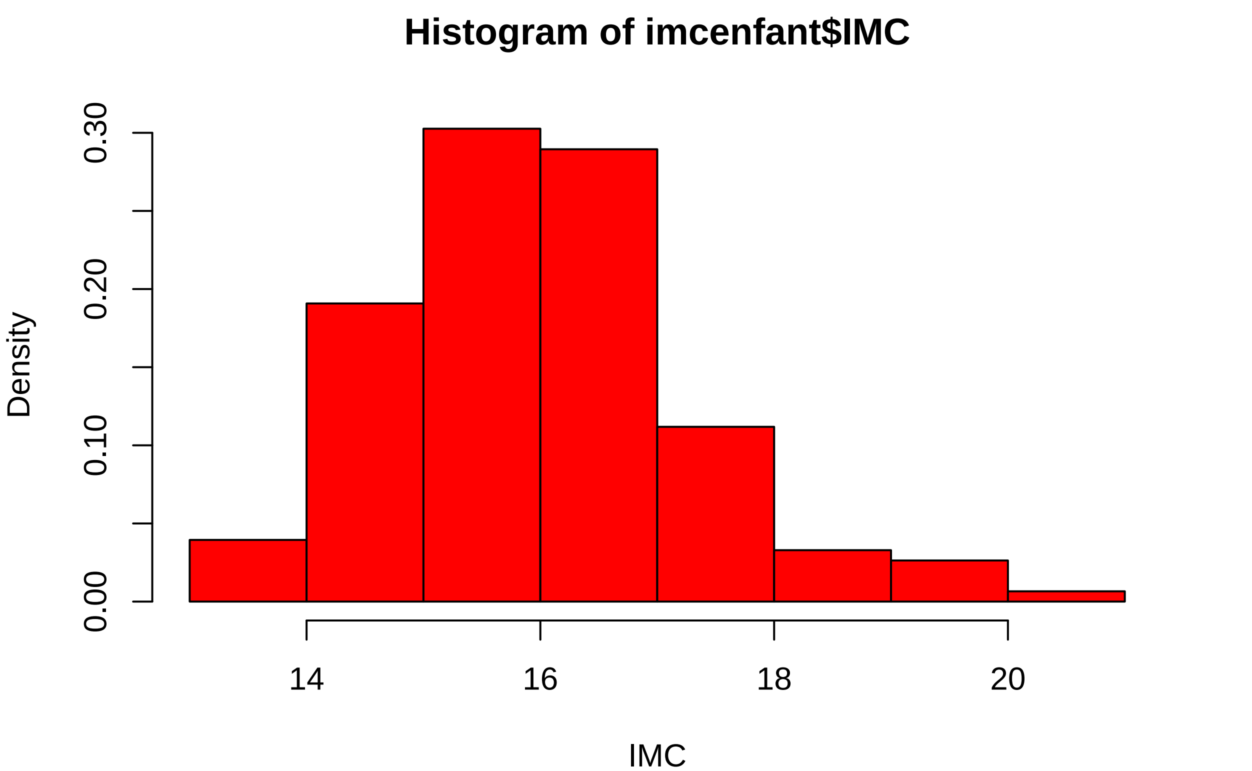
```
## [1] O O O O O N N N
## Levels: N O
```

```r
table(enfants.zep)
```

```
## enfants.zep
## N O
## 3 5
```

5. Représentez la fonction de densité de probabilité de l'IMC chez l'échantillon.

```r
# Utiliser pour repondre la question 4
hist(imcenfant$IMC,freq = FALSE,xlab = "IMC",col="red")
```



6. Est-ce on peut modelé l'IMC avec une variable aléatoire normal?

6.1. Montrer que la différence entre la médian et la moyenne est très petite et alors on peut supposer que les deux sont égal.

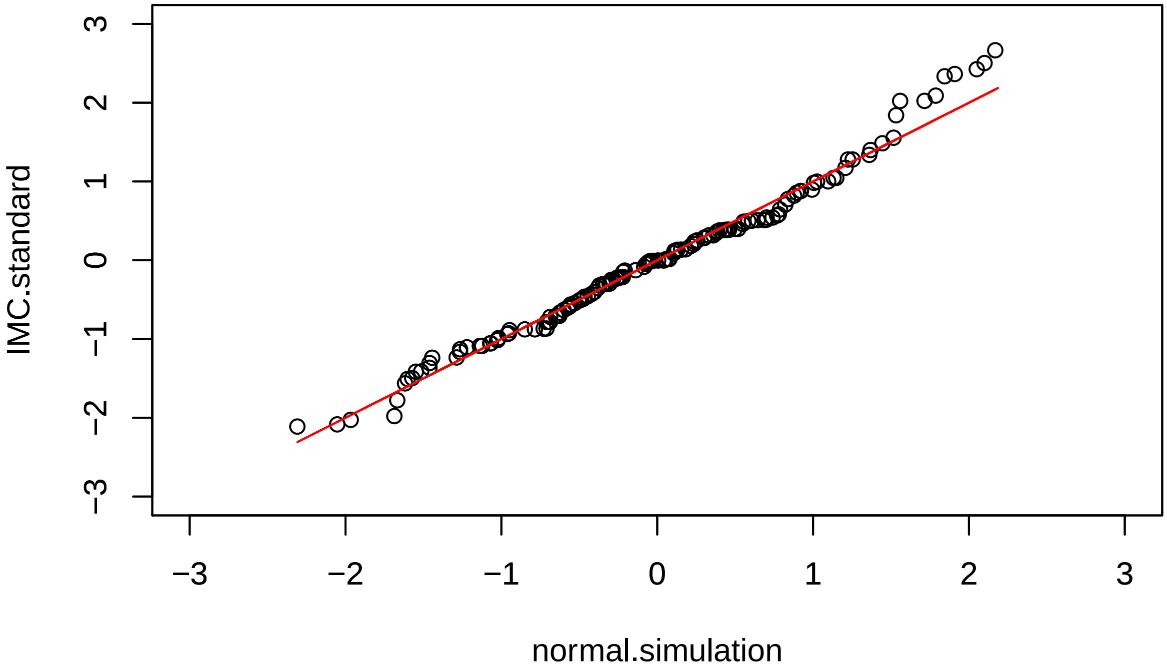
```r
# Utiliser pour repondre la question 6.1
summary(imcenfant$IMC)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.30   15.10   15.98   16.00   16.65   20.29
```

On obtient une moyenne de 16.0049754 et une medianne de 15.9837978, alors la difference est de 0.0211776.

6.2 Utilisez la fonction `qqplot()` pour étudier si l'IMC a le même fonction de distribution que une variable alèatoire normal.

```r
IMC.standard <- (imcenfant$IMC-mean(imcenfant$IMC))/sd(imcenfant$IMC)
set.seed(123)
normal.simulation <- rnorm(length(IMC.standard))
qqplot(normal.simulation,IMC.standard,xlim=c(-3,3),ylim=c(-3,3))
par(new=T)
plot(normal.simulation,normal.simulation,type="l",col="red",xlim=c(-3,3),ylim=c(-3,3),
     xlab = "",ylab = "")
```



6.3 Est-qu'on peut quantifier la différence entre les deux fonctions de distributions?

```r
# Utiliser pour repondre la question 6.3
# On va choisir 10 quantiles
x_maille <- seq(from=0,to=0.9,by=0.1)
#length(x_maille)
#length(imcenfant$IMC)
set.seed(123)
A <- quantile(IMC.standard,x_maille)
B <- quantile(rnorm(length(imcenfant$IMC)),x_maille)
erreur <- sum((A-B)^2)/length(x_maille)
erreur
```

```
## [1] 0.01270941
```

Pour commparer on utilise la variable A qui contient 10 quantiles entre le 0 et le 0.9 de probabilité de la variable `IMC` standarisé et la variable B qui contient les memes valueurs pour un distribution normal. Alors on a

$$A = (A_0, \ldots, A_9) \text{ et } B = (B_0, \ldots, B_9)$$

où $A_i = \texttt{quantile}(\texttt{IMC}, 0.1 * i)$ et $B_i = \texttt{quantile}(\texttt{normal}, 0.1 * i)$ avec $i = 0, 1, \ldots, 9$. Alors on calcule l'erreur parmi la formule:

$$\frac{1}{n}\sum_{i=1}^{n}(A_i - B_i)^2$$

On a un erreur autour de 0.0127094

7. Est-ce que l'IMC chez les enfants dans une école située en zone d'éducation prioritaire est différent de les enfants qui ne sont pas dans une école située en zone d'éducation prioritaire?

7.1 Donnez le résumé statistique de l'IMC chez les enfants dans une école située en zone d'éducation prioritaire.

```
# Utiliser pour repondre la question 7.1
IMC.zepO <- imcenfant$IMC[imcenfant$zep=="O"]
summary(IMC.zepO)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.33   15.10   16.00   16.09   16.70   20.29
```

Il y a dans l'échantillon 111 enfants dans dans une école située en zone d'éducation prioritaire.

7.2 Donnez le résumé statistique de l'IMC chez les enfants que ne sont pas dans une école située en zone d'éducation prioritaire.

```
# Utiliser pour repondre la question 7.2
IMC.zepN <- imcenfant$IMC[imcenfant$zep=="N"]
summary(IMC.zepN)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.30   15.30   15.84   15.76   16.50   17.65
```

Il y a dans l'échantillon 41 enfants qui ne sont pas dans dans une école située en zone d'éducation prioritaire.

8. Quelle est la probabilité de trouver une fille dans dans une école située en zone d'éducation prioritaire?

```
# Utiliser pour repondre la question 8
sexe.zepO <- imcenfant$SEXE[imcenfant$zep=="O"]
sexe.zep1 <- imcenfant$SEXE[imcenfant$zep=="N"]
probabilite.zepO <- length(sexe.zepO)/length(imcenfant$zep)
probabilite.zep1 <- length(sexe.zep1)/length(imcenfant$zep)
table(sexe.zepO)
```

```
## sexe.zepO
##  F  G
## 53 58
```

```
prob.sexe.zepO <- table(sexe.zepO)/length(imcenfant$SEXE)
prob.sexe.zepO[1]
```

```
##         F
## 0.3486842
```

```
prob.sexe.zepO[2]
```

```
##         G
## 0.3815789
```

On a $\Pr(\texttt{zep} = \texttt{O}) = 0.7302632$ et $\Pr(\texttt{zep} = \texttt{N}) = 0.2697368$. Alors,

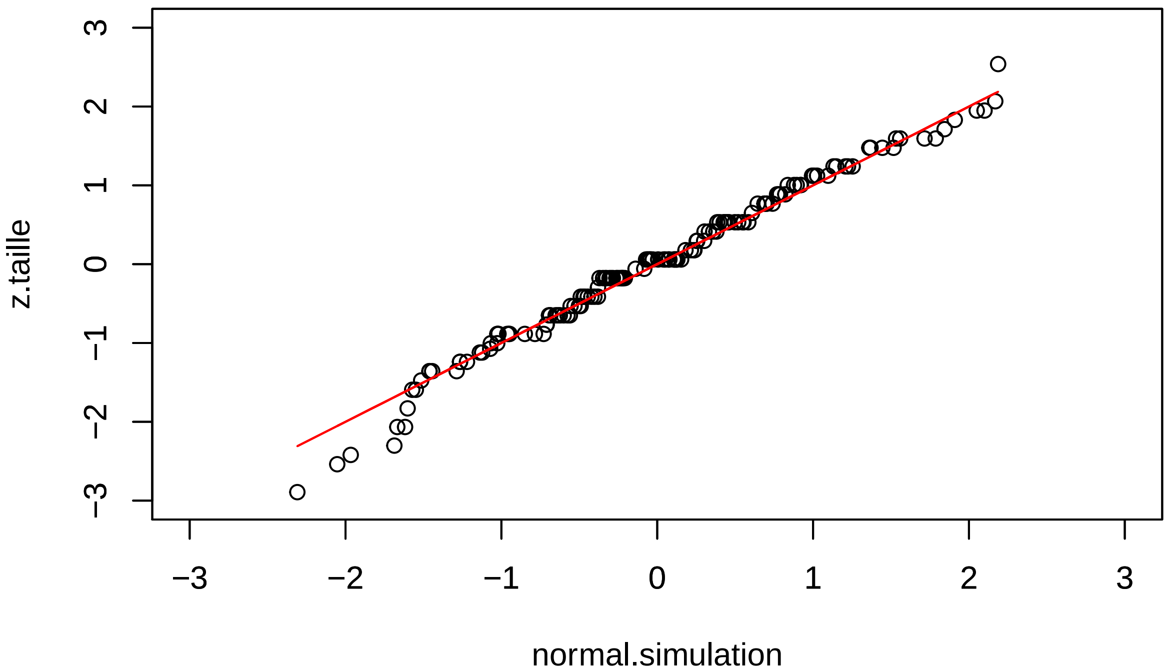
$$\Pr(\texttt{SEXE} = \texttt{F}|\texttt{zep} = \texttt{O}) = \frac{\Pr(\texttt{SEXE} = \texttt{F} \cap \texttt{zep} = \texttt{O}))}{\Pr(\texttt{zep} = \texttt{O})} = \frac{\texttt{prob.sexe.zepO[1]}}{\texttt{probabilite.zepO}}$$

d'ou $\Pr(\texttt{SEXE} = \texttt{F}|\texttt{zep} = \texttt{O}) = 0.4774775$

9. Est-ce qu'on peut modelé la taille des enfants avec un variable aléatoire normal?

```
# Utiliser pour repondre la question 9
z.taille <- (imcenfant$taille-mean(imcenfant$taille))/sd(imcenfant$taille)
set.seed(123)
```

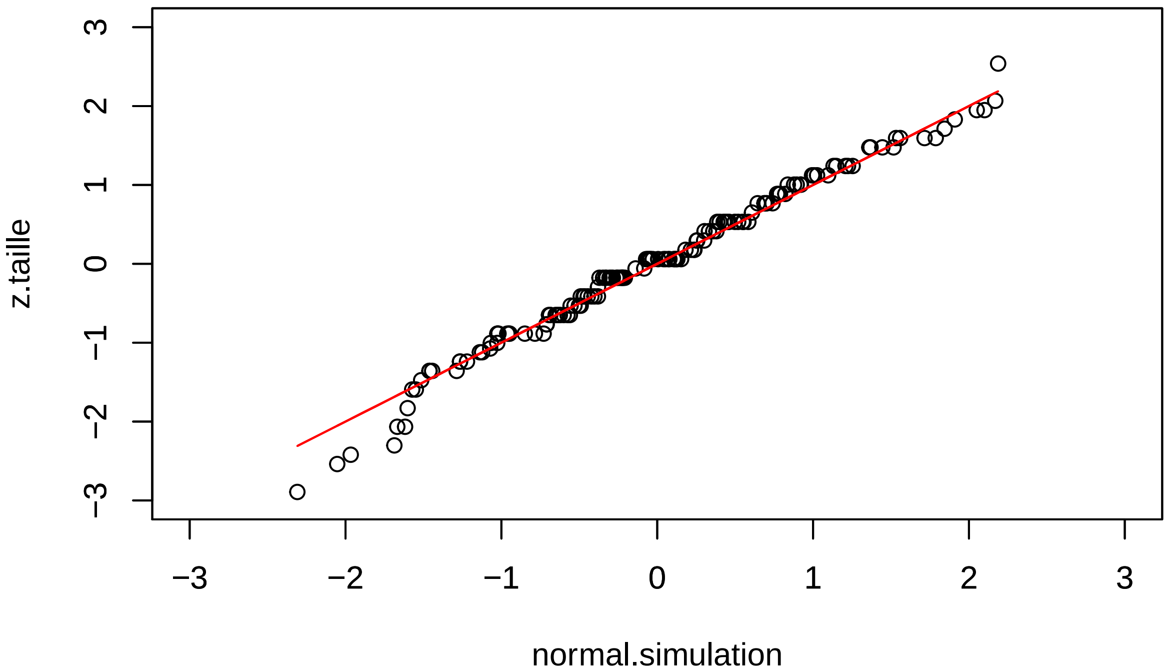
```
normal.simulation <- rnorm(length(z.taille))
qqplot(normal.simulation,z.taille,xlim=c(-3,3),ylim=c(-3,3))
par(new=T)
plot(normal.simulation,normal.simulation,type="l",col="red",xlim=c(-3,3),ylim=c(-3,3),
     xlab = "",ylab = "")
```



Commentaire: La difference absolue entre la moyenne et la medianne est -0.2519737 et la relative -0.002501 est très petite. La distribution de les quantiles est très proche autour la diagonal, alors on peut supposer que la variable `taille` suit un distribution normal.

10. Est-ce qu'on peut modelé le poids des enfants avec un variable aléatoire normal?

```
# Utiliser pour repondre la question 10
z.poids <- (imcenfant$poids-mean(imcenfant$poids))/sd(imcenfant$poids)
set.seed(123)
normal.simulation <- rnorm(length(z.poids))
qqplot(normal.simulation,z.taille,xlim=c(-3,3),ylim=c(-3,3))
par(new=T)
plot(normal.simulation,normal.simulation,type="l",col="red",xlim=c(-3,3),ylim=c(-3,3),
     xlab = "",ylab = "")
```



Commentaire: La difference absolue entre la moyenne et la medianne est 0.2802632 et la relative 0.0172149. La distribution de les quantiles est très proche autour la diagonal, alors on peut supposer que la variable `poids` suit un distribution normal.