TP4

Gautier Poursin

03/05/2020

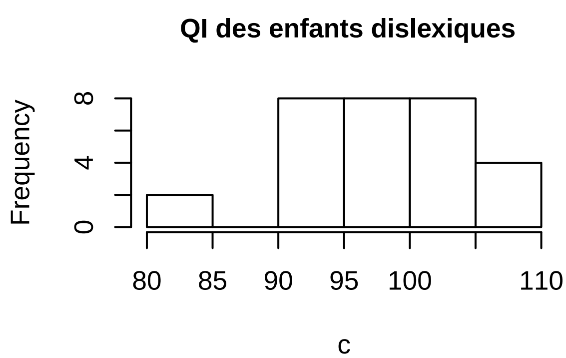
Exercice 1

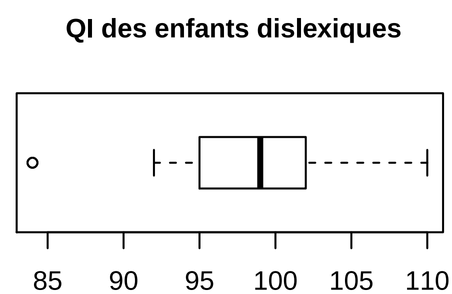
On va commencer par télécharger les variables dans un data frame.

```
data<- read.csv('QI.txt',sep="")
```

On va ensuite s'intéresser aux variables par une étude descriptives de ces variables.

```
## La moyenne est de 98.43333 et la variance de  42.46092 .
```



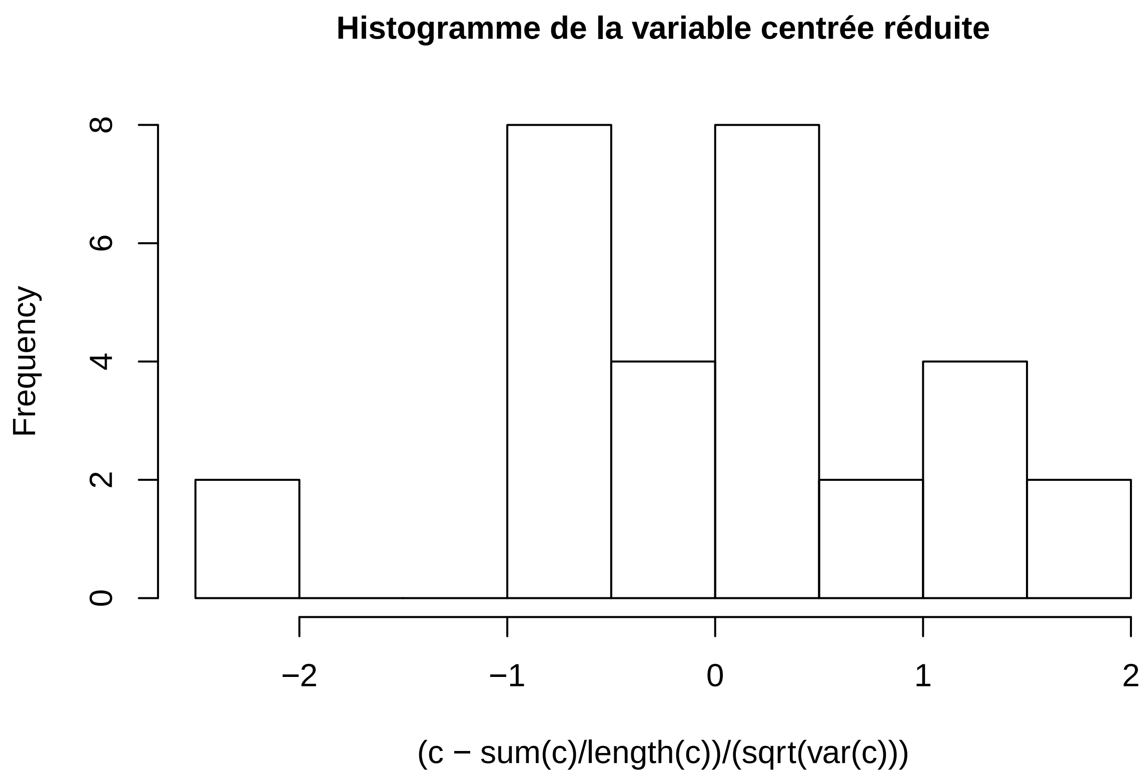


A la vue des résultats, on peut déduire que le QI des personnes dislexiques est similaires aux personnes non dislesiques. En effet, ils possèdent en moyenne un QI de 98,4, qui est très proche de celui des personnes non dislexiques.

3

Tout d'abord, nous allons vérifier si notre échantillon peut suivre une loi gaussienne en le centrant puis en le normalisant. Ensuite nous tracerons l'histogramme associée à cette loi centrée réduite et nous regarderons l'allure de la loi. \ Enfin, nous effectuerons un test statistiques pour la moyenne d'une loi normale avec la variance connue.

```
etude_variable_centree_reduite<-function(c){
  hist((c-sum(c)/length(c))/(sqrt(var(c))),main="Histogramme de la variable centrée réduite",cex.main=1
}
 etude_variable_centree_reduite(data$x)
```



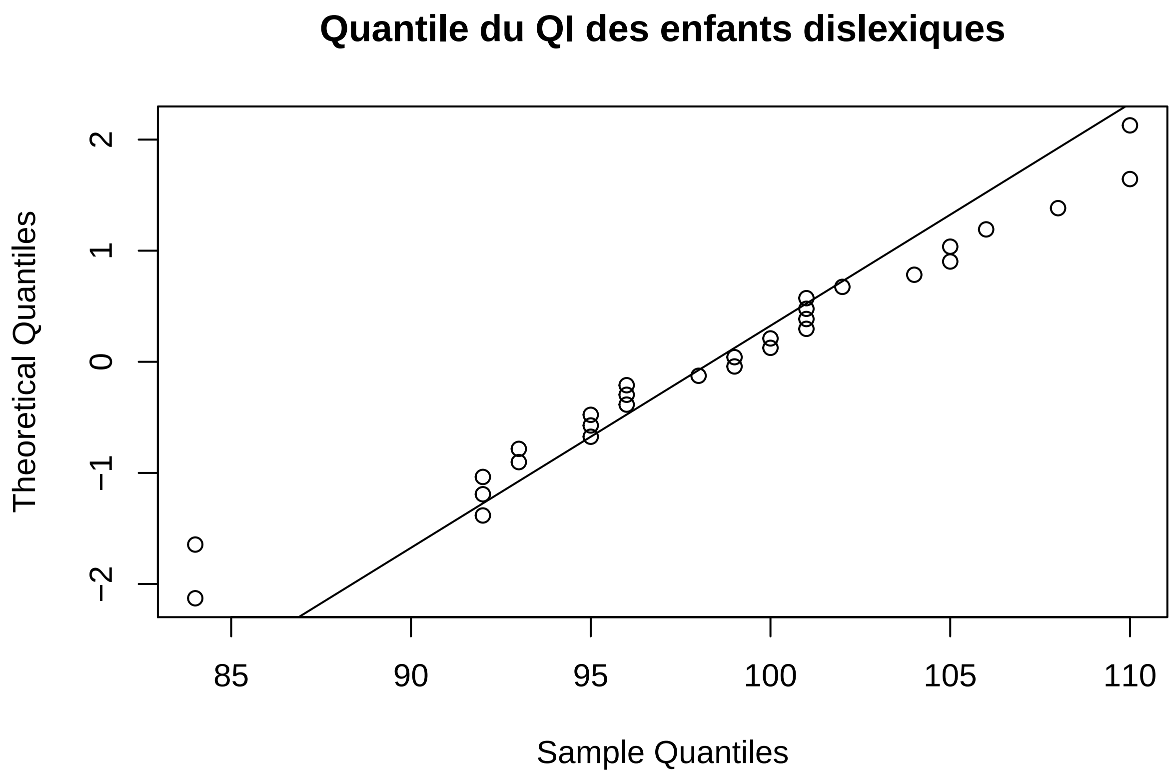
Ce graphique montre qu'il est possible d'affirmer que cette loi suit une loi gaussienne. On a pas réellement la forme d'une loi gaussienne lorsqu'on centre et réduit la loi mais on s'approche de l'allure d'une gaussienne. On peut imaginer qu'avec un nombre conséquent de valeurs, nous aurions bien l'allure d'une gaussienne. \ Pour confirmer cela, on peut effectuer une comparaison de quantiles ou un test de shapiro.

```
shapiro.test(rnorm(100, mean = sum(data$x)/length(data$x), sd = var(data$x)))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rnorm(100, mean = sum(data$x)/length(data$x), sd = var(data$x))
## W = 0.98673, p-value = 0.4199
```

La pvalue ici est assez élevée pour confirmer le fait que l'on suit une loi Gaussienne. \ On peut aussi réaliser un diagramme Quantile/Quantile:

```
qqnorm(data$x, datax=TRUE, main="Quantile du QI des enfants dislexiques")
qqline(data$x,datax=TRUE)
```



Ce diagramme confirme bien le fait que l'on suit une loi normale.Les quantiles sont proches de la droite.

5

On va tout d'abord définir les hypothèses: $H_0 : \mu = \mu_0$, $H_1 : \mu! = \mu_0$ Notre variance est inconnue. On a a statistique de test suivante: $\sigma_n = \frac{sqrt(n)(\hat{X}_n - \mu_0)}{S_n}$. De plus, l'énoncé nous indique $\mu_0 = 100$

```
statistique_test<-function(n,c) {sqrt(n)*(sum(c-100)/length(c))/sd(c)}
statistique_test(30,data$x)
```

```
## [1] -1.316869
```

On va maintenant déterminer la zone de rejet pour cette statistique de test, avec $\alpha = 0.05$. On rejette l'hypothèse si $statistique\_test > F^{-1}_{T_{n-1}}(0.95)$ avec F fonction de répartition d'une loi de Student de degré de liberté n-1.

```
Ka<-function (x,n) {qt(x,n-1)}
Ka(0.95,30)
```

```
## [1] 1.699127
```

On rejette l'hypothèse si $\hat{X}_n > \mu_0 + \frac{S_n}{\sqrt{(n)}} F^{-1}_{T_{n-1}}(1-\alpha)$.

```
Mean<-function(c,n) {sum(c)/n}
Ka1<-function(n,c){100+sd(c)/(n^0.5)*Ka(0.95,30)}
Mean(data$x,30)
```

```
## [1] 98.43333
```

```
Ka1(30,data$x)
```

```
## [1] 102.0214
```

Ici, nous n'avons pas $\hat{X}_n > \mu_0 + \frac{S_n}{\sqrt{(n)}} F^{-1}_{T_{n-1}}(1-\alpha)$ donc on garde l'hypothèse de test!

6

On va utiliser la fonction t.test:

```
t.test(data$x,mu=100,conf.level =0.9)
```

```
##
##  One Sample t-test
##
## data:  data$x
## t = -1.3169, df = 29, p-value = 0.1982
## alternative hypothesis: true mean is not equal to 100
## 90 percent confidence interval:
##   96.4119 100.4548
## sample estimates:
## mean of x
##  98.43333
```

Ici, notre pvalue est supérieure à 0.1, ce qui est nécessaire lorsqu'on souhaite avoir un intevral de confiance au niveau 90%.

```
t.test(data$x,mu=100,conf.level =0.95)
```

```
##
##  One Sample t-test
##
## data:  data$x
## t = -1.3169, df = 29, p-value = 0.1982
## alternative hypothesis: true mean is not equal to 100
```

```
## 95 percent confidence interval:
##   96.00014 100.86652
## sample estimates:
## mean of x
##  98.43333
```

On remarque aussi qu'on obtient immédiatement l'interval de confiance avec cette commande t.test. L'interval de confiance à 95% est plus grand car on souhaite avoir plus de valeurs que dans celui a 90%.

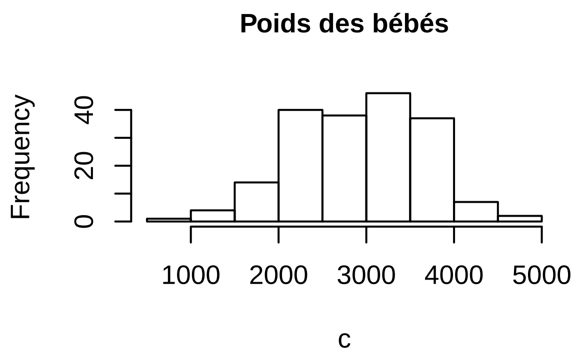
Exercice 2

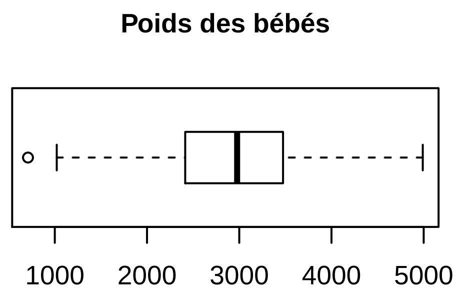
On va commencer par télécharger les variables dans un data frame.

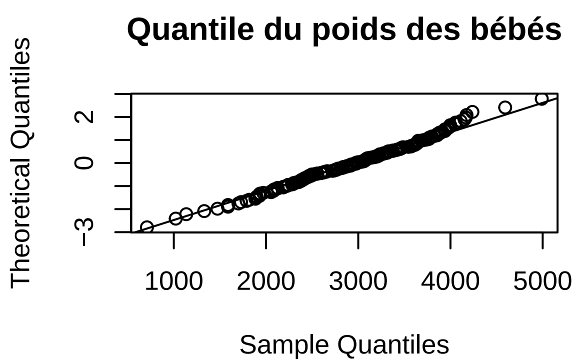
```
birth<- read.csv('birthwt.txt',sep="")

etude_birth <- function(c, nom) {
  m<-sum(c)/length(c)
  cat("La moyenne est de",m, "et la variance de ", var(c),".\n")
  par(mfrow=c(2,2))
  hist(c,main=nom,cex.main=1)
  boxplot(c, main=nom,horizontal = TRUE,cex.main=1)
}
etude_birth(birth$bwt,"Poids des bébés")
```

```
## La moyenne est de 2944.656 et la variance de  531473.7 .
```

```
qqnorm(birth$bwt, datax=TRUE, main="Quantile du poids des bébés")
qqline(birth$bwt,datax=TRUE)
```





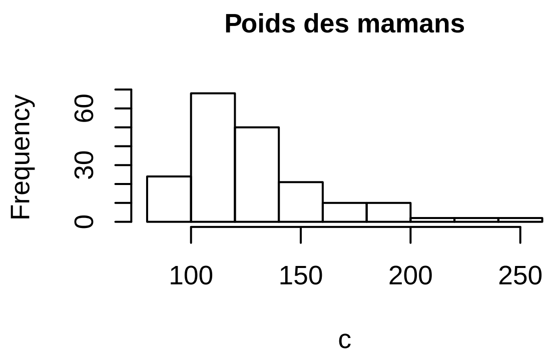


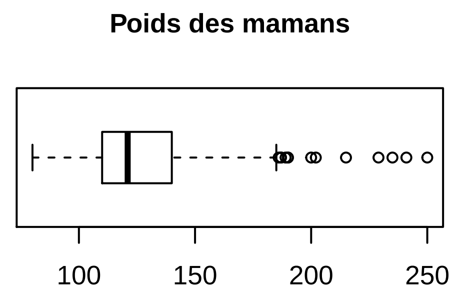
Le poids des bébés semble suivre une loi Normale.

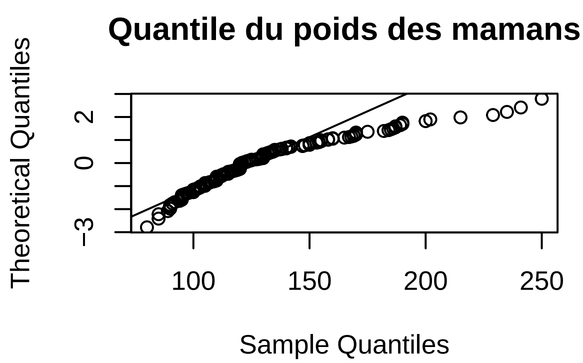
```
etude_birth(birth$lwt,"Poids des mamans")
```

```
## La moyenne est de 129.8148 et la variance de  935.0985 .
```

```
qqnorm(birth$lwt, datax=TRUE, main="Quantile du poids des mamans")
qqline(birth$lwt,datax=TRUE)
```







il est plus difficle d'affirmer que le poids des mamans suit une loi normale, on va effectuer un test de shapiro:

```
shapiro.test(rnorm(130, mean = sum(birth$lwt)/length(birth$lwt), sd = var(birth$lwt)))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rnorm(130, mean = sum(birth$lwt)/length(birth$lwt), sd = var(birth$lwt))
## W = 0.99481, p-value = 0.9208
```

Le test de shapiro nous apporte l'information que c'est bien une loi normale.

2a.

Pour répondre à la question, nous allons mettre en place une série de test, comme pour l'exercice 1.On déterminera la statistique de test puis nous calculerons la zone de rejet afin de déterminer si on garde notre hypothèse ou non.

```
statistique_test<-function(n,c) {sqrt(n)*(sum(c-130)/length(c))/sd(c)}
statistique_test(length(birth$lwt),birth$lwt)
```

```
## [1] -0.08325464
```

On rejette l'hypothèse si $statistique\_test > F^{-1}_{T_{n-1}}(0.95)$ avec F fonction de répartition d'une loi de Student de degré de liberté n-1.

```r
Ka<-function (x,n) {qt(x,n-1)}
Ka(0.95,length(birth$lwt))
```

```
## [1] 1.652999
```

```r
mean<-function(c,n) {sum(c)/n}
Ka2<-function(n,c){130+sd(c)/(n^0.5)*Ka(0.95,length(c))}
mean(birth$lwt,length(birth$lwt))
```

```
## [1] 129.8148
```

```r
Ka2(length(birth$lwt),birth$lwt)
```

```
## [1] 133.6768
```

Ici, la moyenne empirique est inférieure à Ka2 donc nous **conservons l'hypothèse de test!**

2c.

```r
t.test(birth$lwt,mu=130)
```

```
##
##  One Sample t-test
##
## data:  birth$lwt
## t = -0.083255, df = 188, p-value = 0.9337
## alternative hypothesis: true mean is not equal to 130
## 95 percent confidence interval:
##  125.4270 134.2027
## sample estimates:
## mean of x
##  129.8148
```

Les valeurs de ce test sont en rapport avec nos valeurs calculées! En effet, on retrouve bien que la valeur de la statistique est de -0.08. De plus, nous obtenons un interval de confiance à 95%. La forte pvalue confirme l'hypothèse, on conserve notre hypothese.

3.

On s'interesse maintenant à la proportion de femme touchées par l'hypertension lors de leur grossesse. On va déterminer la statistique de test.

```r
statistique_test<-function(n,c) {sqrt(n)*(sum(c-130)/length(c))/sd(c)}
statistique_test(length(birth$ht),birth$ht)
```

```
## [1] -7306.252
```

On rejette l'hypothèse si $statistique\_test > F^{-1}_{T_{n-1}}(0.95)$ avec F fonction de répartition d'une loi de Student de degré de liberté n-1.

```r
Ka<-function (x,n) {qt(x,n-1)}
Ka(0.95,length(birth$lwt))
```

```
## [1] 1.652999
```

```r
mean<-function(c,n) {sum(c)/n}
Ka3<-function(n,c){0.11+sd(c)/(n^0.5)*Ka(0.95,length(c))}
mean(birth$ht,length(birth$ht))
```

```
## [1] 0.06349206
```

```r
Ka3(length(birth$ht),birth$ht)
```

```
## [1] 0.1393974
```

Ici, nous sommes contraint de rejetter l'hypothèse de test! Noux avons $statistique\_test > F^{-1}_{T_{n-1}}(0.95)$.

```r
prop.test(sum(birth$ht),p=0.11,n=189,correct=FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  sum(birth$ht) out of 189, null probability 0.11
## X-squared = 4.1757, df = 1, p-value = 0.04101
## alternative hypothesis: true p is not equal to 0.11
## 95 percent confidence interval:
##  0.03668973 0.10768513
## sample estimates:
##          p
## 0.06349206
```

La faible p-value confirme le rejet de notre hypothèse.

3b.

```r
prop.test(sum(birth$ht[birth$low>0]),n=sum(birth$low>0),p=0.11)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  sum(birth$ht[birth$low > 0]) out of sum(birth$low > 0), null probability 0.11
## X-squared = 1.7313e-05, df = 1, p-value = 0.9967
## alternative hypothesis: true p is not equal to 0.11
## 95 percent confidence interval:
##  0.05300256 0.23536702
## sample estimates:
##         p
## 0.1186441
```

```r
prop.test(sum(birth$ht[birth$low<1]),n=sum(birth$low<1),p=0.11)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  sum(birth$ht[birth$low < 1]) out of sum(birth$low < 1), null probability 0.11
## X-squared = 6.0847, df = 1, p-value = 0.01364
## alternative hypothesis: true p is not equal to 0.11
## 95 percent confidence interval:
##  0.01424520 0.09198853
## sample estimates:
##          p
## 0.03846154
```

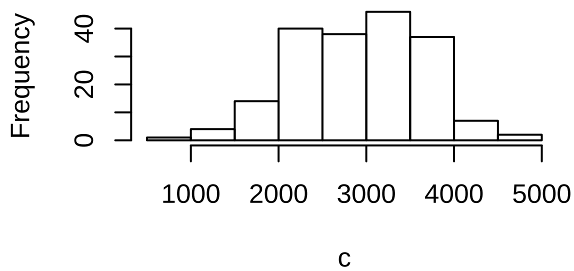
Dans notre premier cas, la pvalue est élevé donc nous confirmons l'hypothèse de test. Lorsqu'on passe au deuxième cas, la pvalue est très faible donc on rejette l'hypothèse de test. ## 4.

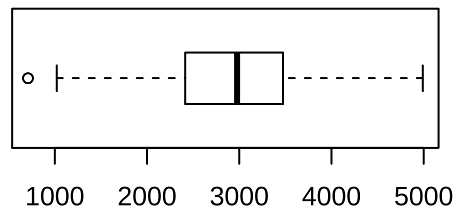
```r
etude_variable(birth$bwt,"")
```

```
## La moyenne est de 2944.656 et la variance de  531473.7 .
```

```r
t.test(birth$bwt,mu=3300,conf.level=0.90)
```

```
##
##  One Sample t-test
##
## data:  birth$bwt
## t = -6.701, df = 188, p-value = 2.339e-10
## alternative hypothesis: true mean is not equal to 3300
## 90 percent confidence interval:
##  2857.000 3032.312
## sample estimates:
## mean of x
##  2944.656
```





Le poids moyen des bébés suit donc lui aussi une loi normale. Cependant, le poids moyen s'éloigne de poids de référence: il s'éloigne de 400g puisque en moyenne, les bébés de l'études pèsent 2.9kg. D'ailleurs, cette très faible valeur de la pvalue confirme le rejet de l'hypothèse de base qui était ce poids moyen des bébés à la naissance. Les bébés ne sont pas représentatifs du poids moyen.

Conclusion

Ce TP m'a permis d'avoir une meilleure vision des tests et surtout de savoir mettre en place des tests informatiquement. J'ai pu aussi remarquer l'importance de la p-value, qui permet de de s'assurer si le rejet ou la conservation d'une hypothèse est une bonne chose ou non. Lorsque la pvalue est faible, il faut rejetter l'hypothèse. Sinon, on conserve notre hypothèse émise.