Examen de PMS 2021 - Nom : Prénom :

Première partie

Question 1.1

Enoncé Calculer les principaux indicateurs statistiques de cet échantillon et commenter les résultats.

```r
lumiere<-c(850,750,900,1040,930,850,950,980,970,880,890,920,930,1050,960,810,1000,950,860,990,780,810,83
n<-length(lumiere)
n
```

Réponse

```
## [1] 24
```

```r
summary(lumiere)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   750.0   850.0   915.0   907.9   962.5  1050.0
```

```r
#ecart type
sd(lumiere)
```

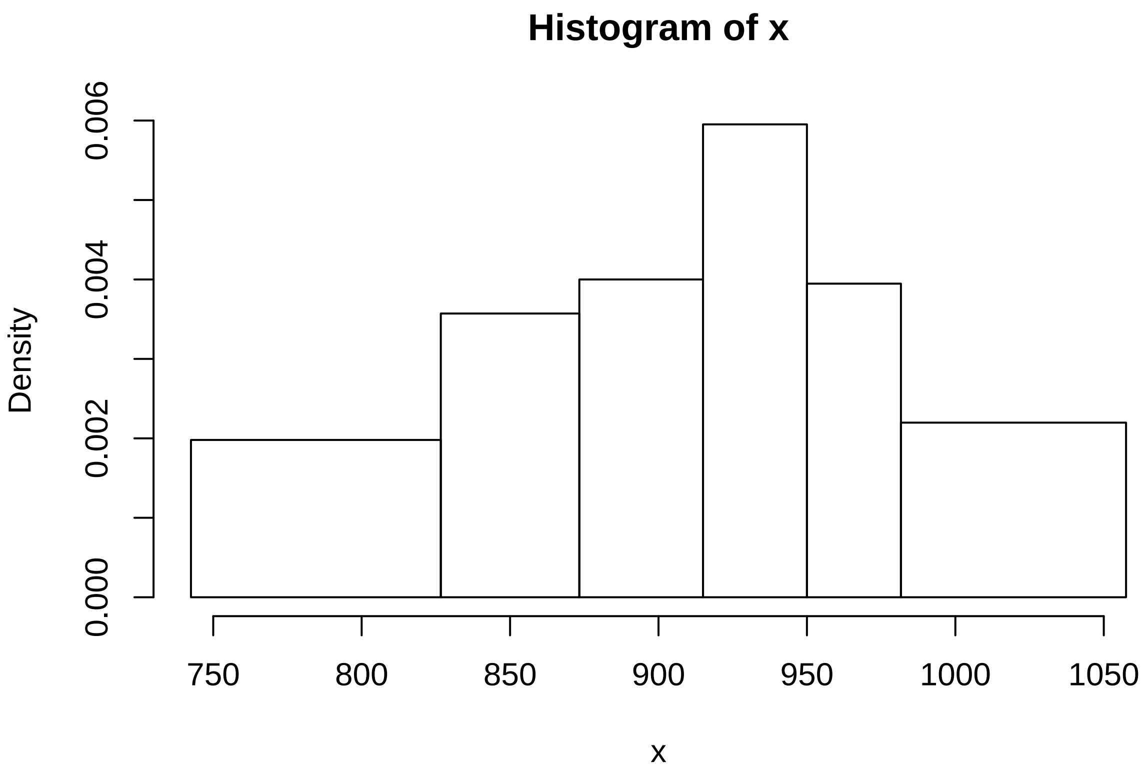
```
## [1] 79.67211
```

On remarque que l'écart type est de 79.67 , or l'écart type représentant l'écart des valeurs à la moyenne ,les valeurs varient très peu par rapport à un ordre de grandeur de 300 000. On remarque également que la moyenne est assez proche de 990. On en déduis donc à priori que l'expérience de Michelson confirme la valeur de Cornus.

Question 1.2

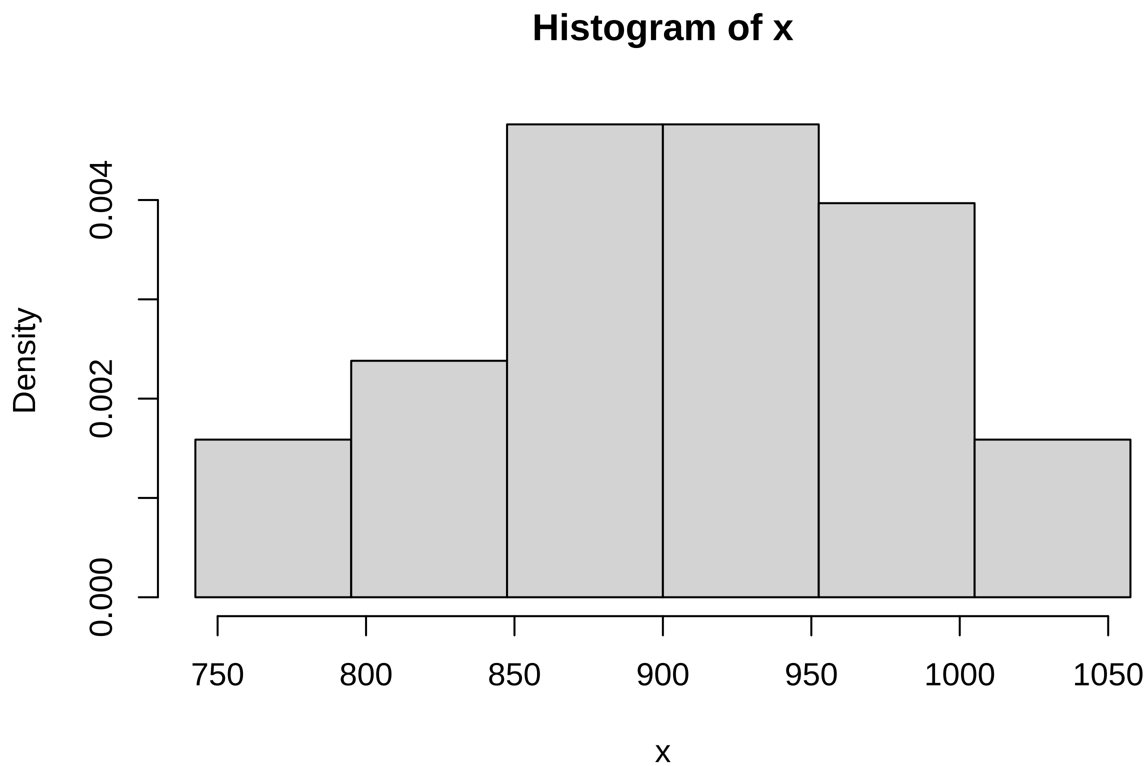
Enoncé A l'aide d'un histogramme et d'un graphe de probabilités, déterminer si l'hypothèse de loi normale est raisonnable.

```r
source("histoeff.R")
histoeff(lumiere)
```



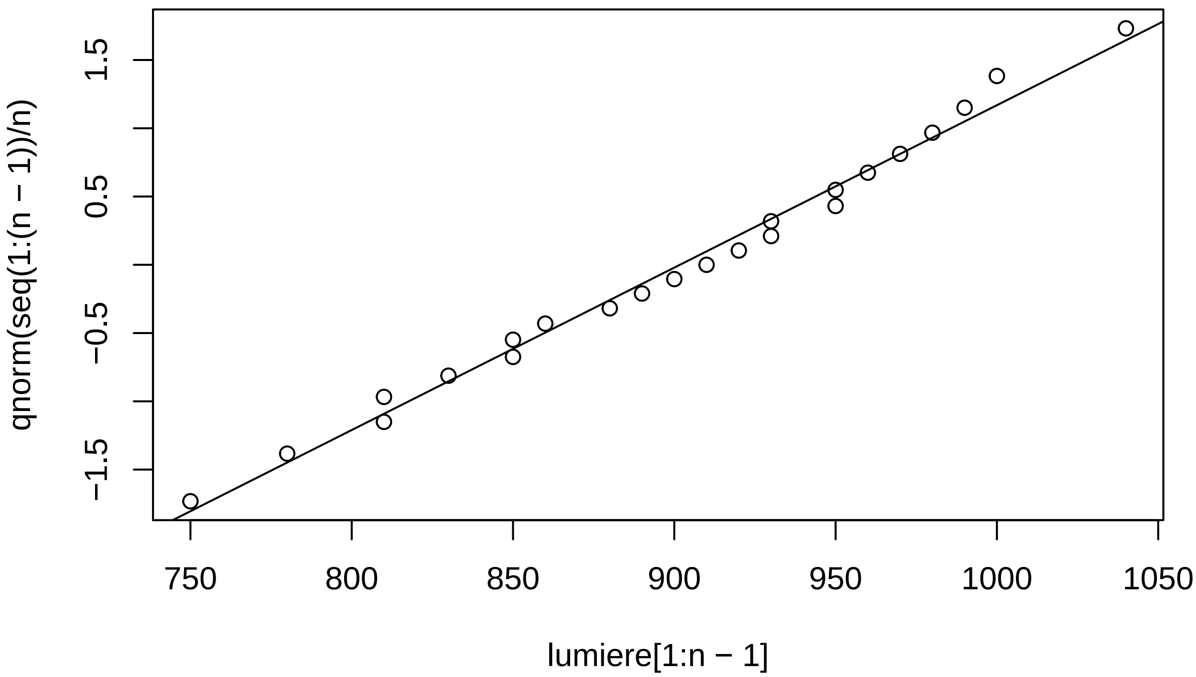
Réponse

```r
source("histolarg.R")
histolarg(lumiere)
```




```
lumiere <- sort(lumiere)

plot(lumiere[1:n-1],qnorm(seq(1:(n-1))/n))
reg<-lm(qnorm(seq(1:(n-1))/n)~lumiere[1:(n-1)])
abline(reg)
```



On trace l'histogramme a classe de même effectif puis l'histogramme a classe de même largeur ce qui nous met sur la voie d'une loi normale.

Afin de vérifier on tracer donc le graphe de probabilité qui est très proche d'une droite. On en conclut que l'hypothèse de loi normale est correcte.

Question 1.3

Enoncé Estimer m, σ^2 et σ à l'aide du graphe de probabilités puis par la méthode du maximum de vraisemblance. Que retenez-vous comme estimations ?

```
sigma_reg<- 1/(reg$coefficients[2])
m_reg <- -sigma_reg * reg$coefficients[1]


cat("On estime sigma par la regression ce qui donne :",sigma_reg, "\n")
```

Réponse

```
## On estime sigma par la regression ce qui donne : 84.0756
```

```r
cat("Ainsi on estime sigma carré par  :",sigma_reg^2, "\n")
```

```
## Ainsi on estime sigma carré par  : 7068.706
```

```r
cat("On estime m par la regression :",m_reg, "\n")
```

```
## On estime m par la regression : 901.7391
```

```r
cat("\n")
```

```r
sigma_carre_MV <- var(lumiere)*(n-1)/n
m_MV <- mean(lumiere)
cat("On estime sigma par la méthode du maximum de vraisemblance :", sqrt(sigma_carre_MV), "\n")
```

```
## On estime sigma par la méthode du maximum de vraisemblance : 77.99461
```

```r
cat("On estime sigma carre par la méthode du maximum de vraisemblance :", sigma_carre_MV, "\n")
```

```
## On estime sigma carre par la méthode du maximum de vraisemblance : 6083.16
```

```r
cat("On estime m par la méthode du maximum de vraisemblance :",m_MV, "\n")
```

```
## On estime m par la méthode du maximum de vraisemblance : 907.9167
```

On garde les estimations obtenus grâce à la méthode du maximum de vraisemblance car les estimateurs de m et de σ^2 par cette méthode sont sans biais contrairement aux estimateurs obtenus grâce à la régression.

Question 1.4

Enoncé Donner des intervalles de confiance de seuil 5% pour m, σ^2 et σ.

```r
#Intervalle de confiance pour la moyenne

alpha<-0.05

a_moy<-mean(lumiere) - (sqrt(var(lumiere))/sqrt(n)* qt(1-alpha/2, n-1)  )
b_moy<-mean(lumiere) + (sqrt(var(lumiere))/sqrt(n)* qt(1-alpha/2, n-1)  )

cat("Intervalle de confiance au seuil 5% pour la moyenne m :\n")
```

Réponse

```
## Intervalle de confiance au seuil 5% pour la moyenne m :
```

```r
cat("[", a_moy, ",", b_moy, "]\n")
```

```
## [ 874.2741 , 941.5592 ]
```

```r
#Intervalle de confiance pour sigma carre et sigma

a<-(n-1)*var(lumiere)/qchisq(1-alpha/2, n-1)
b<-(n-1)*var(lumiere)/qchisq(alpha/2, n-1)

cat("Intervalle de confiance au seuil 5% pour la variance sigma carre :\n")
```

```
## Intervalle de confiance au seuil 5% pour la variance sigma carre :
```

```r
cat("[", a, ",", b, "]\n")
```

```
## [ 3834.364 , 12490.5 ]
```

```r
cat("Intervalle de confiance au seuil 5% pour la variance sigma :\n")
```

```
## Intervalle de confiance au seuil 5% pour la variance sigma :
```

```r
cat("[", sqrt(a), ",", sqrt(b), "]\n")
```

```
## [ 61.92225 , 111.7609 ]
```

Question 1.5

Enoncé Suite à son expérience, Michelson a proposé une estimation de la vitesse de la lumière sous la forme $a \pm b$ km/s. Que proposez-vous pour a et b ?

```r
ecart_intervalle <- (b_moy-a_moy)
moitie_ecart <- ecart_intervalle/2

a <- a_moy + moitie_ecart + 299000
b <- moitie_ecart

cat("On propose a qui vaut :", a, "\n")
```

Réponse

```
## On propose a qui vaut : 299907.9
```

```r
cat("On propose b qui vaut :", b, "\n")
```

```
## On propose b qui vaut : 33.64258
```

Question 1.6

Enoncé Mettre en place un test d'hypothèses permettant de déterminer si la valeur proposée par Cornu est validée par l'expérience de Michelson.

Réponse Afin de mettre en place le test d'hypothèse on pose les hypothèses de la manière suivante : $H_0 = "m = m_0"$ et $H_1 = "m \neq m_0$ avec $m_0 = 299990$ km/s et m vaut la valeur de la vitesse de la lumière. cela donne :

```r
t.test(lumiere ,alternative="two.sided",mu=990 )
```

```
##
##  One Sample t-test
##
## data:  lumiere
## t = -5.0472, df = 23, p-value = 4.149e-05
## alternative hypothesis: true mean is not equal to 990
## 95 percent confidence interval:
##  874.2741 941.5592
## sample estimates:
## mean of x
##  907.9167
```

Ici la p-valeur vaut environs 4.149e-05 donc seulement pour un seuil inférieur à celui-ci on ne rejèterait pas H_0 on prend donc très peu de risques à rejette l'hypothèse H_0.

Question 1.7

Enoncé Aujourd'hui, on sait que la vitesse de la lumière est c=299 792.458$ km/s. Que peut-on en conclure sur l'expérience de Michelson ?

Réponse D'après la question 4 avec les expériences de Michelson on obtient c qui appartient aux intervalles suivants :

```r
alpha = 0.05
c_a_moy<-mean(lumiere) - (sqrt(var(lumiere))/sqrt(n)* qt(1-alpha/2, n-1)  ) + 299000
c_b_moy<-mean(lumiere) + (sqrt(var(lumiere))/sqrt(n)* qt(1-alpha/2, n-1)  ) + 299000

cat("Intervalle de confiance au seuil 5%\n")
```

```
## Intervalle de confiance au seuil 5%
```

```r
cat("[", c_a_moy, ",", c_b_moy, "]\n")
```

```
## [ 299874.3 , 299941.6 ]
```

```r
alpha = 0.5
c_a_moy<-mean(lumiere) - (sqrt(var(lumiere))/sqrt(n)* qt(1-alpha/2, n-1)  ) + 299000
c_b_moy<-mean(lumiere) + (sqrt(var(lumiere))/sqrt(n)* qt(1-alpha/2, n-1)  ) + 299000

cat("Intervalle de confiance au seuil 50%\n")
```

```
## Intervalle de confiance au seuil 50%
```

```r
cat("[", c_a_moy, ",", c_b_moy, "]\n")
```

```
## [ 299896.8 , 299919.1 ]
```

Même en faisant varier le seuil on remarque que la valeur de c n'est jamais dans l'intervalle de confiance, on en conclut que l'expérience de Michelson n'est pas très précise.

Deuxième partie

On observe n données $x_1, \dots, x_n$, que l'on admettra être des réalisations de variables aléatoires $X_1, \dots, X_n$, indépendantes et de même loi de probabilité à valeurs dans $[0, 1]$, définie par sa densité :

$$f(x; \theta) = 2\theta x + 1 - \theta, \qquad \forall x \in [0, 1]$$

où θ est un paramètre appartenant à $[-1, +1]$.

Question 2.1

Enoncé Calculer l'espérance et la variance de cette loi.

Réponse Calculons l'espérance grâce à la densité : $E[X] = \int_0^1 xf(x,\theta)dx = \int_0^1 (2\theta x^2 + (1-\theta)x)dx = \frac{\theta}{6} + \frac{1}{2}$

Calculons E[X^2]:

$E[X^2] = \int_0^1 x^2 f(x,\theta)dx = \int_0^1 (2\theta x^3 + (1-\theta)x^2)dx = \frac{\theta}{6} + \frac{1}{3}$

Calculons la variance:

$Var[X] = E[X^2] - E[X]^2 = \frac{\theta}{6} + \frac{1}{3} - (\frac{\theta}{6} + \frac{1}{2})^2 = \frac{7}{12} - \frac{\theta^2}{36}$

Question 2.2

Enoncé Calculer l'estimateur de θ par la méthode des moments, $\tilde{\theta}_n$. Montrer qu'il est sans biais et convergent en moyenne quadratique.

Réponse D'après les questions précèdentes $E[X] = \frac{\theta}{6} + \frac{1}{3}$ D'où $\theta = 6E[X] - 2$

Ainsi, $\bar{\theta_n} = 6\bar{X_n} - 2$.

Montrons que cet estimateur est sans biais :

$E[\bar{\theta_n}] = E[\frac{6\sum_1^n X_i}{n} - 2]$

Par linérité, $E[\bar{\theta_n}] = \frac{6\sum_1^n E[X_i]}{n} - 2$

Sachant que les X_i suivent la même loi, $E[\bar{\theta_n}] = 6E[X] - 2 = 6(\frac{\theta}{6} + \frac{1}{3}) - 2 = \theta_n$

D'autre part, $EQM(\bar{\theta_n}) = Var[\bar{\theta_n}] - (E[\bar{\theta_n}] - \theta_n)^2 = Var[\bar{\theta_n}]$

$EQM(\bar{\theta_n}) = Var[\bar{\theta_n} = 6\bar{X_n} - 2] = \frac{36Var[\bar{X_n}]}{n^2}$

$EQM(\bar{\theta_n}) = \frac{36Var[X_1]}{n}$ car les X_i sont indépendantes.

Par conséquent comme $Var[X_1]$ est finie, $EQM[\bar{\theta_n}]$ tend vers 0 quand n tend vers $+\infty$.

L'estimateur est donc bien sans biais et converge en moyenne quadratique.

Question 2.3

Enoncé Montrer que l'estimateur de maximum de vraisemblance $\hat{\theta}_n$ n'a pas d'expression explicite. Donner l'équation implicite dont il est la solution.

Réponse

$$\mathcal{L}(\theta; x_1, \dots x_n) = \prod_1^n f(x_i, \theta)$$

Or comme les lois sont identiques et indépendantes, on obtient :

$$\mathcal{L}(\theta; x_1, \dots x_n) = \prod_1^n 2\theta x_i + 1 - \theta$$

Par passage au logarithme:

$$ln(\mathcal{L}(\theta; x_1, \dots x_n)) = \sum_1^n ln(2\theta x_i + 1 - \theta)$$

Dérivons par rapport à θ pour obtenir le maximum:

$$\frac{\partial ln(\mathcal{L}(\theta; x_1, \dots x_n))}{\partial \theta} = \sum_1^n \frac{\partial ln(2\theta x_i + 1 - \theta)}{\partial \theta} = \sum_1^n \frac{2\theta x_i - 1}{2\theta x_i + 1 - \theta}$$

Résolvons l'équation suivante:

$$\frac{\partial ln(\mathcal{L}(\theta; x_1, \dots x_n))}{\partial \theta} = 0$$

Or l'équation $\sum_1^n \frac{2\theta x_i - 1}{2\theta x_i + 1 - \theta} = 0$ est impossible à résoudre.

Par conséquent l'estimateur de maximum de vraisemblance $\bar{\theta_n}$ n'a pas d'expression explicite. Cependant il est solution de l'équation :

$$\frac{\partial ln(\mathcal{L}(\theta; x_1, \dots x_n))}{\partial \theta} = \sum_1^n \frac{2\theta x_i - 1}{2\theta x_i + 1 - \theta} = 0$$

Question 2.4

Enoncé En utilisant le théorème central-limite, donner un intervalle de confiance asymptotique de seuil α pour θ.

Réponse

$$\sqrt{n}|\frac{\bar{X_n} - E[X]}{\sigma(X)}| \leq u_\alpha \Leftrightarrow \sqrt(n)|\frac{\bar{X_n} - (\frac{\theta}{6} + \frac{1}{2})}{\sqrt{\frac{1}{12} - \frac{\theta^2}{36}}}| \leq u_\alpha$$

$$\sqrt{n}|\frac{\bar{X_n} - E[X]}{\sigma(X)}| \leq u_\alpha \Leftrightarrow n|\frac{(\bar{X_n} - (\frac{\theta}{6} + \frac{1}{2}))^2}{\frac{1}{12} - \frac{\theta^2}{36}}| \leq u_\alpha^2$$

$$\sqrt{n}|\frac{\bar{X_n} - E[X]}{\sigma(X)}| \leq u_\alpha \Leftrightarrow n[\bar{X_n}^2 - \frac{\bar{X_n}\theta}{3} - \bar{X_n} + \frac{\bar{\theta_n}^2}{36} + \frac{\bar{\theta_n}}{6} + \frac{1}{16}] \leq u_\alpha^2(\frac{1}{12} - \frac{\theta^2}{36})$$

Le polynome est obtenu est par conséquent θ;

$$\sqrt{n}|\frac{\bar{X_n} - E[X]}{\sigma(X)}| \leq u_\alpha \Leftrightarrow \frac{u_\alpha^2 + n}{36}\theta^2 + \frac{n(2 - \bar{X_n})}{3}\theta + n[\bar{X_n}^2 - \bar{X_n} + \frac{1}{16}] - \frac{u_\alpha \check{s}}{12} \leq 0$$

Calculons le déterminant de ce polynôme :

$$\Delta = (\frac{n(2 - \bar{X_n})}{3})^2 - (\frac{u_\alpha^2 + n}{12}) * (n[\bar{X_n}^2 - \bar{X_n} + \frac{1}{16}] - \frac{u_\alpha \check{s}}{12})$$

On a $\Delta \geq 0$

On a donc les racines:

$$r_1 = \frac{-\frac{n(2-\bar{X_n})}{3} - \sqrt{\Delta}}{\frac{u_\alpha^2+n}{18}}$$

$$r_2 = \frac{-\frac{n(2-\bar{X_n})}{3} + \sqrt{\Delta}}{\frac{u_\alpha^2+n}{18}}$$

Par conséquent l'intervalle de confiance asymptotique de seuil α pour θ est :

$$I = [r_1, r_2] = [\frac{-\frac{n(2-\bar{X_n})}{3} - \sqrt{\Delta}}{\frac{u_\alpha^2+n}{18}}, \frac{-\frac{n(2-\bar{X_n})}{3} + \sqrt{\Delta}}{\frac{u_\alpha^2+n}{18}}]$$