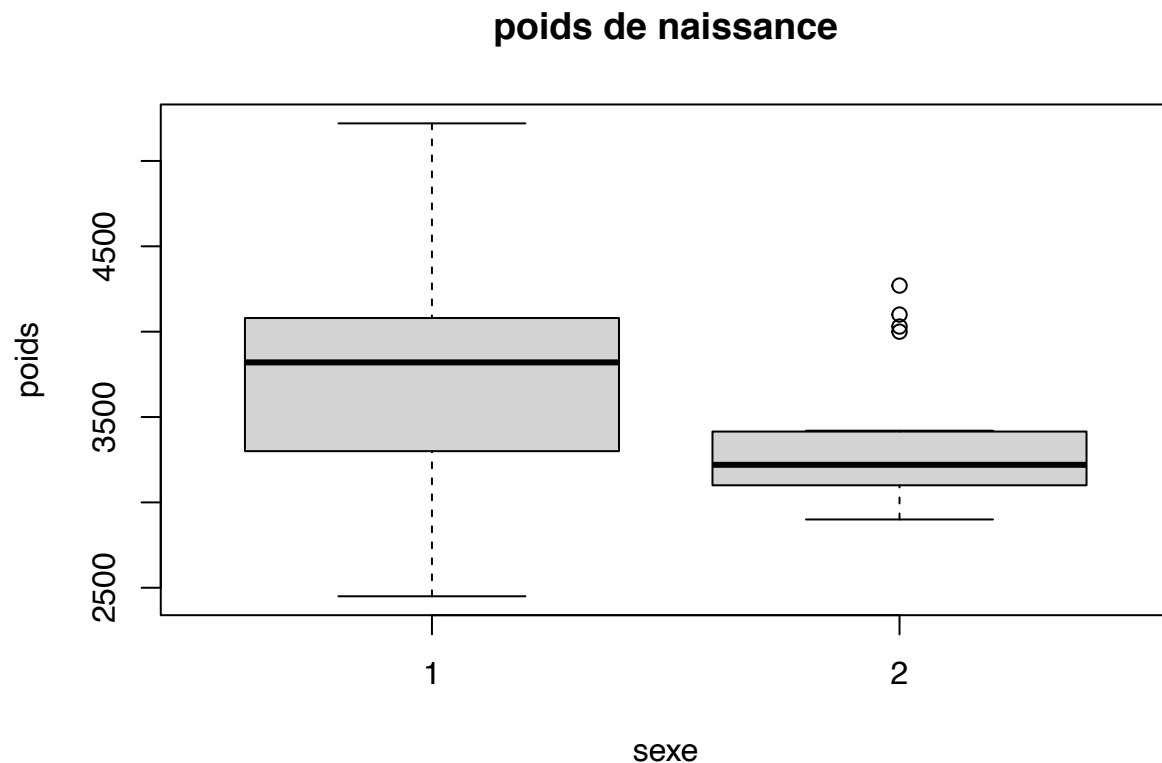


Statistique bayésienne avec R

Exercice sur les poids de naissance

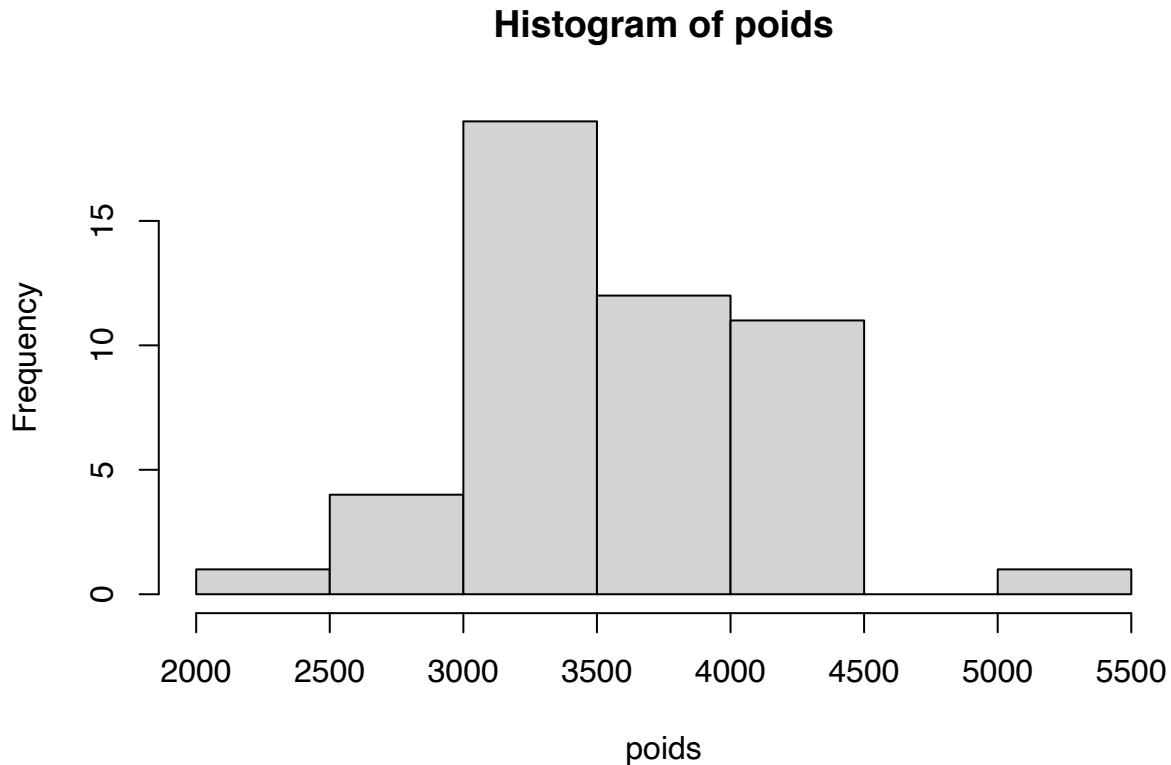
Julien JACQUES

```
data=read.table('Rcode/poidsnaissance.txt',header = T,sep=',',row.names = 1)
data$OBS=NULL
sexe=data$SEXE+1
poids=data$POIDNAIS
boxplot(poids~sexe,main="poids de naissance")
```



L'histogramme du poids de naissance ressemble à peu près à une loi gaussienne, ce qui est confirmé par le test de Shapiro

```
hist(poids)
```



```
shapiro.test(poids)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  poids
## W = 0.96218, p-value = 0.1237
```

Estimation bayésienne

Estimer le poids de naissance moyen

De façon fréquentiste :

```
mean(poids)
```

```
## [1] 3590
```

Les données étant distribuées suivant une loi gaussienne, nous allons choisir un a priori conjugué gaussien. Il nous reste à déterminer les moyennes et variances a priori.

L'histogramme nous donne l'idée d'un a priori gaussien centré en 3250g. Pour l'écart-type, il va traduire la confiance que l'on a dans notre a priori.

Nous avons vu que le MAP est alors donné par :

$$\hat{\theta} = E[\theta|\underline{x}] = \frac{\tau^2 \frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} \left(\frac{\bar{x}}{\frac{\sigma^2}{n}} + \frac{\mu}{\tau^2} \right)$$

dont on peut coder le calcul. En jouant sur la valeur de l'écart-type a priori τ , on pourra jouer sur la confiance en notre a priori et examiner son influence sur le MAP

```
s=sd(poids)
tau=100
n=length(poids)
MAP=(tau^2*s^2/n)/(tau^2+s^2/n)*(mean(poids)/(s^2/n)+3250/(tau^2))
print(MAP)
```

```
## [1] 3461.82
```

Le choix de τ est ici très subjectif, mais on n'a pas le choix car pas d'information supplémentaire. Si dans l'étude Statista2021 on avait pu avoir une idée de l'incertitude sur l'estimation du poids moyen, on aurait pu l'utiliser ici pour choisir τ .

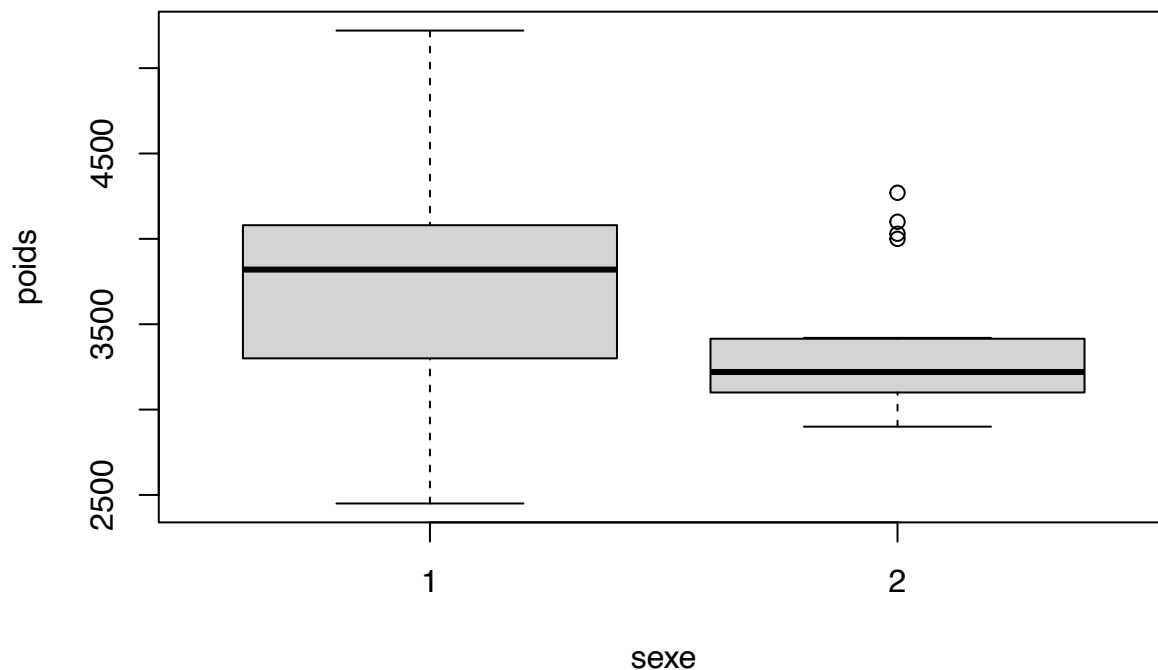
Les calculs sont faits ici *à la main*, nous verrons plus tard comment les faire sous R.

Test d'hypothèse et comparaison bayésienne de modèles

les garçons sont-ils plus lourds que les filles à la naissance ?

De façon fréquentiste c'est significatif,

```
boxplot(poids~sexe)
```



```
t.test(poids~sexe,alternative='greater')
```

```
##
## Welch Two Sample t-test
##
## data:  poids by sexe
## t = 2.4522, df = 45.553, p-value = 0.009047
## alternative hypothesis: true difference in means between group 1 and group 2 is greater than 0
## 95 percent confidence interval:
##  110.0071      Inf
## sample estimates:
## mean in group 1 mean in group 2
```