

## **Problème 1 :**

On entre les données dans R :

```
> rajka=c(10.94, 5.66, 10.88, 10.27, 8.40, 9.55, 14.17, 10.18, 9.36, 9.20,  
12.44, 11.23, 10.74, 10.88, 9.74, 10.99, 10.67, 9.09, 9.75, 7.77, 4.24,  
6.55,10.91, 9.61, 11.84)
```

```
> rubinola=c(20.76, 21.58, 19.51, 20.87, 19.74, 18.11, 22.29, 17.38, 21.70,  
16.88, 18.83, 21.90, 21.27, 21.63, 18.80, 19.30, 18.44, 20.51, 23.34, 17.21,  
17.37, 22.25, 21.30, 20.92, 19.39)
```

Question 1.

Il y a 1 variable quantitative : la quantité de sucre dans les pommes en mg/g de fruit sec.

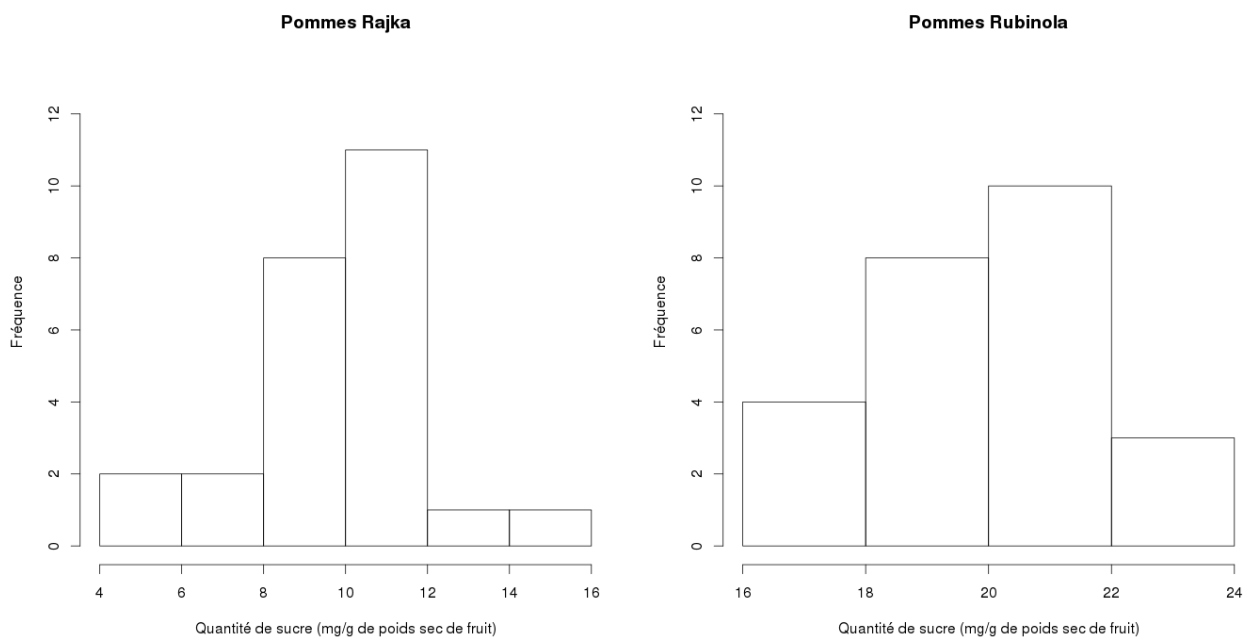
Question 2.

## **Histogramme :**

Commandes R :

```
> par(mfrow=c(1,2))  
> hist(rajka,breaks=4,xlab="Quantité de sucre (mg/g de poids sec de  
fruit)",ylab="Fréquence",main="Pommes Rajka",ylim=c(0,13))  
> hist(rubinola,breaks=4,xlab="Quantité de sucre (mg/g de poids sec de  
fruit)",ylab="Fréquence",main="Pommes Rubinola",ylim=c(0,13))
```

Résultat :



Commentaires :

`par(mfrow=c(1,2))` crée un espace graphique composé d'une ligne et de 2 colonnes. Les commandes suivantes placeront donc les 2 graphiques chacun dans une colonne.

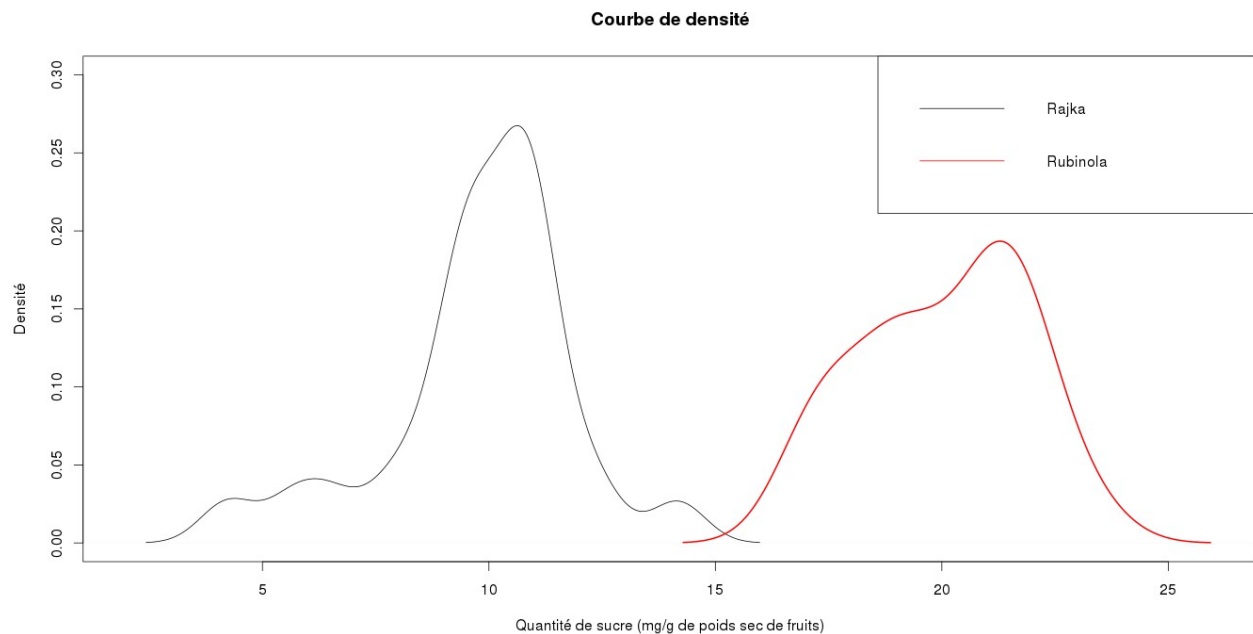
Les commandes suivantes tracent l'histogramme pour chaque pomme, grâce à la commande `hist`. Le premier paramètre correspond aux données, le second paramètre `breaks` définit le nombre de classes. Ici, on utilise `breaks=4` afin d'avoir des classes de longueur 2. `xlab` définit le label de l'axe des abscisses et `ylab` celui de l'axe des ordonnées. Enfin, `ylim` définit le minimum et le maximum de l'axe des abscisses.

## Courbes de densité :

Commandes R :

```
> plot(density(rajka),xlim=c(2,26),ylim=c(0.00,0.30),main="Courbe de
densité",xlab="Quantité de sucre (mg/g de poids sec de fruits)",ylab="Densité")
> lines(density(rubinola),col="red",lwd=2)
> legend("topright",c("Rajka","Rubinola"),col=c("black","red"),lty=c(1,1))
```

Résultat :



Commentaires :

La première commande plot trace la courbe de densité des pommes rajka. Les données brutes sont contenues dans le vecteur rajka. density(rajka) calcule la densité, et ces densités sont envoyées à plot pour être tracées. main correspond au titre du graphique.

La deuxième commande lines ajoute la densité des pommes rubinola au même graphique. Les données sont là encore données par la fonction density, col définit la couleur de la courbe (rouge ici), lwd=2 définit l'épaisseur du trait.

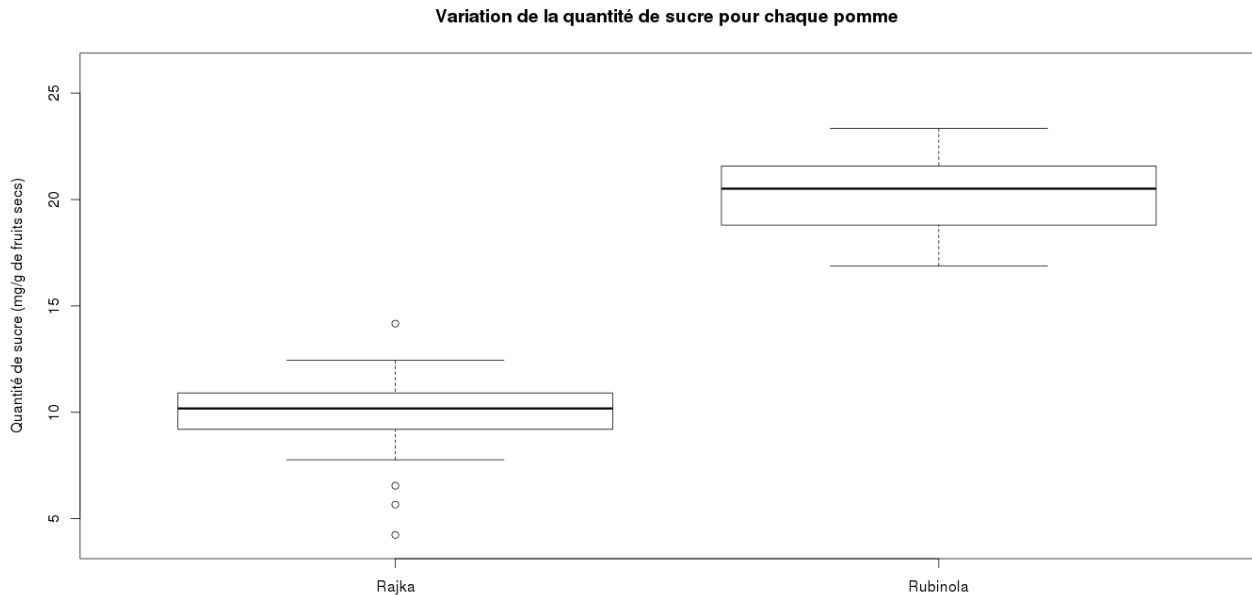
La dernière commande legend place la légende. Le premier paramètre topright place la légende en haut à droite. Ensuite, le premier vecteur contient le nom de chaque trait, le second les couleurs correspondantes et le dernier le type de motif correspondants (ici, 1 pour ligne pleine).

### **Boxplots :**

Commandes :

```
> boxplot(rajka,rubinola,names=c('Rajka','Rubinola'),ylim=c(4,26),  
main="Variation de la quantité de sucre pour chaque pomme",ylab="Quantité de  
sucre (mg/g de fruits secs)")
```

Résultat :



*Question 3.*

### **Pommes Rajka :**

Moyenne :

```
> mean(rajka)  
[1] 9.8024
```

Médiane :

```
> median(rajka)  
[1] 10.18
```

Variance :

```
> var(rajka)  
[1] 4.434094
```

### **Pommes Rubinola :**

Moyenne :

```
> mean(rubinola)  
[1] 20.0512
```

Médiane :

```
> median(rubinola)  
[1] 20.51
```

Variance :

```
> var(rubinola)  
[1] 3.336586
```

### Interprétation :

Pour chaque variété de pomme, on remarque que la médiane et la moyenne sont quasiment identiques. Par contre, la moyenne ou la médiane de la variété rubinola sont 2 fois supérieures à celles de la variété rajka. La variance est proche entre les 2 variétés.

On peut donc penser que les pommes rubinola sont plus sucrées que les pommes rajka.

### *Question 4.*

On utilise le test de student. Ce test suppose que les données suivent une loi normale.

On vérifie que nos données suivent une loi normale grâce au test shapiro-wilk : ce test est utilisé pour n données avec  $n < 200$ . L'hypothèse nulle  $H_0$  est que les données suivent une loi normale.

### Commandes R :

```
> shapiro.test(rajka)
```

Shapiro-Wilk normality test

```
data: rajka  
W = 0.9301, p-value = 0.08736
```

```
> shapiro.test(rubinola)
```

Shapiro-Wilk normality test

```
data: rubinola  
W = 0.9543, p-value = 0.3131
```

### Interprétation :

Pour les deux pommes, on trouve une p-valeur supérieure à 0.05, on ne rejette donc pas  $H_0$ . Les données suivent donc une loi normale, on peut donc maintenant effectuer le t-test.

### Commande R :

```
> t.test(rajka, rubinola, var.equal=T)
```

Two Sample t-test

```
data: rajka and rubinola  
t = -18.3829, df = 48, p-value < 2.2e-16  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -11.369767 -9.127833  
sample estimates:  
mean of x mean of y  
  9.8024  20.0512
```

On utilise `var.equal=T` car les variances sont quasiment identiques.

### Interprétation :

L'hypothèse nulle  $H_0$  est que les moyennes des quantités de sucre dans les deux pommes sont identiques.

On a une p-valeur de  $2.2e-16$ , soit inférieure à 0,05 donc on rejette  $H_0$  : les deux variétés de pommes ont des quantités de sucre significativement différentes.