

## 1. Importation des données

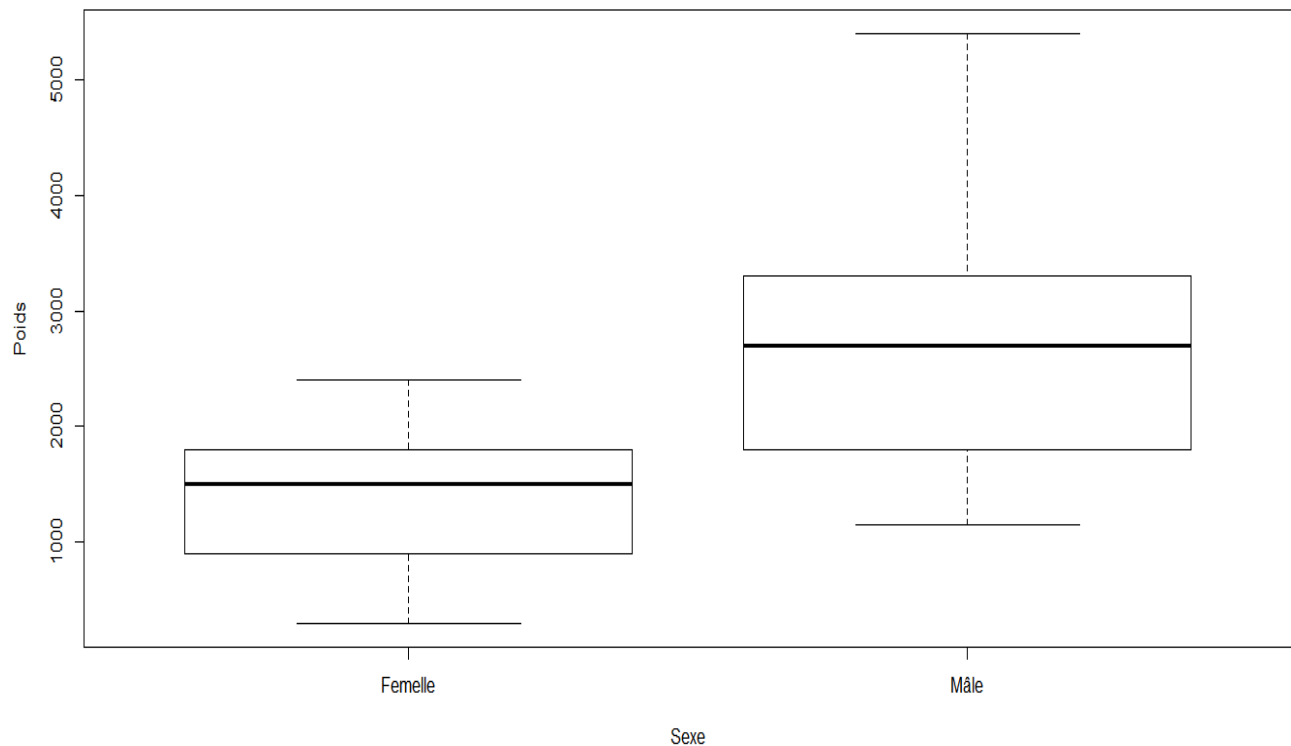
```
> poulpe<-read.csv("C:/Users/Emmanuel Foka/Documents/Analyse de données et RO/poulpe.csv",header=T,sep=";")
> poulpe
  Poids  Sexe
1   300 Femelle
2   700 Femelle
3   850 Femelle
4   900 Femelle
5  1000 Femelle
6  1420 Femelle
7  1500 Femelle
8  1500 Femelle
9  1800 Femelle
10 1800 Femelle
11 1900 Femelle
12 2200 Femelle
13 2400 Femelle
14  1150  Mâle
15  1500  Mâle
16  1700  Mâle
17  1800  Mâle
18  1800  Mâle
19  1850  Mâle
20  2200  Mâle
21  2700  Mâle
22  2900  Mâle
23  3000  Mâle
24  3100  Mâle
25  3500  Mâle
26  3900  Mâle
27  4000  Mâle
28  5400  Mâle
> |
```

```
Poulpe <- read.csv("C:/Users/Emmanuel Foka/Documents/Analyse de données et R
O/poulpe.csv",header=T,sep=";")
```

## 2. Comparons graphiquement les deux sous-populations

La boîte à moustaches des données à l'aide de la commande R permet d comparer graphiquement deux sous-population.

```
boxplot(Poids ~ Sexe, ylab="Poids", xlab="Sexe", data=poulpe)
```



### Interprétation du résultat :

- La valeur médiane du poids des femmes est de 1800 et la valeur du poids de la plupart des femmes est située entre 900 et 1800, mais cela peut chuter entre 300 et 2400.
- La valeur médiane du poids des mâles est de 2700 et la valeur du poids de la plupart des mâles est située entre 1800 et 3300 mais cela chute entre 1150 et 5400

### 3. Estimation des statistiques de base par sous population

La fonction `aggregate()` permet de découper un *data.frame* en sous populations suivant un facteur (spécifié par le paramètre `y`) et d'appliquer une fonction donnée sur chacune de ces sous-populations. Nous résumons le jeu de donnée à l'aide de la fonction `summary()`

```
aggregate(poulpe$Poids,by=list( poulpe$sexe),FUN="summary")
```

```
> aggregate(poulpe$Poids,by=list( poulpe$sexe),FUN="summary")
  Group.1   x.Min. x.1st Qu. x.Median   x.Mean x.3rd Qu.   x.Max.
1  Femelle  300.000   900.000 1500.000 1405.385  1800.000 2400.000
2    Mâle 1150.000  1800.000 2700.000 2700.000  3300.000 5400.000
> |
```

**Interprétation du résultat :** cela permet d'avoir la boîte à moustaches

```
tapply(poulpe$Poids,poulpe$sexe,sd,na.rm=TRUE)
```

```
> tapply(poulpe$Poids,poulpe$sexe,sd,na.rm=TRUE)
  Femelle    Mâle 
621.9943 1158.3547 
> |
```

#### 4. Testons la normalité des données

Le test de Shapiro-Wilk est conçu spécialement pour étudier la non-normalité d'une variable continue X. C'est le test le plus puissant pour tester la normalité d'une distribution.

```
> shapiro.test(poulpe$Poids)

      Shapiro-Wilk normality test

data:  poulpe$Poids
W = 0.93298, p-value = 0.0733

> |
```

**Interprétation résultat :** ce résultat renvoie une p-value non significative car l'échantillon ne suit pas une loi normale

#### 5. Tester l'égalité des variances

Pour Tester l'égalité des variances on utilisera la commande var.test.

```
> var.test(Poids ~ Sexe, conf.level=.95, data=poulpe)

      F test to compare two variances

data:  Poids by Sexe
F = 0.28833, num df = 12, denom df = 14, p-value = 0.03713
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.09452959 0.92444666
sample estimates:
ratio of variances
      0.2883299

> |
```

---

**Interprétation Résultat :** Pour comparer la valeur de deux variances, on se rend compte que la p-value est inférieure à 0.05 donc sexe et poids sont différents significativement.

## 6. Tester l'égalité des moyennes

Pour tester l'égalité des moyennes on utilisera la fonction `t.test` avec l'argument `var.equal=TRUE`. Le test est réalisé au niveau 5% de l'hypothèse  $H_0$  : "Les moyennes sont égales" contre l'alternative  $H_1$  : "Les moyennes sont différentes"

```
t.test(Poids ~ sexe, alternative="two.sided", conf.level=.95, var.equal=TRUE, data=poulpe)
```

```
> t.test(Poids ~ Sexe, alternative="two.sided", conf.level=.95, var.equal=TRUE, data=poulpe)

      Two Sample t-test

data:  Poids by Sexe
t = -3.5992, df = 26, p-value = 0.001317
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2033.9856 -555.2452
sample estimates:
mean in group Femelle      mean in group Mâle
      1405.385              2700.000

> |
```

**Interprétation Résultat :** t est la statistique, df est le degré de liberté et p-value est le degré de significativité du test dont on obtient la valeur moyenne de deux groupes mâles et femelles qui est 1405.385 et 2700.

```
t.test(Poids ~ Sexe, alternative="greater", conf.level=.95,var.equal=TRUE, data=poulpe)
```

```
> t.test(Poids ~ Sexe, alternative="greater", conf.level=.95,var.equal=TRUE, data=poulpe)
```

```
Two Sample t-test
```

```
data: Poids by Sexe
```

```
t = -3.5992, df = 26, p-value = 0.9993
```

```
alternative hypothesis: true difference in means is greater than 0
```

```
95 percent confidence interval:
```

```
-1908.123      Inf
```

```
sample estimates:
```

mean in group Femelle	mean in group Mâle
1405.385	2700.000

```
> |
```