

# TP MISO: comparaison de moyennes

guillemette.marot@univ-lille.fr

Dans ce TP, nous utilisons les variables `ronfle` et `age` d'un jeu de données sur les ronflements. La description du jeu de données initial est disponible à l'url <http://www.info.univ-angers.fr/~gh/Datasets/ronfle.htm>.

```
setwd("/home/marot/Documents/Cours/ULille/bio-informatique et omics/TP")
ronflement=read.csv("ronflement.csv", header=TRUE, sep=";", dec=".")
head(ronflement)
dim(ronflement)
```

1. Vérifiez que la variable `ronflement` est considérée comme une variable qualitative. Si nécessaire, attribuez le bon type et remplacez les labels "1/0" par "oui/non". Vérifiez que l'âge est traitée comme une variable quantitative.

```
str(ronflement)
ronfle <- as.factor(ronflement$ronfle)
levels(ronfle)<-c("non", "oui")
age <-as.numeric(ronflement$age)
```

2. Tracer un histogramme de l'âge chez les ronfleurs et chez les non ronfleurs.

```
par(mfrow=c(2,2))
hist(age[which(ronfle=="oui")],nclass=20,main="Age des ronfleurs",xlab="age",col="blue")
hist(age[which(ronfle=="non")],nclass=20,main="Age des non ronfleurs",xlab="age",col="blue")
```

3. La distribution de l'âge semble t-elle suivre une distribution normale?
4. On cherche à savoir si le ronflement est lié à l'âge. Pour cela, on utilise le t-test, test de comparaison de moyennes. Avait-on besoin de vérifier la normalité de la distribution? Pourquoi?

Pour la suite, on note  $Y$  la variable ronflement et  $X$  la variable âge. On sait qu'il y a 35 ronfleurs et 65 non ronfleurs. On donne

$$\sum_{i/y_i=0} x_i = 3267 \quad \sum_{i/y_i=1} x_i = 1960 \quad \sum_{i/y_i=0} x_i^2 = 172021 \quad \sum_{i/y_i=1} x_i^2 = 114054$$

5. Calculer la moyenne et l'écart-type non biaisé ("standard deviation") de l'âge dans chacun des deux groupes (ronfleurs/non ronfleurs). Utilisez les fonction `by`, `sum`, `mean` et `sd` pour retrouver les résultats de ces calculs.
6. Que devez vous tester avant de choisir entre un test de student et un test de Welch? Faites le test, à la fois à la main et en utilisant la commande `var.test`.
7. Dans le cadre d'un test de Student, combien de degrés de liberté aura la loi suivie par la statistique de test? Calculer à la fois à la main et avec la commande `t.test` les valeurs prises par la statistique de student et la statistique de Welch.