Chapitre 2

Test de Student-Fisher

On considère deux échantillons de taille respective n_1 et n_2, correspondant aux réalisations de 2 variables indépendantes X_1 et X_2 de moyenne et de variance respectives μ_1, μ_2 et σ_1,σ_2. On s'intéresse dans ce chapitre à la question suivante : peut on affirmer que la moyenne et la variance de X_1 et X_2 sont égales à partir du calculs des moyenne et des variances empiriques ? Ceci revient à tester les hypothèses :

$$\begin{cases} H_0 : \mu_1 = \mu_2 \text{ et } \sigma_1 = \sigma_2 \\ H_1 : \mu_1 \neq \mu_2 \text{ et } \sigma_1 \neq \sigma_2 \end{cases}$$

Le test de Student-Fisher se déroule en deux étapes : on teste dans un premier temps l'égalité des variances à l'aide du test de Fisher-Snedecor. Si l'on accepte l'hypothèse d'égalité, on teste alors l'égalité des moyennes à l'aide du test de Student en supposant que $\sigma_1 = \sigma_2$.

Rappel de cours sur ces 2 tests...

Test de Student-Fisher et R

Pour pouvoir mettre en place ce test sous **R**, nous aurons besoin de deux fonctions : var.test et t.test pour respectivement les tests de Fisher et Student. Ces dernières sont contenues dans le package **stats** chargé par défaut dans la plupart des distributions (library(help="stats") pour plus de détails).

La fonction var.test prend comme arguments deux vecteurs de même taille correspondant aux deux échantillons observés dont la variance est à tester. Il est également utile de renseigner les options suivantes :

- ratio : la valeurs du rapport à tester entre σ_1 et σ_2. Par défaut, vaut 1.
- alternative : une chaîne de caractère spécifiant le type d'hypothèse alternative à choisir parmi *two.sided* (par défaut) pour un test bilatéral, *greater* ou *less*.
- conf.level : renseigne le niveau de confiance pour l'intervalle de confiance affiché en sortie. Le niveau par défaut est 0.95.

Supposons par exemple que l'on dispose des deux échantillons :

```
> x <- c(23.7,21.8,20.6,22.4,21,21,20.2,18.8)
> y <- c(24.8,22.75,22.52,22.8,20.7,23,23.1,22.96)
```

Pour tester l'égalité des deux variances, on utilise :

```
> var.test(x,y)

        F test to compare two variances

data:  x and y
F = 1.7649, num df = 7, denom df = 7, p-value = 0.4711
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
 0.3533447 8.8156382
sample estimates:
ratio of variances
          1.764925
```

La fonction renvoie la valeur de la statistique de test, le nombre de degré de liberté du numérateur et du dénominateur, la p-valeur, l'intervalle de confiance, etc. Comme pour les tests vu dans le chapitre précédant, il est possible de récupérer l'ensemble des informations calculée dans cette fonction. Il suffit d'affecter la fonction à une variable :

```
> X <- var.test(x,y)
```

et de sélectionner les valeurs souhaitées : statistic pour la valeur de la statistique, parameter pour les degrés de liberté, p.value pour la p-valeur ou encore conf.int pour l'intervalle de confiance. Exemple :

```
> X$p.value
[1] 0.4711269
```

Concernant la manipulation de la loi de Fisher, on notera l'existence des commandes :
- pf(t,r_1,r_2) : donne la valeur de $P(F \leq t)$ pour F suivant une loi de Fisher de paramètres r_1 et r_2.
- qf(q,r_1,r_2) : valeur de t telle que $P(F \leq t) = q$.

Intéressons nous à présent à la mise en place du test de Student (dans l'éventualité ou le test précédant inciterait à conclure que les variances des deux échantillons sont égales). La fonction t.test prend comme argument les deux échantillons à tester. Les options sont essentiellement les mêmes que pour le test de Fisher. Il est cependant nécessaire de préciser que les variances sont supposées égales en donnant la valeur TRUE à la variable var.equal. Dans le cas contraire, c'est une variante du test de Student qui est utilisée.

Reprenons l'exemple précédant. Au vu de la p-valeur, il est raisonnable d'accepter l'hypothèse d'égalité des deux variances. Pour tester l'égalité des deux moyennes, on utilise :

```
> t.test(x,y,var.equal=TRUE)

        Two Sample t-test

data:  x and y
t = -2.5129, df = 14, p-value = 0.02484
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.0420811 -0.2404189
sample estimates:
mean of x mean of y
 21.18750  22.82875
```

On accepte ainsi l'hypothèse d'égalité des moyennes au niveau 1% mais pas au niveau 5%. Il est encore une fois possible d'accéder séparément à l'ensemble des informations fournies par cette fonction.