Fiche TD avec le logiciel R : tdr31p

Comparaisons de deux moyennes avec le test paramétrique de Student (test t)

D. Chessel, A.B. Dufour & J.R. Lobry

Cette fiche donne un exemple simple, complet et reproductible d'un test de comparaison de deux moyennes avec le test t de STUDENT. Elle suit un mode opératoire classique où l'on cherche à vérifier que les conditions d'application (suffisantes mais non nécessaires) du test sont réunies. Elle utilise volontairement une syntaxe de R suggérant que des généralisations de l'approche sont possibles. Deux exercices d'application sont proposés.

Contents

1 Le jeu de données **2**

2 Les données sont-elles issues d'une distribution normale ? **3**
2.1 Ce qu'il ne faut pas faire 3
2.2 Ce qu'il convient de faire 4

3 Les variances sont-elles égales ? **5**

4 Les échantillons sont-ils appariés ? **5**

5 Le test de Student (test t) **6**
5.1 Ce qu'il ne faut pas faire 6
5.2 Ce qu'il convient de faire 6

6 Exercices **7**
6.1 Hauteur des arbres de 27 peuplements forestiers 7
6.2 Hauteur de 12 arbres avant et après abattage 8

Références **9**

1 Le jeu de données

La variable mesurée est la longueur de la mâchoire inférieure (en mm) de 10 chacals mâles et 10 chacals femelles (*Canis aureus*) conservés au British Museum [3]. La variable mesurée diffère-t-elle entre les sexes dans cette espèce ? On range les données dans le tableau `chac` et on ajoute une variable qualitative binaire pour noter le sexe des individus.

```
mal <- c(120, 107, 110, 116, 114, 111, 113, 117, 114, 112)
fem <- c(110, 111, 107, 108, 110, 105, 107, 106, 111, 111)
mâch <- c(mal, fem)
chac <- data.frame(mâch)
chac$plan <- gl(n = 2, k = 10, lab = c("mâle", "femelle"))
head(chac)
```

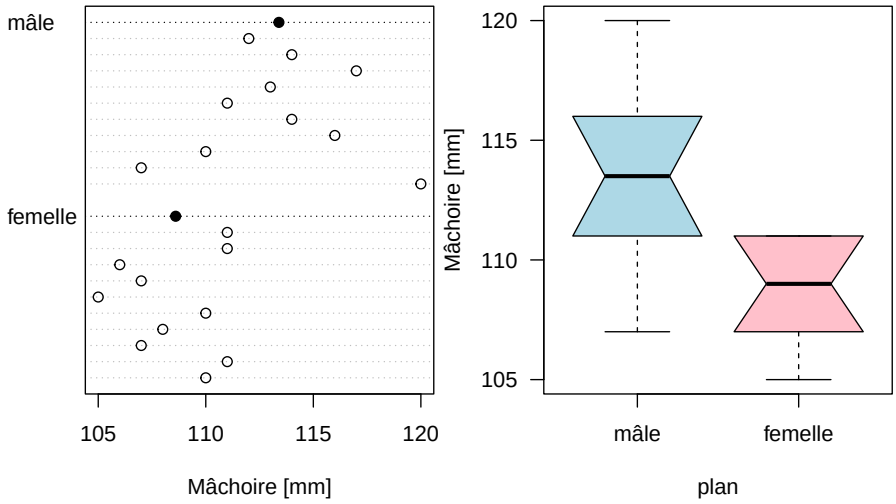
```
  mâch plan
1  120 mâle
2  107 mâle
3  110 mâle
4  116 mâle
5  114 mâle
6  111 mâle
```

```
tail(chac)
```

```
   mâch    plan
15  110 femelle
16  105 femelle
17  107 femelle
18  106 femelle
19  111 femelle
20  111 femelle
```

On va utiliser un mode opératoire bien répandu. On teste l'hypothèse de normalité par groupe. Si elle n'est pas rejetée, on teste l'égalité des variances entre groupes. Si elle n'est pas rejetée, on fait un test t pour comparer les moyennes. Pourquoi pas ? Ces indications de procédure rigide peuvent simplifier la vie du débutant. En fait, tous ces éléments doivent s'intégrer dans une analyse qui permet une interprétation. La première chose à faire *impérativement* est de regarder les données.

```
moy <- with(chac, tapply(mâch, plan, mean))
par(mfrow = c(1, 2), mar = c(5, 4, 0, 0) + 0.1)
with(chac, dotchart(mâch, groups = plan, gdata = moy,
  gpch = 19, xlab = "Mâchoire [mm]"))
boxplot(mâch~plan, chac,
  col = c("lightblue", "pink"), notch = TRUE, las = 1,
  ylab = "Mâchoire [mm]")
```



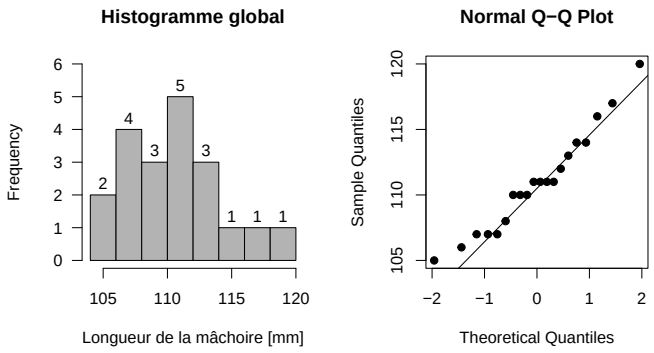
On constate que la mâchoire des mâles est en moyenne plus longue que celle des femelles. Les encoches des boites à moustaches nous indiquent que l'on est à la limite de la significativité. Un test d'hypothèse serait le bienvenu. Mais les conditions d'utilisation sont-elles remplies ? Déroulons le mode opératoire classique.

2 Les données sont-elles issues d'une distribution normale ?

2.1 Ce qu'il ne faut pas faire

On peut être tenté d'étudier la normalité de l'échantillon tout entier.

```
par(mfrow = c(1, 2))
with(chac, {
  hist(mâch, nclass=8, las = 1, col = grey(0.7), label = T, ylim = c(0, 6),
  main = "Histogramme global", xlab = "Longueur de la mâchoire [mm]")
  qqnorm(mâch, pch = 19)
  qqline(mâch)
})
```



C'est une erreur très répandue. En effet, chercher à savoir si l'échantillon total provient d'une loi normale, c'est supposer que le problème posé –la différence des moyennes– est résolu, puisque la moyenne estimée par l'échantillon total est commune aux deux échantillons censés provenir d'une loi unique. Un test t suppose que chacune des lois dont sont tirés les deux échantillons est normale, pas que l'échantillon tout entier est tiré d'une loi normale. Ce qu'il ne faut pas faire est donc :

```
with(chac, shapiro.test(mâch)) # Ne pas faire
```

```
        Shapiro-Wilk normality test
data:  mâch
W = 0.95929, p-value = 0.5298
```

Le résultat est sans ambiguïté : rien ne permet de penser que le mélange des deux échantillons ne provient pas d'une loi unique, donc qu'ils ont même moyenne. On ne demande pas aux deux échantillons d'avoir même moyenne pour tester l'hypothèse qu'ils ont même moyenne. Mais l'indication apportée par ce test n'est pas nulle. Les moyennes ne sont pas énormément différentes (si elles le sont) et les lois d'origine ne sont pas franchement anormales.

```
shapiro.test(c(rnorm(n = 10, mean = 1, sd = 1), rnorm(n = 10, mean = 10, sd = 1)))
```

```
        Shapiro-Wilk normality test
data:  c(rnorm(n = 10, mean = 1, sd = 1), rnorm(n = 10, mean = 10, sd = 1))
W = 0.77299, p-value = 0.0003511
```

Un mélange d'échantillons de lois normales de moyennes très différentes ne passe pas pour un échantillon d'une loi normale.

2.2 Ce qu'il convient de faire

On utilise un test de normalité[1], comme celui de Shapiro et Wilk [4], séparément pour chaque échantillon :

```
with(chac, tapply(mâch, plan, shapiro.test))
```

```
$mâle
        Shapiro-Wilk normality test

data:  X[[i]]
W = 0.99374, p-value = 0.9995


$femelle
        Shapiro-Wilk normality test

data:  X[[i]]
W = 0.88085, p-value = 0.1335
```

Avec un risque de première espèce 5 %, on ne rejette pas l'hypothèse de normalité. Notez cependant que les échantillons sont de petite taille et donc le test peu puissant. Quand on considère la valeur-p pour l'échantillon des mâles on pourrait *a contrario* suspecter les données d'être un peu trop jolies, sélectionnées pour faire de beaux exercices...

[1] Voir la section 2 du document « Risques, puissance et robustesse » pour plus de détails sur les tests de normalités possibles : `https://esb.univ-lyon1.fr/pdf/tdr36.pdf`

3 Les variances sont-elles égales ?

On teste l'égalité de deux ou plusieurs variances (homoscédasticité), pour des lois normales, avec le test de Bartlett [1]. Il se calcule avec la fonction `bartlett.test()` dont le cas particulier, pour $k = 2$ échantillons, est la fonction `var.test()`.

```
bartlett.test(mâch~plan, data = chac)
```

```
        Bartlett test of homogeneity of variances
data:  mâch by plan
Bartlett's K-squared = 1.9942, df = 1, p-value = 0.1579
```

```
var.test(mâch~plan, data = chac)
```

```
        F test to compare two variances
data:  mâch by plan
F = 2.681, num df = 9, denom df = 9, p-value = 0.1579
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.665931 10.793829
sample estimates:
ratio of variances
          2.681034
```

Avec un risque de première espèce de 5 %, on ne peut donc pas rejeter l'hypothèse d'égalité des variances.

4 Les échantillons sont-ils appariés ?

On parle de données appariées quand on mesure la même variable sur le même individu (au sens statistique du terme). Par exemple, une glycémie chez un patient avant et après une prise orale de glucose. Dans notre exemple, il n'y a pas à hésiter puisque nous avons d'un côté des mâles et de l'autre des femelles, il ne peut donc pas s'agir des mêmes individus, nous avons des données non-appariées. Les tests de R utilisent un paramètre logique `paired` pour indiquer si les échantillons sont appariés ou non. Ce paramètre est faux (`FALSE`) par défaut, donc si on ne précise rien les données sont considérées comme non-appariées.

La distinction entre données appariées et non-appariées n'est pas toujours aussi évidente. Supposez par exemple que l'on ait mesuré l'empan de la main dominante et de la main non-dominante pour les étudiants d'un groupe de TD. On dispose donc pour chaque étudiant de deux mesures. On veut tester l'hypothèse de l'égalité de l'empan entre les deux mains. Doit-on considérer que les données sont appariées ? Une première tentation serait de dire oui puisque nous avons deux mesures pour chaque individu (étudiant), mais les individus (statistiques) ne sont par les étudiants mais les mains dominantes et non dominantes. Si nous avions mesuré à la place la taille et la masse des étudiants il serait évident qu'il n'y a pas de sens à apparier les échantillons. Donc on pencherait pour des données non-appariées. Cependant, l'hypothèse nulle que nous voulons tester est celle de la symétrie des deux mains : si nous étions parfaitement symétriques l'empan de nos deux mains serait égal. Ce que l'on approche avec la main dominante et la main non dominante c'est l'empan d'une main idéale sous l'hypothèse de symétrie. On doit donc considérer que l'on a mesuré deux fois la même chose chez le même individu dans deux conditions différentes (le fait d'être dominant ou non pour la main). Les données sont appariées.

5 Le test de Student (test t)

5.1 Ce qu'il ne faut pas faire

```
t.test(chac[chac$plan == "mâle", "mâch"], chac[chac$plan == "femelle", "mâch"], paired = TRUE, data = chac) #
```

```
        Paired t-test
data:  chac[chac$plan == "mâle", "mâch"] and chac[chac$plan == "femelle", "mâch"]
t = 3.417, df = 9, p-value = 0.007665
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 1.622226 7.977774
sample estimates:
mean difference
            4.8
```

Il serait idiot d'apparier les échantillons dans notre cas puisqu'ils ne le sont pas. Il est peu probable de faire cette erreur puisque `paired` est faux par défaut.

```
t.test(mâch~plan, data = chac) # Ne pas faire
```

```
        Welch Two Sample t-test
data:  mâch by plan
t = 3.4843, df = 14.894, p-value = 0.00336
alternative hypothesis: true difference in means between group mâle and group femelle is not equal to 0
95 percent confidence interval:
 1.861895 7.738105
sample estimates:
   mean in group mâle mean in group femelle
                113.4                 108.6
```

C'est une erreur plus facile à commettre parce que la fonction `t.test()` considère par défaut que les variances sont inégales alors que nous avons vu dans notre exemple que l'on ne pouvait pas rejeter l'hypothèse d'homoscédasticité.

5.2 Ce qu'il convient de faire

```
t.test(mâch~plan, var.equal = TRUE, data = chac)
```

```
        Two Sample t-test
data:  mâch by plan
t = 3.4843, df = 18, p-value = 0.002647
alternative hypothesis: true difference in means between group mâle and group femelle is not equal to 0
95 percent confidence interval:
 1.905773 7.694227
sample estimates:
   mean in group mâle mean in group femelle
                113.4                 108.6
```

On indique avec `var.equal = TRUE` que les deux variances sont considérées comme égales. Mais alors pourquoi nous embêter à tester l'hypothèse d'égalité des variances puisque la fonction `t.test()` est pafaitement capable de gérer l'hétéroscédasticité ? Comparez les valeurs-p, vous pouvez constater que l'on a gagné en puissance. À noter, quand il y a homoscédasticité, le test est strictement identique à une analyse de variance à un facteur binaire :

```
anova(lm(mâch~plan, data = chac))
```

```
Analysis of Variance Table
Response: mâch
          Df Sum Sq Mean Sq F value   Pr(>F)
plan       1  115.2 115.200   12.14 0.002647 **
Residuals 18  170.8   9.489
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

On peut également utiliser la fonction `aov()` qui accepte directement des objets de la classe `formula` :

```
summary(aov(mâch~plan, data = chac))
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
plan         1  115.2  115.20   12.14 0.00265 **
Residuals   18  170.8    9.49
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

6 Exercices

Des générations de praticiens ont utilisé les ouvrages de Pierre Dagnelie[2] pour vérifier leurs procédures. La précision de l'ouvrage de P. Dagnelie est légendaire. On peut l'utiliser pour vérifier qu'on manipule correctement le logiciel. Voici quelques exercices issus de [2].

6.1 Hauteur des arbres de 27 peuplements forestiers

On compare la hauteur (en mètre) des arbres de 27 peuplements forestiers appartenant à deux types différents (*op. cit.* p.25). Les données sont les suivantes :

```
arbre <- list()
arbre$hau <- c(23.4,24.4,24.6,24.9,25.0,26.2,26.3,26.8,26.8,26.9,27.0,27.6,27.7,
  22.5,22.9,23.7,24.0,24.4,24.5,25.3,26.0,26.2,26.4,26.7,26.9,27.4,28.5)
arbre$plan <- factor(rep(c("T1", "T2"), c(13, 14)))
arbre <- as.data.frame(arbre)
head(arbre)
```

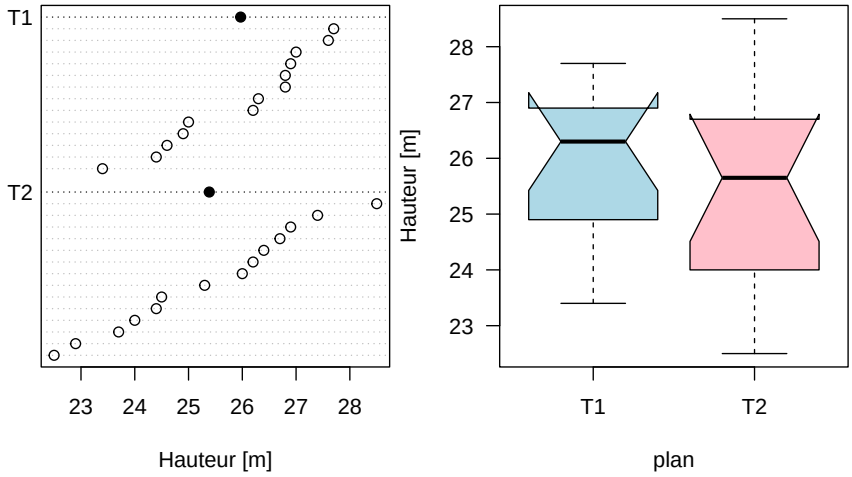
```
   hau plan
1 23.4   T1
2 24.4   T1
3 24.6   T1
4 24.9   T1
5 25.0   T1
6 26.2   T1
```

```
tail(arbre)
```

```
    hau plan
22 26.2   T2
23 26.4   T2
24 26.7   T2
25 26.9   T2
26 27.4   T2
27 28.5   T2
```

Testez s'il y a une différence significative de hauteur entre les deux types d'arbres. Seuls les résultats sont donnés ici, à vous de trouver le code R correspondant.

[2] http://www.dagnelie.be/




```
$T1
        Shapiro-Wilk normality test

data:  X[[i]]
W = 0.92201, p-value = 0.267


$T2

        Shapiro-Wilk normality test

data:  X[[i]]
W = 0.97278, p-value = 0.9115
        Bartlett test of homogeneity of variances
data:  hau by plan
Bartlett's K-squared = 0.83582, df = 1, p-value = 0.3606
        Two Sample t-test
data:  hau by plan
t = 0.95416, df = 25, p-value = 0.3491
alternative hypothesis: true difference in means between group T1 and group T2 is not equal to 0
95 percent confidence interval:
 -0.6759945  1.8430275
sample estimates:
mean in group T1 mean in group T2
        25.96923         25.38571
```

6.2 Hauteur de 12 arbres avant et après abattage

Des arbres ont été mesurés debouts puis abattus (*op. cit.* p. 387). Les données sont les suivantes :

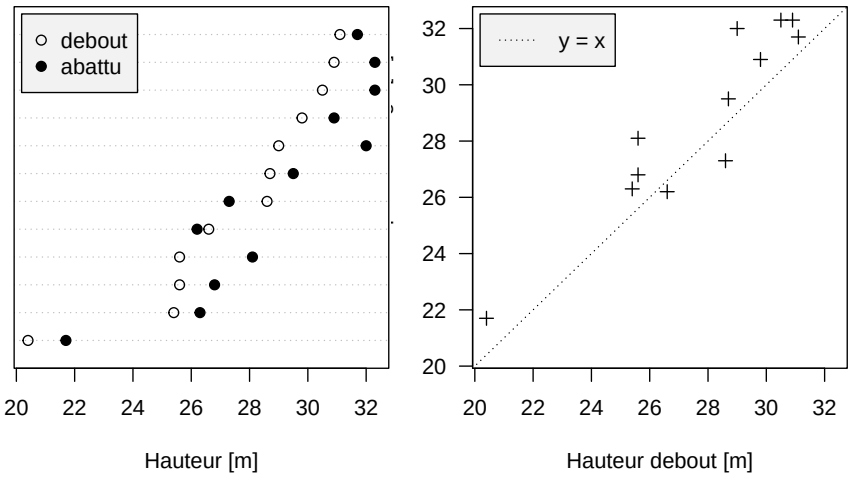
```
debout <- c(20.4,25.4,25.6,25.6,26.6,28.6,28.7,29.0,29.8,30.5,30.9,31.1)
abattu <- c(21.7,26.3,26.8,28.1,26.2,27.3,29.5,32.0,30.9,32.3,32.3,31.7)
ada <- data.frame(hau = c(debout, abattu))
ada$plan <- gl(2, 12, labels = c("debout", "abattu"))
head(ada)
```

```
   hau   plan
1 20.4 debout
2 25.4 debout
3 25.6 debout
4 25.6 debout
5 26.6 debout
6 28.6 debout
```

```
tail(ada)
```

```
   hau   plan
19 29.5 abattu
20 32.0 abattu
21 30.9 abattu
22 32.3 abattu
23 32.3 abattu
24 31.7 abattu
```

Testez s'il y a une différence significative de hauteur entre les arbres mesurés debout ou après abattage. Seuls les résultats sont donnés ici, à vous de trouver le code R correspondant.




```
$debout
        Shapiro-Wilk normality test

data:  X[[i]]
W = 0.89146, p-value = 0.1231


$abattu

        Shapiro-Wilk normality test

data:  X[[i]]
W = 0.90147, p-value = 0.1657

        Bartlett test of homogeneity of variances
data:  hau by plan
Bartlett's K-squared = 0.03031, df = 1, p-value = 0.8618

        Paired t-test
data:  ada[ada$plan == "debout", "hau"] and ada[ada$plan == "abattu", "hau"]
t = -3.2343, df = 11, p-value = 0.007954
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -1.8065536 -0.3434464
sample estimates:
mean difference
         -1.075
```

La célèbre mais néanmoins apocryphe expression « Il est plus grand mort que vivant ! » prêtée au roi de France Henri III suite à l'assassinat du duc Henri de Guise s'applique-t-elle également à la hauteur des arbres avant et après abattage ? Quelle mesure privilégeriez-vous entre la donnée avant et après abattage ? Pourquoi ?

References

[1] M.S. Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London - Series A*, 160:268–282, 1937.

[2] P. Dagnelie. *Théories et méthodes statistiques. Applications agronomiques. II Les méthodes de l'inférence statistique.* Editions L. Duculot, Gembloux, 1970.

[3] D.J. Hand, F. Daly, A.D. Lunn, K.J. McConway, and E. Ostrowski. *A handbook of small data sets.* Chapman & Hall, London, 1994.

[4] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, (52):591–611, 1965.