Fiche TD avec le logiciel R : tdr31np

Comparaisons de deux moyennes avec le test non-paramétrique de Wilcoxon-Mann-Whitney

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 non-paramétrique de Wilcoxon-Mann-Whitney (cas particulier du test de Kruskal-Wallis quand on n'a que deux échantillons). Deux exercices d'application sont proposés.

Table des matières

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*) [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)
```

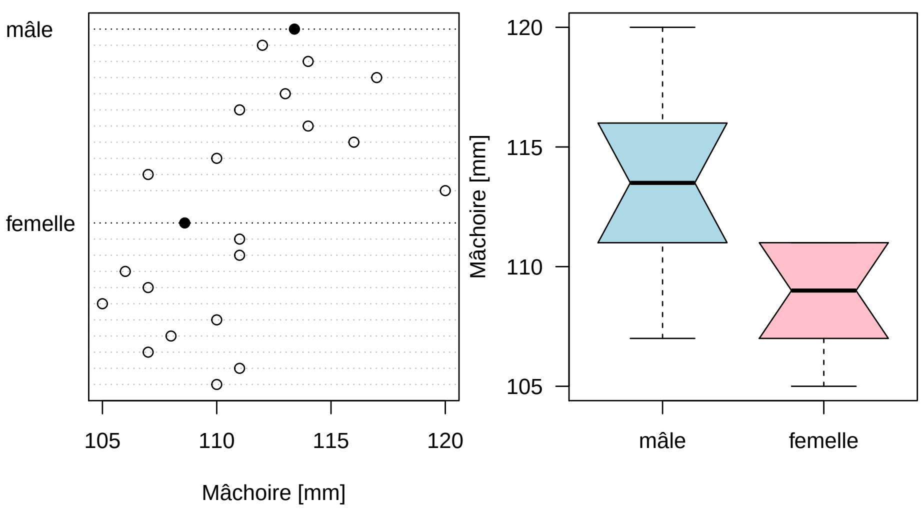
```
  mâch plan
1  120 mâle
2  107 mâle
3  110 mâle
4  116 mâle
5  114 mâle
6  111 mâle
```

```
tail(chac)
```

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

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.

2 Les échantillons sont-ils appariés ?

On parle de données appariées quand on mesure la même chose sur le même individu (au sens statistique du terme) comme 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.

3 Le test de Wilcoxon-Mann-Whitney

3.1 Données non appariées

Le test de Wilcoxon-Mann-Whitney [9, 8] est implémenté dans la fonction `wilcox.test()` :

```
wilcox.test(mâch~plan, data = chac)
       Wilcoxon rank sum test with continuity correction
data:  mâch by plan
W = 87.5, p-value = 0.004845
alternative hypothesis: true location shift is not equal to 0
```

On a ici un message d'avertissement « *cannot compute exact p-value with ties* » signifiant que la valeur-p ne peut pas être calculée exactement en présence d'ex æquo. La documentation de la fonction `wilcox.test()` nous suggère d'utiliser la fonction `wilcox_test()` du paquet `coin` [4, 5] :

```
library(coin)
wilcox_test(mâch~plan, data = chac)
       Asymptotic Wilcoxon-Mann-Whitney Test
data:  mâch by plan (mâle, femelle)
Z = 2.8552, p-value = 0.004301
alternative hypothesis: true mu is not equal to 0
```

Une autre solution si on ne désire pas installer le paquet `coin` est d'utiliser la fonction `kruskal.test()` qui sait gérer les ex æquo du paquet standard `stats` puisque le test de Wilcoxon-Mann-Whitney est un cas particulier du test de Kruskal-Wallis [6, 7] quand on n'a que deux échantillons.

```
kruskal.test(mâch~plan, data = chac)
       Kruskal-Wallis rank sum test
data:  mâch by plan
Kruskal-Wallis chi-squared = 8.1522, df = 1, p-value = 0.004301
```

Avec un risque de première espèce de 5 %, on rejette donc l'hypothèse nulle. Notez que l'on a perdu en puissance par rapport au test paramétrique de Student [2] :

```
t.test(mâch~plan, data = chac, var.equal = TRUE)
       Two Sample t-test
data:  mâch by plan
t = 3.4843, df = 18, p-value = 0.002647
alternative hypothesis: true difference in means 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
```

3.2 Données appariées

Nos données ne sont pas appariées mais on fait comme si pour illustrer les commandes R utilisables dans ce cas de figure. Une première solution est de préciser comme dans la fonction `t.test()` que l'argument logique `paired` est vrai. Avec un premier argument de type `formula` on a alors :

```
wilcox.test(mâch~plan, data = chac, paired = TRUE)

        Wilcoxon signed rank test with continuity correction
data:  mâch by plan
V = 50.5, p-value = 0.02157
alternative hypothesis: true location shift is not equal to 0
```

On peut également, toujours en précisant que l'argument logique `paired` est vrai, fournir directement les deux vecteurs en premier et deuxième argument :

```
wilcox.test(mal, fem, paired = TRUE)

        Wilcoxon signed rank test with continuity correction
data:  mal and fem
V = 50.5, p-value = 0.02157
alternative hypothesis: true location shift is not equal to 0
```

On peut encore faire « à la main » le test des rangs signés de Wilcoxon :

```
wilcox.test(mal - fem)

        Wilcoxon signed rank test with continuity correction
data:  mal - fem
V = 50.5, p-value = 0.02157
alternative hypothesis: true location is not equal to 0
```

Pour gérer les ex æquo on peut utiliser la fonction `wilcoxsign_test()` du paquet `coin` [4, 5] avec la syntaxe suivante :

```
wilcoxsign_test(mal~fem)

        Asymptotic Wilcoxon-Pratt Signed-Rank Test
data:  y by x (pos, neg)
         stratified by block
Z = 2.349, p-value = 0.01883
alternative hypothesis: true mu is not equal to 0
```

On ne peut pas, comme dans le cas des données non-appariées, utiliser le test de Kruskal-Wallis avec la fonction `kruskal.test()` du paquet standard `stats` car elle contrôle fort logiquement qu'il y a au moins deux échantillons.

4 Exercices

Des générations de praticiens ont utilisé les ouvrages de Pierre Dagnelie [1] 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 [1].

1. `http://www.dagnelie.be/`

4.1 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 :

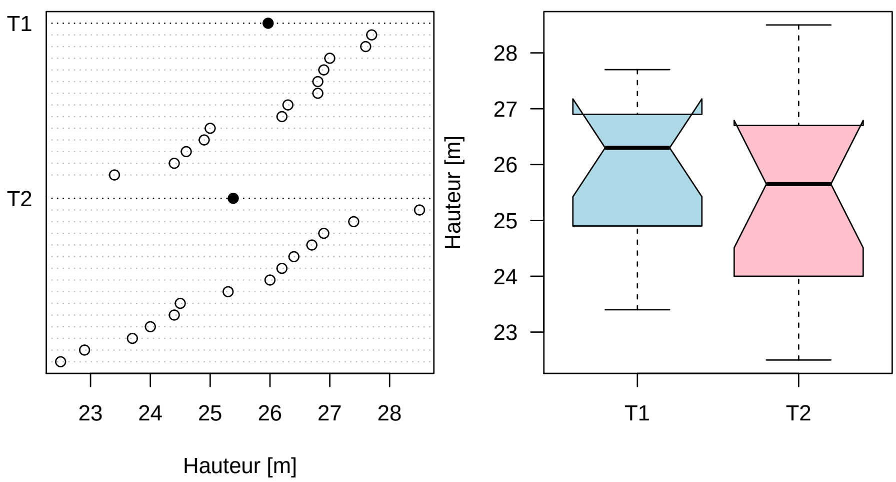
```
arbre <- list()
arbre$hau <- c(23.4,24.4,24.6,24.9,25.0,26.2,26.3,26.8,26.8,26.9,27.0,27.6,27.7,
  22.5,22.9,23.7,24.0,24.4,24.5,25.3,26.0,26.2,26.4,26.7,26.9,27.4,28.5)
arbre$plan <- factor(rep(c("T1", "T2"), c(13, 14)))
arbre <- as.data.frame(arbre)
head(arbre)
```

```
   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 avec un test non-paramétrique et comparez avec les résultats d'un test paramétrique. Seuls les résultats sont donnés ici, à vous de trouver le code R correspondant pour les représentations graphiques et les tests proprement dit.




```
        Kruskal-Wallis rank sum test
data:  hau by plan
Kruskal-Wallis chi-squared = 1.1936, df = 1, p-value = 0.2746

        Two Sample t-test
data:  hau by plan
t = 0.95416, df = 25, p-value = 0.3491
alternative hypothesis: true difference in means 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
```

4.2 Hauteur de 12 arbres avant et après abattage

DES arbres ont été mesurés debout puis abattu (*op. cit.* p. 387). Les données sont les suivantes :

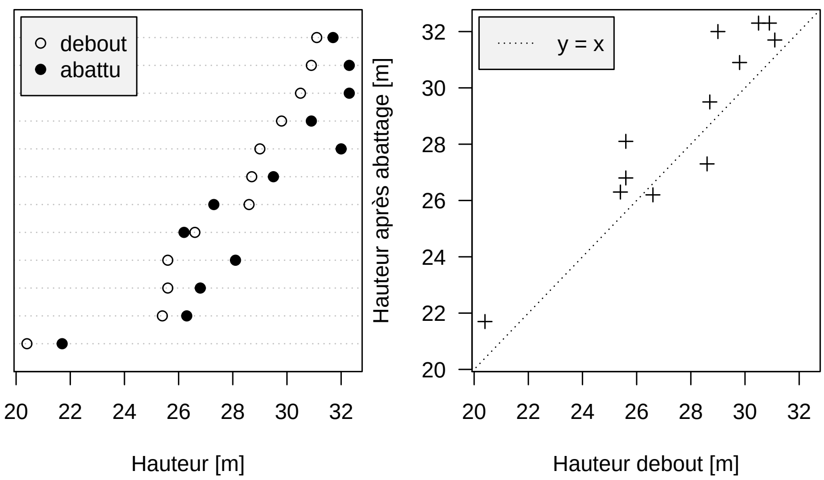
```
debout <- c(20.4,25.4,25.6,25.6,26.6,28.6,28.7,29.0,29.8,30.5,30.9,31.1)
abattu <- c(21.7,26.3,26.8,28.1,26.2,27.3,29.5,32.0,30.9,32.3,32.3,31.7)
ada <- data.frame(hau = c(debout, abattu))
ada$plan <- gl(2, 12, labels = c("debout", "abattu"))
head(ada)
```

```
   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 avec un test non-paramétrique et comparez avec les résultats d'un test paramétrique. Seuls les résultats sont donnés ici, à vous de trouver le code R correspondant pour les représentations graphiques et les tests proprement dit.




```
        Asymptotic Wilcoxon-Pratt Signed-Rank Test
data:  y by x (pos, neg)
         stratified by block
Z = -2.3935, p-value = 0.01669
alternative hypothesis: true mu is not equal to 0


        Paired t-test
data:  hau by plan
t = -3.2343, df = 11, p-value = 0.007954
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.8065536 -0.3434464
sample estimates:
mean of the differences
                 -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 ?

Références

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

[2] Student [Gosset, W.S.]. The probable error of a mean. *Biometrika*, 6 :1–25, 1908.

[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] T. Hothorn, K. Hornik, M.A. van de Wiel, and A. Zeileis. A lego system for conditional inference. *The American Statistician*, 60 :257–263, 2006.

[5] T. Hothorn, K. Hornik, M.A. van de Wiel, and A. Zeileis. Implementing a class of permutation tests : The coin package. *Journal of Statistical Software*, 28 :1–23, 2008.

[6] W.H. Kruskal and W.A. Wallis. Use of ranks on one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260) :583–621, 1952.

[7] W.H. Kruskal and W.A. Wallis. Errata : Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 48(264) :907–911, 1953.

[8] H.B. Mann and D.R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1) :50–60, 1947.

[9] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6) :80–83, 1945.