

Comparaisons de moyennes

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

Le test de Shapiro et Wilks. Le test de Bartlett. Normalité et variances égales. Non normalité. La loi de Wilcoxon. Exercices : choisir un test.

Table des matières

1	Normalité et variances égales	2
1.1	Ce qu'il ne faut pas faire	2
1.2	Le test de Shapiro et Wilks	3
1.3	Les variances sont-elles égales?	6
1.4	Le test t	9
2	Non normalité	11
3	La loi de Wilcoxon	12
4	Exercices : choisir un test	14
4.1	Silex	14
4.2	Fécondité	15
4.3	Alcool	15
4.4	Filtres	15
4.5	Force	15
4.6	Les illustrations de P. Dagnelie	16
	Références	18

1 Normalité et variances égales

La variable mesurée est la longueur de la mâchoire inférieure (en mm) de 10 chacals mâles et 10 chacals femelles (*Canis Aureus*) conservées au British Museum [4]. La variable mesurée diffère t'elle entre les sexes dans cette espèce ?

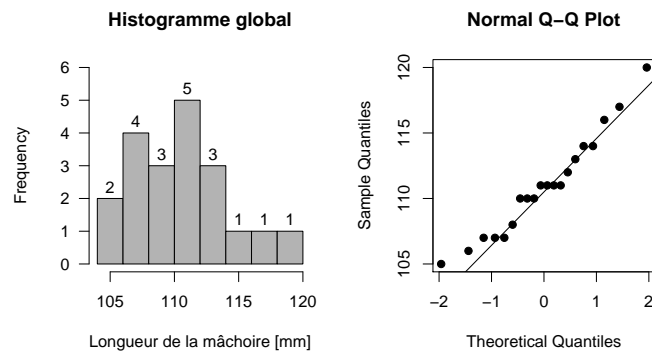
```
(mal <- c(120, 107, 110, 116, 114, 111, 113, 117, 114, 112))
[1] 120 107 110 116 114 111 113 117 114 112
(fem <- c(110, 111, 107, 108, 110, 105, 107, 106, 111, 111))
[1] 110 111 107 108 110 105 107 106 111 111
```

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

1.1 Ce qu'il ne faut pas faire

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

```
tot <- c(mal, fem)
par(mfrow = c(1, 2))
hist(tot, nclass = 8, las = 1, col = grey(0.7), label = T, ylim = c(0,
  6), main = "Histogramme global", xlab = "Longueur de la mâchoire [mm]")
qqnorm(tot, pch = 19)
qqline(tot)
```



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 :

```
shapiro.test(tot)
  Shapiro-Wilk normality test
data:  tot
W = 0.9593, 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 = 0.8271, p-value = 0.002259
```

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

1.2 Le test de Shapiro et Wilks

La documentation, dans une version antérieure de \mathbb{R} de la fonction disait (dans la version 2.4.1 elle est très simplifiée) :

The Shapiro-Wilk Test For Normality

The Shapiro-Wilk test, proposed in 1965, calculates a W statistic that tests whether a random sample, x_1, x_2, \dots, x_n comes from (specifically) a normal distribution. Small values of W are evidence of departure from normality and percentage points for the W statistic, obtained via Monte Carlo simulations, were reproduced by Pearson and Hartley (1972, Table 16). This test has done very well in comparison studies with other goodness of fit tests ...

Elle donnait une indication sur le calcul et les sources bibliographiques. Quelques explications s'imposent. La statistique W est calculée par :

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

où $x_{(i)}$ est la i^{me} valeur observée par ordre croissant et a_i est une constante générée à partir des moyennes, variances et covariances des statistiques d'ordre d'un échantillon d'une loi normale de taille n (voir Pearson and Hartley (1972, Table 15)). Pour plus d'information, recommande cette fiche de documentation maintenant simplifiée, le lecteur est renvoyé à l'article original (Shapiro and Wilk (1965)[9]) et aux tables de Pearson and Hartley (1972)[7].

Il n'est pas question de calculer à la main la statistique en question. Le calcul techniquement est bien trop compliqué. Mais on peut séparer ce qui revient à l'expertise du numéricien qui a mis au point les procédures et à la compétence de l'utilisateur qui peut quand même espérer savoir de quoi ça parle !

Un bon exercice consiste à chercher comment se fait le calcul. Faire `shapiro.test` au clavier : on obtient le source de la fonction en \mathbb{R} . Repérer la ligne essentielle :

```
...
sw <- .C(R_swilk, init = FALSE, as.single(x), n, n1 = as.integer(n),
as.integer(n2), a = single(n2), w = double(1), pw = double(1),
ifault = integer(1))
...
```

Cela signifie que la fonction en R utilise la fonction en C qui s'appelle `swilk`. Le lecteur concerné ira alors sur le serveur CRAN récupérer les sources :



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

R Sources

Official releases

Official releases of R source code are most likely what you want if you are looking for the sources of R (both Unix and Windows).

- The latest release: [R 2.11 tar.gz](#)
- Changes to the previous version are documented in the file [NEWS](#) (also contained in the sources).
- Older releases are available [here](#).

Snapshots

Snapshots of R sources for development and patched release versions can be obtained as gzipped and bzipipped tar files from <ftp://ftp.stat.math.ethz.ch/Software/R/>.

Il faut ensuite les décompresser et chercher le fichier `swilk.c`. On y trouve :

```
/* swilk.f -- translated by f2c (version 19980913).
 * ----- and produced by f2c-clean,v 1.8 --- and hand polished: M.Maechler
 */
```

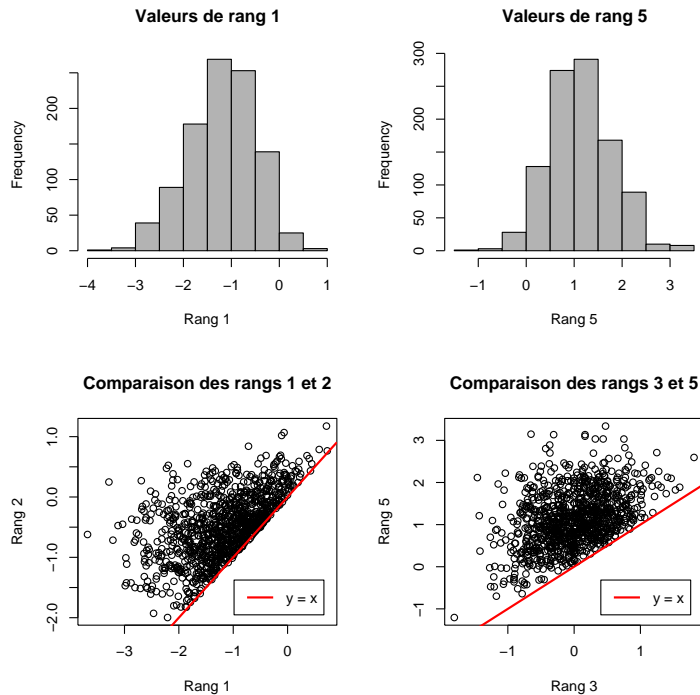
C'est du Fortran traduit en C et mis au point par Martin Mächler.

```
...
const static float g[2] = { -2.273f,.459f };
const static float
  c1[6] = { 0.f,.221157f,-.147981f,-2.07119f, 4.434685f, -2.706056f },
  c2[6] = { 0.f,.042981f,-.293762f,-1.752461f,5.682633f, -3.582633f };
const static float c3[4] = { .544f,-.39978f,.025054f,-6.714e-4f };
...
/* Calculate W statistic as squared correlation
  between data and coefficients */

sa /= *n1;
sx /= *n1;
ssa = ssx = sax = zero;
...
```

A chacun son métier. Mais l'essentiel est perceptible. W mesure la corrélation (c'est un carré de corrélation) entre les données ordonnées et les statistiques d'ordre d'une loi normale. De quoi s'agit-il ?

```
experience <- function(n = 5) sort(rnorm(n))
w <- replicate(n = 1000, experience())
par(mfrow = c(2, 2), lend = "butt")
hist(w[1, ], main = "Valeurs de rang 1", col = grey(0.7), xlab = "Rang 1")
hist(w[5, ], main = "Valeurs de rang 5", col = grey(0.7), xlab = "Rang 5")
plot(w[1, ], w[2, ], xlab = "Rang 1", ylab = "Rang 2", main = "Comparaison des rangs 1 et 2")
abline(c(0, 1), col = "red", lwd = 2)
legend("bottomright", inset = 0.05, lwd = 2, col = "red", legend = "y = x")
plot(w[1, ], w[5, ], xlab = "Rang 1", ylab = "Rang 5", main = "Comparaison des rangs 1 et 5")
abline(c(0, 1), col = "red", lwd = 2)
legend("bottomright", inset = 0.05, lwd = 2, col = "red", legend = "y = x")
```



```

apply(w, 1, mean)
[1] -1.188143518 -0.483335460 0.003697301 0.467600269 1.141998726
apply(w, 1, sd)
[1] 0.6912629 0.5251316 0.5201412 0.5359551 0.6542447
round(cor(t(w)), dig = 2)
      [,1] [,2] [,3] [,4] [,5]
[1,] 1.00 0.55 0.37 0.25 0.11
[2,] 0.55 1.00 0.66 0.44 0.21
[3,] 0.37 0.66 1.00 0.69 0.42
[4,] 0.25 0.44 0.69 1.00 0.58
[5,] 0.11 0.21 0.42 0.58 1.00

```

Nous avons fait l'expérience suivante. 1000 fois, un échantillon de taille $n = 5$ d'une loi normale de moyenne 0 et variance 1 est tiré. Chaque fois l'échantillon est rangé par ordre croissant. Le plus petit est la statistique d'ordre de rang 1. Il a une moyenne et une variance. Le suivant est évidemment dépendant du premier. Il est forcément plus grand. Il a une moyenne et une variance. Les cinq statistiques d'ordre ont une matrice de corrélation. Pour continuer cette expérience, on utilise autant de points que dans notre échantillon :

```

w <- replicate(n = 1000, experience(length(tot)))
moy <- rowMeans(w)
a1 <- (cor(sort(tot), moy))^2
a2 <- (cor(sort(tot), qnorm(ppoints(20))))^2
a3 <- shapiro.test(tot)$statistic
c(a1, a2, a3)

```

W

```

0.9597676 0.9608445 0.9592896

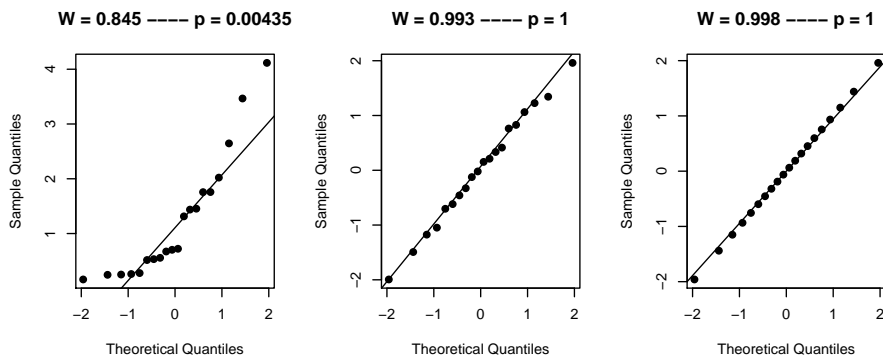
```

On retiendra que, en première approximation le test de Shapiro et Wilks porte sur la corrélation au carré qu'on voit sur un qqplot. Seule les nécessités

techniques pour le calcul des niveaux de signification font que la statistique n'est pas exactement celle-là. La probabilité critique est celle d'un test à gauche. La corrélation est toujours très forte, la question est toujours "l'est-elle assez?" La probabilité critique est la probabilité pour que la statistique soit inférieure ou égale à l'observation.

```
monplot <- function(x) {
  test <- shapiro.test(x)
  qqnorm(x, main = paste("W =", signif(test$statistic, 3), "----",
    "p =", signif(test$p.value, 3)), pch = 19)
  qqline(x)
}
par(mfrow = c(1, 3))
monplot(rexp(20))
monplot(rnorm(20))
monplot(qnorm(seq(0.025, 0.975, by = 0.05)))
```

A gauche, un échantillon d'une loi exponentielle, pas normal du tout! Le test est très significatif à *gauche*. Au centre un échantillon d'une loi normale qui apparaît comme normal! A droite un échantillon trop normal pour être honnête! La statistique est trop élevée et le test est très significatif à *droite*.



```
shapiro.test(mal)
      Shapiro-Wilk normality test
data:  mal
W = 0.9937, p-value = 0.9995
shapiro.test(fem)
      Shapiro-Wilk normality test
data:  fem
W = 0.8809, p-value = 0.1335
```

Votre conclusion ?

1.3 Les variances sont-elles égales ?

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

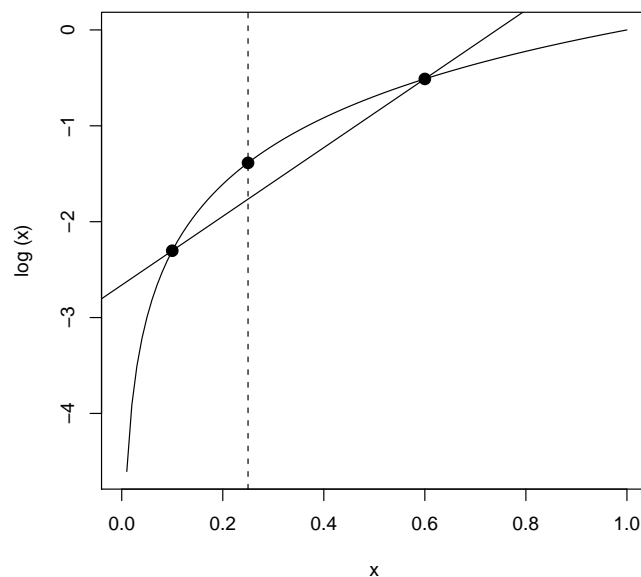
```
plan <- gl(n = 2, k = 10, lab = c("mal", "fem"))
bartlett.test(tot, plan)
      Bartlett test of homogeneity of variances
data:  tot and plan
Bartlett's K-squared = 1.9942, df = 1, p-value = 0.1579
```

```

var.test(mal, fem)
      F test to compare two variances
data:  mal and fem
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

```

Le calcul de ce test est simple et peut se faire manuellement (ce n'est évidemment pas une raison pour le faire à la main). Mais ceci rend plus facile d'accès la signification du calcul, laquelle n'est pas simpliste. Pour mesurer la différence de deux variances on pourrait penser que calculer la différence ou le rapport suffit. Il n'en est rien, parce qu'on ne connaît pas la loi, sous l'hypothèse de normalité et d'égalité de variance, de la différence des variances observées. On utilise alors une fonction qui augmente avec cette différence et qui permet de calculer cette loi. La fonction logarithme est convexe :



La différence observée entre deux échantillons, par exemple 0.1 et 0.6, ne représente en elle-même pas grand-chose. Si la mesure porte sur des échantillons de taille (10, 10) ou (1000, 1000), cela n'a pas du tout la même signification. De même si on a affaire à des échantillons de taille (3, 47), la situation est très différente du cas (25, 25). Le test utilise la différence entre le logarithme de la moyenne et la moyenne des logarithmes. Le premier est sur la courbe et le second est sur la corde. Le premier est toujours plus grand que le second. Cette moyenne utilise les poids $n_1 - 1$ et $n_2 - 1$ si ces quantités sont les effectifs respectifs des deux échantillons. Le test porte sur :

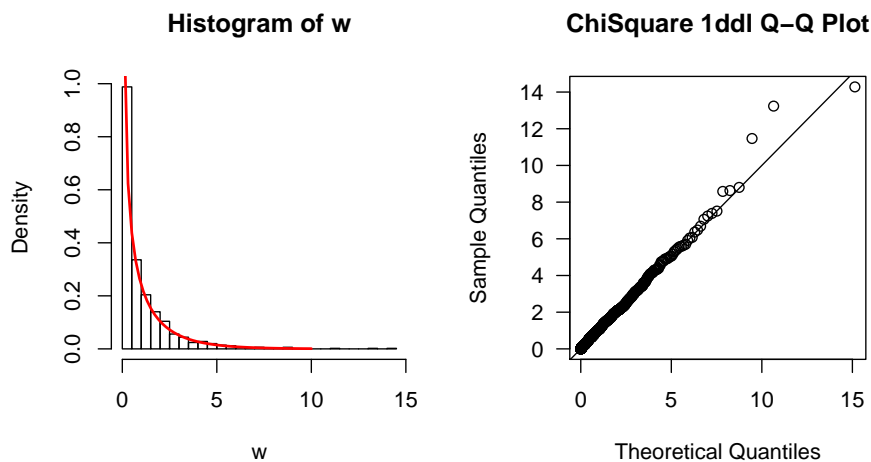
$$v_1 = \frac{\sum (x_{1j} - \bar{x}_1)^2}{(n_1 - 1)} \quad v_2 = \frac{\sum (x_{2j} - \bar{x}_2)^2}{(n_2 - 1)}$$

$$\alpha_1 = \frac{(n_1 - 1)}{(n_1 + n_2 - 2)} \quad \alpha_2 = \frac{(n_2 - 1)}{(n_1 + n_2 - 2)}$$

$$M = (n_1 + n_2 - 2) [\log(\alpha_1 v_1 + \alpha_2 v_2) - \alpha_1 \log(v_1) - \alpha_2 \log(v_2)]$$

Cette fonction augmente avec les différences entre les deux estimations de variance, elle augmente avec la taille des échantillons et elle augmente avec l'équilibre entre les effectifs des échantillons. M.S. Bartlett a démontré que sa distribution d'échantillonnage converge vers une loi Khi^2 [1]. Pour s'en convaincre, on fera une simulation du type :

```
f1 <- function(k) {
  x <- rnorm(11, 1, 5)
  y <- rnorm(21, 3, 5)
  vx <- var(x)
  vy <- var(y)
  vmoy <- (10 * vx + 20 * vy)/30
  w <- 30 * log(vmoy) - 10 * log(vx) - 20 * log(vy)
  return(w)
}
par(mfrow = c(1, 2))
w <- unlist(lapply(1:1000, f1))
hist(w, proba = T, nclass = 50)
lines(x0 <- seq(0.1, 10, le = 50), dchisq(x0, 1), lwd = 2, col = "red")
qqplot(qchisq(ppoints(n = 5000), df = 1), sort(w), main = "ChiSquare 1ddl Q-Q Plot",
  xlab = "Theoretical Quantiles", ylab = "Sample Quantiles", las = 1)
abline(0, 1)
```



On améliore encore l'approximation en divisant la statistique par :

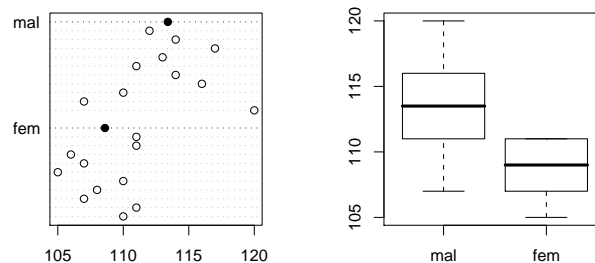
$$1 + \frac{1}{3} \left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_1 + n_2} \right)$$

Il s'agit là de détail technique qui ne change pas l'interprétation du test. Le test est réputé pour être sensible à la non normalité (ce qui n'est pas ici un défaut). Votre conclusion dans le cas étudié ?

1.4 Le test t

On a respecté un schéma classique 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 simplifie la vie du débutant ? En fait, tous ces éléments doivent s'intégrer dans une analyse qui permet une interprétation. Il serait temps de regarder les données !

```
moy <- tapply(tot, plan, mean)
par(mfrow = c(1, 2))
dotchart(tot, gr = plan, gdata = moy, gpch = 19)
boxplot(split(tot, plan))
```



Lire la fiche de documentation de la fonction `t.test`, identifier les paramètres et noter ceux qui posent des problèmes.

```
TMF <- t.test(mal, fem, var.equal = TRUE)
ANOVAMF <- anova(lm(tot ~ plan))
TMF

Two Sample t-test
data: mal and fem
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 of x mean of y
 113.4    108.6

ANOVAMF
Analysis of Variance Table
Response: tot
      Df Sum Sq Mean Sq F value Pr(>F)
plan    1 115.200  115.200  12.140 0.002647 **
Residuals 18 170.800    9.489
---
Signif. codes:  0
```

Le résultat comprend plusieurs éléments. La statistique du test t vaut 3.484. Avec `var.equal = TRUE`, Les deux variances sont considérées comme égales. Le test est strictement à une analyse de variance à un facteur à deux modalités. On estime la variance commune par :

$$\left. \begin{array}{l} SCE_1 = \sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 \\ SCE_2 = \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2 \end{array} \right\} \frac{SCE_1 + SCE_2}{\sigma^2} \rightarrow \chi_{n_1-1+n_2-1}^2$$

Estimer cette variance commune.

```
sum((mal - mean(mal))^2) + sum((fem - mean(fem))^2)
[1] 170.8
v <- (sum((mal - mean(mal))^2) + sum((fem - mean(fem))^2))/(10 +
10 - 2)
```

Retrouver cette valeur de 9.489 dans l'analyse de variance. Calculer la différence des moyennes des échantillons :

```
md <- mean(mal) - mean(fem)
```

Cette valeur vaut 4.8. Calculer la variance estimée de la différence des moyennes :

$$\left. \begin{array}{l} \bar{X}_1 \rightarrow N\left(\mu_1, \frac{\sigma^2}{n_1}\right) \\ \bar{X}_2 \rightarrow N\left(\mu_2, \frac{\sigma^2}{n_2}\right) \end{array} \right\} \bar{X}_1 - \bar{X}_2 \rightarrow N\left(\mu_1 - \mu_2, \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

```
vd <- v * (1/10 + 1/10)
vd
[1] 1.897778
```

Tester l'hypothèse nulle "la différence des moyennes des populations est égale à 20"

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightarrow N(0, 1) \quad \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightarrow t(n_1 + n_2 - 2)$$

```
(md - 20)/sqrt(vd)
[1] -11.03369
pt((md - 20)/sqrt(vd), 18)
[1] 9.627822e-10
```

C'est impossible que la vraie différence soit aussi grande. Tester l'hypothèse nulle "la différence des moyennes des populations est égale à 10"

```
pt((md - 10)/sqrt(vd), 18)
[1] 0.000693747
```

C'est encore très significatif. Tester l'hypothèse nulle "la différence des moyennes des populations est égale à 1" :

```
pt((md - 1)/sqrt(vd), 18)
[1] 0.9935312
1 - pt((md - 1)/sqrt(vd), 18)
[1] 0.006468819
```

C'est toujours très significatif mais de l'autre côté. Tester l'hypothèse nulle "la différence des moyennes des populations est égale à 1.906". On trouve 0.975 : c'est juste non significatif à 2.5%.

Tester l'hypothèse nulle "la différence des moyennes des populations est égale à 7.694". On trouve 0.02501 : c'est juste non significatif à 2.5%, de l'autre côté.

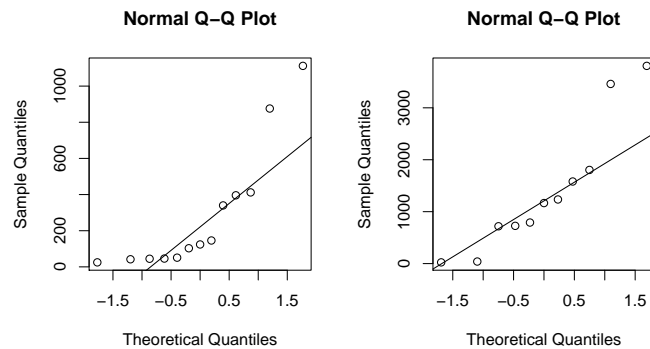
Rassembler toutes les valeurs qui donnent un test bilatéral significatif au seuil de 5%. On y trouve toutes les valeurs plus grandes que 7.694 et toutes les valeurs plus petites que 1.906.

On dit que [1.906,7.694] est l'intervalle de confiance de la différence des moyennes au seuil de confiance de 95%. Non seulement on sait qu'elle n'est pas nulle, mais on peut donner une indication sur l'écart acceptable.

2 Non normalité

La variable mesurée est le temps de survie (en jours) de 13 patients atteints d'un cancer de l'estomac et de 11 patients atteint d'un cancer du poumon [2]. La question posée est "le temps de survie varie-t-il avec le type de cancer?".

```
y1 <- c(124, 42, 25, 45, 412, 51, 1112, 46, 103, 876, 146, 340,
        396)
y2 <- c(1235, 24, 1581, 1166, 40, 727, 3808, 791, 1804, 3460, 719)
par(mfrow = c(1, 2))
qqnorm(y1)
qqline(y1)
qqnorm(y2)
qqline(y2)
shapiro.test(y1)
      Shapiro-Wilk normality test
data:  y1
W = 0.7547, p-value = 0.002075
shapiro.test(y2)
      Shapiro-Wilk normality test
data:  y2
W = 0.8686, p-value = 0.07431
```



Calculer les moyennes des échantillons. Observer que la non-normalité est évidente et que les tests utilisant cette hypothèse sont hors d'usage.

```
wilcox.test(y1, y2)
      Wilcoxon rank sum test
data:  y1 and y2
W = 31, p-value = 0.01836
alternative hypothesis: true location shift is not equal to 0
```

Lire attentivement le fiche de documentation de la fonction.

```
help(wilcox.test)
```

Retrouver le détail des calculs dans la fiche de cours dans :

<http://pbil.univ-lyon1.fr/R/cours/bs3.pdf>

Essayer d'obtenir le source de la fonction

```
wilcox.test
```

Devant cette difficulté, essayer de connaître les fonctions associées :

```
methods(wilcox.test)
```

Alors, y aurait-il des sources inaccessibles dans \mathbb{R} ? Non! Mais la curiosité peut mener loin.

```
getAnywhere(wilcox.test.default)
```

La fonction est riche : elle peut faire le test exact, les tests unilatéraux, introduire des corrections pour augmenter la précision et donner des intervalles de confiance :

```
t.test(mal, fem)$conf.int
[1] 1.861895 7.738105
attr(,"conf.level")
[1] 0.95

wilcox.test(mal, fem, conf.int = T)$conf.int
[1] 1.999999 7.999950
attr(,"conf.level")
[1] 0.95
```

En cas de normalité, les résultats sont très voisins alors que les techniques sont très différentes. Conclure sur l'exemple traité.

3 La loi de Wilcoxon

Tirer au hasard 2 des 5 premiers entiers. Quelle est la valeur par défaut du paramètre `replace` ?

```
sample(1:5, 2)
[1] 5 3
```

Recommencer quelques tirages (ça ne coûte pas cher!). Calculer SR, dite "la somme des rangs", c'est à dire la somme des nombres tirés. Recommencer (ça ne coûte toujours pas plus cher!).

```
sum(sample(1:5, 2))
[1] 9
```

Appeler m le nombre d'entiers tirés et n le nombre d'entiers restés. Le minimum de la somme des rangs est obtenu quand on a tiré les m premiers parmi les $n+m$, soit :

$$\min = 1 + \dots + m = \frac{m(m+1)}{2}$$

Le maximum est obtenu quand on a tiré les m derniers parmi les $n+m$, donc qu'on a tiré les n premiers parmi les $n+m$. Comme la somme est constante :

$$\max = \frac{(m+n)(m+n+1)}{2} - \frac{n(n+1)}{2}$$

La statistique de Wilcoxon est $SR - \min$. Elle varie de 0 à :

$$\max - \min = \frac{(m+n)(m+n+1)}{2} - \frac{n(n+1)}{2} - \frac{m(m+1)}{2} = nm$$

Wilcoxon package:stats R Documentation

Distribution of the Wilcoxon Rank Sum Statistic

Description:

Density, distribution function, quantile function and random generation for the distribution of the Wilcoxon rank sum statistic obtained from samples with size 'm' and 'n', respectively.

Usage:

```
dwilcox(x, m, n, log = FALSE)
pwilcox(q, m, n, lower.tail = TRUE, log.p = FALSE)
qwilcox(p, m, n, lower.tail = TRUE, log.p = FALSE)
rwilcox(nn, m, n)
```

Arguments:

x, q: vector of quantiles.

p: vector of probabilities.

nn: number of observations. If 'length(nn) > 1', the length is taken to be the number required.

m, n: numbers of observations in the first and second sample, respectively. Can be vectors of positive integers.

log, log.p: logical; if TRUE, probabilities p are given as log(p).

lower.tail: logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.

Details:

This distribution is obtained as follows. Let 'x' and 'y' be two random, independent samples of size 'm' and 'n'. Then the Wilcoxon rank sum statistic is the number of all pairs '(x[i], y[j])' for which 'y[j]' is not greater than 'x[i]'. This statistic takes values between '0' and 'm * n', and its mean and variance are 'm * n / 2' and 'm * n * (m + n + 1) / 12', respectively.

If any of the first three arguments are vectors, the recycling rule is used to do the calculations for all combinations of the three up to the length of the longest vector.

Value:

'dwilcox' gives the density, 'pwilcox' gives the distribution function, 'qwilcox' gives the quantile function, and 'rwilcox' generates random deviates.

Note:

S-PLUS uses a different (but equivalent) definition of the Wilcoxon statistic: see 'wilcox.test' for details.

Author(s):

Kurt Hornik

Dans la section `details`, la définition n'est pas exactement celle de la fiche BS3 mais est très voisine. Vérifier par une simulation :

```
sim1 <- function(x = 0, m, n) {
  a = sample(1:(n + m), m, rep = F)
  a = sum(a)
  a = a - m * (m + 1)/2
  return(a)
}
sim1(m = 2, n = 3)
[1] 3
```

Créer le vecteur `w` des 1000 simulations de la fonction `sim1`. Comparer avec la loi :

```
w <- sapply(1:1000, sim1, m = 2, n = 3)
table(w)
w
 0  1  2  3  4  5  6
104 106 189 189 213 101 98
dwilcox(0:6, 2, 3)
[1] 0.1 0.1 0.2 0.2 0.2 0.1 0.1
```

Tabuler les cas possibles ($\binom{5}{2} = 10$ cas possibles) et retrouver la loi. Autre exemple ($\binom{6}{3} = 20$ cas possibles) :

```
dwilcox(0:9, 3, 3)
[1] 0.05 0.05 0.10 0.15 0.15 0.15 0.15 0.10 0.05 0.05
```

Afficher la documentation de `wilcox.test` et trouver la *p-value* –d'abord **sans le logiciel**, vérifier ensuite– du test :

```
wilcox.test(c(2.3, 3.6, 8.1, 3.4), c(12.1, 9.7, 9.6, 8.3), alt = "1")
```

Trouver la *p-value* –d'abord **sans le logiciel**, vérifier ensuite– du test :

```
wilcox.test(c(2.3, 3.6, 8.1, 3.4), c(12.1, 9.7, 9.6, 8.3))
```

Trouver la *p-value* –d'abord **sans le logiciel**, vérifier ensuite– du test :

```
wilcox.test(c(2.3, 3.6, 8.1, 3.4), c(12.1, 9.7, 9.6, 8.3), alt = "g")
```

N.B. *pour les filous!* $\frac{1}{70}$ vaut 0.01429

4 Exercices : choisir un test

4.1 Silex

On a mesuré la dureté du 12 silex provenant de deux régions A et B, et on a classé les silex du moins dur au plus dur :

<i>Rang</i>	1	2	3	4	5	6	7	8	9	10	11	12
<i>Origine</i>	A	A	A	B	A	A	B	A	B	B	B	B

La dureté est-elle la même dans les deux régions ?

4.2 Fécondité

On a étudié[6] la fécondité d'une guêpe parasite (*Diadromus pulchellus*) en fonction de la présence d'un hôte au stade cocon (*Acrolepia assectella*). Deux lots A et B de parasites sont formés. Aux parasites du lot A, on présente un hôte les jours (du 1^{er} au 35^{me} jour), à ceux du lot B à partir du 6^{me} jour seulement (du 6^{me} au 35^{me} jour). Les nombre d'œufs pondus par chaque insecte sont les suivantes :

Lot A (13 insectes) : 98, 84, 63, 75, 84, 66, 56, 48, 109, 85, 95, 106
 Lot B (18 insectes) : 124, 83, 75, 123, 105, 108, 155, 128, 56, 72, 96, 45, 71, 45, 73, 60, 89, 83

Commenter les résultats.

4.3 Alcool

On a étudié[5] le temps de réaction nécessaire pour arrêter une automobile chez des sujets sous l'influence de trois onces d'alcool (environ 0,09 l). On mesure le temps de réaction (en 100^{me} de seconde) avant et après l'ingestion d'alcool.

<i>Cobaye</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>Avant</i>	33	29	26	23	21	36	27	38	22	33	42	35	22	39	37
<i>Après</i>	46	41	37	37	30	43	38	47	33	42	54	48	33	54	50

Conclusion ?

4.4 Filtres

Dans une étude sur le traitement des eaux usées, l'efficacité de deux filtres, l'un en fibre de verre et l'autre en papier filtre Whatman 40 a été testée[8]. Sur des prélèvements de 200 millilitres d'eau provenant d'une usine de pâtes à papier, la quantité de solides en suspension retenus par les deux filtres a été mesurée. L'efficacité du filtre de verre est-elle supérieure à celle du papier filtre ?

<i>Prélèvement</i>	1	2	3	4	5	6	7	8
<i>Verre</i>	68	86	91	77	100	138	190	110
<i>Papier</i>	64	75	72	63	96	120	184	96

4.5 Force

On mesure la force statique par dynamométrie manuelle (exprimée en kg) chez 10 enfants atteint de trisomie 21. Un premier test est réalisé en septembre. Puis, sur une période de six mois, ces enfants essaient de développer, sous forme de jeux, leur force statique. Un second test est réalisé en février. Peut-on dire que le programme suivi par ces enfants a permis d'améliorer leur résultat ?

33	42.5	54	60	61	68	68	69	72	86
38	45	52	63	61	75	66	70	81	90

4.6 Les illustrations de P. Dagnelie

Des générations de praticiens ont utilisé les ouvrages de P. Dagnelie

<http://www.dagnelie.be/>

pour vérifier leurs procédures. Voici quelques exemples[3].

- ★ On compare la hauteur des arbres de 27 peuplements forestiers appartenant à deux types différents (en m, *op. cit.* p.25). Planter les données pour obtenir l'information suivante :

```
arbre$hau
[1] 23.4 24.4 24.6 24.9 25.0 26.2 26.3 26.8 26.8 26.9 27.0 27.6 27.7 22.5 22.9 23.7
[17] 24.0 24.4 24.5 25.3 26.0 26.2 26.4 26.7 26.9 27.4 28.5
arbre$type
[1] T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2
Levels: T1 T2
```

Comparer les deux groupes par le test *t* (*op. cit.* p. 132), le test de la médiane (*op. cit.* p. 381) et le test de Wilcoxon (*op. cit.* p. 384). Pour le premier, utiliser :

```
table(arbre$hau > median(arbre$hau), arbre$type)
```

- ★ Des arbres ont été mesurés debouts puis abattus. Commenter en utilisant le test des paires de Wilcoxon (*op. cit.* p. 387)

```
debout <- c(20.4, 25.4, 25.6, 25.6, 26.6, 28.6, 28.7, 29, 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, 30.9,
32.3, 32.3, 31.7)
```

- ★ Etendre les données sur les peuplements forestiers pour obtenir :

```
arbre$hau
[1] 23.4 24.4 24.6 24.9 25.0 26.2 26.3 26.8 26.8 26.9 27.0 27.6 27.7 22.5 22.9 23.7
[17] 24.0 24.4 24.5 25.3 26.0 26.2 26.4 26.7 26.9 27.4 28.5 19.9 21.1 21.2 22.1 22.5
[33] 23.6 24.5 24.6 26.2 26.7
arbre$type
[1] T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T1 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2 T2
[28] T3 T3 T3 T3 T3 T3 T3 T3 T3 T3
Levels: T1 T2 T3
```

Consulter la documentation de `kruskal.test` et comparer les trois groupes par le test de Kruskal et Wallis (*op. cit.* p. 392).

- ★ On a appliqué cinq traitements fongicides (en ligne dans la matrice pomme) à une culture de pomme de terre. Le champ expérimental a été divisé en 4 blocs homogènes (en colonne dans la matrice pomme), chaque bloc étant divisé en cinq parcelles recevant un traitement différent. L'observation est le rendement moyen en grammes par pied. Planter les données pour obtenir la matrice (voir `matrix`, `col`, `row`).

```
pomme
      [,1] [,2] [,3] [,4]
[1,] 527 604 606 533
[2,] 633 600 650 567
[3,] 642 708 662 504
[4,] 623 550 562 667
[5,] 377 408 500 333
```

Quel est la fonction des paramètres `row` et `col` et quel est l'intérêt des fonctions `row` et `col`. Voir `friedman.test` et utiliser le test de Friedman (*op. cit.* p. 394) pour tester l'effet traitement. On devra trouver :

```
Friedman rank sum test
data: pomme.vec, traitement and bloc
Friedman chi-squared = 9.8, df = 4, p-value = 0.04393
```

- ★ On pourra revenir à cet exercice après avoir vu la fonction `lm` pour faire une comparaison avec l'analyse de variance pour tester l'effet bloc "sachant traitement" (*op. cit.* p. 183-184) :

```
Analysis of Variance Table
Response: pomme.vec
      Df Sum Sq Mean Sq F value Pr(>F)
traitement 4 133419 33355 9.5763 0.001026 **
bloc       3 14987 4996 1.4343 0.281402
```

```
Residuals 12 41797 3483
```

```
---  
Signif. codes: 0
```

La précision de l'ouvrage de P. Dagnelie est légendaire. On peut l'utiliser pour vérifier qu'on manipule correctement le logiciel. Compléter votre catalogue personnel.

Références

- [1] M.S. Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London - Series A*, 160 :268–282, 1937.
- [2] E. Cameron and L. Pauling. Supplemental ascorbate in the supportive treatment of cancer : re-evaluation of prolongation of survival times in terminal human cancer. *Proceeding of the National Academy of Sciences of the USA*, 75 :4538–4542, 1978.
- [3] P. Dagnelie. *Théories et méthodes statistiques. Applications agronomiques. II Les méthodes de l'inférence statistique*. Editions L. Duculot, Gembloux, 1970.
- [4] D.J. Hand, F. Daly, A.D. Lunn, K.J. McConway, and E. Ostrowski. *A handbook of small data sets*. Chapman & Hall, London, 1994.
- [5] D.E. Honkle, W. Wiersma, and S.G. Jurs. *Applied Statistics for the Behavioral Sciences*. Houghton Mifflin, 1988.
- [6] V. Labeyrie. Contribution à l'étude de la dynamique des populations d'insectes. influence de l'hôte *Acrolepia assectella* z. sur la multiplication d'un hyménoptère ichneumonidae (*Diadromus* sp.). *Entomophaga*, Mémoire hors-série 1, 1960.
- [7] E.S. Pearson and H.O. Hartley. *Biometrika tables for statisticians, Vol. 2*. Cambridge University Press, Cambridge, 1972.
- [8] B. Scherrer. *Biostatistique*. Gaëtan Morin éditeur, 1984.
- [9] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, (52) :591–611, 1965.