Fiche TD avec le logiciel R : tdr36

Risques, puissance et robustesse

A.B. Dufour

L'objectif de cette fiche est d'appréhender les notions de risques, puissance et robustesse des tests statistiques à travers des exemples et des simulations.

Contents

1 Les concepts de base

1.1 Situation : Préférences manuelles chez les grands singes

Un trait connu de l'espèce humaine est la latéralisation manuelle avec une nette dominance pour la main droite dans les actions motrices. Pour beaucoup, cette asymétrie est spécifique à l'homme. Mais des études récentes ont montré que pour certaines tâches, il y a asymétrie également chez les grands singes.
Les données analysées sont issues de l'article de Hopkins *et al* [3]. La latéralité a été observée chez 774 grands singes (536 chimpanzés, 76 gorilles, 118 bonobos et 47 orangs outans).

Afin de déterminer la latéralité d'un singe, Hopkins propose le protocole suivant [2].

1. Prendre des tubes en PVC de longueurs variant entre 24 et 31 cm et de 4 cm de diamètre
2. Mettre du beurre de cacahuète à l'intérieur du tube sur environ 7 cm

Une expérience est réalisée si le singe prend le tube dans une main et retire, dans le même temps, le beurre de cacahuète avec l'autre main. La main utilisée pour prélever du beurre de cacahuète est enregistrée chaque fois que le sujet a introduit le doigt dans le tube et porté le doigt à sa bouche. Les observations continuent jusqu'à ce que le singe se désintéresse du tube. L'expérience est répétée deux fois avec 24 jours d'intervalle.
L'ensemble des expériences, pour un sujet, suit une loi binomiale. Si la valeur standardisée est inférieure ou égale à -1.64, le singe est déclaré gaucher ; si la valeur est supérieure ou égale à 1.64, le singe est déclaré droitier ; entre les deux, il est déclaré sans préférence manuelle.

Voici les résultats obtenus chez les adultes.

Espèce	gauche	sans préférence	droite
Orang-Outan	16	6	7
Gorille	7	10	23
Chimpanzé	95	66	163
Bonobo	20	4	28
Total	138	86	221

On considère l'ensemble des grands singes et on se place sous l'hypothèse que la latéralisation ne concerne que l'espèce humaine. Il y a donc autant de grands singes gauchers que de grands singes droitiers.

H_0 : La proportion de gauchers chez les grands singes est identique à 0.5.
H_1 : La proportion de gauchers chez les grands singes est différente de 0.5.

1.2 Comparaison d'une proportion observée à une proportion théorique

La démarche consiste ensuite à tester cette hypothèse. Hopkins a réalisé une étude qui conduit à la proportion de gauchers suivante :

```
pgauche <- 138/359
pgauche
```

```
[1] 0.3844011
```

Notez que nous avons choisi de mettre de côté les singes sans préférence manuelle. La proportion de gauchers dans la population p est inconnue. On l'estime à partir de l'échantillon $\hat{p} = f$ et on calcule son intervalle de confiance à 0.95.

$$f \pm u_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}}$$

```
pgauche <- 138/(138+221)
pgauche
```

```
[1] 0.3844011
```

```
(pgauche)+qnorm(0.025)*sqrt((pgauche*(1-pgauche))/359)
```

```
[1] 0.3340809
```

```
(pgauche)+qnorm(0.975)*sqrt((pgauche*(1-pgauche))/359)
```

```
[1] 0.4347213
```

Cet intervalle de confiance s'obtient, sous R lorsqu'on applique le test de comparaison d'une proportion observée à une proportion théorique.

```
prop.test(138,359,p=0.5, correct=FALSE) -> res
res$estimate
```

```
        p
0.3844011
```

```
res$conf.int
```

```
[1] 0.3355569 0.4356930
attr(,"conf.level")
[1] 0.95
```

Si la proportion estimée est la même, on note une différence pour les bornes inférieure et supérieure de l'intervalle. La réponse est apportée par Edwin Wilson [8] et Robert Newcombe [4].

> Really the chance that the true probability p lies outside a specified range is either 0 or 1; for p actually lies within that range or does not. It is the observed rate p_0 which has a greater or less chance of lying within a certain interval of the true rate p. [Wilson, 1927]

Le raisonnement devient le suivant. Il existe une certaine proportion p. Son écart-type est $\sigma = \sqrt{\frac{pq}{n}}$. La probabilité qu'une observation aussi 'mauvaise' que f se réalise, lorsque f est en dehors des limites $p - u_{1-\alpha/2}$ et $p + u_{1-\alpha/2}$ est inférieure ou égale à α.
Ceci se traduit par l'équation :

$$(f-p)^2 = \frac{u_{1-\alpha/2}p(1-p)}{n}$$

forme quadratique en p. La solution est alors :

$$p = \frac{f + u^2_{1-\alpha/2}/2n}{1 + u^2_{1-\alpha/2}/n} \pm \frac{\sqrt{f(1-f)(u^2_{1-\alpha/2}/n) + (u^2_{1-\alpha/2}/n)^2/4}}{1 + (u^2_{1-\alpha/2}/n)}$$

ce qui s'écrit encore de la façon suivante :

$$\frac{2nf + u^2_{1-\alpha/2} \pm u_{1-\alpha/2}\sqrt{u^2_{1-\alpha/2} + 4nf(1-f)}}{2(n + u^2_{1-\alpha/2})}$$

```
ualpha <- qnorm(0.975)
ualpha2 <- qnorm(0.975)^2
(2*359*pgauche+ualpha2-ualpha*sqrt(ualpha2+(4*359*pgauche*(1-pgauche))))/(2*(359+ualpha2))
```

```
[1] 0.3355569
```

```
(2*359*pgauche+ualpha2+ualpha*sqrt(ualpha2+(4*359*pgauche*(1-pgauche))))/(2*(359+ualpha2))
```

```
[1] 0.435693
```

Voilà un problème réglé. Mais revenons au test d'adéquation de la proportion observée à la proportion théorique. Le résultat observé par Hopkins infirme-t-il notre hypothèse de non latéralisation des grands singes ?
Les tests sont fondés sur le schéma suivant :

Hypothèse H_0	vraie	fausse
acceptée	$1-\alpha$	β
rejetée	α	$1-\beta$

Le risque de première espèce α est la probabilité de rejeter H_0 quand H_0 est vraie.
Le risque de deuxième espèce β est la probabilité d'accepter H_0 quand H_0 est fausse.
La puissance d'un test est la probabilité de rejeter H_0 alors que H_0 est fausse. C'est $1-\beta$.
Les résultats du test de comparaison d'une proportion observée à une proportion théorique sont calculés.

```
(res <- prop.test(138, 359, p=0.5, correct=FALSE))
```

```
        1-sample proportions test without continuity correction
data:  138 out of 359, null probability 0.5
X-squared = 19.189, df = 1, p-value = 1.184e-05
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3355569 0.4356930
sample estimates:
        p
0.3844011
```

1.2.1 Risque de première espèce

L'interprétation du test peut se faire autour de deux raisonnements.

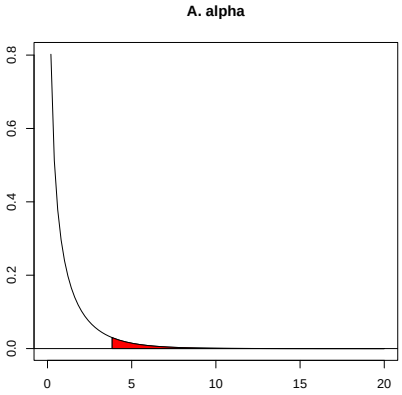
1. Raisonnement A.
   On fixe le risque de première espèce. Par exemple, $\alpha = 0.05$. On calcule la valeur du chi-deux à un degré de liberté associée à α.

   ```
   qchisq(0.95,1)
   ```

   ```
   [1] 3.841459
   ```

Puis on compare cette valeur théorique à la valeur calculée par le test 19.1894.

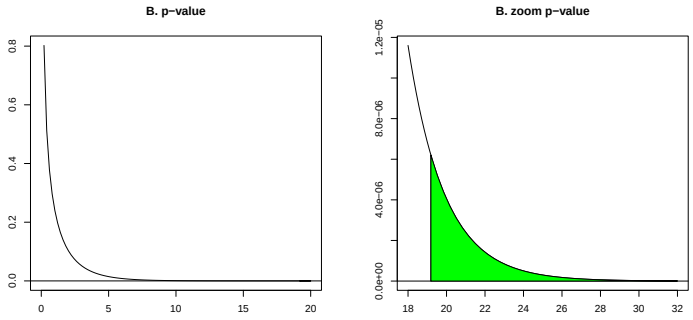
```
valx <- seq(0,20, le=100)
plot(valx,dchisq(valx,1),type="l",xlab="",ylab="", main="A. alpha")
abline(h=0)
segments(qchisq(0.95,1),0,qchisq(0.95,1),dchisq(qchisq(0.95,1),1))
intsup <- seq(qchisq(0.95,1),20,le=40)
axpxsup <- c(qchisq(0.95,1),intsup,20)
axpysup <- c(0,dchisq(intsup,1),0)
polygon(axpxsup,axpysup,col="red")
```



2. Raisonnement B.
   On calcule la probabilité de dépasser la valeur calculée à partir de l'échantillon (dite p-value) et on compare cette valeur au risque α

```
1-pchisq(19.1894,1)
```

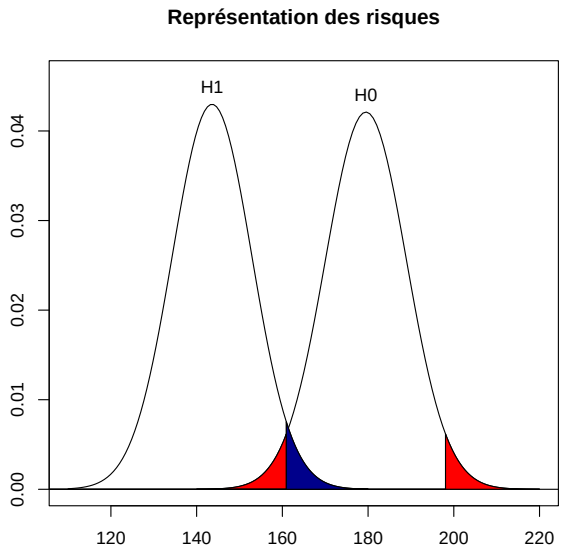
```
[1] 1.183689e-05
```



1.2.2 Risque de deuxième espèce et puissance

On ne peut le calculer que si l'on connait l'hypothèse alternative.
H_0 : $p = 0.50$
H_1 : $p = 0.40$

```
valx <- seq(110,220, le=30)
valy <- seq(0,0.046,le=30)
moynorm1 <- 359*0.40
varnorm1 <- 359*0.40*0.60
valx1 <- seq(110,180,le=100)
moynorm0 <- 359*0.50
varnorm0 <- 359*0.50*0.50
valx0 <- seq(140,220,le=100)
plot(valx,valy,type="n",xlab="",ylab="",main="Représentation des risques")
lines(valx1,dnorm(valx1,moynorm1,sqrt(varnorm1)))
abline(h=0)
lines(valx0,dnorm(valx0,moynorm0,sqrt(varnorm0)))
text(moynorm1,0.045,"H1")
text(moynorm0,0.044,"H0")
#
# Première espèce
#
x1 <-moynorm0-qnorm(0.975)*sqrt(varnorm0) ; x2 <- moynorm0+qnorm(0.975)*sqrt(varnorm0)
segments(x1,0,x1,dnorm(x1,moynorm1,sqrt(varnorm1)))
segments(x2,0,x2,dnorm(x2,moynorm0,sqrt(varnorm0)))
intinf <- seq(140,x1,le=30)
intsup <- seq(x2,220,le=30)
axpxinf <-c(140,intinf,x1)
axpyinf <- c(0,dnorm(intinf,moynorm0,sqrt(varnorm0)),0)
axpxsup <- c(x2,intsup,0)
axpysup <- c(0,dnorm(intsup,moynorm0,sqrt(varnorm0)),0)
polygon(axpxinf,axpyinf,col="red")
polygon(axpxsup,axpysup,col="red")
#
# Deuxième espèce
#
intsup2 <- seq(x1,180,le=30)
axpxsup2 <-c(x1,intsup2,x1)
axpysup2 <- c(0,dnorm(intsup2,moynorm1,sqrt(varnorm1)),0)
polygon(axpxsup2,axpysup2,col="darkblue")
```



Dans notre exemple, le risque de première espèce a été fixé à $\alpha = 0.05$. Le risque de deuxième espèce β vaut :

```
1-pnorm(x1, moynorm1, sqrt(varnorm1))
```
```
[1] 0.03093539
```

La puissance du test $1-\beta$ est :

```
pnorm(x1, moynorm1, sqrt(varnorm1))
```
```
[1] 0.9690646
```

Exercice
Prendre quelques valeurs théoriques, comprises entre 0.15 et 0.45, pour l'hypothèse alternative H_1. Calculer les puissances de test associées et les stocker dans un vecteur.
Représenter graphiquement les résultats et conclure.

1.3 Comparaison de deux proportions observées

Dans l'étude précédente, nous avons regroupé les 4 espèces de grands singes. Le calcul des pourcentages de gauchers montre que ce n'est pas une bonne idée car il y a, sur ces échantillons, de grandes variabilités entre les effectifs et entre les proportions.

```
gauche <- c(16,7,95,20)
droite <- c(7,23,163,28)
gauche+droite
```
```
[1]  23  30 258  48
```
```
gauche/(gauche+droite)
```
```
[1] 0.6956522 0.2333333 0.3682171 0.4166667
```

Nous allons nous intéresser à deux espèces : les gorilles (groupe 1) et les bonobos (groupe 4).

1.3.1 Rappel sur le test de comparaison de deux proportions

Soient p_1 et p_2 les proportions théoriques de gauchers chez les gorilles et les bonobos.
On extrait de chacune des populations des échantillons de taille n_1 et n_2. On observe le nombre de gauchers k_1 et k_2. On calcule les proportions f_1 et f_2 de gauchers dans ces échantillons.
H_0 : $p_1 = p_2$
H_1 : $p_1 \neq p_2$
La valeur de la statistique du test est $z = \dfrac{f_1 - f_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ avec $\hat{p} = \frac{k_1+k_2}{n_1+n_2}$
La statistique du test Z suit une loi normale centrée réduite.
La statistique du test Z^2 suit une loi du Chi-Deux à un degré de liberté.

Dans l'exemple traité, on obtient

```
prop.test(c(7,20),c(30,48), correct=FALSE)
```
```
        2-sample test for equality of proportions without continuity correction
data:  c(7, 20) out of c(30, 48)
X-squared = 2.7416, df = 1, p-value = 0.09777
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.38914467  0.02247801
sample estimates:
   prop 1    prop 2
0.2333333 0.4166667
```

Exercice.
1. Vérifier la concordance entre le rappel du test et les résultats de R.
2. Quelle conclusion donner à cette étude ?

1.3.2 Risques, puissance et effectif

Risques, puissance et taille d'échantillon sont liés. Le risque de deuxième espèce dépend de l'écart de proportion que l'on souhaite détecter. Si l'écart est faible, on a peu de chance de le détecter.
Lors de la mise en place d'une étude, il faut rechercher un compromis entre les risques, la puissance et l'effectif. C'est l'objet de la fonction `power.prop.test`

```
args(power.prop.test)
```

```
function (n = NULL, p1 = NULL, p2 = NULL, sig.level = 0.05, power = NULL,
    alternative = c("two.sided", "one.sided"), strict = FALSE,
    tol = .Machine$double.eps^0.25)
NULL
```

Elle permet de résoudre les question suivantes :

- ⋆ Connaissant les proportions p_1, p_2 et la puissance du test souhaitée, quel est l'effectif optimum dans chacun des groupes ? (Remarquez qu'ici, on ne traite que $n_1 = n_2 = n$)
  Exemple. On connait les deux proportions 0.23 et 0.42, combien d'individus doit-on échantillonner dans chaque population pour espérer détecter une différence entre les deux groupes avec une puissance de 0.85 ?

```
power.prop.test(p1=0.23,p2=0.42,power=0.85)
```

```
     Two-sample comparison of proportions power calculation
              n = 107.5577
             p1 = 0.23
             p2 = 0.42
      sig.level = 0.05
          power = 0.85
    alternative = two.sided

NOTE: n is number in *each* group
```

- ⋆ Connaissant la taille des échantillons n, les proportions p_1 et p_2, quelle est la puissance du test ?
  Exemple. On connait les deux proportions 0.23 et 0.42, pour une même taille $n = 30$. Quelle est la puissance associée à ce test ?

```
power.prop.test(n=30, p1=0.23,p2=0.42)
```

```
     Two-sample comparison of proportions power calculation
              n = 30
             p1 = 0.23
             p2 = 0.42
      sig.level = 0.05
          power = 0.3456426
    alternative = two.sided

NOTE: n is number in *each* group
```

⋆ Connaissant la taille des échantillons n, la proportion p_1 dans un échantillon, quelle devrait être la proportion p_2 dans l'échantillon pour détecter une différence entre les deux groupes avec une puissance donnée ?
Exemple. On connait la taille des échantillons $n = 30$, la proportion dans le premier échantillon $p_1 = 0.23$, quelle devrait être la proportion p_2 pour avoir une puissance de 0.55 ?

```
power.prop.test(n=30, p1=0.23,power=0.55)

     Two-sample comparison of proportions power calculation
              n = 30
             p1 = 0.23
             p2 = 0.487726
      sig.level = 0.05
          power = 0.55
    alternative = two.sided

NOTE: n is number in *each* group
```

1.3.3 Exercice.

1. Faire varier la puissance du test de 0.05 à 0.95 (de 0.05 en 0.05) et donner la taille des effectifs obtenue pour $p_1 = 0.23$, $p_2 = 0.42$ fixées. Conclure.
2. Faire varier la taille des échantillons de 10 à 100 (de 20 en 20) pour $p_1 = 0.23$, $p_2 = 0.42$ fixées et donner la puissance du test. Conclure.
3. Sachant que $n = 48$, $p_2 = 0.42$, quelle proportion devrait-on observée dans le premier échantillon pour obtenir une puissance du test de 0.55 ?

2 Robustesse autour des tests liés à la normalité

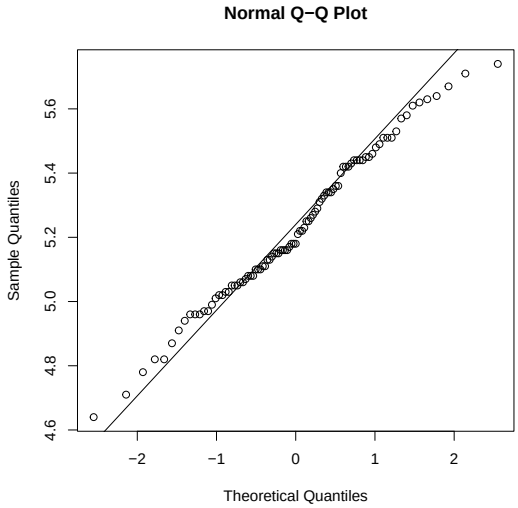
2.1 Situation : Longueur du thorax des punaises d'eau

Les données nous ont été communiquées par Ch. Klingenberg (`http://www.flywings.org.uk/`) lors d'un atelier consacré à la morphométrie (MNHN, 1996). On connait la longueur du thorax (`tho`), le segment de la première antenne (`ant`), la longueur du fémur de la patte postérieure (`mf`), la longueur du fémur de la patte du milieu (`hf`) ainsi que la provenance géographique de 93 punaises d'eau.

```
gerris <- read.table("http://pbil.univ-lyon1.fr/R/donnees/gerris.txt",h=T)
names(gerris)
```

```
[1] "tho"   "ant"   "mf"    "hf"    "group"
```

```
longthorax <- gerris$tho
qqnorm(longthorax)
qqline(longthorax)
```

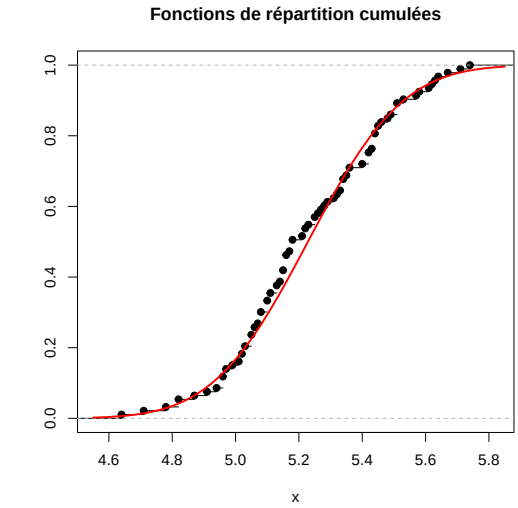


L'objectif est de vérifier la normalité de la longueur du thorax sur l'ensemble des punaises d'eau.
H_0 : La distribution observée suit une loi de Gauss
H_1 : La distribution observée ne suit pas une loi de Gauss.

2.2 Quelques Rappels sur les principaux tests

Cinq tests sont présentés : le très connu test de Shapiro-Wilks et quatre autres fondés sur la fonction de répartition empirique cumulée (Kolmogorov-Smirnov, Lilliefors, Cramer-Von Mises et Anderson Darling).

```
plot(ecdf(longthorax),ylab="",main="Fonctions de répartition cumulées")
valx <- seq(4.55,5.85, le=50)
lines(valx,pnorm(valx,mean=mean(longthorax),sd=sd(longthorax)), col="red", lwd=2)
```



Les n valeurs de la variable aléatoire X sont classées par ordre croissant : $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. On note $z_i = F(x_{(i)})$ les valeurs de la fonction de répartition empirique.

2.2.1 Test de Kolmogorov-Smirnov

On calcule les différences entre la fonction de répartition emprique et la fonction de répartition théorique (définie par H_0). La moyenne et la variance de la distribution théorique sont considérées comme connues.

$$D^+ = max_{1 \leq i \leq n} \left[\frac{i}{n} - z_i \right] \quad \text{et} \quad D^- = max_{1 \leq i \leq n} \left[z_i - \frac{i-1}{n} \right]$$

La statistique de Kolmogorov-Smirnov est $D = max(D^+, D^-)$.

```
ks.test(longthorax,"pnorm",mean=mean(longthorax), sd=sd(longthorax))
```

```
        One-sample Kolmogorov-Smirnov test
data:  longthorax
D = 0.086724, p-value = 0.4863
alternative hypothesis: two-sided
```

2.2.2 Test de Lilliefors

La statistique du test est identique à celle de Kolmogorov-Smirnov mais la moyenne et la variance de la distribution théorique sont considérées comme inconnues et estimées à partir des informations sur l'échantillon.

$$D^+ = max_{1 \leq i \leq n} \left[\frac{i}{n} - z_i \right] \quad \text{et} \quad D^- = max_{1 \leq i \leq n} \left[z_i - \frac{i-1}{n} \right]$$

La statistique de Lilliefors est $D = max(D^+, D^-)$.

```
library(nortest)
lillie.test(longthorax)
```

```
        Lilliefors (Kolmogorov-Smirnov) normality test
data:  longthorax
D = 0.086724, p-value = 0.08111
```

2.2.3 Test de Cramer- Von Mises

La statistique du test est :

$$W^2 = \sum_{i=1}^{n} \left(z_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}$$

```
cvm.test(longthorax)
```

```
        Cramer-von Mises normality test
data:  longthorax
W = 0.08301, p-value = 0.1869
```

2.2.4 Test de Anderson-Darling

La statistique du test est :

$$A^2 = -\frac{\sum_{i=1}^{n}(2i-1)\left[ln(z_i) + ln(1-z_{n+1-i})\right]}{n} - n$$

```
ad.test(longthorax)
```

```
        Anderson-Darling normality test
data:  longthorax
A = 0.43943, p-value = 0.2867
```

2.2.5 Test de Shapiro-Wilks

La statistique du test est :

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

où $[n/2]$ est la partie entière du rapport $n/2$ et $x_{(i)}$ une valeur de la distribution X ordonnée.
a_i représente une constante générée à partir de la moyenne et de la matrice de variance-covariance des quantiles d'un échantillon de taille n suivant une loi normale [5].

```
shapiro.test(longthorax)
```

```
        Shapiro-Wilk normality test
data:  longthorax
W = 0.98767, p-value = 0.5374
```

2.3 Simulations

L'idée est de réaliser des mélanges de lois normales et de regarder les résultats des tests de normalité (cf tdr22.pdf 'Graphes quantiles-quantiles').

Avant de commencer les simulations, bien s'approprier le contenu de la fonction `simulmixnor`.

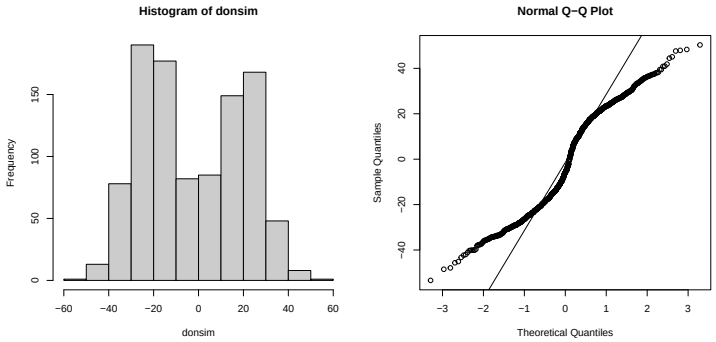
```
simulmixnor <- function(n = 100, p = 0.5, m1 = -1, sd1 = 1, m2 = 2,
sd2 = 2) {
n1 <- rbinom(1, n, p)
x1 <- rnorm(n1, m = m1, sd = sd1)
x2 <- rnorm(n - n1, m = m2, sd = sd2)
c(x1, x2)
}
```

2.3.1 Variances identiques dans les deux distributions

Dans ce paragraphe, on se place dans le cas où les variances des lois normales échantillonnées sont toutes égales à 100. Les simulations portent alors uniquement sur les moyennes. Pour chacune des situations proposées, réaliser les cinq tests de normalité et conclure.

Situation 1 : moyennes à -20 et 20.

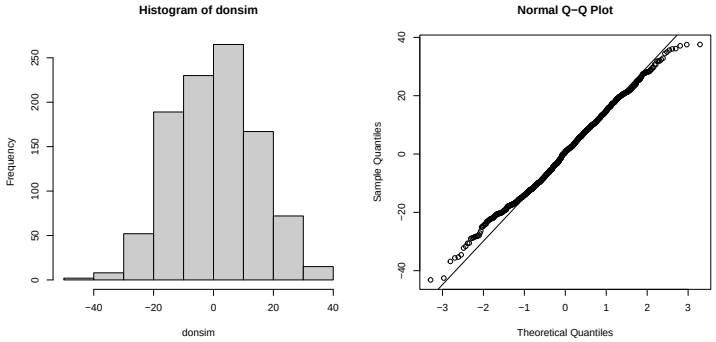
```
par(mfrow = c(1, 2))
donsim <- simulmixnor(1000, 0.5, -20, 10, 20, 10)
hist(donsim, col = grey(0.8))
qqnorm(donsim)
qqline(donsim)
```



Test	Valeur de la Statistique	p-value
Kolmogorov-Smirnov		
Lilliefors		
Cramer-Von Mises		
Anderson-Darling		
Shapiro-Wilks		

Situation 2 : moyennes à -10 et 10.

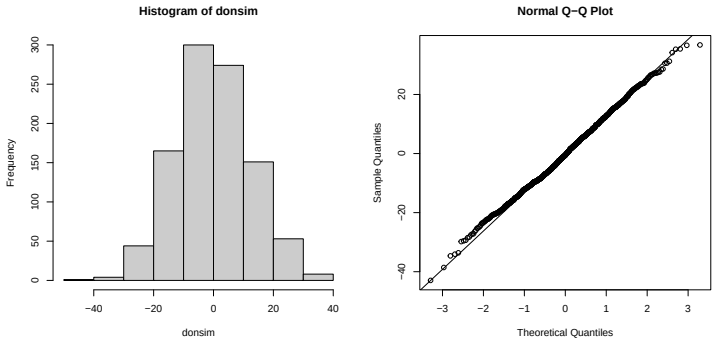
```
par(mfrow = c(1, 2))
donsim <- simulmixnor(1000, 0.5, -10, 10, 10, 10)
hist(donsim, col = grey(0.8))
qqnorm(donsim)
qqline(donsim)
```



Test	Valeur de la Statistique	p-value
Kolmogorov-Smirnov		
Lilliefors		
Cramer-Von Mises		
Anderson-Darling		
Shapiro-Wilks		

Situation 3 : moyennes à -8 et 8.

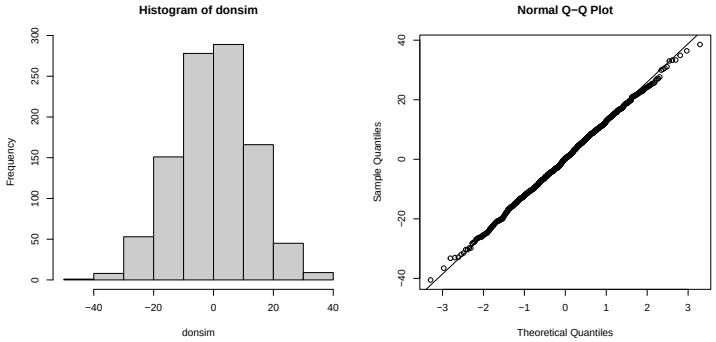
```
par(mfrow = c(1, 2))
donsim <- simulmixnor(1000, 0.5, -8, 10, 8, 10)
hist(donsim, col = grey(0.8))
qqnorm(donsim)
qqline(donsim)
```



Test	Valeur de la Statistique	p-value
Kolmogorov-Smirnov		
Lilliefors		
Cramer-Von Mises		
Anderson-Darling		
Shapiro-Wilks		

Situation 4 : moyennes à -7 et 7.

```
par(mfrow = c(1, 2))
donsim <- simulmixnor(1000, 0.5, -7, 10, 7, 10)
hist(donsim, col = grey(0.8))
qqnorm(donsim)
qqline(donsim)
```



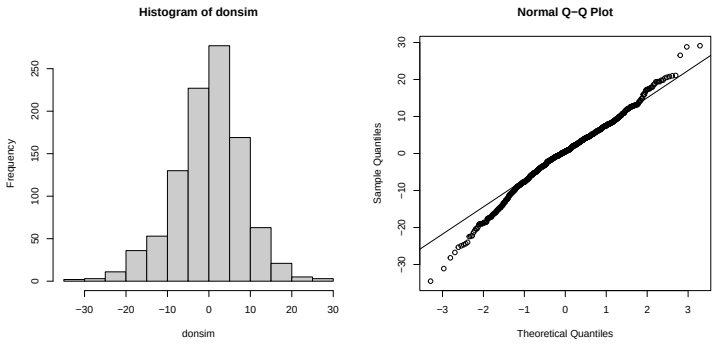
Test	Valeur de la Statistique	p-value
Kolmogorov-Smirnov		
Lilliefors		
Cramer-Von Mises		
Anderson-Darling		
Shapiro-Wilks		

2.3.2 Variances différentes dans les deux distributions

Dans ce paragraphe, on se place dans le cas où les moyennes des deux lois normales échantillonnées sont fixées à -2 et 2. Les simulations portent alors uniquement sur les variances. Pour chacune des situations proposées, réaliser les cinq tests de normalité et conclure.

Situation 1 : écarts-type à 10 et 5.

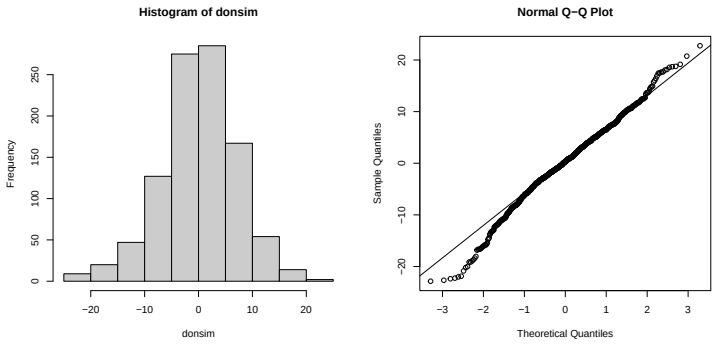
```
par(mfrow = c(1, 2))
donsim <- simulmixnor(1000, 0.5, -2, 10, 2, 5)
hist(donsim, col = grey(0.8))
qqnorm(donsim)
qqline(donsim)
```



Test	Valeur de la Statistique	p-value
Kolmogorov-Smirnov		
Lilliefors		
Cramer-Von Mises		
Anderson-Darling		
Shapiro-Wilks		

Situation 2 : écarts-type à 8 et 5.

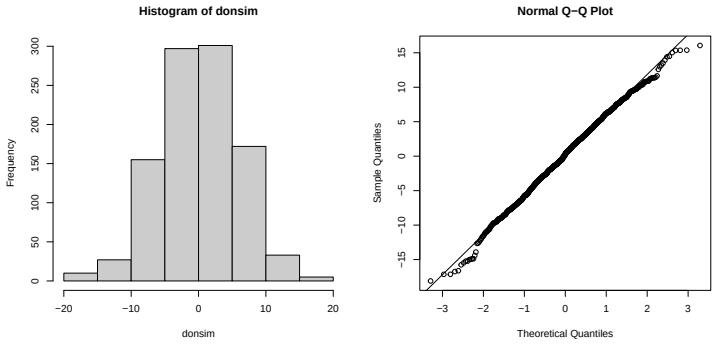
```
par(mfrow = c(1, 2))
donsim <- simulmixnor(1000, 0.5, -2, 8, 2, 5)
hist(donsim, col = grey(0.8))
qqnorm(donsim)
qqline(donsim)
```



Test	Valeur de la Statistique	p-value
Kolmogorov-Smirnov		
Lilliefors		
Cramer-Von Mises		
Anderson-Darling		
Shapiro-Wilks		

Situation 3 : écarts-type à 6 et 5.

```
par(mfrow = c(1, 2))
donsim <- simulmixnor(1000, 0.5, -2, 6, 2, 5)
hist(donsim, col = grey(0.8))
qqnorm(donsim)
qqline(donsim)
```



Test	Valeur de la Statistique	p-value
Kolmogorov-Smirnov		
Lilliefors		
Cramer-Von Mises		
Anderson-Darling		
Shapiro-Wilks		

2.3.3 Exercice.

Réaliser des simulations en faisant varier effectifs et variances. Qu'en est-il de la robustesse ? c'est-à-dire de la violation de l'hypothèse d'homoscédasticité et de la fluctuation des effectifs ai sein de chacun des groupes.

2.4 Commentaire général [7]

Lorsque la fonction de répartition théorique est continue et ses paramètres connus, les tests basés sur la fonction de répartition cumulée sont les plus puissants. Dans le cas où la moyenne et la variance de la loi normale sont inconnues, le test de Kolmogorov-Smirnov est le plus pauvre de tous. Les tests de Anderson-Darling et de Cramer-Von Mises sont très bons. Le test de Shapiro-Wilks lorsque $n > 50$ est le plus compétitif.
Mais il faut rester prudent. En effet, de faibles améliorations sur les statistiques

des tests existants peuvent faire varier les puissances selon les hypothèses alternatives proposées. Tout dépend du coefficient d'asymétrie (en anglais, `skewness` que l'on trouve par exemple dans la librairie `e1071`) et des queues de distribution.

3 Robustesse et tests liés aux comparaisons de variances

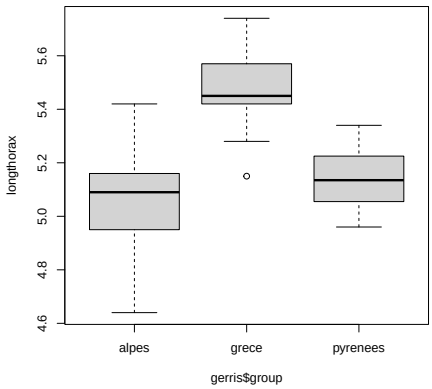
3.1 Situation : Longueur du thorax des punaises d'eau en fonction des zones géographiques

En reprenant les données précédentes, on constate que les punaises proviennent de trois zones géographiques : les Alpes, la Grèce, les Pyrénées.

```
summary(gerris$group)
```

```
  Length     Class      Mode
      93 character character
```

```
boxplot(longthorax~gerris$group)
```



L'objectif est de vérifier l'homoscédasticité de la longueur du thorax en fonction des trois zones géographiques.
H_0 Les variances sont identiques : $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$.
H_1 Au moins une des variances est différente des autres.
On note g le nombre de groupes, n le nombre total d'individus, n_j le nombre d'individus dans le groupe j.

$$n = \sum_{j=1}^{g} n_j$$

3.2 Quelques Rappels sur les principaux tests

3.2.1 Test de Bartlett

On note $\widehat{\sigma^2}$ la variance totale estimée et $\widehat{\sigma_j^2}$ les variances intra groupes.

La statistique de Bartlett est :

$$\chi^2 = \frac{(n-g)ln\widehat{\sigma^2} - \sum_{j=1}^{g}(n_j-1)ln\widehat{\sigma_j^2}}{1+\frac{1}{3(g-1)}\left(\sum_{j=1}^{g}\frac{1}{n_j-1}-\frac{1}{n-g}\right)}$$

```
bartlett.test(longthorax, gerris$group)
```

```
        Bartlett test of homogeneity of variances
data:  longthorax and gerris$group
Bartlett's K-squared = 6.5415, df = 2, p-value = 0.03798
```

3.2.2 Test de Levene

On note $z_{ij} = |y_{ij} - \bar{y}_{i\bullet}|$.
La statistique de Levene est :

$$W = \frac{(n-g)\sum_{j=1}^{g} n_j \left(z_j - \bar{z}_{\bullet\bullet}\right)^2}{(g-1)\sum_{j=1}^{g}\sum_{i=1}^{n_j}\left(z_{ij} - \bar{z}_j\right)^2}$$

```
library(car)
leveneTest(longthorax, gerris$group)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  2.1674 0.1204
      90
```

3.3 Simulations

On considère la variable longueur du thorax comme gaussienne. On fixe les moyennes de chacun des groupes à 5, 5.5 et 5.1. Pour chaque situation, réaliser les deux tests de variance et tirer une conclusion globale.

3.3.1 Effectifs identiques dans chacun des groupes

Situation 1 : Petits effectifs à $n = 10$

```
n <- 10
var1 <- 0.03
var2 <- 0.03
var3 <- 0.03
groupe <- gl(3,n)
lt1 <- rnorm(n, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

```
n <- 10
var1 <- 0.03
var2 <- 0.02
var3 <- 0.01
groupe <- gl(3,n)
lt1 <- rnorm(n, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

Situation 2 : Grands effectifs à $n = 30$

```
n <- 30
var1 <- 0.03
var2 <- 0.03
var3 <- 0.03
groupe <- gl(3,n)
lt1 <- rnorm(n, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

```
n <- 30
var1 <- 0.03
var2 <- 0.02
var3 <- 0.01
groupe <- gl(3,n)
lt1 <- rnorm(n, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

3.3.2 Effectifs différents dans chacun des groupes

Situation 1 : Petits effectifs, variances constantes.

```
valn <- 5:10
n1 <- sample(valn,1)
n2 <- sample(valn,1)
n3 <- sample(valn,1)
n1;n2;n3
```

```
[1] 9
```

```
[1] 5
```

```
[1] 10
```

```
var1 <- 0.03
var2 <- 0.03
var3 <- 0.03
groupe <- factor(rep(1:3,c(n1,n2,n3)))
lt1 <- rnorm(n1, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n2, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n3, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

Situation 2 : Grands effectifs, variances constantes.

```
valn <- 30:50
n1 <- sample(valn,1)
n2 <- sample(valn,1)
n3 <- sample(valn,1)
n1;n2;n3
```

```
[1] 32
```

```
[1] 38
```

```
[1] 38
```

```
var1 <- 0.03
var2 <- 0.03
var3 <- 0.03
groupe <- factor(rep(1:3,c(n1,n2,n3)))
lt1 <- rnorm(n1, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n2, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n3, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

Situation 3 : Petits effectifs, variances différentes.

```
valn <- 5:10
n1 <- sample(valn,1)
n2 <- sample(valn,1)
n3 <- sample(valn,1)
n1;n2;n3
```

```
[1] 7
```

```
[1] 9
```

```
[1] 10
```

```
var1 <- 0.03
var2 <- 0.02
var3 <- 0.01
groupe <- factor(rep(1:3,c(n1,n2,n3)))
lt1 <- rnorm(n1, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n2, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n3, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

Situation 4 : Grands effectifs, variances différentes.

```
valn <- 3:50
n1 <- sample(valn,1)
n2 <- sample(valn,1)
n3 <- sample(valn,1)
n1;n2;n3
```

```
[1] 24
```

```
[1] 33
```

```
[1] 16
```

```
var1 <- 0.03
var2 <- 0.02
var3 <- 0.01
groupe <- factor(rep(1:3,c(n1,n2,n3)))
lt1 <- rnorm(n1, mean=5,sd=sqrt(var1))
lt2 <- rnorm(n2, mean=5.5, sd=sqrt(var2))
lt3 <- rnorm(n3, mean=5.1, sd=sqrt(var3))
lt <- c(lt1,lt2,lt3)
```

Test	Valeur de la Statistique	p-value
Bartlett		
Levene		

3.3.3 Exercice.

Réaliser d'autres simulations en faisant varier les effectifs et les variances. Qu'en est-il de la robustesse ?

3.4 Commentaire général [1] [6]

Il existe trois formes pour le test de Levène. La première est basée sur la moyenne (définition ci-dessus), la seconde sur la médiane, la troisième sur le dixième percentile. Quand il y a égalité entre les deux variances de deux populations, sous l'hypothèse de normalité, le test de Levène basée sur la médiane se révèle conservateur de H_0 pour les petits échantillons.
Lorsque les queues des distributions sont grandes, le test classique de Fisher (rapport des deux variances) rejettent trop souvent l'hypothèse nulle. Des trois procédures de Levène, celle basée sur le dixième percentile se révèle la plus robuste.
Si la distribution est asymétrique, il vaut mieux utiliser le test de Levène sur la médiane.

Quoi qu'il en soit, la discussion ci-dessus reste ouverte car le choix dépend de la taille de la distribution, de la forme des queues de la distribution et de l'asymétrie. Il n'y a pas de test optimum donc pas de gagnant.

4 Exercice ouvert

Dans l'étude sur les punaises d'eau, choisissez une des variables morphologiques ainsi que deux groupes géographiques.

1. Vérifiez la normalité ou non de la variable dans chacun des deux groupes.
2. Vérifiez l'hypothèse d'homoscédasticité ou non c'est-à-dire l'égalité entre les variances de la variable pour les deux groupes.
3. Réalisez une test de comparaison de moyennes à l'aide du test de Student (`t.test`).
4. Il existe, pour ce test, une fonction permettant d'étudier la puissance : `power.t.test`. Les paramètres de cette fonction sont l'effectif `n` commun aux deux échantillons, l'écart vrai entre les deux moyennes des populations noté `delta` ainsi que l'écart-type associé à cette différence `sd`, le risque de première espèce α et la puissance du test $1 - \beta$. En vous inspirant des résultats de l'étude concrète, réalisez des simulations (comme nous l'avons fait pour la comparaison de proportions) et discutez les résultats obtenus.

References

[1] M.B. Brown and A.B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974.

[2] W.D. Hopkins. Hand preferences for a coordinated bimanual task in 110 chimpanzees (*Pan troglodytes*): croos-sectional analysis. *Journal of Comparative Psychology*, 109(3):291–297, 1995.

[3] W.D. Hopkins, K.A. Phillips, A. Bania, S.E. Calcutt, M. Gardner, J. Russell, J. Schaeffer, E.V. Lonsdorf, S.R. Ross, and S.J. Schapiro. Hand preferences for coordinated bimanual actions in 777 great apes: implications for the evolution of handedness in hominins. *Journal of Human Evolution*, 60:605–611, 2011.

[4] R.G. Newcombe. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine*, 17:857–872, 1998.

[5] R. Rakotomala. Tests de normalité. techniques empiriques et tests statistiques. Université Lumière Lyon 2.

[6] L.P. Rivest. Bartlett's, cochran's and hartley's tests of variances are liberal when the underlying distribution is long-tailed. *Journal of the American Statistical Association*, 81(393):124–128, 1986.

[7] M.A. Stephens. Edf statistics for ggodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.

[8] E.B. Wilson. Probable inference, the law of succession and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.