

Initiation à la statistique avec R, code et compléments

chapitre 7

Frédéric Bertrand et Myriam Maumy-Bertrand

11 décembre 2018

```
#Chapitre 7
require(BioStatR)

## Loading required package: BioStatR

#page 299
gaz<-c(52.0,60.2,68.8,46.8,62.2,53.5,50.9,44.9,73.2,60.4,61.9,
      67.8,30.5,52.5,40.4,29.6,58.3,62.6,53.6,64.6,54.4,53.8,49.8,
      57.4,63.1,53.4,59.4,48.6,40.7,51.9)
shapiro.test(gaz)

##
## Shapiro-Wilk normality test
##
## data:  gaz
## W = 0.95988, p-value = 0.3077

length(gaz)

## [1] 30

#page 300
(z<-(sqrt(30)*(mean(gaz)-50))/10)

## [1] 2.322344

qnorm(0.95)

## [1] 1.644854

if(!("TeachingDemos" %in% rownames(installed.packages()))){
  install.packages("TeachingDemos")}

#page 301
library(TeachingDemos)
z.test(gaz,mu=50,sd=10,alternative="greater",conf.level=0.95)

##
## One Sample z-test
##
## data:  gaz
## z = 2.3223, n = 30.0000, Std. Dev. = 10.0000, Std. Dev. of the
## sample mean = 1.8257, p-value = 0.01011
## alternative hypothesis: true mean is greater than 50
## 95 percent confidence interval:
##  51.23692      Inf
## sample estimates:
## mean of gaz
##      54.24
```

```
#page 302
glycine<-subset(Mesures,subset=(Mesures$espece=="glycine blanche"))
shapiro.test(glycine$taille)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  glycine$taille
## W = 0.97977, p-value = 0.4906
```

```
#page 303
length(glycine$taille)
```

```
## [1] 54
```

```
t.test(glycine$taille,mu=15)
```

```
##
##  One Sample t-test
##
## data:  glycine$taille
## t = -0.50666, df = 53, p-value = 0.6145
## alternative hypothesis: true mean is not equal to 15
## 95 percent confidence interval:
##  13.87050 15.67395
## sample estimates:
## mean of x
##  14.77222
```

```
power.t.test(n=54,delta=mean(glycine$taille)-15,
sd=sd(glycine$taille),type="one.sample",alternative="two.sided")
```

```
##
##      One-sample t test power calculation
##
##              n = 54
##          delta = 0.2277778
##             sd = 3.303652
##      sig.level = 0.05
##         power = 0.07181315
##      alternative = two.sided
```

```
#page 303
power.t.test(power=.8,delta=mean(glycine$taille)-15,
sd=sd(glycine$taille),type="one.sample",alternative="two.sided")
```

```
##
##      One-sample t test power calculation
##
##              n = 1653.023
##          delta = 0.2277778
##             sd = 3.303652
##      sig.level = 0.05
##         power = 0.8
##      alternative = two.sided
```

```
#page 306
pesee<-c(2.53,1.51,1.52,1.44,4.32,2.36,2.41,2.06,1.57,1.68,
  3.09,0.54,2.32,0.19,2.66,2.20,1.04,1.02,0.74,1.01,
  0.35,2.42,2.66,1.11,0.56,1.75,1.51,3.80,2.22,2.88)
shapiro.test(pesee)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  pesee
## W = 0.97163, p-value = 0.5848
```

```
length(pesee)
```

```
## [1] 30
```

```
((length(pesee)-1)*var(pesee))/4
```

```
## [1] 7.135268
```

```
#page 307
qchisq(0.95,29)
```

```
## [1] 42.55697
```

```
library(TeachingDemos)
sigma.test(pesee,sigma=2,alternative="greater")
```

```
##
##  One sample Chi-squared test for variance
##
## data:  pesee
## X-squared = 7.1353, df = 29, p-value = 1
## alternative hypothesis: true variance is greater than 4
## 95 percent confidence interval:
##  0.6706556      Inf
## sample estimates:
## var of pesee
##  0.9841748
```

```
if(!("OneTwoSamples" %in% rownames(installed.packages()))){
  install.packages("OneTwoSamples")}
library(OneTwoSamples)
var_test1(pesee,sigma2=4)
```

```
##          var df   chisq2      P_value
## 1 0.9841748 29 7.135268 2.227029e-05
```

```
#page 309
binom.test(507,988,0.5)
```

```
##
##  Exact binomial test
##
## data:  507 and 988
## number of successes = 507, number of trials = 988, p-value =
## 0.4264
## alternative hypothesis: true probability of success is not equal to 0.5
```

```
## 95 percent confidence interval:
## 0.4814855 0.5447516
## sample estimates:
## probability of success
## 0.5131579
```

#page 316

```
pipit<-c(17.0,16.9,16.9,17.3,16.8,16.8,17.0,16.5,16.9,16.5,
17.0,17.0,16.8,17.0,16.9,17.0,17.0,17.3,16.8,17.1,16.9,16.8,
17.1,17.0,17.1,17.2,16.7,16.6,17.2,17.0,17.0)
fauvette<-c(16.0,16.1,16.3,16.5,16.2,15.2,15.6,15.6,16.6,16.0,
16.2,16.8,16.0,17.0,17.9,16.0,16.4,16.3,16.9,17.1,17.0,16.1,
16.5,16.5,16.1,16.5,17.9,16.5,16.7,16.8)
shapiro.test(pipit)
```

```
##
## Shapiro-Wilk normality test
##
## data: pipit
## W = 0.94442, p-value = 0.1094
```

```
length(pipit)
```

```
## [1] 31
```

```
shapiro.test(fauvette)
```

```
##
## Shapiro-Wilk normality test
##
## data: fauvette
## W = 0.94926, p-value = 0.1615
```

```
length(fauvette)
```

```
## [1] 30
```

#page 317

```
var.test(pipit,fauvette)
```

```
##
## F test to compare two variances
##
## data: pipit and fauvette
## F = 0.11195, num df = 30, denom df = 29, p-value = 4.767e-08
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.05350369 0.23314689
## sample estimates:
## ratio of variances
## 0.1119467
```

```
t.test(pipit,fauvette,var.equal=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: pipit and fauvette
```

```
## t = 4.3599, df = 35.224, p-value = 0.0001078
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.2647656 0.7259871
## sample estimates:
## mean of x mean of y
## 16.93871 16.44333
```

```
t.test(pipit,fauvette)
```

```
##
## Welch Two Sample t-test
##
## data: pipit and fauvette
## t = 4.3599, df = 35.224, p-value = 0.0001078
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.2647656 0.7259871
## sample estimates:
## mean of x mean of y
## 16.93871 16.44333
```

```
#page 324
```

```
#Problème 7.1
```

```
#2)
```

```
glycines<-subset(Mesures,subset=(Mesures$espece=="glycine violette"
|Mesures$espece=="glycine blanche"))
glycines$espece<-factor(glycines$espece)
tapply(glycines$taille,glycines$espece,summary)
```

```
## $`glycine blanche`
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.20  12.32   14.60   14.77   16.88   21.10
##
## $`glycine violette`
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.80  10.68   12.00   13.77   14.65   27.00
```

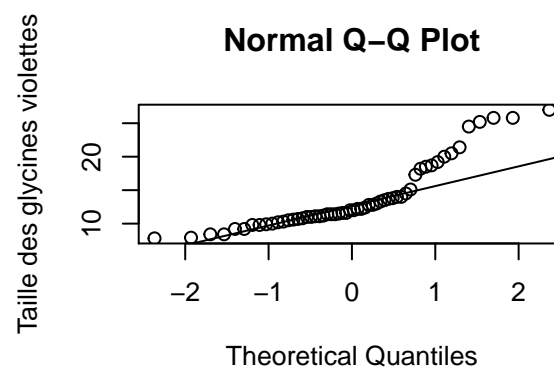
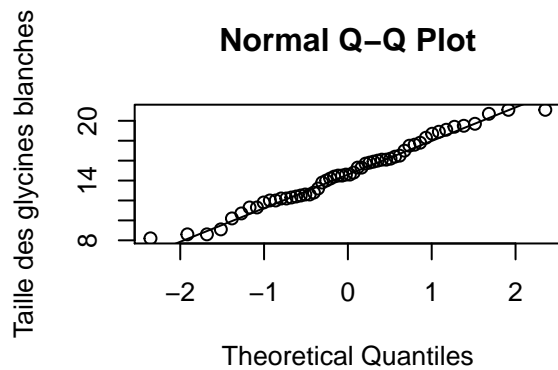
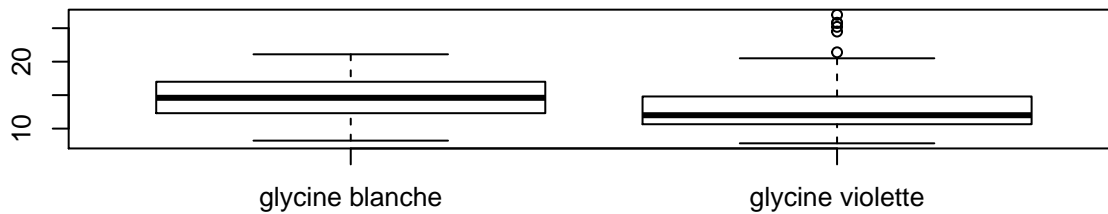
```
tapply(glycines$taille,glycines$espece,sd)
```

```
## glycine blanche glycine violette
##      3.303652      4.937306
```

```
#page 325
```

```
#4)
```

```
layout(matrix(c(1,2,1,3),nrow=2,ncol=2,byrow=F))
boxplot(taille~espece,data=glycines)
glycine_blanche<-glycines[glycines$espece=="glycine blanche",]
qqnorm(glycine_blanche$taille,ylab="Taille des glycines blanches")
qqline(glycine_blanche$taille)
glycine_violette<-glycines[glycines$espece=="glycine violette",]
qqnorm(glycine_violette$taille,ylab="Taille des glycines violettes")
qqline(glycine_violette$taille)
```



#Page 326

#7)

```
wilcox.test(taille~espece,data=glycines,conf.int=TRUE)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  taille by espece
## W = 1947.5, p-value = 0.009289
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  0.5000023 3.2999868
## sample estimates:
## difference in location
##          1.899959
```

#Page 329

#Exercice 7.1

#1)

```
jus_orange=c(8.2,9.4,9.6,9.7,10.0,14.5,15.2,16.1,17.6,21.5,14.0,13.8,
12.8,15.0,9.5,10.9,12.4,14.7,10.7,11.1,13.8,13.1,8.6,13.9,15.2,13.6,13.4,
12.3,15.2,11.2,19.6,7.8,14.1,12.5,14.1,17.6,13.5,12.4,12.6,14.6,15.5,11.6,
11.8,12.9,8.1,11.8,18.7,12.6,16.0,15.8,17.2,16.4,11.2,10.2,13.6,13.2,15.9,
9.8,8.8,12.0)
acide_ascorbique=c(4.2,5.2,5.8,6.4,7.0,7.3,10.1,11.2,11.3,11.5,7.1,9.8,
5.3,4.8,11.9,10.1,12.5,14.6,4.9,9.7,7.0,3.8,5.0,9.3,8.7,8.7,8.7,9.5,2.5,
6.6,13.6,6.6,9.4,12.1,13.1,4.1,12.1,8.8,7.0,7.5)
```

#2)

```
shapiro.test(jus_orange)
```

```
##
## Shapiro-Wilk normality test
```

```

##
## data:  jus_orange
## W = 0.98357, p-value = 0.5962
length(jus_orange)

## [1] 60
#Page 330
shapiro.test(acide_ascorbique)

##
##  Shapiro-Wilk normality test
##
## data:  acide_ascorbique
## W = 0.97781, p-value = 0.6087
length(acide_ascorbique)

## [1] 40
#4)
var.test(jus_orange,acide_ascorbique)

##
## F test to compare two variances
##
## data:  jus_orange and acide_ascorbique
## F = 0.93325, num df = 59, denom df = 39, p-value = 0.798
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5139606 1.6370265
## sample estimates:
## ratio of variances
##      0.9332472
#Page 331
t.test(jus_orange,acide_ascorbique,alternative="greater",var.equal=TRUE)

##
## Two Sample t-test
##
## data:  jus_orange and acide_ascorbique
## t = 7.8835, df = 98, p-value = 2.246e-12
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.771844      Inf
## sample estimates:
## mean of x mean of y
## 13.14833  8.37000
#Exercice 7.2
#1)
avnt<-c(15,18,17,20,21,18,17,15,19,16,19,17,19,15,14,16,21,20,21,18,17,17,
17,15,17,18,16,10,17,18,14,15,15,17,17,20,17)
aprs<-c(12,16,17,18,17,15,18,14,16,18,20,16,15,17,18,16,15,14,11,13,13,15,
14,15,19,14,16,14,14,15,19,19,16,19,15,17,16)
mode(avnt)

```

```

## [1] "numeric"
#Page 332
mode(aprs)

## [1] "numeric"
length(avnt)

## [1] 37
length(aprs)

## [1] 37
#2)
diff<-aprs-avnt
diff

## [1] -3 -2 0 -2 -4 -3 1 -1 -3 2 1 -1 -4 2 4 0 -6
## [18] -6 -10 -5 -4 -2 -3 0 2 -4 0 4 -3 -3 5 4 1 2
## [35] -2 -3 -1
#4)
shapiro.test(diff)

##
## Shapiro-Wilk normality test
##
## data: diff
## W = 0.97103, p-value = 0.4369
#Page 333
length(diff)

## [1] 37
#5)
t.test(diff)

##
## One Sample t-test
##
## data: diff
## t = -2.3931, df = 36, p-value = 0.02205
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -2.3467942 -0.1937463
## sample estimates:
## mean of x
## -1.27027
#Problème 7.1
#page 334
glycines<-subset(Mesures,subset=(Mesures$espece=="glycine violette"|Mesures$espece=="glycine blanche"))
glycines$espece<-factor(glycines$espece)
#2)
tapply(glycines$taille,glycines$espece,summary)

## $`glycine blanche`

```



```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      8.20  12.32   14.60   14.77  16.88   21.10
##
## $`glycine violette`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      7.80  10.68   12.00   13.77  14.65   27.00
```

```
tapply(glycines$taille,glycines$espece,sd)
```

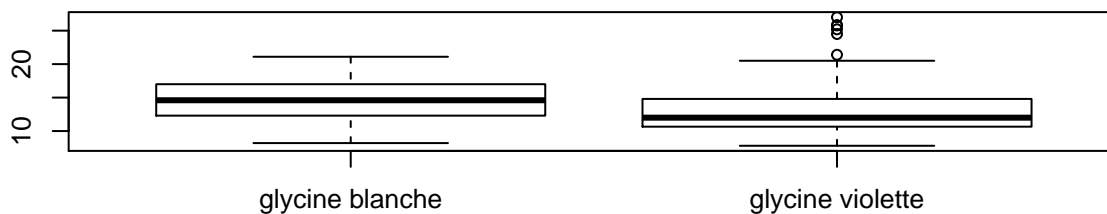
```
##      glycine blanche glycine violette
##           3.303652           4.937306
```

```
#page 335
```

```
#4)
```

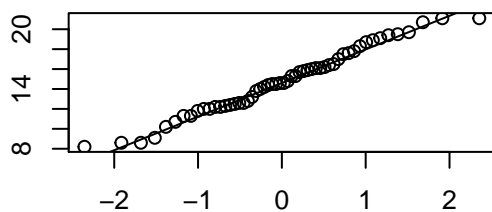
```
layout(matrix(c(1,2,1,3),nrow=2,ncol=2,byrow=F))
boxplot(taille~espece,data=glycines,main="Boîtes à moustaches")
glycine_blanche<-glycines[glycines$espece=="glycine blanche",]
qqnorm(glycine_blanche$taille,ylab="Taille des glycines blanches")
qqline(glycine_blanche$taille)
glycine_violette<-glycines[glycines$espece=="glycine violette",]
qqnorm(glycine_violette$taille,ylab="Taille des glycines violettes")
qqline(glycine_violette$taille)
```

Boîtes à moustaches



Taille des glycines blanches

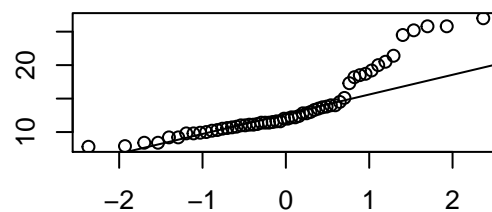
Normal Q-Q Plot



Theoretical Quantiles

Taille des glycines violettes

Normal Q-Q Plot



Theoretical Quantiles

```
#page 336
```

```
#6)
```

```
tapply(glycines$taille,glycines$espece,shapiro.test)
```

```
## $`glycine blanche`
##
##      Shapiro-Wilk normality test
##
```

```

## data:  X[[i]]
## W = 0.97977, p-value = 0.4906
##
##
## `$glycine violette`
##
## Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.83629, p-value = 2.386e-06
tapply(glycines$taille,glycines$espece,length)

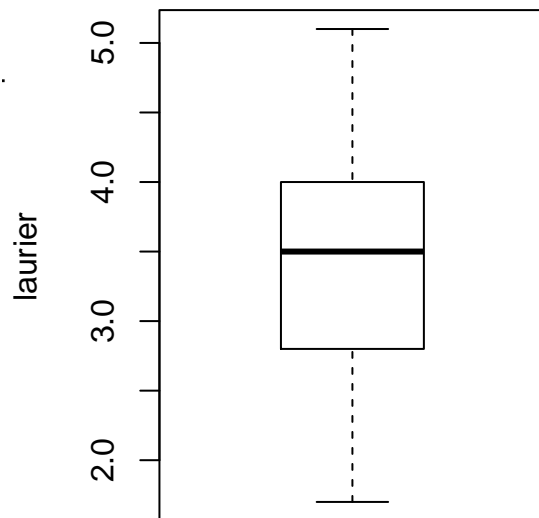
## glycine blanche glycine violette
##           54           56
#page 337
#8)
wilcox.test(taille~espece,data=glycines,conf.int=TRUE)

##
## Wilcoxon rank sum test with continuity correction
##
## data:  taille by espece
## W = 1947.5, p-value = 0.009289
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  0.5000023 3.2999868
## sample estimates:
## difference in location
##           1.899959
#Problème 7.2
#1)
lauriers<-subset(Mesures5,subset=(Mesures5$espece=="laurier rose"))
#2)
str(lauriers)

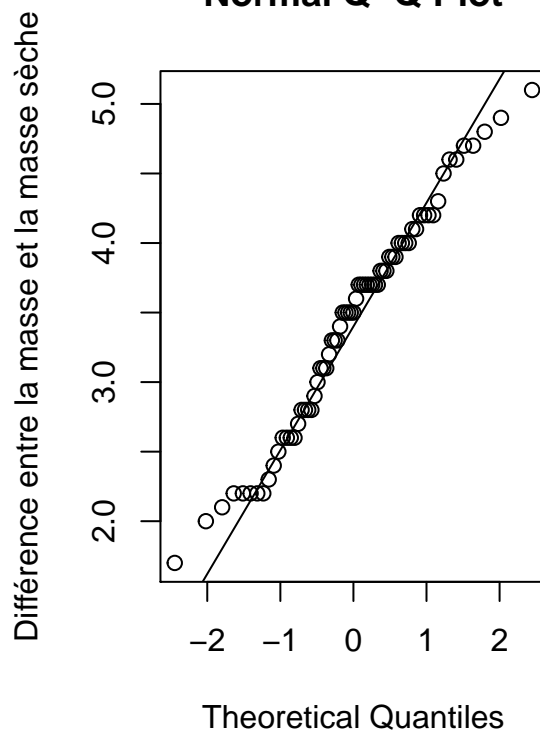
## 'data.frame':  72 obs. of  5 variables:
## $ masse      : num  4.9 6.2 4 3.3 4.8 5.6 4.5 6.3 4.2 3.9 ...
## $ taille     : num  15.3 15.9 15 11 15.1 15.6 15.3 18.4 14 12.6 ...
## $ graines    : int   NA NA NA NA NA NA NA NA NA NA ...
## $ masse_sec  : num   1.2 1.1 0.7 NA 0.8 1 0.7 1.6 0.8 1.1 ...
## $ espece     : Factor w/ 4 levels "bignone","glycine blanche",...: 4 4 4 4 4 4 4 4 4 4 ...
#page 338
#3)
la_masse<-lauriers$masse
la_masse_sec<-lauriers$masse_sec
diff_laurier<-(la_masse-la_masse_sec)
#4)
layout(matrix(c(1,2),nrow=1,ncol=2,byrow=F))
boxplot(diff_laurier,ylab="Différence entre la masse et la masse sèche pour une graine de
laurier",main="Boîte à moustaches")
abline(h=0, lty=2)
qqnorm(diff_laurier,ylab="Différence entre la masse et la masse sèche")
qqline(diff_laurier)

```

Boîte à moustaches



Normal Q-Q Plot



```
#page 339
#6)
shapiro.test(diff_laurier)

##
##  Shapiro-Wilk normality test
##
## data:  diff_laurier
## W = 0.97636, p-value = 0.2149
```

```
length(diff_laurier)
```

```
## [1] 72
```

```
#7)
t.test(diff_laurier)
```

```
##
##  One Sample t-test
##
## data:  diff_laurier
## t = 35.342, df = 68, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.240851 3.628714
## sample estimates:
## mean of x
##  3.434783
```

```
#page 340
#9)
wilcox.test(diff_laurier)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: diff_laurier
## V = 2415, p-value = 5.28e-13
## alternative hypothesis: true location is not equal to 0
t.test(lauriers$masse,lauriers$masse_sec,paired=TRUE)

##
## Paired t-test
##
## data: lauriers$masse and lauriers$masse_sec
## t = 35.342, df = 68, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 3.240851 3.628714
## sample estimates:
## mean of the differences
## 3.434783
wilcox.test(lauriers$masse,lauriers$masse_sec,paired=TRUE)

##
## Wilcoxon signed rank test with continuity correction
##
## data: lauriers$masse and lauriers$masse_sec
## V = 2415, p-value = 5.28e-13
## alternative hypothesis: true location shift is not equal to 0
```