Tests dhypothèses sur les données (Etudiants qui n'ont jamais suivi de MOOC) :

Résumé des données :

## 'data.frame': 19 obs. of 25 variables:

## $ X : int 1 2 3 4 5 6 7 8 9 10 ...

## $ a\_suivi : Factor w/ 1 level "Jamais": 1 1 1 1 1 1 1 1 1 1 ...

## $ performer\_travail\_projet\_1: Factor w/ 4 levels "D'accord","Désaccord",..: 3 1 3 3 3 4 1 4 4 1 ...

## $ motivation\_certificat\_1 : Factor w/ 4 levels "D'accord","Désaccord",..: 1 1 3 1 1 1 4 1 4 1 ...

## $ trouver\_emploi\_1 : Factor w/ 4 levels "D'accord","Désaccord",..: 4 2 3 2 2 3 1 3 4 3 ...

## $ mooc\_inclu\_etude\_1 : Factor w/ 2 levels "Non","Oui": 2 2 2 2 2 2 1 2 2 1 ...

## $ edt\_chargee\_1 : Factor w/ 2 levels "Non","Oui": 1 1 1 2 2 1 2 2 1 2 ...

## $ plateformes\_connues : Factor w/ 9 levels "Coursera","Coursera, Edx",..: 8 8 1 4 7 5 3 6 8 2 ...

## $ aide\_financiere\_1 : Factor w/ 2 levels "Non","Oui": 1 1 1 2 1 1 2 1 2 2 ...

## $ prix\_1 : int NA 40 NA 50 0 NA 20 20 5 5 ...

## $ outils\_exercices\_1 : Factor w/ 1 level "Oui": 1 1 1 1 1 1 1 1 1 1 ...

## $ sciences\_donnees\_1 : Factor w/ 3 levels "Inintéressant",..: 2 3 3 2 2 2 2 3 2 2 ...

## $ informatique\_1 : Factor w/ 3 levels "Inintéressant",..: 2 2 2 3 2 2 2 3 2 2 ...

## $ sciences\_physiques\_1 : Factor w/ 3 levels "Inintéressant",..: 3 2 2 3 2 2 2 2 2 1 ...

## $ business\_1 : Factor w/ 3 levels "Inintéressant",..: 2 2 2 2 2 2 1 3 2 3 ...

## $ arts\_1 : Factor w/ 3 levels "Inintéressant",..: 3 3 3 1 2 3 2 1 2 2 ...

## $ langues\_1 : Factor w/ 2 levels "Interessant",..: 1 1 1 1 1 1 1 2 1 1 ...

## $ sciences\_sociales\_1 : Factor w/ 3 levels "Inintéressant",..: 3 2 3 1 2 2 3 3 2 2 ...

## $ sexe : Factor w/ 2 levels "Femme","Homme": 2 2 2 2 1 2 2 2 2 2 ...

## $ pays : Factor w/ 6 levels "Autre","Côte d'ivoire",..: 2 3 6 1 3 3 3 3 2 4 ...

## $ age : int 20 23 32 28 21 20 22 20 35 21 ...

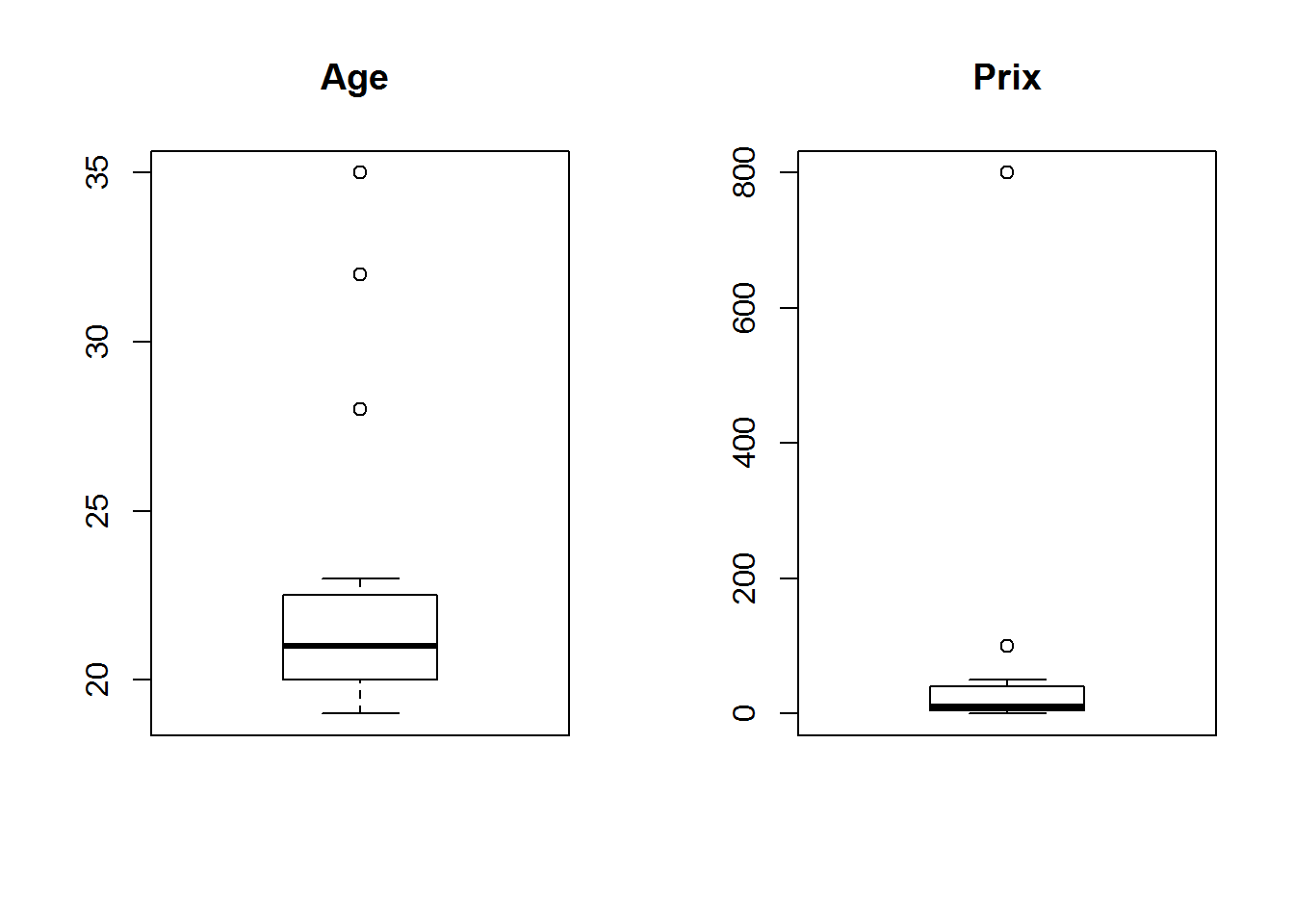
## $ est\_etudiant : Factor w/ 1 level "Oui": 1 1 1 1 1 1 1 1 1 1 ...

## $ niveau\_actuel : Factor w/ 5 levels "BAC +2","BAC +3",..: 4 4 5 4 2 1 3 2 4 3 ...

## $ formation : Factor w/ 4 levels "FIG","FIP","Master of science",..: 1 1 4 3 1 2 1 2 4 1 ...

## $ background : Factor w/ 3 levels "DUT","Prépa",..: 2 2 3 3 3 1 3 1 3 2 ...

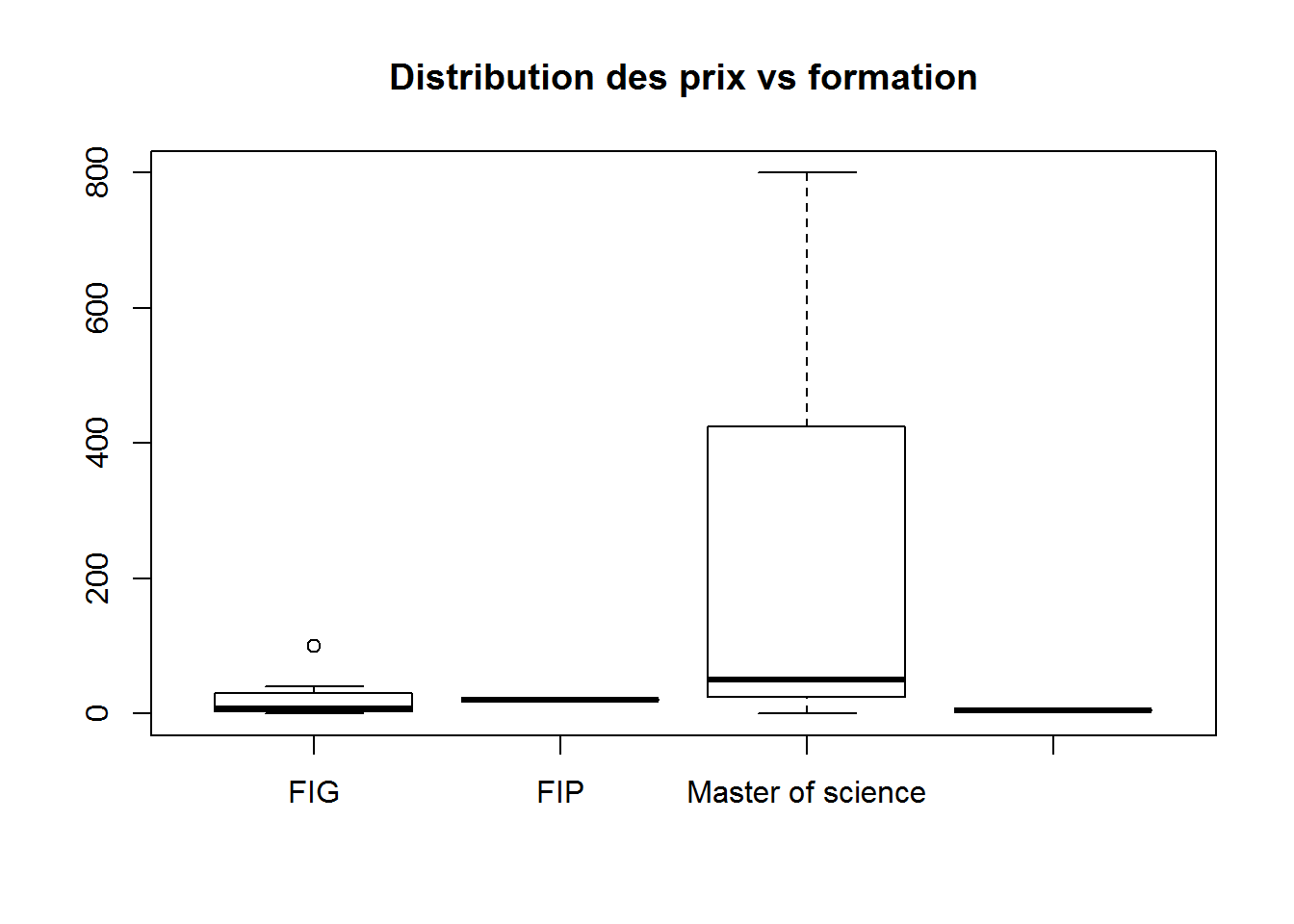
Boxplots des variables numériques (Age et Prix que les étudiants sont prêts à payer):

****

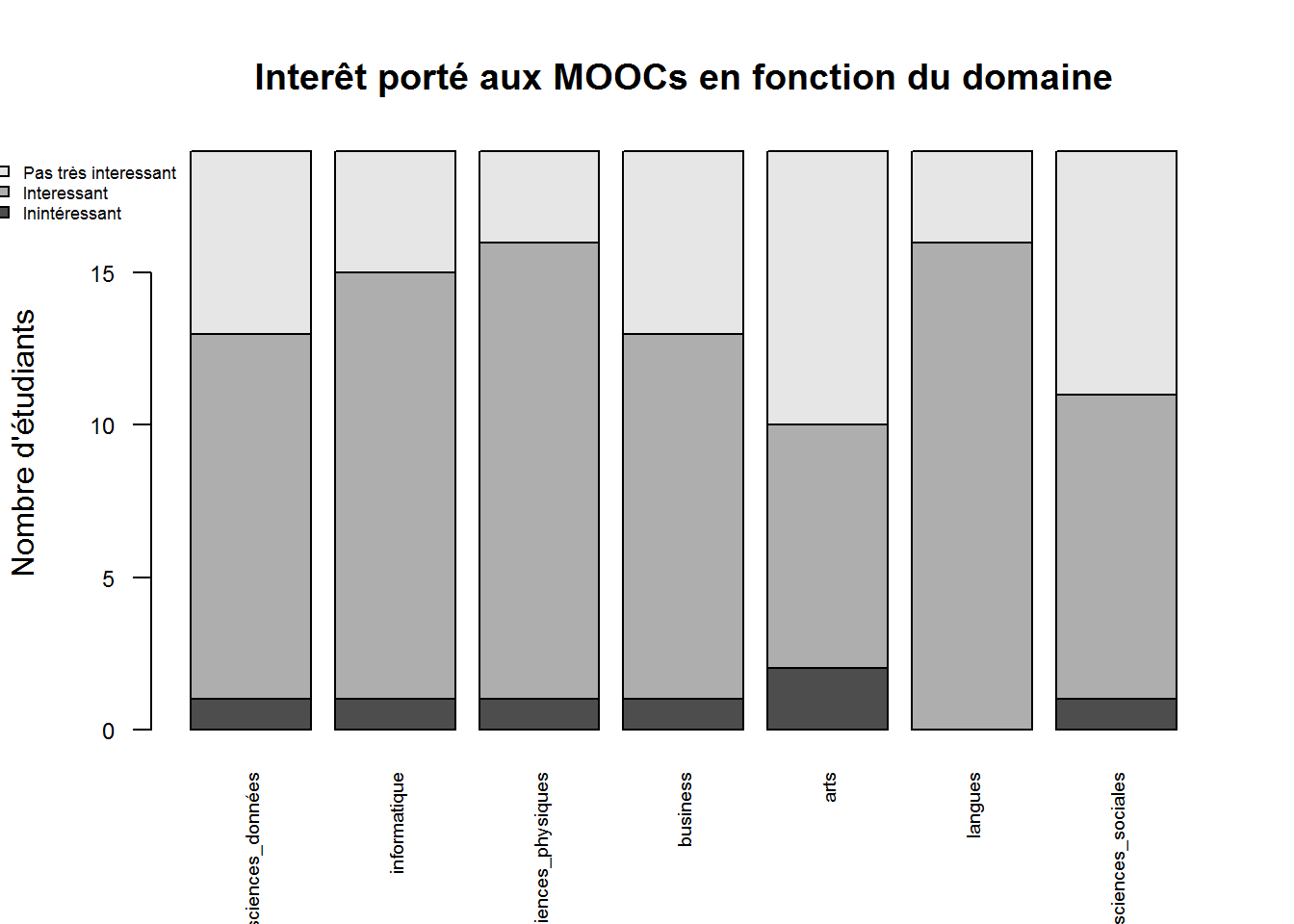
Commentaire : les deux variables présentent des outliers

Explication : Pour l’âge, la plupart des personnes ayant répondu au sondage sont des FIG (57%) ce qui signifie que l’âge est en général est inférieur à 25 ans

**Distribution du prix que les étudiants sont prêts à payer pour un MOOC :**

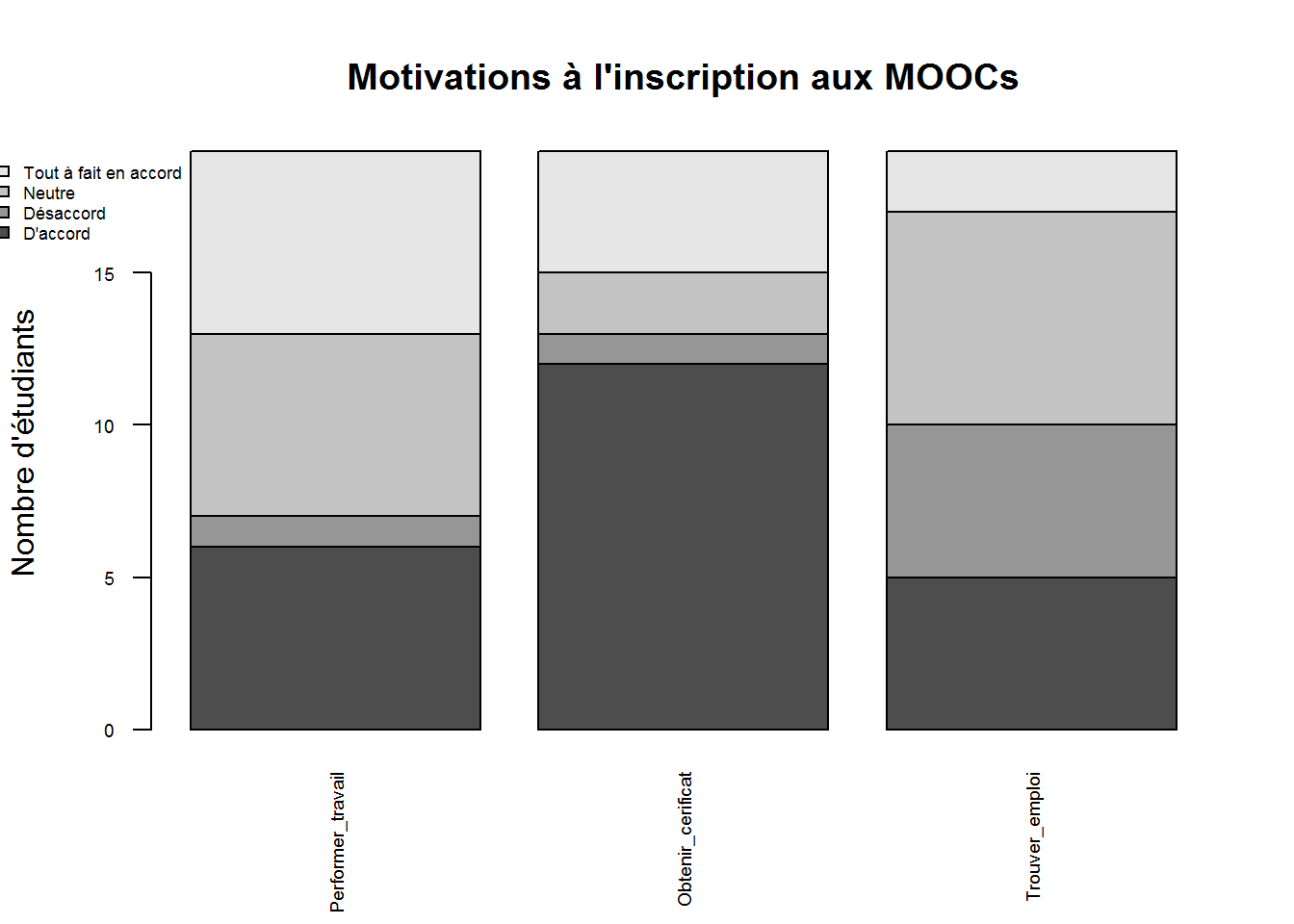
****

**Commentaire :**  En général, pour ceux qui n’ont jamais suivi de moocs, les étudiants en Master of science sont prêt à payer plus que les étudiants des autres formations

**Représentation des domaines auxquels l’intérêt des étudiants est porté :**

**Commentaire :**  Les domaines les plus intéressants sont les langues, sciences / info et business

**Les motivations pour faires les Moocs :**

****

**Commentaire :**  Le facteur le plus motivant pour faire un MOOC est l’obtention du certificat

**Hypothèse1**

1)-le facteur motivant (performer son travail et projet) ne dépend pas du sexe

table(jamais$performer\_travail\_projet\_1,jamais$sexe)

##

## Femme Homme

## D'accord 3 3

## Désaccord 0 1

## Neutre 3 3

## Tout à fait en accord 1 5

chisq\_sexe1 <- chisq.test(jamais$performer\_travail\_projet\_1,jamais$sexe)

## Warning in chisq.test(jamais$performer\_travail\_projet\_1, jamais$sexe): Chi-

## squared approximation may be incorrect

chisq\_sexe1

##

## Pearson's Chi-squared test

##

## data: jamais$performer\_travail\_projet\_1 and jamais$sexe

## X-squared = 2.5258, df = 3, p-value = 0.4706

Avec un risque de 10%, on garde l’hypothèse nulle, à savoir ce premier facteur motivant ne dépend pas du sexe

2)-le facteur motivant (motivation\_certificat\_1) ne dépend pas du sexe

table(jamais$motivation\_certificat\_1,jamais$sexe)

## Femme Homme

## D'accord 4 8

## Désaccord 1 0

## Neutre 1 1

## Tout à fait en accord 1 3

chisq\_sexe2 <- chisq.test(jamais$motivation\_certificat\_1,jamais$sexe)

## Warning in chisq.test(jamais$motivation\_certificat\_1, jamais$sexe): Chi-

## squared approximation may be incorrect

chisq\_sexe2

##

## Pearson's Chi-squared test

##

## data: jamais$motivation\_certificat\_1 and jamais$sexe

## X-squared = 2.1677, df = 3, p-value = 0.5383

Avec un risque de 10%, on garde l’hypothèse nulle, à savoir ce deuxième facteur motivant ne dépend pas du sexe

3)-le facteur motivant (trouver un emploi) ne dépend pas du sexe:

table(jamais$trouver\_emploi\_1,jamais$sexe)

##

## Femme Homme

## D'accord 3 2

## Désaccord 2 3

## Neutre 2 5

## Tout à fait en accord 0 2

chisq\_sexe3 <- chisq.test(jamais$trouver\_emploi\_1,jamais$sexe)

## Warning in chisq.test(jamais$trouver\_emploi\_1, jamais$sexe): Chi-squared

## approximation may be incorrect

chisq\_sexe3

##

## Pearson's Chi-squared test

##

## data: jamais$trouver\_emploi\_1 and jamais$sexe

## X-squared = 2.5463, df = 3, p-value = 0.467

Avec un risque de 10%, on garde l’hypothèse nulle, à savoir ce troisième facteur motivant ne dépend pas du sexe

**Commentaire**: On peut dire que les motivations à s’inscrire aux MOOCs ne dépend pas du sexe

**Hypothèse2:**

1)-le facteur motivant (performer son travail et projet) ne dépend pas de la formation

table(jamais$performer\_travail\_projet\_1,jamais$formation)

##

## FIG FIP Master of science Master spécialisé

## D'accord 6 0 0 0

## Désaccord 0 0 1 0

## Neutre 3 0 2 1

## Tout à fait en accord 2 2 1 1

chisq\_form1 <- chisq.test(jamais$performer\_travail\_projet\_1,jamais$formation)

## Warning in chisq.test(jamais$performer\_travail\_projet\_1, jamais$formation):

## Chi-squared approximation may be incorrect

chisq\_form1

## Pearson's Chi-squared test

##

## data: jamais$performer\_travail\_projet\_1 and jamais$formation

## X-squared = 13.314, df = 9, p-value = 0.1489

Avec un risque de 10%, on garde l’hypothèse nulle, à savoir ce premier facteur motivant ne dépend pas de la formation

2)-le facteur motivant (obtenir certificat) ne dépend pas de la formation

table(jamais$motivation\_certificat\_1,jamais$formation)

## FIG FIP Master of science Master spécialisé

## D'accord 7 2 3 0

## Désaccord 1 0 0 0

## Neutre 1 0 0 1

## Tout à fait en accord 2 0 1 1

chisq\_form2 <- chisq.test(jamais$motivation\_certificat\_1,jamais$formation)

## Warning in chisq.test(jamais$motivation\_certificat\_1, jamais$formation):

## Chi-squared approximation may be incorrect

chisq\_form2

##

## Pearson's Chi-squared test

##

## data: jamais$motivation\_certificat\_1 and jamais$formation

## X-squared = 7.4129, df = 9, p-value = 0.5942

Avec un risque de 10%, on garde l’hypothèse nulle, à savoir ce deuxième facteur motivant ne dépend pas de la formation

3)-le facteur motivant (trouver emploi) ne dépend pas de la formation

table(jamais$trouver\_emploi\_1,jamais$formation)

##

## FIG FIP Master of science Master spécialisé

## D'accord 4 0 1 0

## Désaccord 2 0 3 0

## Neutre 4 2 0 1

## Tout à fait en accord 1 0 0 1

chisq\_form3 <- chisq.test(jamais$trouver\_emploi\_1,jamais$formation)

## Warning in chisq.test(jamais$trouver\_emploi\_1, jamais$formation): Chi-

## squared approximation may be incorrect

chisq\_form3

##

## Pearson's Chi-squared test

##

## data: jamais$trouver\_emploi\_1 and jamais$formation

## X-squared = 13.756, df = 9, p-value = 0.1313

Avec un risque de 10%, on garde l’hypothèse nulle, à savoir ce deuxième facteur motivant ne dépend pas de la formation

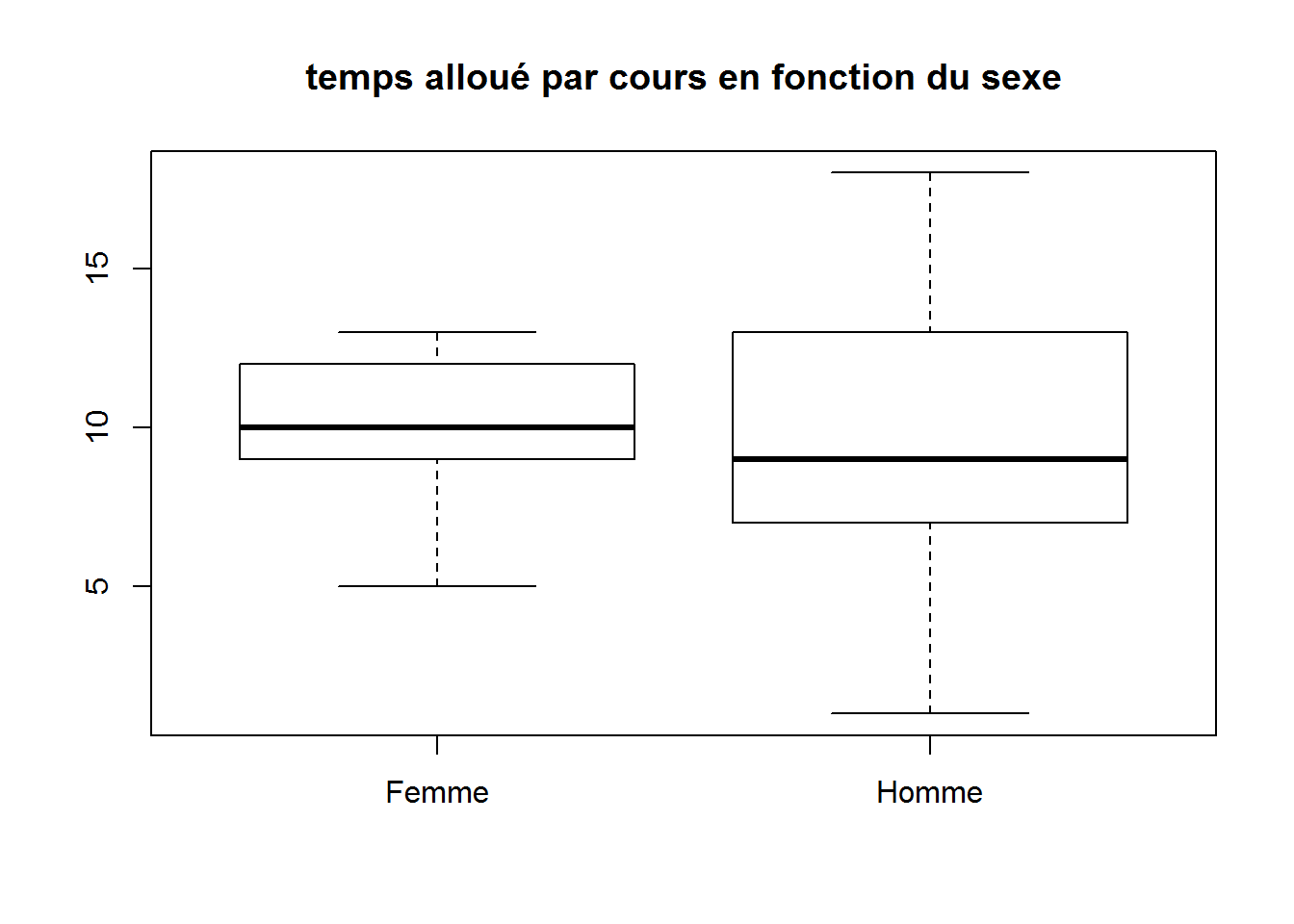
**Commentaire:** On peut dire que les motivations à s’inscrire aux moocs ne dépend pas de la formation

Tests d'hypothèses sur le temps alloué à un cours en moyenne (Etudiants ayant suivi des moocs) :

**Quelles sont les populations ayant passé plus de temps sur les moocs?**

On souhaite comparer le temps alloué en moyenne pour un mooc en fonction de la catégorie du lecteur. Le sexe a-t-il une influence sur ce temps ? Et qu’en est-il de la nationalité du sondé ? Ou encore de l’année d’étude de l’élève ?

**Filles Vs Garçons :**



temps\_sexe<-suivi[,c("temps\_alloue\_cours","sexe")]

res<-ddply(temps\_sexe,~sexe,summarise,mean=mean(temps\_alloue\_cours),sd=sd(temps\_alloue\_cours))

res

## sexe mean sd

## 1 Femme 9.781250 2.445989

## 2 Homme 9.534247 4.031412

Les variances ne sont pas égales, donc allons voir si les distributions sont normales ou non ?

**Pour les hommes :**

hommes<-filter(temps\_sexe,sexe=="Homme")

shapiro.test(hommes$temps\_alloue\_cours)

##

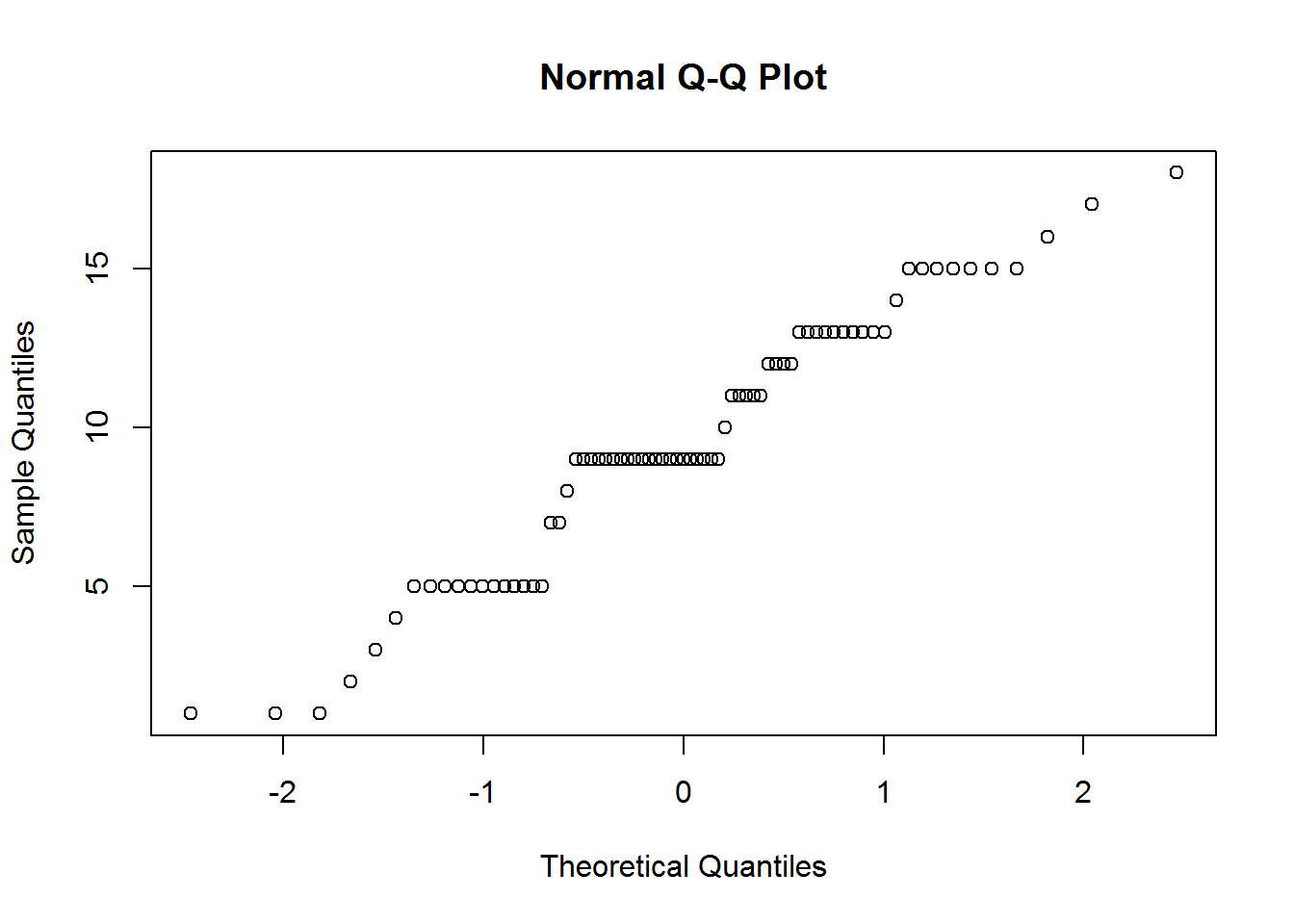
## Shapiro-Wilk normality test

##

## data: hommes$temps\_alloue\_cours

## W = 0.95853, p-value = 0.01713

qqnorm(hommes$temps\_alloue\_cours)

****

Pour cette p-value et risque de 10%, on conclut que La distribution du temps alloué en moyenne par cours des hommes n’est pas normale.

**Pour les femmes :**

femmes<-filter(temps\_sexe,sexe=="Femme")

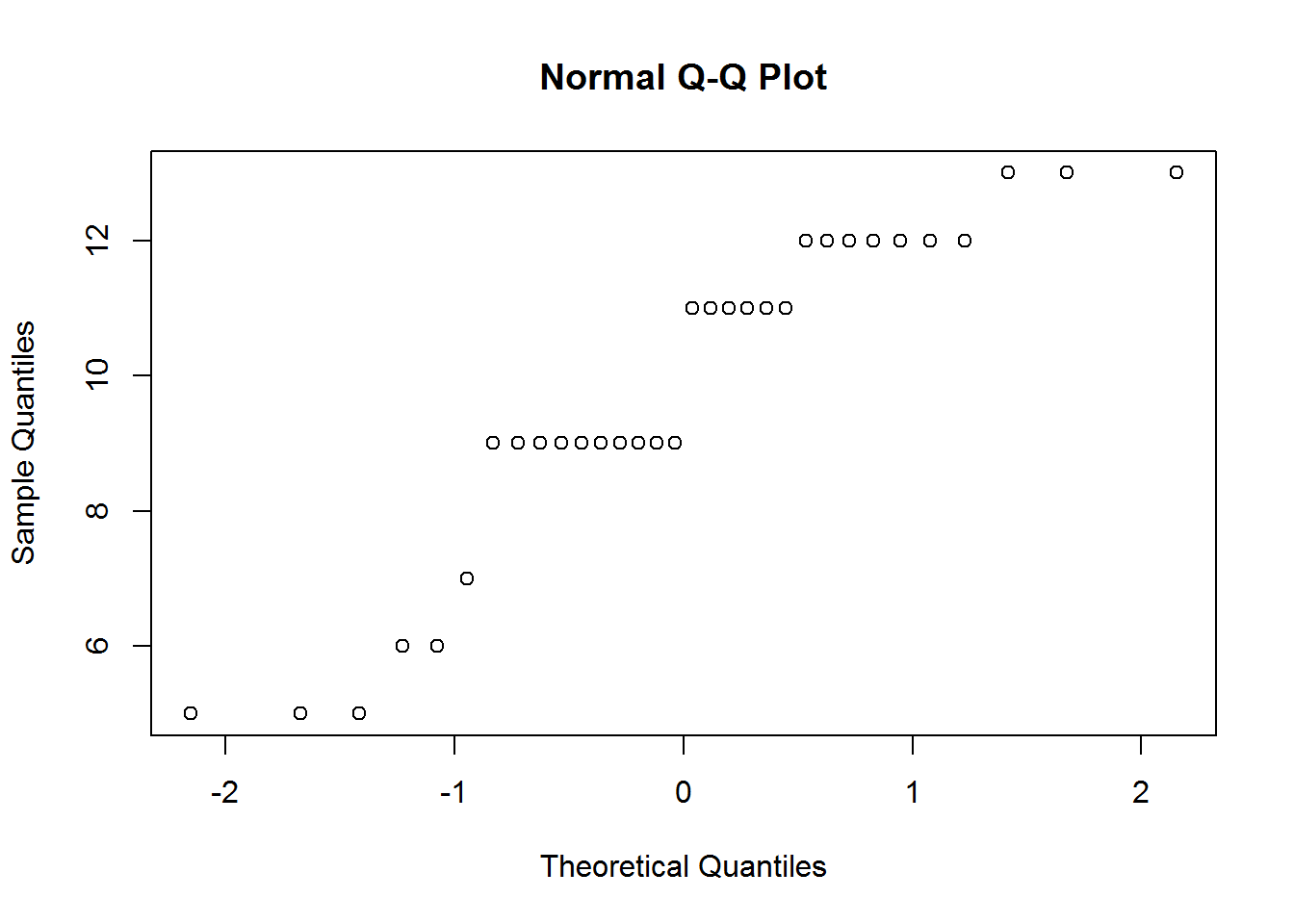
shapiro.test(femmes$temps\_alloue\_cours)

## Shapiro-Wilk normality test

## data: femmes$temps\_alloue\_cours

## W = 0.88893, p-value = 0.003275

qqnorm(femmes$temps\_alloue\_cours)

****

Même conclusion qu’avant. Donc on doit utiliser un test de Wilcoxon qui est non paramétrique:

wilcox.test(femmes$temps\_alloue\_cours,hommes$temps\_alloue\_cours)

## Wilcoxon rank sum test with continuity correction

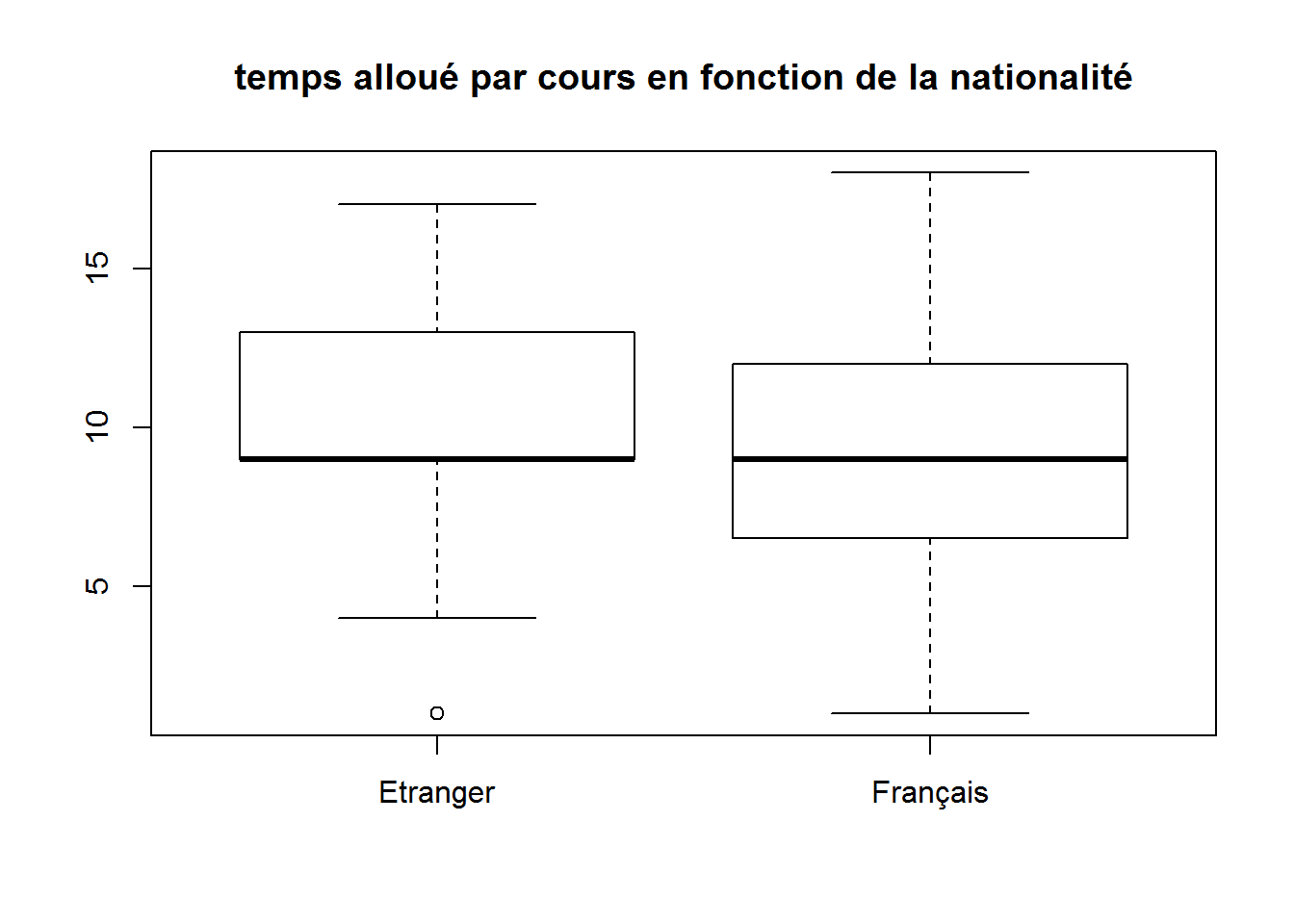
## data: femmes$temps\_alloue\_cours and hommes$temps\_alloue\_cours

## W = 1200, p-value = 0.8235

## alternative hypothesis: true location shift is not equal to 0

Avec cette p-value et pour un risque de 10 %, on accepte H0 donc pas de différence de temps alloué entre les hommes et les femmes

**Nationalité Français Vs Etrangers :**

****

temps\_nat<-suivi[,c("temps\_alloue\_cours","Fr\_vs\_Etranger")]

res2<-ddply(temps\_nat,~Fr\_vs\_Etranger,summarise,mean=mean(temps\_alloue\_cours),sd=sd(temps\_alloue\_cours))

res2

## Fr\_vs\_Etranger mean sd

## 1 Etranger 9.809524 3.637462

## 2 Français 9.476190 3.618310

Les variances sont égales, donc allons voir si les distributions sont normales ou non ? Pour les français :

fra<-filter(temps\_nat,Fr\_vs\_Etranger=="Français")

shapiro.test(fra$temps\_alloue\_cours)

##

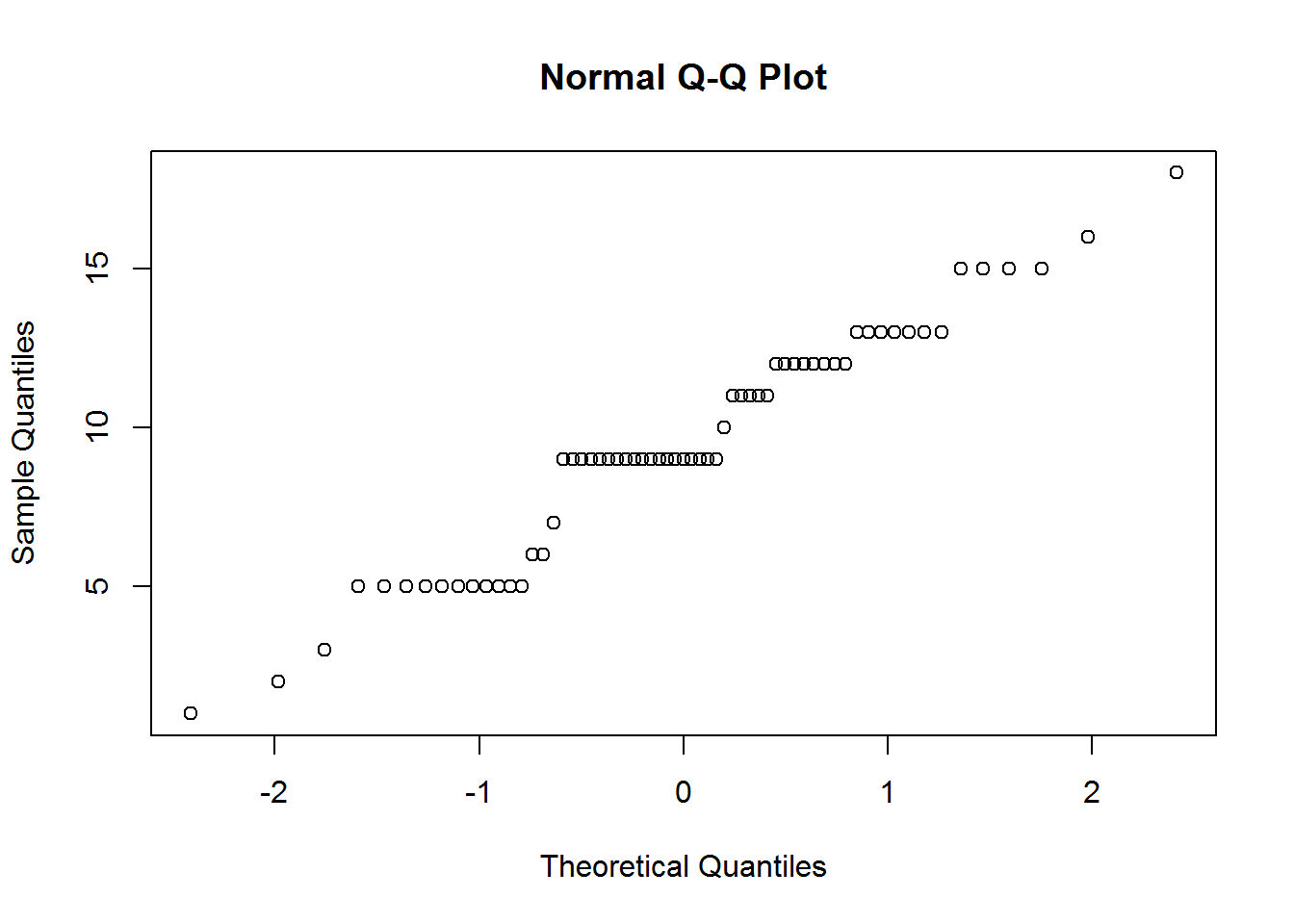
## Shapiro-Wilk normality test

##

## data: fra$temps\_alloue\_cours

## W = 0.95913, p-value = 0.03515

qqnorm(fra$temps\_alloue\_cours)



Pour cette p-value, on conclut que La distribution du temps alloué en moyenne par cours des français n’est pas tout à fait normale et vu la courbe qui est proche d’une droite. On procède de même pour les Etrangers :

etr<-filter(temps\_nat,Fr\_vs\_Etranger=="Etranger")

shapiro.test(etr$temps\_alloue\_cours)

##

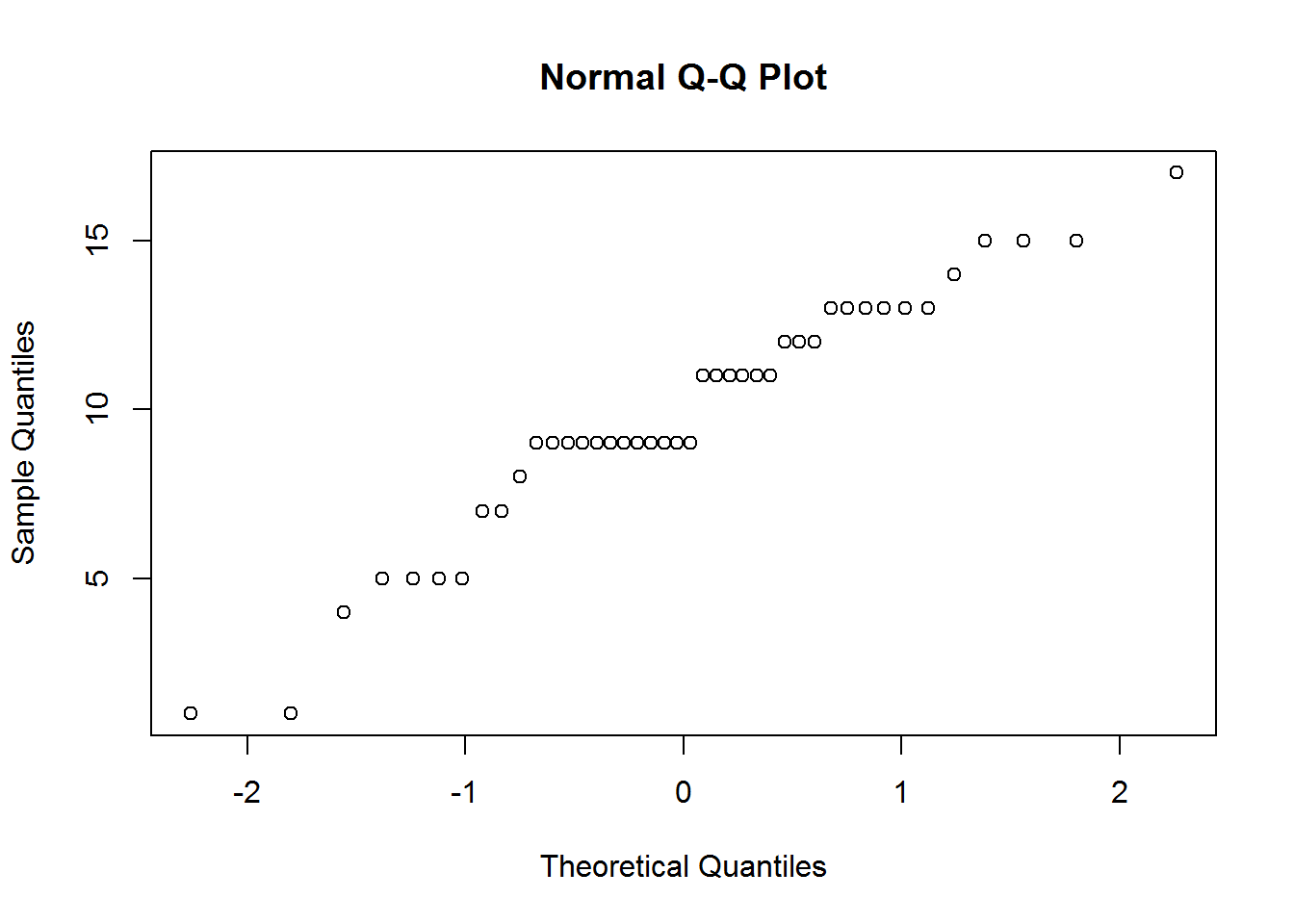
## Shapiro-Wilk normality test

##

## data: etr$temps\_alloue\_cours

## W = 0.95443, p-value = 0.09314

qqnorm(etr$temps\_alloue\_cours)



Pour cette p-value et un risque de 10%, on conclut que La distribution du temps alloué en moyenne par cours des étrangers est presque normale.

**On effectue un test de Student :**

t.test(etr$temps\_alloue\_cours,fra$temps\_alloue\_cours)

##

## Welch Two Sample t-test

##

## data: etr$temps\_alloue\_cours and fra$temps\_alloue\_cours

## t = 0.46099, df = 87.698, p-value = 0.6459

## alternative hypothesis: true difference in means is not equal to 0

## 95 percent confidence interval:

## -1.103697 1.770364

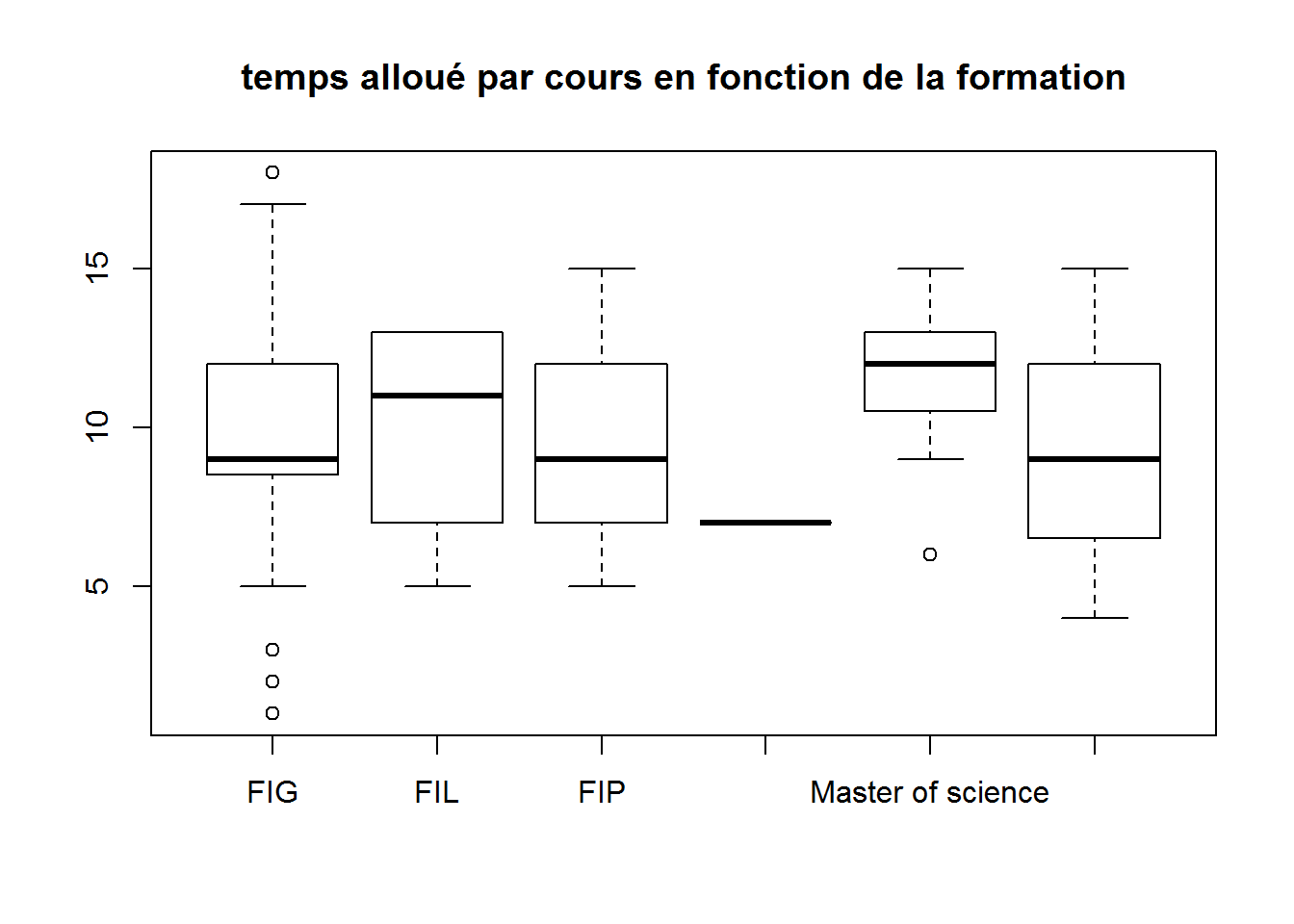
## sample estimates:

## mean of x mean of y

## 9.809524 9.476190

**Conclusion :** D’après ce test, on garde H0 donc on conclut qu’il n y a pas de différence de temps alloué entre les français et les étrangers

**Formation FIG vs FIP vs FIL vs Master of science vs Master Spé :**

****

temps\_form<-suivi[,c("temps\_alloue\_cours","formation")]

res3<-ddply(temps\_form,~formation,summarise,mean=mean(temps\_alloue\_cours),sd=sd(temps\_alloue\_cours))

## formation mean sd

## 1 FIG 9.482759 3.601880

## 2 FIL 10.000000 3.829708

## 3 FIP 9.666667 5.033223

## 5 Master of science 11.428571 2.992053

## 6 Master spécialisé 9.333333 5.507571

Les variances ne sont pas égales, voyons les distributions :

fig<-filter(temps\_form,formation=="FIG")

shapiro.test(fig$temps\_alloue\_cours)

## Shapiro-Wilk normality test

## data: fig$temps\_alloue\_cours

## W = 0.95658, p-value = 0.005292

La distribution du temps n’est pas normale pour les FIG

fip<-filter(temps\_form,formation=="FIP")

shapiro.test(fip$temps\_alloue\_cours)

## Shapiro-Wilk normality test

## data: fip$temps\_alloue\_cours

## W = 0.98684, p-value = 0.7804

La distribution du temps est normale pour les FIP

fil<-filter(temps\_form,formation=="FIL")

shapiro.test(fil$temps\_alloue\_cours)

## Shapiro-Wilk normality test

## data: fil$temps\_alloue\_cours

## W = 0.86337, p-value = 0.2725

La distribution du temps est normale pour les FIL

msc<-filter(temps\_form,formation=="Master of science")

shapiro.test(msc$temps\_alloue\_cours)

## Shapiro-Wilk normality test

## data: msc$temps\_alloue\_cours

## W = 0.90214, p-value = 0.3441

La distribution du temps est normale pour les Master of science

mssp<-filter(temps\_form,formation=="Master spécialisé")

shapiro.test(mssp$temps\_alloue\_cours)

## Shapiro-Wilk normality test

## data: mssp$temps\_alloue\_cours

## W = 0.99725, p-value = 0.8999

La distribution du temps est normale pour les Master spécialisé

**Tests de wilcoxon sur le temps alloué pour les étudiants selon formation :**

wilcox.test(fig$temps\_alloue\_cours,fip$temps\_alloue\_cours)

##

## Wilcoxon rank sum test with continuity correction

##

## data: fig$temps\_alloue\_cours and fip$temps\_alloue\_cours

## W = 132, p-value = 0.9817

## alternative hypothesis: true location shift is not equal to 0

Avec un risque de 10%, on accepte H0, les deux formations (fig et fip) ont le même temps alloué au mooc.

wilcox.test(fip$temps\_alloue\_cours,fil$temps\_alloue\_cours)

## Warning in wilcox.test.default(fip$temps\_alloue\_cours, fil

## $temps\_alloue\_cours): cannot compute exact p-value with ties

##

## Wilcoxon rank sum test with continuity correction

##

## data: fip$temps\_alloue\_cours and fil$temps\_alloue\_cours

## W = 6, p-value = 1

## alternative hypothesis: true location shift is not equal to 0

Avec un risque de 5%, on accepte H0, les deux formations (fip et fil) ont le même temps alloué au mooc.

wilcox.test(fig$temps\_alloue\_cours,fil$temps\_alloue\_cours)

## Wilcoxon rank sum test with continuity correction

## data: fig$temps\_alloue\_cours and fil$temps\_alloue\_cours

## W = 152, p-value = 0.6714

## alternative hypothesis: true location shift is not equal to 0

Avec un risque de 10%, on accepte H0, les deux formations (fig et fil) ont le même temps alloué au mooc.

wilcox.test(fig$temps\_alloue\_cours,msc$temps\_alloue\_cours)

## Wilcoxon rank sum test with continuity correction

## data: fig$temps\_alloue\_cours and msc$temps\_alloue\_cours

## W = 193, p-value = 0.1038

## alternative hypothesis: true location shift is not equal to 0

Avec un risque de 10%, on accepte H0, les deux formations (fig et Master of science) ont le même temps alloué au mooc.

wilcox.test(fip$temps\_alloue\_cours,msc$temps\_alloue\_cours)

## Warning in wilcox.test.default(fip$temps\_alloue\_cours, msc

## $temps\_alloue\_cours): cannot compute exact p-value with ties

## Wilcoxon rank sum test with continuity correction

## data: fip$temps\_alloue\_cours and msc$temps\_alloue\_cours

## W = 8, p-value = 0.6445

## alternative hypothesis: true location shift is not equal to 0

Avec un risque de 10%, on accepte H0, les deux formations (fip et Master of science) ont le même temps alloué au mooc.

wilcox.test(msc$temps\_alloue\_cours,mssp$temps\_alloue\_cours)

## Warning in wilcox.test.default(msc$temps\_alloue\_cours, mssp

## $temps\_alloue\_cours): cannot compute exact p-value with ties

## Wilcoxon rank sum test with continuity correction

## data: msc$temps\_alloue\_cours and mssp$temps\_alloue\_cours

## W = 13, p-value = 0.6445

## alternative hypothesis: true location shift is not equal to 0

Avec un risque de 10%, on accepte H0, les deux formations (Mastr of science et Master spé) ont le même temps alloué au mooc.

wilcox.test(fig$temps\_alloue\_cours,mssp$temps\_alloue\_cours)

##

## Wilcoxon rank sum test with continuity correction

##

## data: fig$temps\_alloue\_cours and mssp$temps\_alloue\_cours

## W = 138.5, p-value = 0.8636

## alternative hypothesis: true location shift is not equal to 0

Avec un risque de 10%, on accepte H0, les deux formations (fig et Master spé) ont le même temps alloué au mooc.

**Conclusion :** pas de différences entre les différentes formation en ce qui concerne le temps alloué à un mooc en moyenne.