Begimai Konushbaeva INF-4

Etude des deux échantillons appariés

La base de données étudiée est composée des deux échantillons, où le premier représente les notes d’étudiants d’une certaine classe en maths et le deuxième représente les notes de cette classe en français
Chaque note peut varier entre 0 et 100
Ces deux échantillon sont donc appariés car les deux notes appartiennent à une personne On veut trouver si les étudiants ont généralement plus des competences dans une des 2 matières.

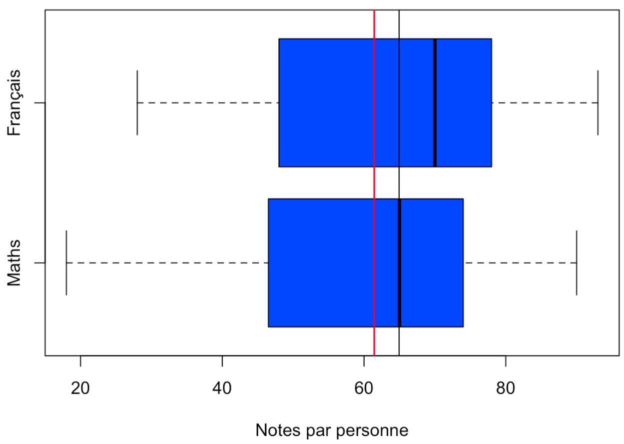
Statistiques du premiere échantillon:

```
     Maths
Min.   :18.00
1st Qu.:46.50
Median :65.00
Mean   :61.42
3rd Qu.:74.00
Max.   :90.00
```

Statistiques du deuxième échantillon:

```
   Français
Min.   :28.00
1st Qu.:48.00
Median :70.00
Mean   :64.95
3rd Qu.:78.00
Max.   :93.00
```

Boxplot:



Le debut de la box pour chaque échantillon commence par le premier quartile et finit par le troisième quartile:

q1=46.50 q2=74.00

Pour le premier échantillon et q1=48.00 q2=78.00
Pour le deuxième échantillon

Les bornes des lignes sont les valeurs minimale et maximale respectivement de chaque échantillon

Donc, 18 et 90 pour le premier échantillon (Maths), 28 et 93 pour le deuxième (Français) respectivement.

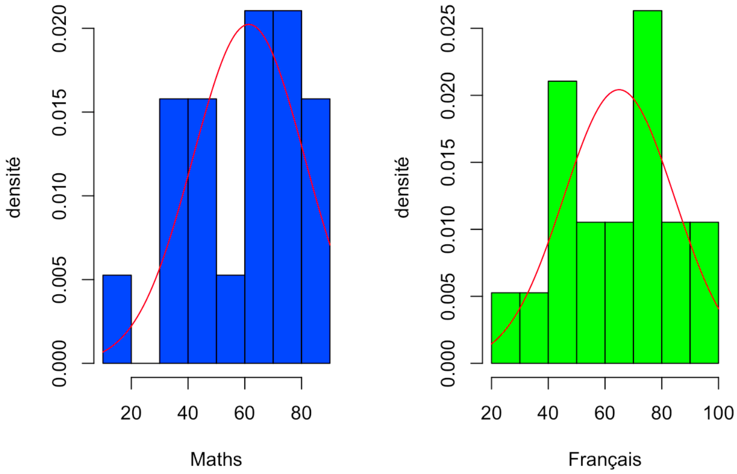
On y ajoute la verticale de la moyenne du 1er échantillon en rouge et du deuxième en noir.

On en peut déduire que la performance académique d'étudiants est quelque peu élevé en français qu’en mathématiques.

La ligne a l'intérieur de notre fonction indique la médiane de l’échantillon étudié m1=65 et m2=70

Histogrammes:

On trace l’histogramme de chaque échantillon et on y ajoute la densité de la loi normale en



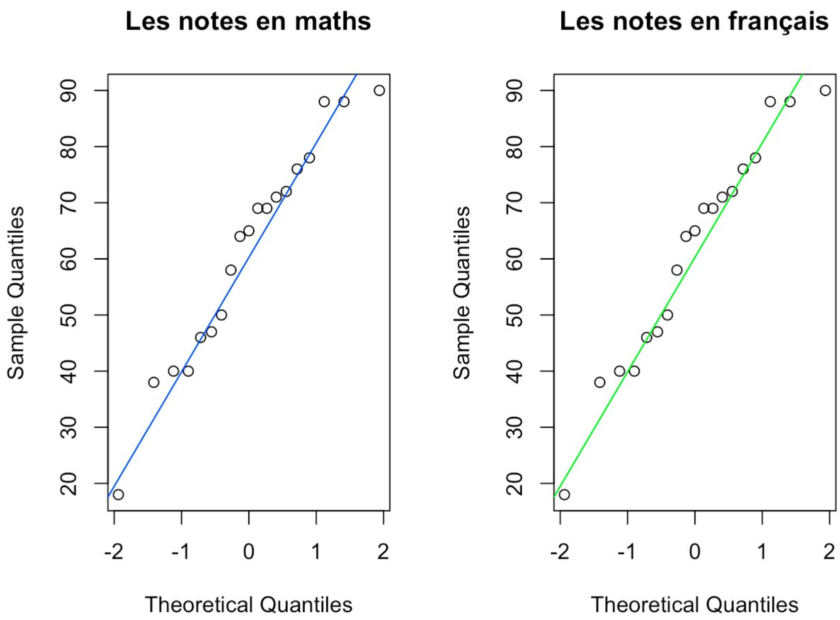
rouge.
D’où on peut conclure qu’il y a plus de stabilité de performance en français, vu qu’il y a un certain nombre d’étudiants avec le même niveau.
Par contre, sur l’histogramme de l’échantillon des notes en maths il y a une grande différence de niveau.
Les histogrammes s’approchent de la densité de la loi normale.

QQplots:

On peut voir que les points se déclinent légèrement de la ligne qui corresponde a la droite de Henry y=mu*x+sigma

Les deux échantillons donc s’approchent de la loi Normale, ils sont donc gaussiens.

$N(\mu;\sigma^2)$



Conclusion: Comme on peut voir sur les histogrammes et les qqplots les deux échantillon étudié sont gaussiens, on peut donc construire les intervalles de confiances pour ces statistiques et calculer leurs estimations.

Estimations et intervalles de confiance

L'estimation de l’espérance du premier échantillon (notes en math) : mû = 61.42105

L’intervalle de confiance eu niveau de confiance égale a 95% : [51.91983; 70.92227]
Pour obtenir une precision +-12:
On calcul l’intervalle de confiance au niveau de confiance 80%, on obtient l’intervalle: [55.40449; 67.43762]

L'estimation de l’espérance du deuxième échantillon (notes en français) : mû = 64.94737

Intervalle de confiance au niveau de confiance 95%: [55.53190 ; 74.36284]
Pour obtenir une precision +-12:
On calcul l’intervalle de confiance au niveau 80%: [58.98510 70.90963]

L’estimation de l’écart-type du premier échantillon (notes en maths) : sigma1 = 19.7127

Intervalle de confiance ai niveau 95% : [14.89517 ; 29.15163]

L’estimation de l’écart-type du deuxième échantillon (notes en français): sigma2 = 19.53479

Intervalle de confiance au niveau 95%: [14.76074 ; 28.88853]

Tests statistiques

Test de la moyenne sur le premier échantillon:

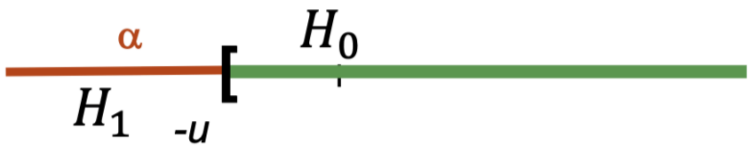
Le professeur de maths assure que la moyenne de la classe en maths est au moins 70

On veut tester si la moyenne des notes en maths est égale ou inférieur a 70

Donc, on effectue le test suivant:

mu0=70

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$



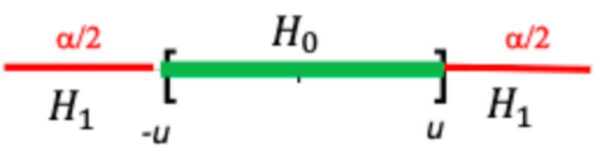

```
> t.test(x1, mu=70, alternative="less")

        One Sample t-test

data:  x1
t = -1.897, df = 18, p-value = 0.037
alternative hypothesis: true mean is less than 70
95 percent confidence interval:
     -Inf 69.26319
sample estimates:
mean of x
 61.42105
```

Donc on rejette H0/ accepte H1 avec un risque de se tromper 3,7%

Test de la moyenne sur le deuxième échantillon:

On veut tester si la moyenne est égale a 65
Donc, on effectue le test suivant:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$




```
> t.test(x2, mu=65, alternative = "two.sided")

        One Sample t-test

data:  x2
t = -0.011744, df = 18, p-value = 0.9908
alternative hypothesis: true mean is not equal to 65
95 percent confidence interval:
 55.53190 74.36284
sample estimates:
mean of x
 64.94737
```

On en déduit que on accepte H0/ rejette
H1 avec un risque de se tromper < 1%

Comparaison des deux échantillon

On veut savoir si la moyenne en français est plus élevé que en maths
Notons mud la difference entre les deux moyennes
On effectue le test suivant:

$$\begin{cases} H_0 : \mu_d = 0 \\ H_1 : \mu_d \neq 0 \quad (ou > 0) \end{cases}$$

```
> t.test(x2,x1,alternative="greater",paired=TRUE)

	Paired t-test

data:  x2 and x1
t = 2.322, df = 18, p-value = 0.01608
alternative hypothesis: true difference in means is greater
 than 0
95 percent confidence interval:
 0.8929085       Inf
sample estimates:
mean of the differences
               3.526316
```

Donc on rejette H0/ accepte H1 avec un risque de se tromper 1,6%
La moyenne en français est légèrement plus élevée que la moyenne en maths.

Effectuons un test sur la difference des moyennes autre que 0
On note D=x2-x1 (x2 — notes en français, x1 — notes en maths)
Où mu0=10
On suppose que la difference des moyenne entre les deux échantillon est noté 10 par le professeur Sur le nouveau échantillon D on effectue un test statistique suivant:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

```
> t.test(D,mu=10,alternative="two.sided")

	One Sample t-test

data:  D
t = -4.2628, df = 18, p-value = 0.0004681
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
 0.3357849 6.7168466
sample estimates:
mean of x
 3.526316
```

Donc, on rejette H0/accepte H1 avec un risque < 0,1% de se tromper
La difference entre les deux échantillons n'est pas égale a 10

Conclusion : de tous les tests effectués sur ces deux échantillon on peut déduire que la performance académique en français est légèrement plus élevée qu'en mathématiques. On peut bien le voir sur le test de comparaison des moyenne des deux échantillon.