Groupe: INF3 Nom prénom: Collomb Sullivan

DM2 Projet Individuel : Différence de temps de repas pris entre le midi et le soir

Introduction :

Les données ont été récolté auprès d'une seul population de 15 personnes Françaises à l'aide d'un formulaire réalisé sur Google Form. Le but était d'obtenir des valeurs continues, cependant aux vues des résultats obtenues nous avons donc recueilli des variables discrètes. Les deux échantillons de données sont exprimés dans une même unité qui est le temps (en minute). Les deux variables discrètes appariées observée sont :
-Le temps pris pour manger le midi
-Le temps pris pour manger le soir.
La problématique proposée sur ces données est la suivante : Prend-t-on plus de temps pour manger le soir plutôt que le midi ?

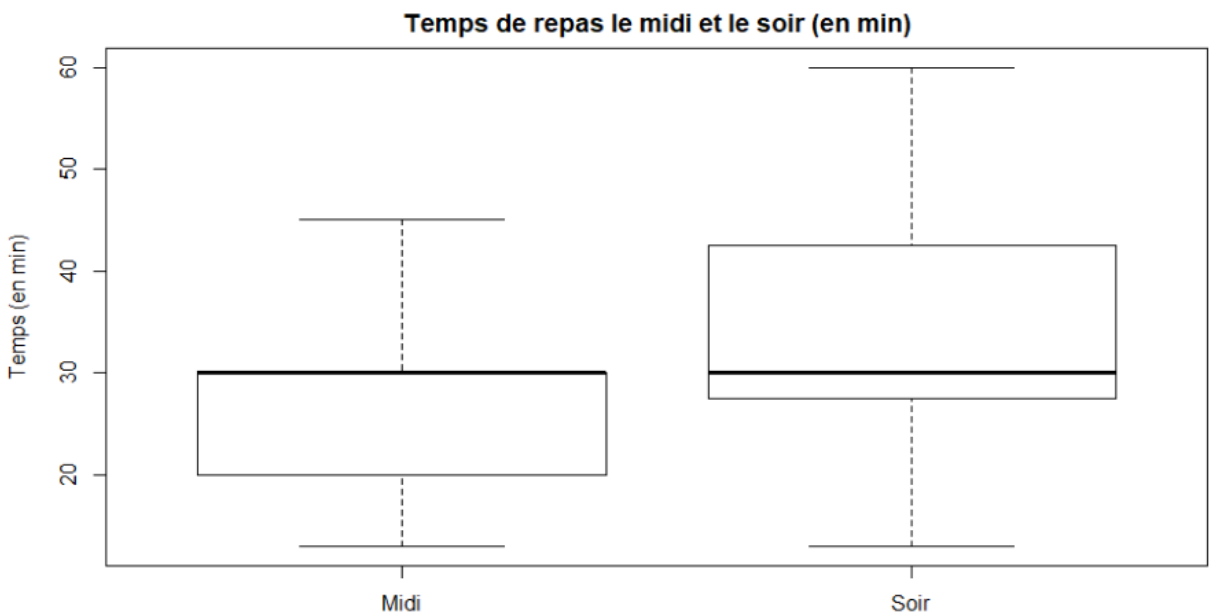
Analyse descriptive :

Afin d'éviter une trop grande répétition nous simplifierons l'appellation de l'échantillon des résultats du temps pris pour manger le midi par « résultats du midi », et pour l'échantillon des résultats de temps pris pour manger le soir par « résultats du soir ».
Résultats de la fonction summary() appliquée à chaque échantillon (en minute) :

	Résultats du midi	**Résultats du soir**
Valeur minimale	13	13
1er quartile	20	27,5
Médiane	30	30
Moyenne	27,87	33,2
3ème quartile	30	42,5
Valeur maximale	45	60

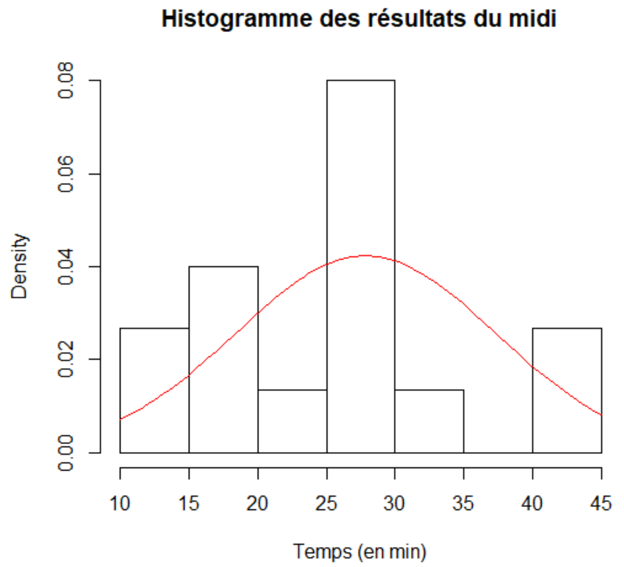
L'écart-type pour les résultats du midi est de environ 9,433 alors que pour les résultats du soir on obtient un écart-type plus grand avec une valeur de 11,833. On observe donc que les résultats du midi sont plus homogènes que les résultats du soir.

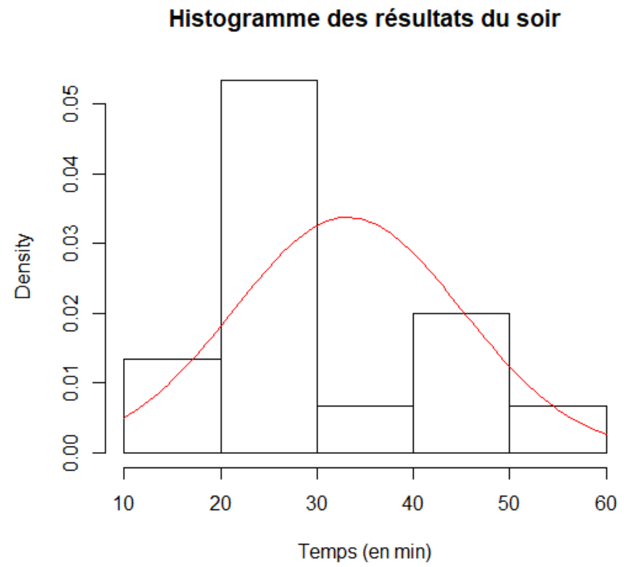
Voici une représentation graphique des résultats obtenu :



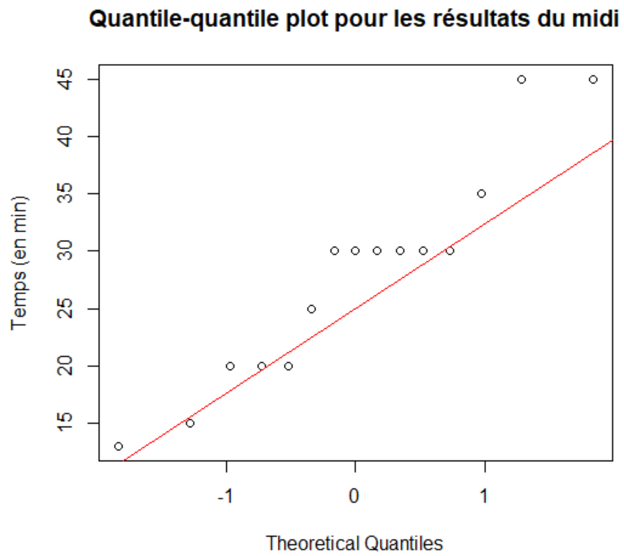
Groupe: INF3 Nom prénom: Collomb Sullivan

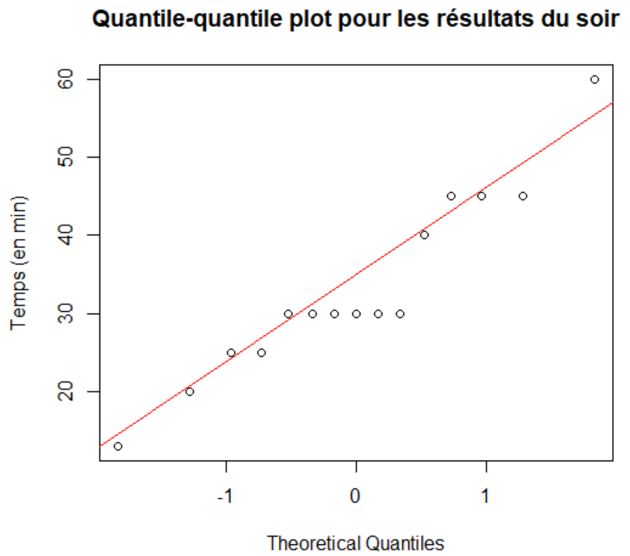
Le graphique ci-dessus utilisé est une boîte à moustaches (fonction boxplot() en langage R), nous permettant de bien mettre en évidence la différence entre la médiane, 1er et 3ème quartile des deux boîte à moustaches.





Les graphiques suivant représentés ci-dessus représentent chacun l'histogramme de leurs résultats, ainsi qu'une courbe représentant leur densité Gaussienne.





Enfin on observe ci-dessus les deux derniers graphiques quantile-quantile avec pour chacun leurs droites passant par leurs 1er et 3ème quartiles. Ils permettent de les comparer à un loi normale.

Les résultats numériques obtenus ainsi que ceux présent sur la boîte à moustaches permettent de constater une différence de temps pris pour manger entre le midi et le soir, avec tout d'abord des repas plus long en termes de temps pour le midi que pour le soir en moyenne, et en termes de valeurs maximales. Cependant la médianes des deux reste identique. Les graphiques représentants les histogrammes nous permettent de voir que aucune des 2 distributions ne suivent de loi normale. La courbe des 1er et 3ème quartiles sur les graphiques quantile-quantile correspond bien avec une majorité des résultats

Groupe: INF3 Nom prénom: Collomb Sullivan

Analyse inférentielle :

Estimation et intervalle de confiance :

Pour l'intervalle de confiance en fonction de la moyenne on observe que 95% des données observées sont comprises dans l'intervalle. Avec un intervalle compris entre 22,6 minutes et 33,1 minutes pour les résultats du midi, ainsi qu'un intervalle pour les résultats du soir compris entre 26,6 et 39,8, on observe bien que les valeurs obtenus ne fluctuent pas grandement. On peut alors en conclure que chaque données qui viendrait a être rajoutée pourrait alors avoir un grand impact sur les données statistiques obtenus jusqu'ici. Pour ne pas impacter ces données il ne faudrait pas obtenir de résultats compris dans les 5% qui ne sont pas compris par l'intervalle de confiance. La deuxième intervalle calculée est celle de l'écart-type. On obtient pour les résultats du midi un intervalle compris entre 47 et 221, et pour les résultats du soir un intervalle compris entre 75 et 348. On peut alors observer que les résultats obtenus entre ces deux intervalles sont très proches. Au vue des estimations et des intervalles de confiance nous pouvons aisément conclure que la différence obtenues entre les résultats du midi et du soir ne sont pas significatifs pour l'instant.

Tests sur échantillons :

On effectue un test sur les données pour la modalités du midi, en prenant $\mu = 27$ car cela correspond à la moyenne observé pour cette modalité. Nous allons effectuer un test de Student sur la moyenne.
H0 : $\mu= 27$ vs H1 : $\mu \neq 27$.
Si la p-value > 0,05 alors on ne rejette pas H0. Et si la p-value < 0.05 alors on rejette H0.
On obtient une p-value = 0,7273, donc nous avons p-value > 0,05 alors on ne rejette pas H0 et $\mu = 27$.
Pour le second test on effectue un test sur les données pour la modalités du soir, en prenant $\mu = 33$ car cela correspond à la moyenne observé pour cette modalité. Nous allons effectuer un test de Student sur la moyenne.
H0 : $\mu= 33$ vs H1 : $\mu \neq 33$.
Si la p-value > 0,05 alors on ne rejette pas H0. Et si la p-value < 0.05 alors on rejette H0.
On obtient une p-value = 0,9487, donc nous avons p-value > 0,05 alors on ne rejette pas H0 et $\mu = 33$.

Tests sur deux échantillons :

Pour les tests sur les deux échantillons obtenues nous allons effectuer respectivement un test de Student afin de comparer leurs moyennes et un test de Fischer pur comparer leurs variances. Pour le test de Student nous testons si les modalités suivantes :
H0 : $\mu_{midi} = \mu_{soir}$ vs H1 : $\mu_{midi} \neq \mu_{soir}$
On obtient une p-value = 0,1837, la condition ou l'on rejette H0 étant d'avoir une p-value < 0,05, on ne rejette alors pas H0. Pour le test du Fischer nous testons les modalités suivantes :
H0 : $\sigma_{midi} = \sigma_{soir}$ vs H1 : $\sigma_{midi} \neq \sigma_{soir}$
On obtient une p-value = 0,4066, la condition ou l'on rejette H0 étant d'avoir une p-value < 0,05, comme pour le test de Student ici on ne rejette pas H0.
Ces tests nous permettent de conclure que les tests statistiques effectués sur les moyennes et les variances ne nous permettent pas de démontrer qu'il y a une différence significative entres les résultats du midi et du soir.

Conclusion :

Les résultats statistique obtenues lors de l'analyse descriptive nous laisse entrevoir une différence entre les résultats du midi et les résultats du soir. Les graphiques montrent bien des différences même si ces différences ne sont pas si évidentes. Cependant ces conjectures obtenues lors de l'observation des graphique de l'analyse descriptive sont démentie grâce aux résultats obtenues lors de l'analyse inférentielle. Elles nous permettent, grâce aux résultats obtenues sur les différents tests, qu'il n'existe pas de différence significative entre les résultats du midi et les résultats du soir.

Groupe: INF3 Nom prénom: Collomb Sullivan

Annexe :

Données :

	Midi	Soir
1	30	45
2	30	30
3	30	30
4	30	30
5	45	60
6	35	45
7	30	30
8	30	45
9	20	40
10	13	13
11	20	25
12	25	25
13	45	30
14	15	20
15	20	30

Groupe: INF3 Nom prénom: Collomb Sullivan

Script R :

```r
### Création des variables avec les données récoltées (en min)
setwd("~/GitHub/L2_INFO/STA401/TP/Projet")
df <- read.table('INF3-Collomb-Sullivan.csv', header=TRUE, sep=',')
midi <- df$Midi
soir <- df$Soir

### Résumé numérique
summary(df)
sd(midi)
sd(soir)

### Graphique
# Boxplot
boxplot(midi,soir, main="Temps de repas le midi et le soir (en min)", ylab="Temps (en min)", names=c("Midi","Soir"))

# Hist
par(mfrow=c(1,2))
hist(midi, main = "Histogramme des résultats du midi", probability =  TRUE)
curve(dnorm(x,mean(midi),sd(midi)), add = T, col = "red")
hist(soir, main = "Histogramme des résultats du soir", probability =  TRUE)
curve(dnorm(x,mean(soir),sd(soir)), add = T, col = "red")

# Qqnorm
par(mfrow=c(1,2))
qqnorm(midi, main = "Quantile-quantile plot pour les résultats du midi", ylab = "Temps (en min)")
qqline(midi,col="red")
qqnorm(soir, main = "Quantile-quantile plot pour les résultats du soir", ylab = "Temps (en min)")
qqline(soir,col="red")

### Estimation et intervalle de confiance
## Sur la moyenne
t.test(midi)
t.test(soir)

## Sur l'ecart type
# Midi :
n=length(midi)
sx <- sd(midi)
estimate <- sx**2
c=qchisq(1-(0.05/2),df=n-1)
d=qchisq((0.05/2),df=n-1)
C=(n - 1)*estimate/c
D=(n - 1)*estimate/d
paste('L intervalle de confiance recherche est IC(95%)=[',C,D,']')

# Soir
n=length(soir)
sx <- sd(soir)
estimate <- sx**2
c=qchisq(1-(0.05/2),df=n-1)
d=qchisq((0.05/2),df=n-1)
C=(n - 1)*estimate/c
D=(n - 1)*estimate/d
paste('L intervalle de confiance recherche est IC(95%)=[',C,D,']')
### Tests sur un échantillon
# Test de la moyenne avec mu0=27
t.test(midi, mu=27)
#  H0 : mu(midi) = 27 VS H1 : mu(midi) != 27

# Test de la moyenne avec mu0=33
t.test(soir, mu=33)
#  H0 : mu(midi) = 33 VS H1 : mu(midi) != 33

### Tests sur deux échantillons-
## Comparaison des moyennes de deux groupes : test de Student
t.test(midi, soir)
# H0 : mu(midi) = mu(soir) VS mu(midi) != mu(soir)

## Comparaison des variances de deux groupes : test de Fisher
# H0 : sigma(midi) = sigma(soir) VS H1 : sigma(midi) != sigma(soir)
var.test(midi,soir)
```