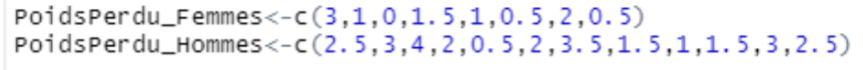
# TP2- Statistique avec R studio

Exercice 1:

1)



2)

shapiro.test(poidsF)

var.test(poidsF,poidsH)

t.test(poidsH,poidsF,alternative="greater",paired =FALSE,var.equal=TRUE)

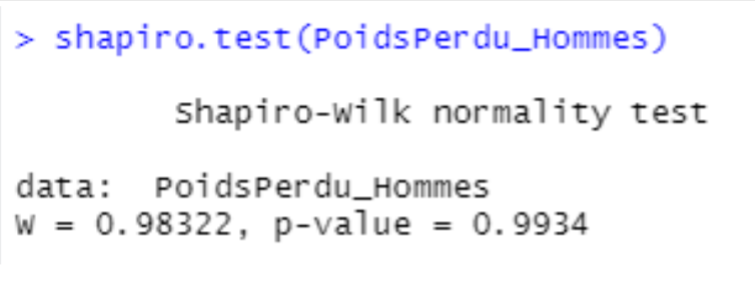
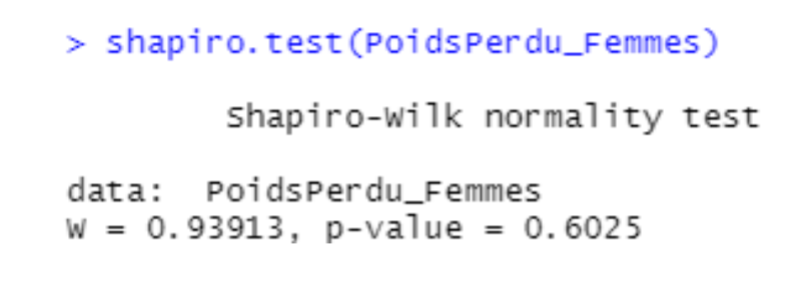
On émet les hypothèses :

H0 : La distribution suit une loi normale

H1 : La distribution ne suit pas une loi normale

α = 5%

**On effectue un test de shapiro :**



Ici, p-value > 0,05. Donc on admet que les données suivent une loi normale au risque de 5% avec une erreur de seconde espèce.

Pour montrer que le régime est plus efficace pour les hommes que pour les femmes. On va utiliser un test unilatéral

On veut comparer des variances.

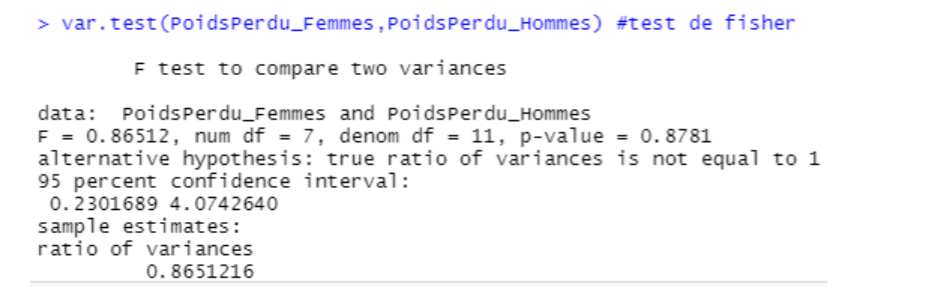
**On effectue un test de Fisher**:

On émet les hypothèses :

H0 : les variances sont les mêmes

H1 : les variances sont différentes

α=5%



La p-value > 0,05 donc on ne peut pas dire que les variances sont différentes. On admet H0 et on rejette H1.

3) et 4) On teste si la moyenne des hommes est supérieure à la moyenne des femmes. C’est donc un test unilatéral

On admet l’hypothèse de normalité et d’égalité des variances.

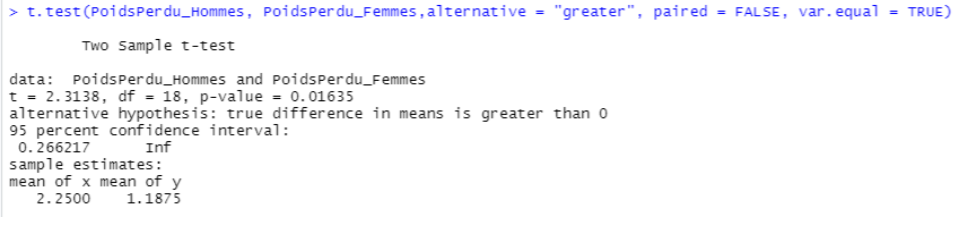
**On effectue le test de Student :**

On émet les hypothèses :

H0 : Les moyennes des régimes pour les hommes et les femmes sont les mêmes (µf = µh)

H1 : Le régime est plus efficace pour les hommes que pour les femmes (µh > µf)

α=5%

5)

p-value < 0,05 donc on rejette avec un risque de 5% l'hypothèse H0 et on admet H1, c’est-à-dire que le régime semble plus efficace chez les hommes que les femmes

Exercice 2



1) On souhaite comparer les moyennes des précipitations des différentes zones.

On considère qu’il s’agit d’une comparaison de deux échantillons appariés

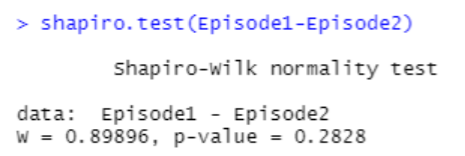
2) On émet les hypothèses :

H0 : La série de données suit une loi normale

H1 : La série de données ne suit pas une loi normale

α=5%

**On vérifie qu’ils suivent une loi normale**



La p-value > 0,05 donc on peut dire que les deux séries suivent une distribution normale

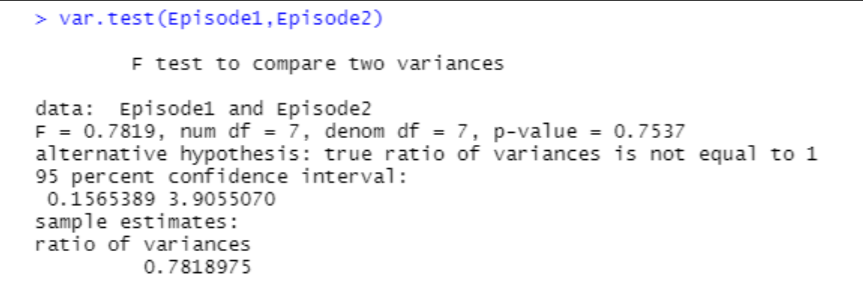
3) On émet les hypothèses :

H0 : les variances sont les mêmes

H1 : les variances sont différentes

α=5%

**On effectue un test de Fisher :**



La p-value > 0,05 donc on ne peut pas dire que les variances sont significativement différentes. On admet H0

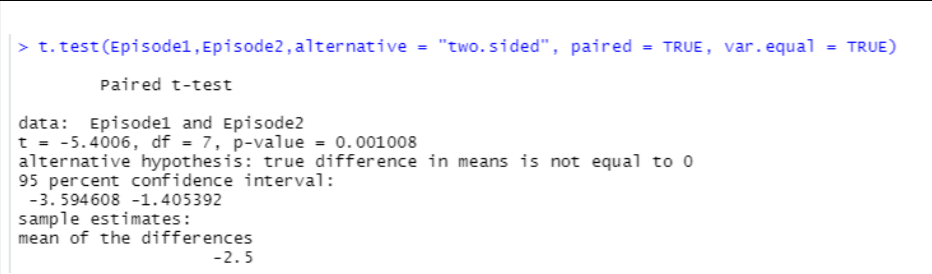
4) **On réalise une comparaison des moyennes :**

H0 : Il n’y a pas de différence entre les moyennes de précipitations pour les zones

H1 : Il y a une différence entre les zones

α=5%

**On réalise un test de Student :**



p-value < 5% donc on rejette H0, et on accepte H1 avec un risque de 5% On a donc une différence entre les deux zones.

Exercice 3 :



1. On compare un rendement horaire de deux machines avec des échantillons non appariés.

On émet les hypothèses suivantes :

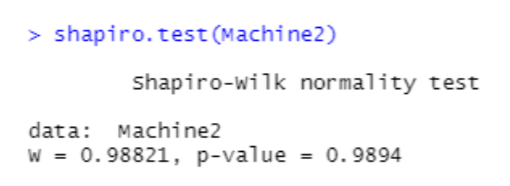
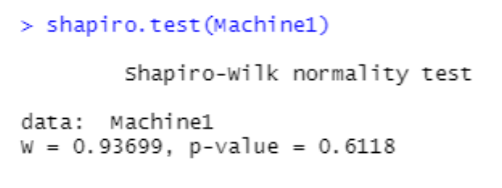
H0 : La série de données suit une loi normale

H1 : La série de données ne suit pas une loi normale

α=5%

On vérifie que les données suivent une loi normale :

**On réalise un test de Shapiro pour les deux séries de données :**



Les p-value > 0,05 donc suivent une distribution normale

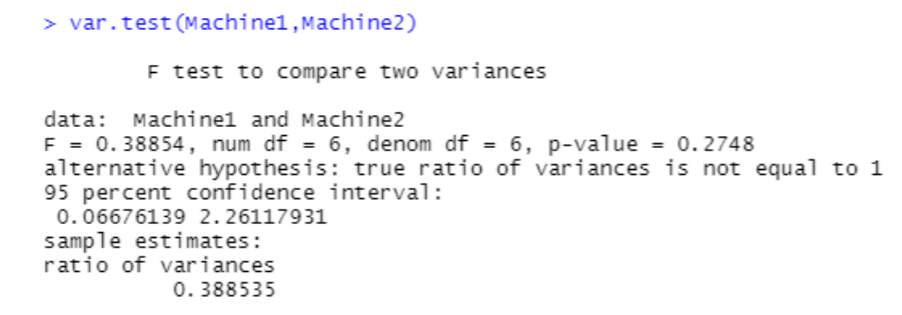
b)

On émet les hypothèses suivantes :

H0 : les variances sont les mêmes

H1 : les variances sont différentes α=5%

**On réalise un test de Fisher :**



p-value > 5% donc on accepte H0. Les variances sont donc les mêmes.

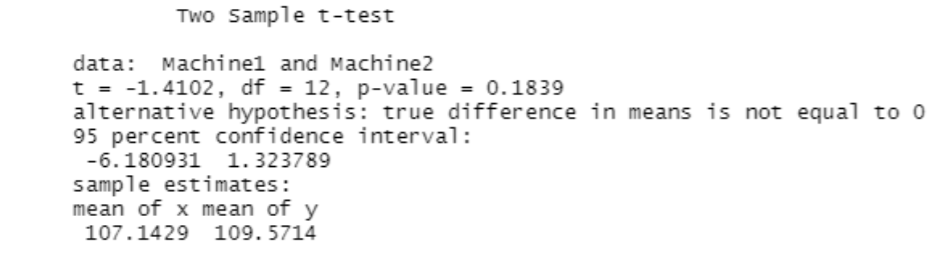
**On réalise un test de Student :**

H0 : Il n’y a pas de différence entre les moyennes de rendement

H1 : Il y a une différence

α=5%





p-value > 5% donc on accepte H0 avec un risque alpha de 5%. Il admet qu’il n’y ai de différence de moyenne sur le rendement des deux machines.

Exercice 4 :

Enoncé des hypothèses

Les variables sont indépendantes

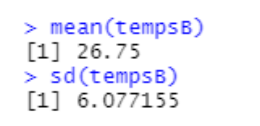
Les variances sont égales

α=5%

1. On crée un vecteur tempsB



1. Moyenne et écart-type de la série :

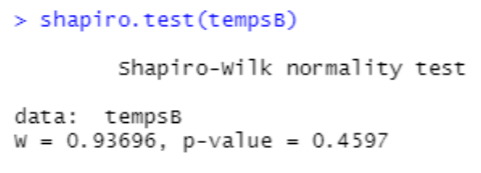


1. Condition de la normalité de la distribution :

H0 : La série de données suit une loi normale

H1 : La série de données ne suit pas une loi normale

α=5%



p-value > 0,05 donc les données suit une distribution normale

On utilise le test de Student pour comparer les 2 moyennes. On va donc utiliser un test unilatéral.

**On fait une comparaison des moyennes.**

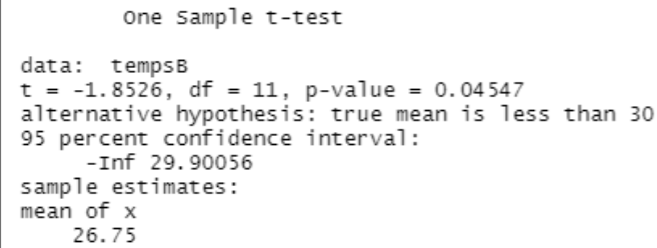
On émet les hypothèses :

H0 : Les médicaments A et B ont la même efficacité :µA = µB

H1 : Le médicament B est plus efficace que le médicament A : µB > µA

α=5%





On admet H1 au risque de 5% car p-value < 5% . On admet donc que le médicament B a une meilleure efficacité que le médicament A.

PARTIE 2

Etude 1

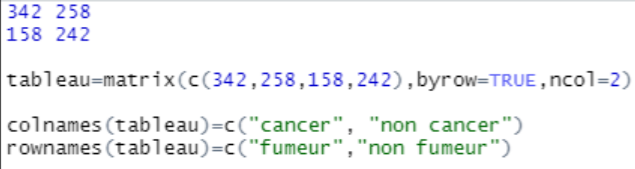
Nous allons réaliser le test du Khi-deux d’indépendance avec les hypothèses suivantes :

H0 : Les caractères sont indépendants

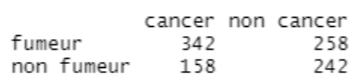
H1 : La série de données ne suit pas une loi normale

α=5%

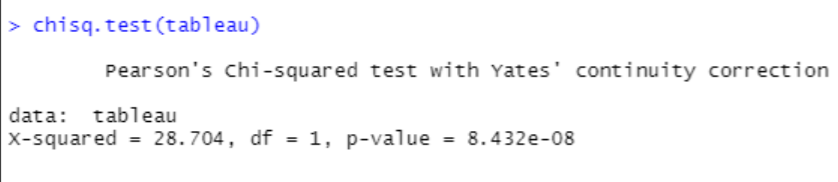
On crée un tableau de contingence :



Quand on rentre « tableau » dans la console on obtient:



On effectue le test de Khi deux :



La p-value est très inférieur à α = 0.05, donc on ne peut pas considérer que les variables sont liées.

Etude 2

Exercice 1 : Enseignement des TICE dans les lycées

2)

> tableaudecontingence <- table(TPRAPP\_TICE)

On a des valeurs « NA » donc on sélectionne l’option pour les masquer et donc qu’ils ne soient pas pris en compte lors du test du chi-2

3)

Enoncé des hypothèses :

H0 : Indépendance de la fréquence d’utilisation d’un PC et de la discipline

H1 : Dépendance de la fréquence d’utilisation d’un PC et de la discipline

α=5%

> chisq.test(tableaudecontingence)

Pearson's Chi-squared test

data: tableaudecontingence

X-squared = 22.18, df = 4, p-value = 0.0001845

On admet H1 au risque de 5%. (p-value = 0.0001845). On observe donc une dépendance entre la fréquence d’utilisation d’un PC et de la discipline. La suite des questions permettra de comprendre cette constatation.

4)

> khideux1<-chisq.test(tableaudecontingence)

> khideux1$observed

Discipline

Frequence\_utilisation\_PC autres hg langues lettres

plus\_d\_une\_fois\_par\_sem 35 10 22 15

une\_fois\_par\_sem\_ou\_moins 6 10 20 19

Discipline

Frequence\_utilisation\_PC math

plus\_d\_une\_fois\_par\_sem 10

une\_fois\_par\_sem\_ou\_moins 19

> khideux1$expected

Discipline

Frequence\_utilisation\_PC autres hg

plus\_d\_une\_fois\_par\_sem 22.72289 11.084337

une\_fois\_par\_sem\_ou\_moins 18.27711 8.915663

Discipline

Frequence\_utilisation\_PC langues lettres math

plus\_d\_une\_fois\_par\_sem 23.27711 18.84337 16.07229

une\_fois\_par\_sem\_ou\_moins 18.72289 15.15663 12.92771

> khideux1$residuals

Discipline

Frequence\_utilisation\_PC autres hg

plus\_d\_une\_fois\_par\_sem 2.5755162 -0.3256938

une\_fois\_par\_sem\_ou\_moins -2.8717217 0.3631513

Discipline

Frequence\_utilisation\_PC langues lettres

plus\_d\_une\_fois\_par\_sem -0.2647057 -0.8853872

une\_fois\_par\_sem\_ou\_moins 0.2951490 0.9872140

Discipline

Frequence\_utilisation\_PC math

plus\_d\_une\_fois\_par\_sem -1.5146545

une\_fois\_par\_sem\_ou\_moins 1.6888521

On remarque que plus les résidus s’éloignent de 0, plus l’écart à l’indépendance est marqué par rapport aux autres données. Ici, les écarts à l’indépendance sont plus importants dans les disciplines « autres » et « math ». De plus, les écarts à l’indépendance sont moins importants pour « hg » et « langues ».

On peut donc en conclure que la dépendance entre la fréquence d’utilisation du pc et la matière est forte pour autre, à contrario, l’indépendance est marquée pour les langues.

Dans « autres », il y a une forte utilisation d’un pc soit plus d’une fois par semaine, d’où la dépendance. En revanche, c’est le contraire pour « langues », il y a une très faible utilisation d’un pc, soit une fois par semaine, ou moins, d’où l’indépendance.

ETUDE 3 :

1. On va utiliser un test du Chi 2 d’homogénéité.

Comparaison entre plusieurs distributions.

**Test du chi deux d’homogénéité :**

1. Enoncé des hypothèses :

H0 : Distributions similaires des séries

H1 : Distributions différentes des séries

α=5%

> vague1 <- c(0.075,0.125,0.35,0.3,0.15)

> vague2 <- c(15,40,100,55,27)

> chisq.test(vague2,p=vague1)

Chi-squared test for given probabilities

data: vague2 X-squared = 13.273, df = 4, p-value = 0.01001

On admet H1 au risque de 5%. (p-value = 0.01001). Il y a donc une distribution différente des variables, les virus ont évolué et donc il va y avoir une adaptation au niveau des hôpitaux pour soigner les patients.