

Protocole de TP 04 :
Test de Student
(Validation et application)

Notre objectif dans cette TP c'est

- 1) Savoir comment faire une comparaison entre deux échantillons indépendants.
- 2) Comment faire le test d'homogénéité, et aussi d'établir l'efficacité entre deux échantillons.
- 3) Comment valider les conditions du test de Student.

En utilisant pour cette raison les techniques du test de comparaison des moyennes, par application du logiciel SPSS.

I) Comparaison entre deux échantillons indépendants :

On suppose qu'on a deux échantillons indépendants, avec N_1 et N_2 sont l'effectifs respectivement pour échantillon 1 et 2.

De plus soient \bar{x}_1 et \bar{x}_2 les moyennes respectivement pour échantillon 1 et 2, aussi σ_1 et σ_2 l'écart-types respectivement pour échantillon 1 et 2.

On utilise ce type du test lorsqu'on a deux échantillons, et on veut mettre en évidence l'homogénéité entre deux échantillons, et de faire la comparaison entre les moyennes des deux échantillons.

Pour cela on doit prendre un exemple d'explication :

Exemple 1

Dans des études d'anesthésie, voulant comparer l'effet de deux somnifères, on a noté les durées de sommeil qui ont suivi les injections d'une dose bien définie. Les durées étant exprimées en minutes:

Somnifère 01	170	175	187	180	190	165	175	174	173	181		
Somnifère 02	155	160	164	150	160	159	154	156	160	167	153	158

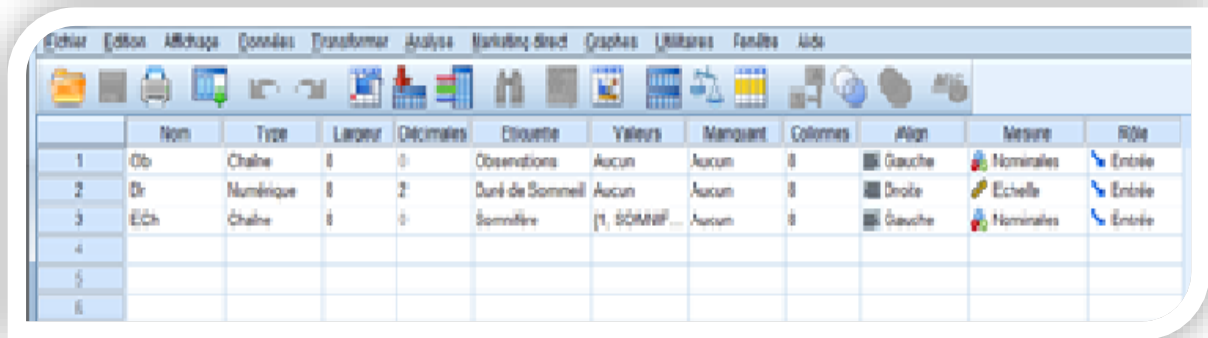
- 1) Déterminer l'objectif pour cette expérience.
- 2) Déterminer la variable qualitative qui exprime les deux échantillons, et la variable quantitative a mesuré.
- 3) Déterminer l'hypothèse nulle et alternative.
- 4) Tracer le tableau de de la statistique descriptive pour l'effectif, moyennes ainsi que l'écart-types.
- 5) Avec un risque de signification de 5%, que peut-on dire pour cette comparaison?

Solution

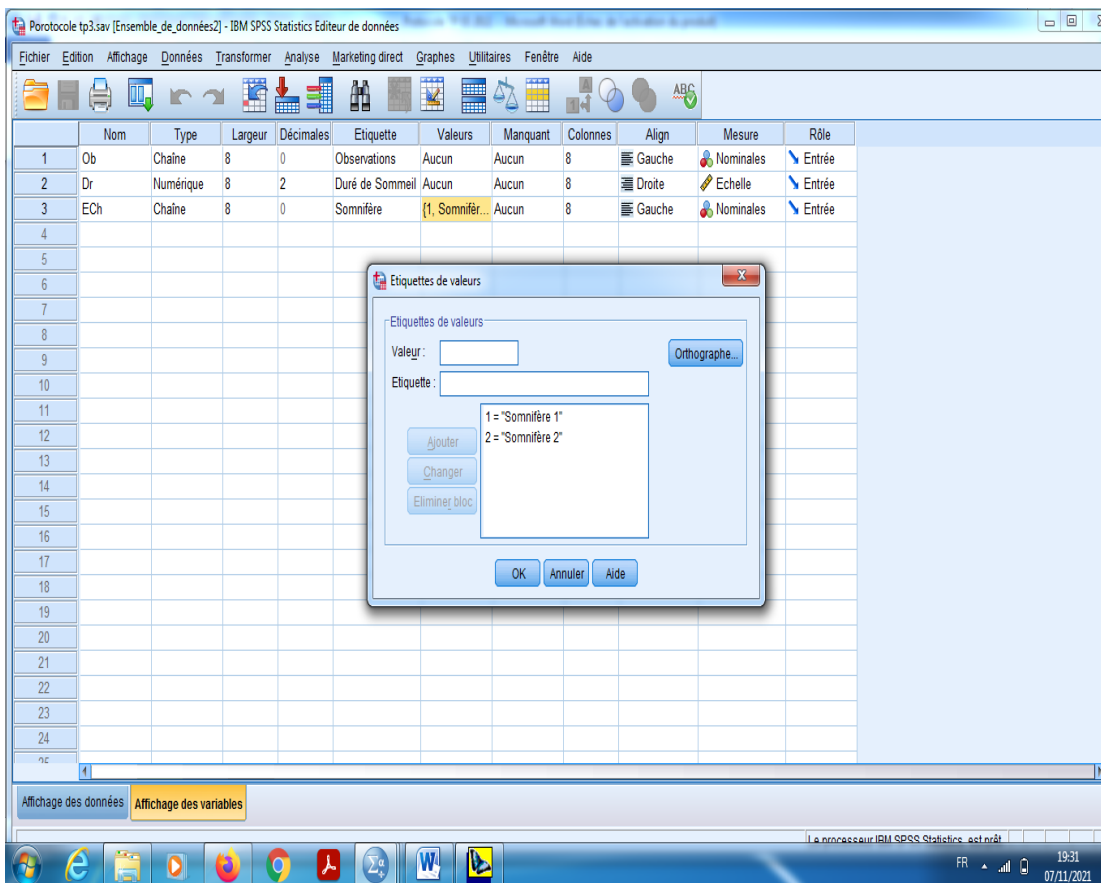
Tout d'abord il faut saisir ces données dans logiciel SPSS, on a déjà vu les techniques dans TP 1.

Nous suivant les étapes suivantes :

- a) Il faut définir dans la barre en bas « Affichage des variables » : les variables qualitative et quantitative suivantes : observations, la durée de sommeil, et les somnifères.



- b) On choisit les modalités pour la variable qualitative qui représente deux échantillons de somnifères dans « valeurs », on peut prendre comme un exemple (le code 1 pour la variable somnifère 1, et le code 2 pour la somnifère 2).



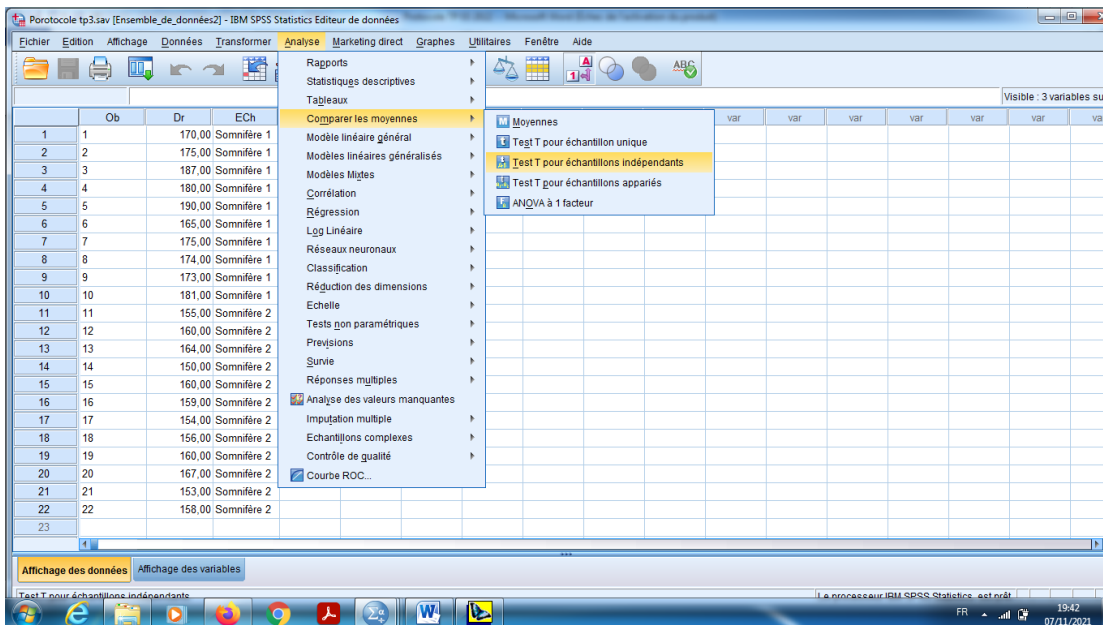
- c) On introduit les données dans la barre « Affichage des données » par SPSS.

	Ob	Dr	ECh	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1	1	170,00	1														
2	2	175,00	1														
3	3	187,00	1														
4	4	180,00	1														
5	5	190,00	1														
6	6	165,00	1														
7	7	175,00	1														
8	8	174,00	1														
9	9	173,00	1														
10	10	181,00	1														
11	11	155,00	2														
12	12	160,00	2														
13	13	164,00	2														
14	14	150,00	2														
15	15	160,00	2														
16	16	159,00	2														
17	17	154,00	2														
18	18	156,00	2														
19	19	160,00	2														
20	20	167,00	2														
21	21	153,00	2														
22	22	158,00	2														
23																	

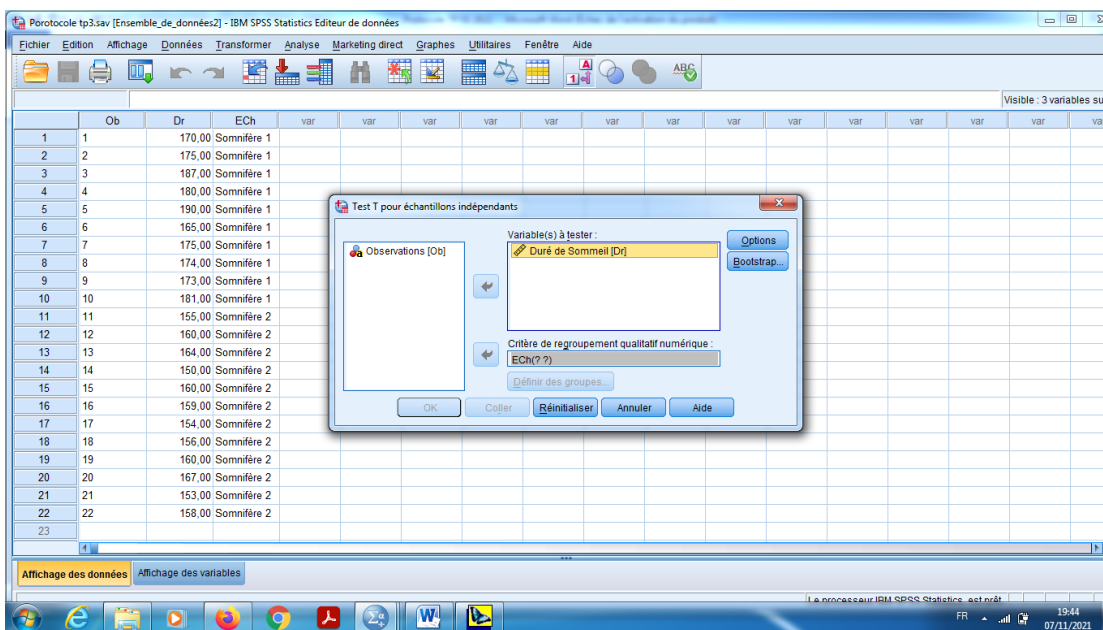
d) En cliquant sur le bouton « Etiquettes des valeurs » pour visualiser le codage des échantillons.

	Ob	Dr	ECh	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1	1	170,00	Somnifère 1														
2	2	175,00	Somnifère 1														
3	3	187,00	Somnifère 1														
4	4	180,00	Somnifère 1														
5	5	190,00	Somnifère 1														
6	6	165,00	Somnifère 1														
7	7	175,00	Somnifère 1														
8	8	174,00	Somnifère 1														
9	9	173,00	Somnifère 1														
10	10	181,00	Somnifère 1														
11	11	155,00	Somnifère 2														
12	12	160,00	Somnifère 2														
13	13	164,00	Somnifère 2														
14	14	150,00	Somnifère 2														
15	15	160,00	Somnifère 2														
16	16	159,00	Somnifère 2														
17	17	154,00	Somnifère 2														
18	18	156,00	Somnifère 2														
19	19	160,00	Somnifère 2														
20	20	167,00	Somnifère 2														
21	21	153,00	Somnifère 2														
22	22	158,00	Somnifère 2														
23																	

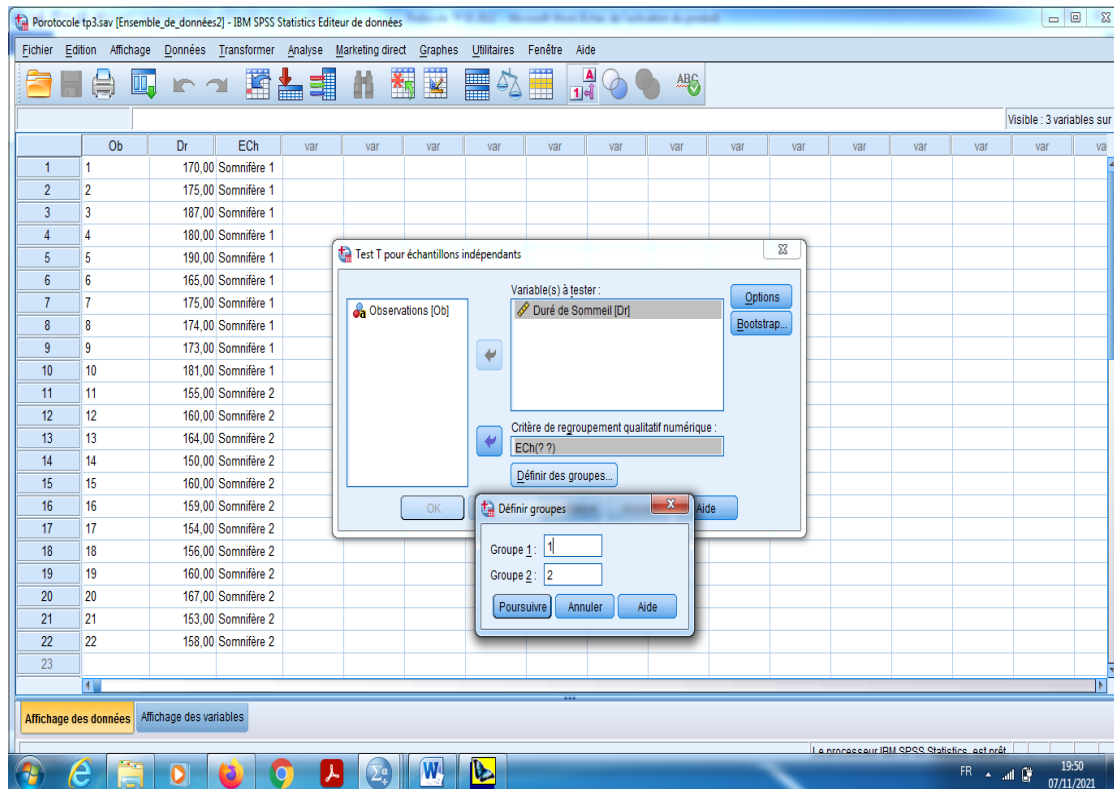
e) Pour obtenir la comparaison entre deux échantillons de somnifères, en cliquant sur le bouton « Analyse » qui se trouve dans la barre des outils, et choisir « Comparer les moyenne », et puis « Test T pour échantillons indépendants ».



- f) On choisit dans « la variable à tester » la variable Durée de Sommeil.
- g) On choisit la variable qualitative (Echantillons=somnifères) dans la barre « Critère des regroupements qualitatif numérique ».
- h) On choisit aussi sur « Définir les groupes », et on prend pour le somnifère 1 (groupe 1) le code 1, et on prend pour le somnifère 2 (groupe 2) le code 2.



- i) Dans le Bouton « Options » on garde le pourcentage pour l'intervalle de confiance le seuil de signification 95%, en cliquant sur « poursuivre ».



En fin en cliquant sur « Ok ».

Interprétation des résultats

1) Notre objectif est de comparer les deux types des somnifères, pour cela on utilise le teste de comparaison entre les moyennes (Test d'échantillons indépendants).

2) Proposition d'hypothèses :

Hypothèse nulle, H0 :

« Il n'y a pas une différence significative entre les moyennes de somnifères $\bar{x}_1 = \bar{x}_2$ »,

« les deux échantillons sont significativement homogènes »

« Il n'y a pas effet du type d'anesthésie sur la durée de sommeil ».

Hypothèse alternative, H1 :

« Il y a une différence significative entre les moyennes de somnifères $\bar{x}_1 = \bar{x}_2$ »,

« les deux échantillons ne sont pas significativement homogènes »

« Il y a effet du type d'anesthésie sur la durée de sommeil ».

Règle de la décision :

Pour la décision, on utilise souvent la règle suivante

Si Signification (bilatérale) inférieure à 0,05, alors on rejette H0 (Il y a une différence significative).

Si Signification (bilatérale) supérieure à 0,05, alors on accepte H0 (Pas différence significative).

Dans la table 1 «_Statistique de groupe» on remarque qu'il y a 10 observations dans la somnifère 1 et 12 pour la somnifère 2, et les moyennes 177 pour le somnifère 1 et 158 pour le somnifère 2, de plus pour l'écart-type : 7,60117 et 4,74821 respectivement pour la somnifère 1 et 2.

Test-t

[Ensemble_de_données2] F:\L3 Biologie\TPs 2021-2022\Porotocole tp3.sav

Statistiques de groupe

Somnifère	N	Moyenne	Ecart-type	Etreur standard moyenne
Duré de Sommeil SOMNIFERE 1	10	177,0000	7,60117	2,40370
Somnifère 2	12	158,0000	4,74821	1,37069

Test d'échantillons indépendants

		Test de Levene sur l'égalité des variances		Test-t pour égalité des moyennes						
		F	Sig.	t	ddl	Sig. (bilatérale)	Différence moyenne	Différence écart-type	Intervalle de confiance 95% de la différence	
									Inférieure	Supérieure
Duré de Sommeil	Hypothèse de variances égales	2,408	,136	7,161	20	,000	19,00000	2,65330	13,46531	24,53469
	Hypothèse de variances inégales			6,867	14,546	,000	19,00000	2,76705	13,08611	24,91389

DATASET ACTIVATE Ensemble_de_données2.
DATASET CLOSE Ensemble_de_données1.
DATASET ACTIVATE Ensemble_de_données2.
SAVE OUTFILE='F:\L3 Biologie\TPs 2021-2022\Porotocole tp3.sav'
/COMPRESSED.
NEW FILE.

On s'intéresse seulement sur (Test T pour l'égalité des moyennes).

Comme l'échantillon de petites tailles, alors on utilise la statistique

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{\sigma_1} + \frac{1}{\sigma_2}}}$$

Avec S c'est la variance pondéré.

Alors Statistique t=7,161.

DDL=n1+n2-2=20.

Signification (bilatérale)=0,000.

Pour cela on peut voir dans les tableaux obtenues dans les résultats de SPSS.

Signification (bilatérale)=0,000<0,05. Donc on rejette H_0 , alors il y a une différence significative entre la moyenne des somnifères, les deux échantillons ne sont pas homogènes, c'est-à-dire que le type d'anesthésie a un effet sur la durée de sommeil.

D'autre part, on peut conclure aussi que la moyenne se trouve dans un intervalle qui s'appelle « Intervalle de confiance de la différence », par la forme suivante

$Dm \in [13,46; 24,53]$, alors il est clair que les deux bornes **sont positives**, donc on en conclure que la différence entre les deux somnifères (Somnifère 1-Somnifère 2) est positive, c'est – à –dire que avec un taux de 95% la Somnifère 1 est plus efficace que Somnifère 2.

Remarque :

Pour $Dm \in [A; B]$, on distinct trois cas possibles :

- 1) Si A, B sont positives, alors $Dm > 0$, alors $Dm = m_1 - m_2 > 0$, alors $m_1 > m_2$, d'où anesthésie de la Somnifère 1 est plus efficace que celle de la Somnifère 2.
- 2) Si A, B sont négatives, alors $Dm < 0$, alors $Dm = m_1 - m_2 < 0$, alors $m_1 < m_2$, d'où anesthésie Somnifère 2 est plus efficace que celle de la Somnifère 1.
- 3) Si A, négative et B positives, alors $Dm = 0$, alors $Dm = m_1 - m_2 = 0$, alors $m_1 = m_2$, d'où Somnifère 1 est homogène avec Somnifère 2 (On accepte H_0).

II) Validation du test de Student

Pour appliquer le test de comparaison entre deux échantillons (test de Student), il faut que les conditions suivantes soient valides :

- a) Une seule variable quantitative mesurée, et une variable qualitative avec deux modalités (pour test de Student).
- b) La distribution soit gaussienne (c'est-à-dire que la variable quantitative à mesuré suit la loi de Gauss).
- c) L'échantillon est homogène pour la variable quantitative. (toutes les valeurs de la variable sont proches de la moyenne), on ne peut pas trouver des valeurs plus loin que la moyenne.

Pour cela notre objectif dans cette TP est de savoir comment vérifier la condition de normalité (c'est la condition de validité d'un test de comparaison des moyennes (comme déjà vu dans TP4), et aussi de vérifier l'homogénéité d'échantillon.

En utilisant pour cette raison les techniques du test de normalité de Shapiro-Wilk (le cas des petits échantillons) ou bien test de normalité de Kolmogorov-Smirnov (le cas des grands échantillons), par utilisation du logiciel SPSS.

II-1) Test de normalité (Shapiro-Wilk, Kolmogorov-Smirnov)

On utilise ce type du test lorsqu'on a une variable quantitative.

Pour cela on doit prendre un exemple d'explication :

Exemple 1

Dans des études d'anesthésie, voulant tester la normalité de la distribution ainsi l'homogénéité (la durée de somnifères), on a noté les durées de sommeil qui ont suivi les injections d'une dose bien définie. Les durées étant exprimées en minutes :

Somnifère 01	170	175	187	180	190	165	175	174	173	181		
Somnifère 02	155	160	164	150	160	159	154	156	160	167	153	158

- 1) Déterminer l'objectif pour cette expérience.
- 2) Déterminer la variable qualitative qui exprime les deux échantillons, et la variable quantitative a mesuré.
- 3) Déterminer l'hypothèse nulle et alternative pour la normalité et pour l'homogénéité.
- 4) Tracer le tableau de de la statistique descriptive.
- 5) Avec un risque de signification 6%, que peut-on dire sur la normalité de la variable quantitative ? et sur l'homogénéité sur les modalités de la variable quantitative ?

Remarque :

Pour test de normalité, on utilise test de Kolmogorov-Smirnov lorsque la taille des échantillons est très élevée ($n > 30$).

On utilise test de Shapiro-Wilk lorsque la taille des échantillons est assez petit ($n < 30$).

Remarque:

Pour la décision, on utilise souvent la règle suivante :

Si Signification (bilatérale) inférieure à $\alpha\%$, alors on rejette H_0 .

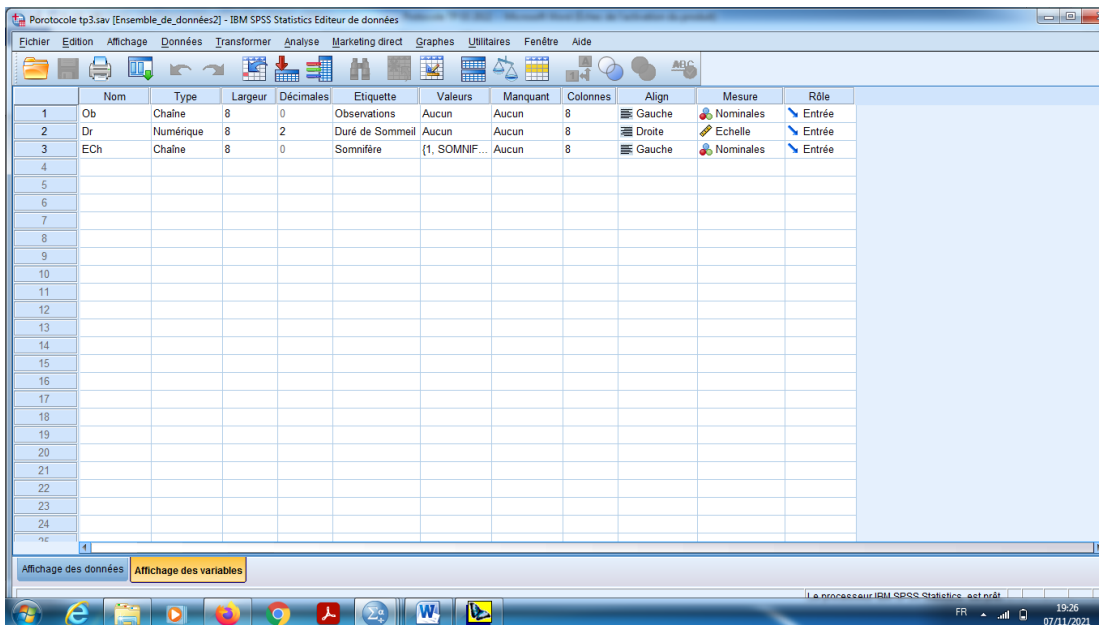
Si Signification (bilatérale) supérieure à $\alpha\%$, alors on accepte H_0 .

Solution d'exemple

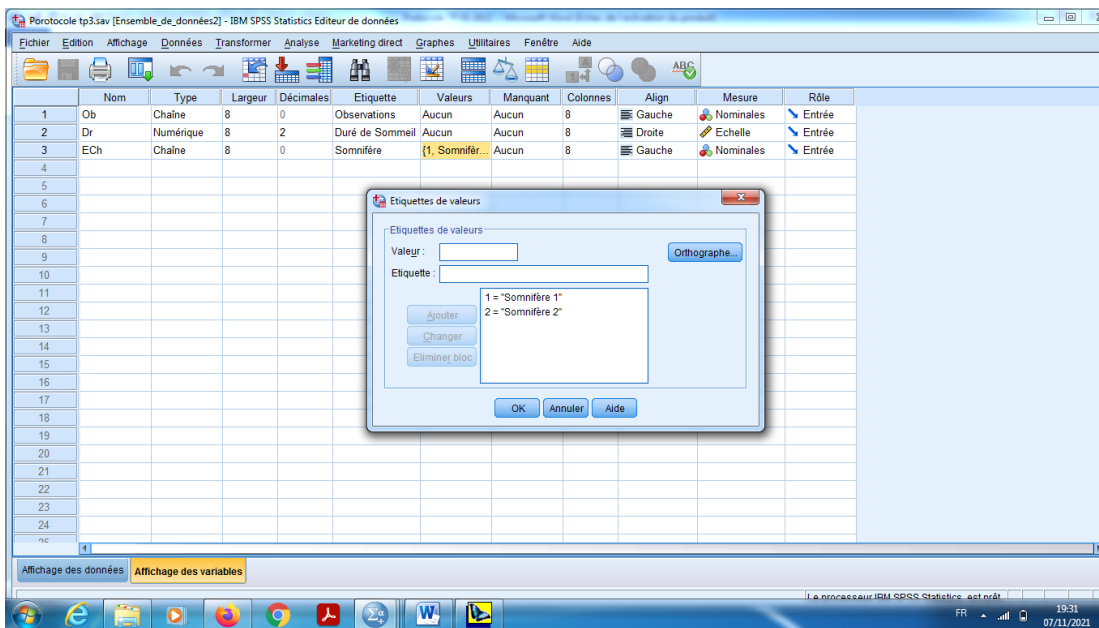
Tout d'abord il faut saisir ces données dans logiciel SPSS, on a déjà vu les techniques dans TP 1.

En suivant les étapes suivantes :

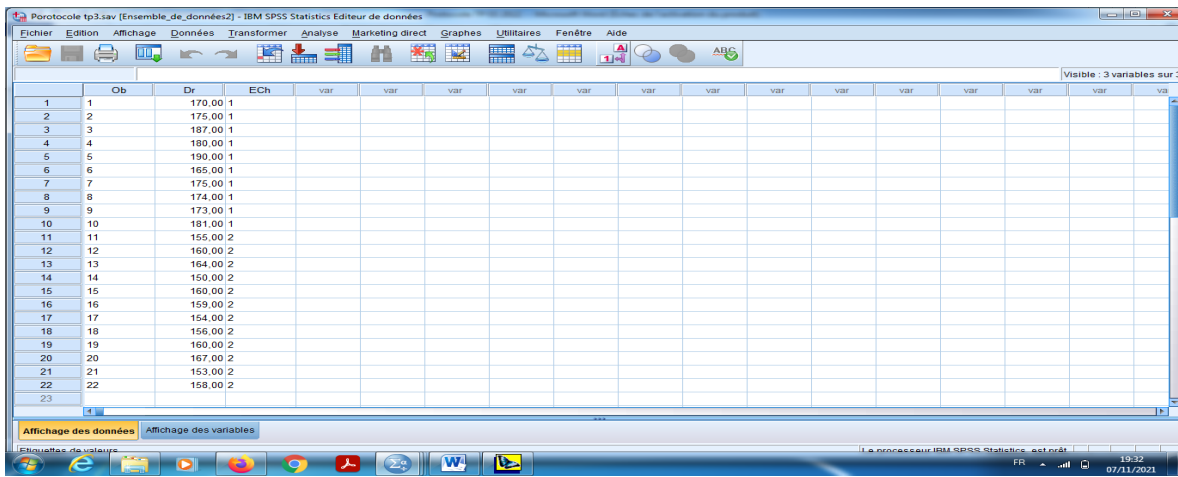
- a) Il faut définir dans la barre en bas « Affichage des variables » : les variables qualitative et quantitative suivantes : observations, la durée de sommeil, et les somnifères.



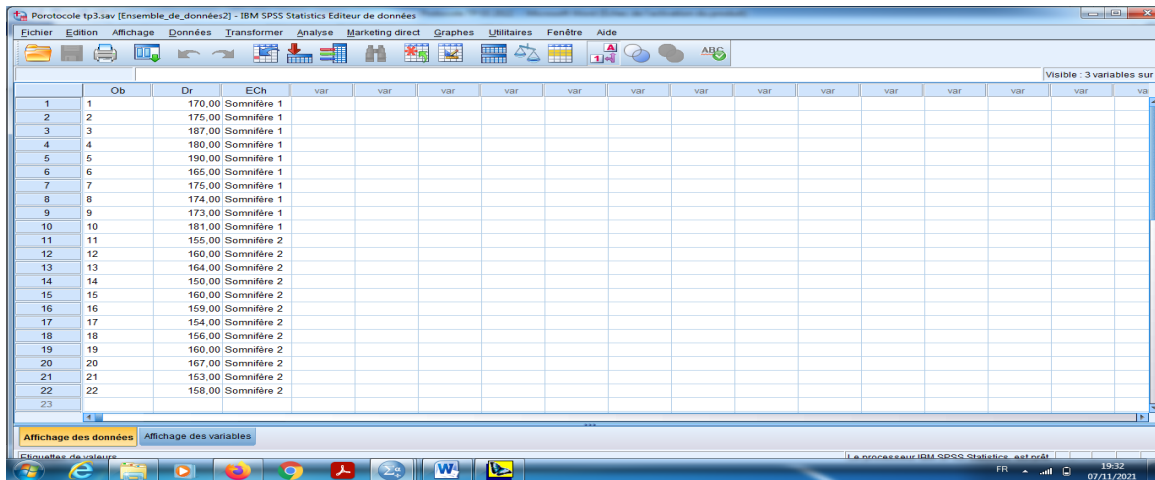
b) On choisit les modalités pour la variable qualitative qui représente les deux échantillons de somnifères dans « valeurs », on peut prendre comme un exemple (la valeur 1 pour somnifère 1 et la valeur 2 pour somnifère 2).



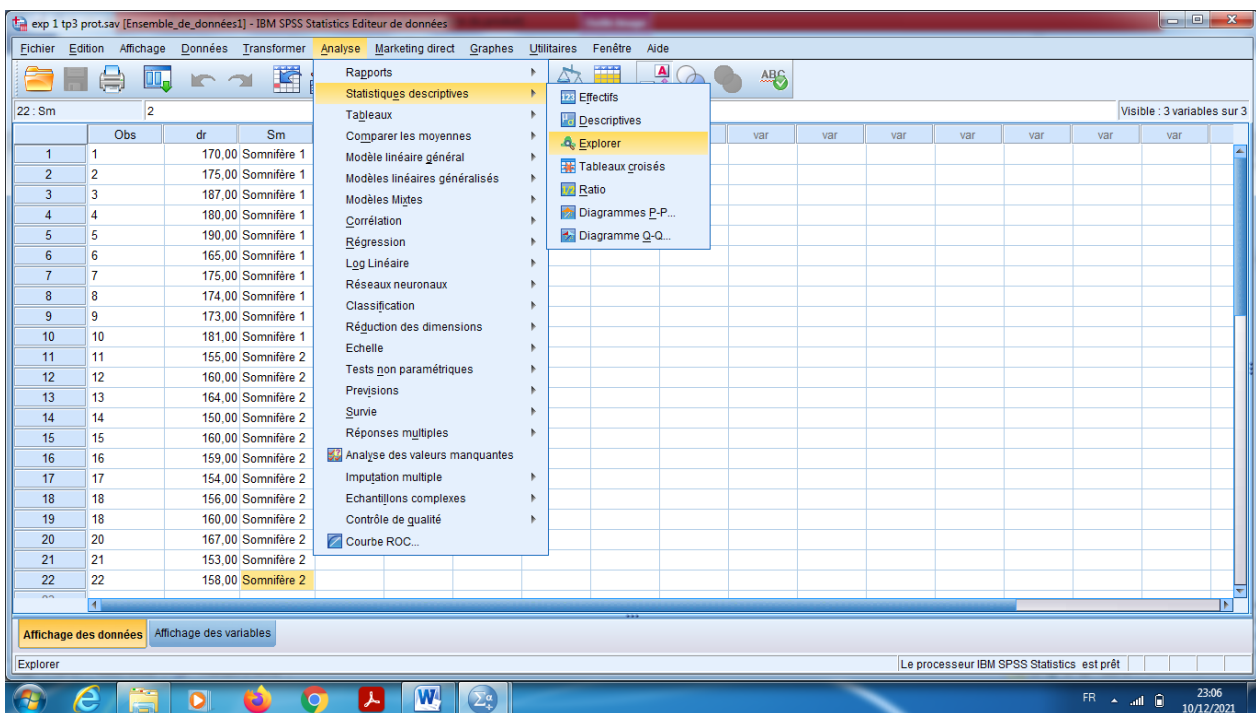
c) On saisit les données dans la barre « Affichage des données ».



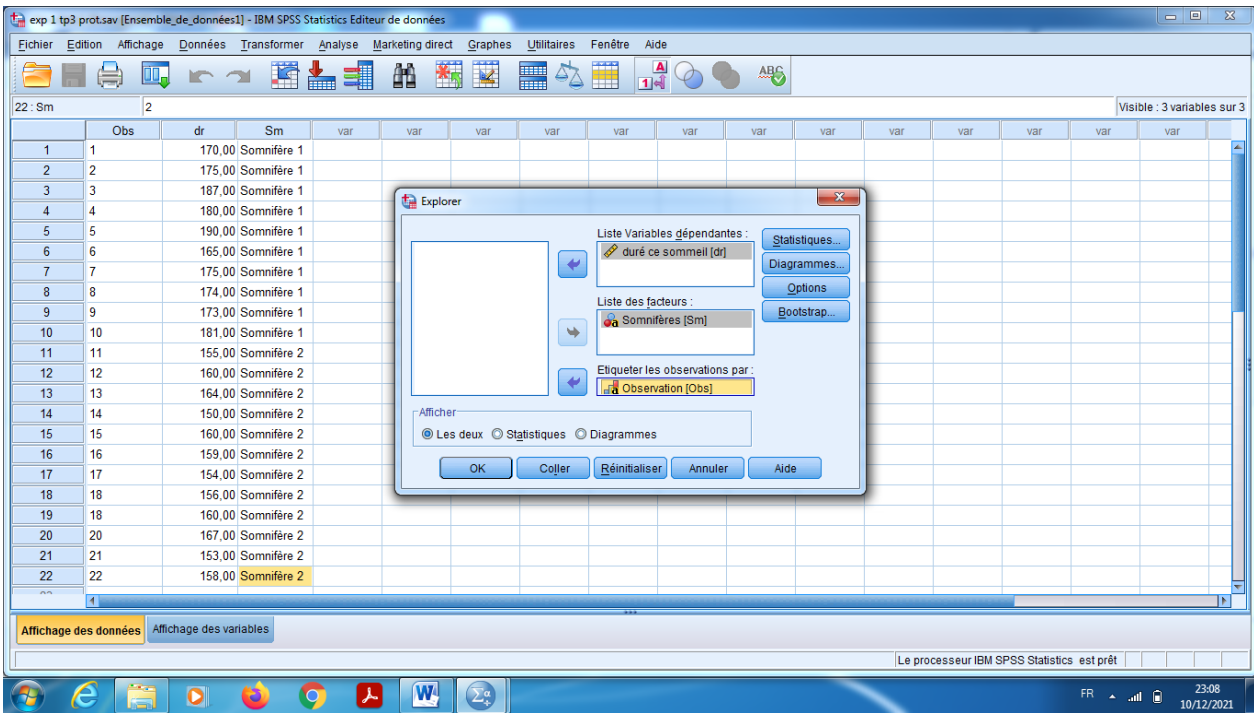
d) En cliquant sur le bouton « Etiquettes des valeurs » pour visualiser le codage des échantillons.



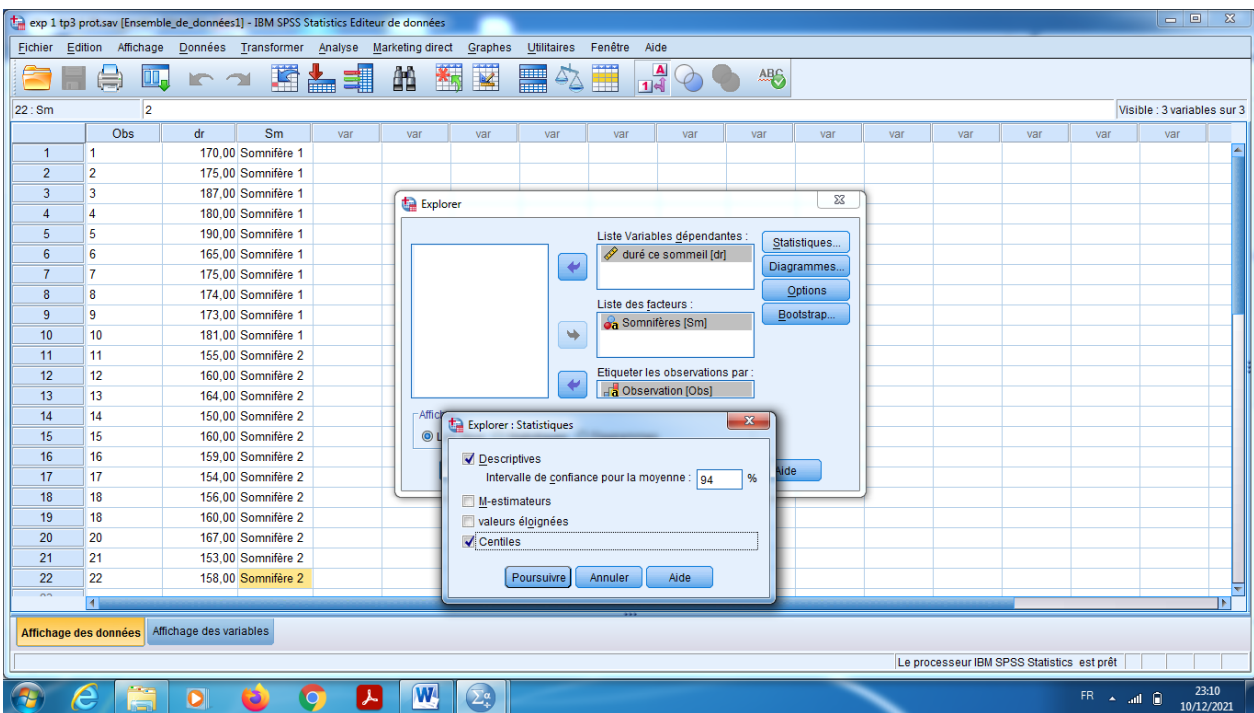
e) En suit, pour obtenir le test de normalité (c'est-à-dire que la variable quantitative à mesuré suit la loi de Gauss), en cliquant sur le bouton « Analyse » qui se trouve dans la barre des outils, et choisir « Statistiques descriptives », et puis « Explorer ».



- f) On choisit dans « liste variables dépendantes» la variable quantitative mesurable « Durée de Sommeil».
- g) On choisit dans la barre « liste des facteurs» la variable qualitative (Echantillons=somnifères)



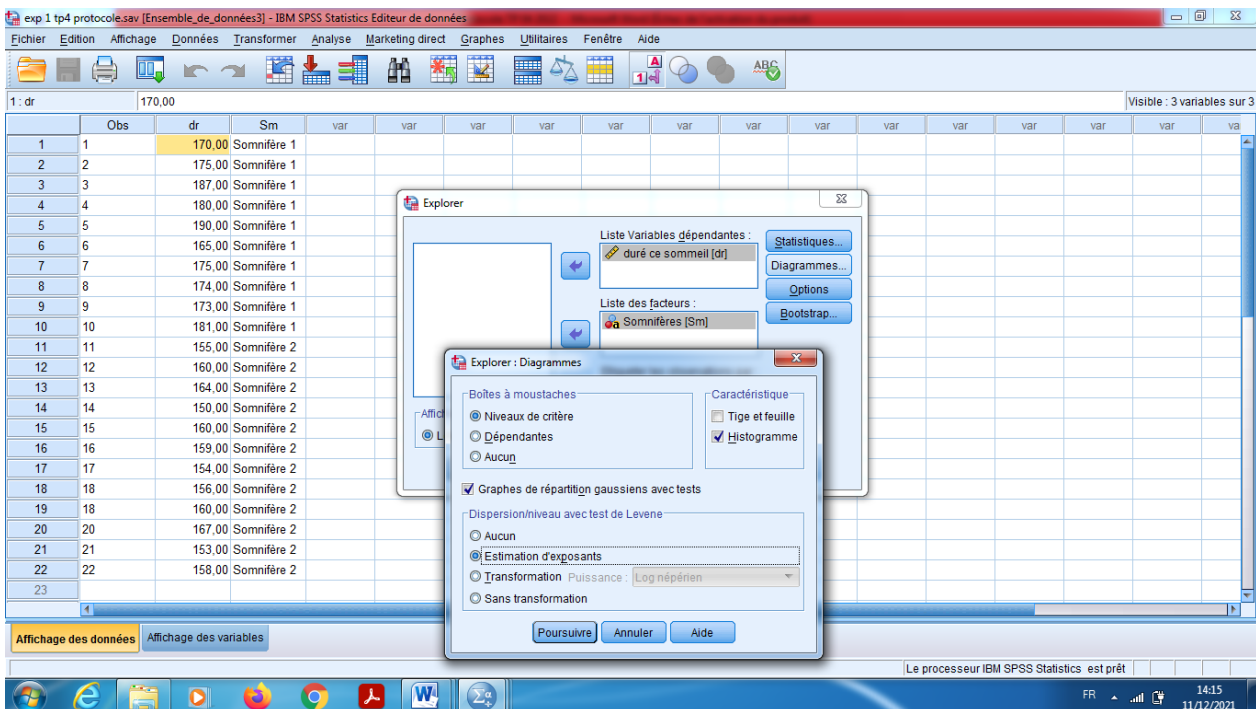
- h) Dans le Bouton « Statistiques » on peut changer le pourcentage pour l'intervalle de confiance le seuil de signification 94%.
- i) On choisit « Centiles » pour visualiser les quartiles (Q1, Q2, Q3).
- j) En cliquant sur « poursuivre ».



- k) Sur le bouton « Diagrammes », on garde dans « Boîtes à moustaches » le niveau de critère, et dans « Caractéristiques » en cliquant sur le choix Histogramme, et aussi on coche sur le choix « Graphes de répartitions gaussiennes avec tests » pour validé test de

normalité, en fin sur «Dispersion /Niveau avec test de Levene » on prend le choix « Estimation d'exposants » pour validé test d'homogénéité.

l) En cliquant sur « poursuivre ».



En fin en cliquant sur « Ok ».

Interprétation des résultats

1) Notre objectif est de valider les conditions de comparaison entre les deux types des somnifères (Test d'échantillons indépendants) qui s'appelle test de normalité c'est à dire que la variable quantitative (Duré de sommeil) est Gaussienne, ainsi que le test d'homogénéité d'échantillon pour la variation.

3) Proposition d'hypothèses :

a) Pour la normalité :

Hypothèse nulle, H0 : « la distribution est gaussienne».

Hypothèse alternative, H1 : « la distribution n'est pas gaussienne».

b) Pour l'homogénéité

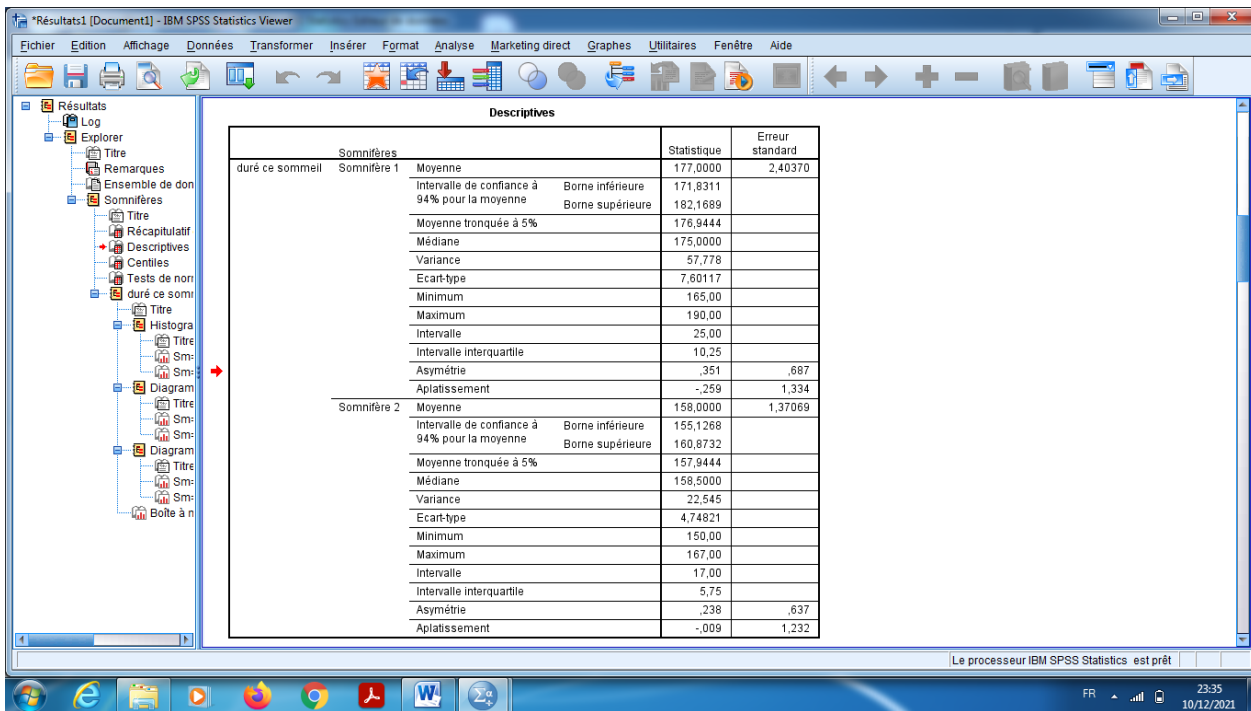
Hypothèse nulle, H0 : « la distribution est Homogène pour la variation».

Hypothèse alternative, H1 : « la distribution n'est pas homogène pour la variation».

4) Dans la table 1 «Statistique descriptive» on remarque que 'il y a 10 observations dans les somnifères 1, et 12 observations pour la somnifère 2.

Dans la table 3 : Les moyennes 177 pour le somnifère 1 et 158 pour la somnifère 2, de plus pour l'écart-type c'est 7,60117 et 4,74821 respectivement pour la somnifère 1 et 2.

Pour les médianes 175 et 158,5 respectivement pour le somnifère 1 et 2.



5) **Pour le test de normalité**, il faut choisir le test de Shapiro-Wilk.

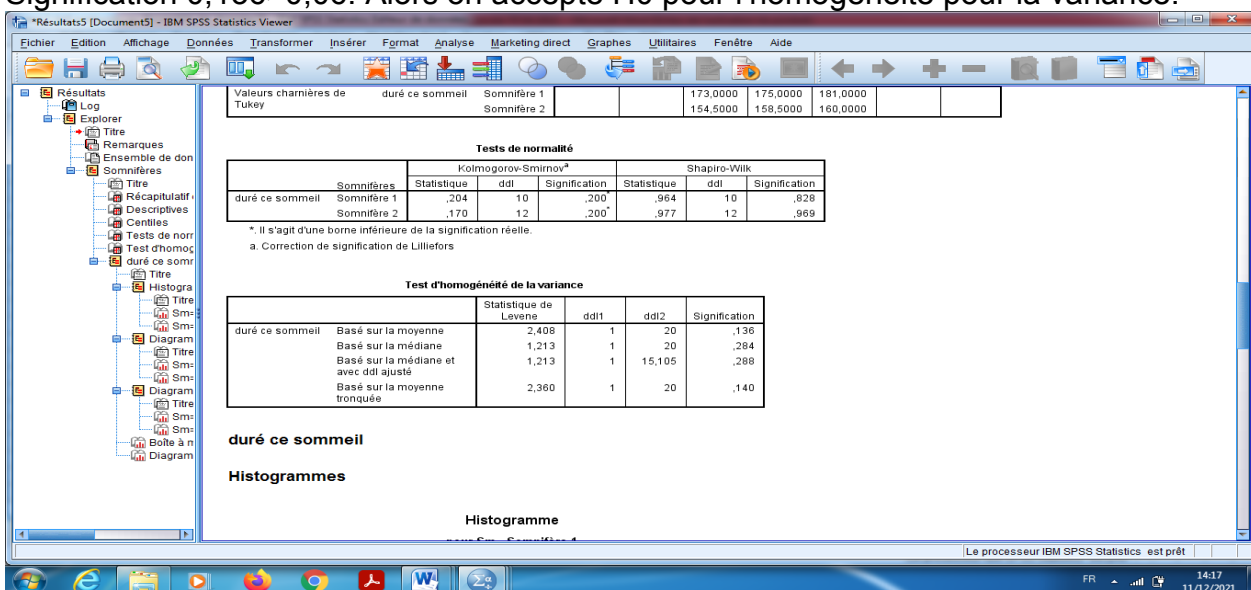
On remarque que dans somnifère 1 ($Sig=0,828 > 0,06$) on accepte H_0 , alors la distribution pour échantillon 1 est gaussienne.

D'autre part dans somnifère 2 ($Sig=0,969 > 0,06$) on accepte H_0 , alors la distribution pour échantillon 2 est gaussienne.

II-2) Test d'homogénéité de la variation

Mais pour l'homogénéité de la variance pour la variable quantitative « Durée de Sommeil », on peut établir à partir de la deuxième table que la durée de sommeil qui basé sur la moyenne que la distribution est homogène, c'est-à-dire qu'il n ya pas des valeurs critiques pour cette variable, (échantillon est homogène par rapport à la variable durée).

Signification $0,136 > 0,06$. Alors on accepte H_0 pour l'homogénéité pour la variance.

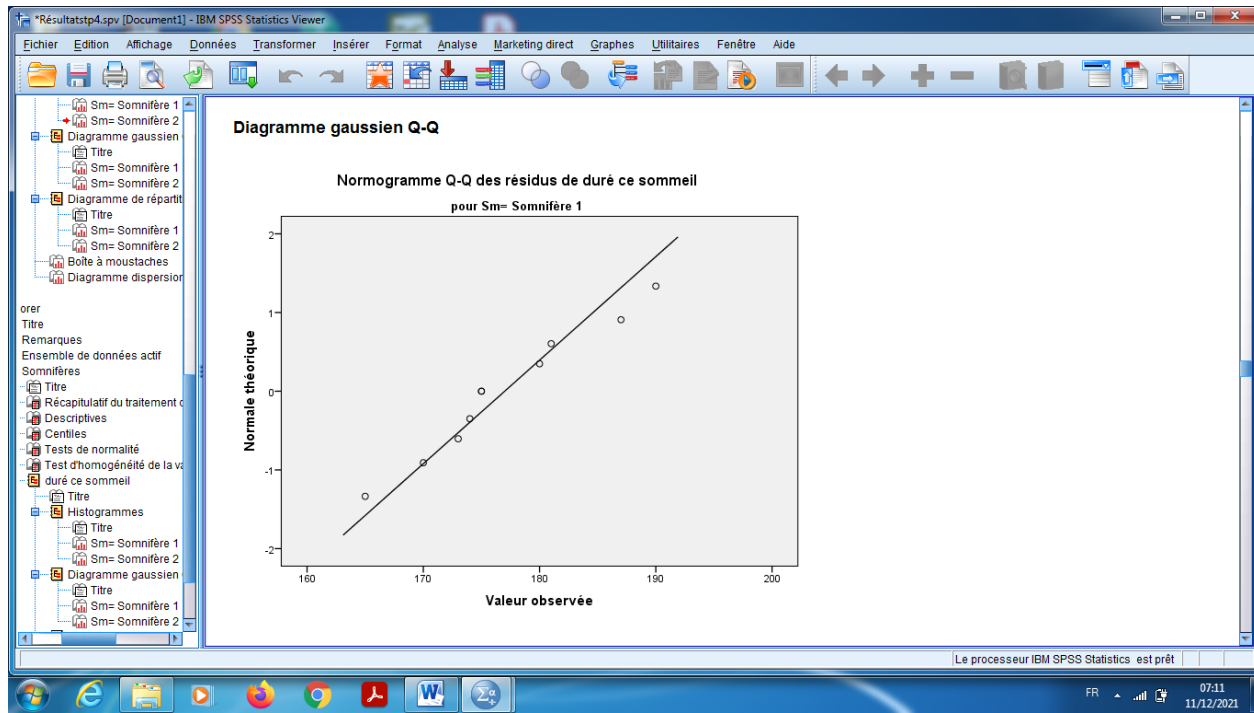


II-3) Test de normalité graphiquement

D'autre part on peut vérifier la validité du test de Student (test de comparaison entre les moyennes), en utilisant quelques type des graphes.

Diagramme Q-Q gaussien

Pour Somnifère 1 :



Pour Somnifère 1 : On compare la droite de normalité avec nuage des points, on remarque facilement que les pluparts des points se trouve au voisinage de la droite.

Conclusion :

D'après de tout ce qui précède, on conclut que la distribution du paramètre quantitative « Durée de sommeil » est gaussien au taux de confiance 94%, de plus l'échantillon est homogène pour la variation, alors on peut appliquer le test de comparaison des moyennes.