

Protocole de TP 05 :

Validation du test de Student

Test de normalité (Shapiro-Wilk ou de Kolmogorov-Smirnov)

Test d'homogénéité de la variance

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 soit 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.

I) Test de normalité

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 la 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 0,06, alors on rejette H_0 .

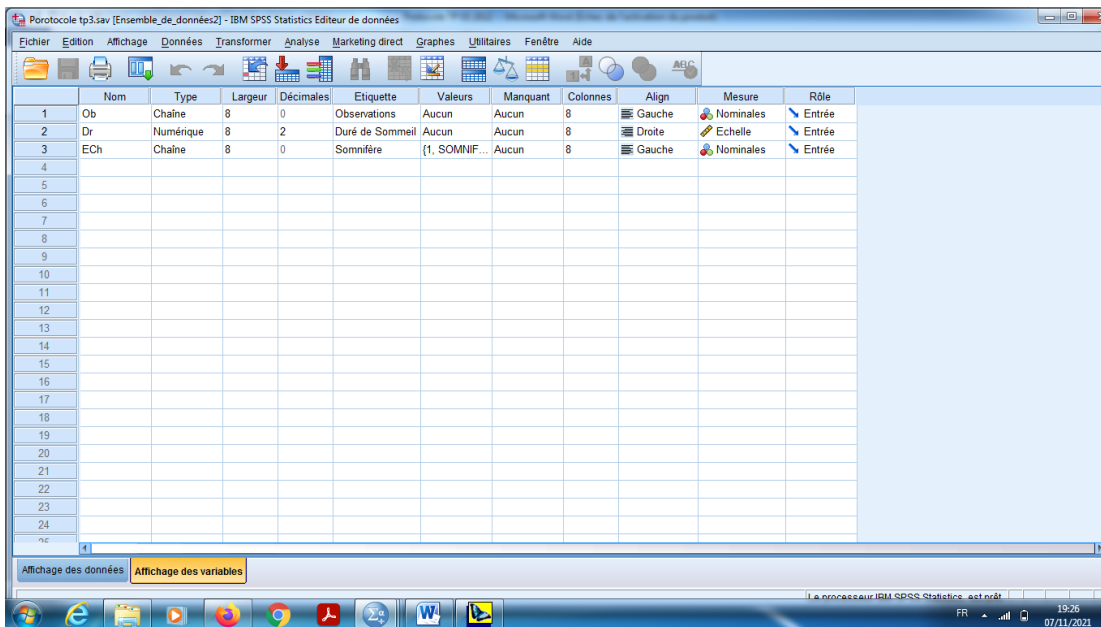
Si Signification (bilatérale) supérieure à 0,06, alors on accepte H_0 .

Solution d'exemple

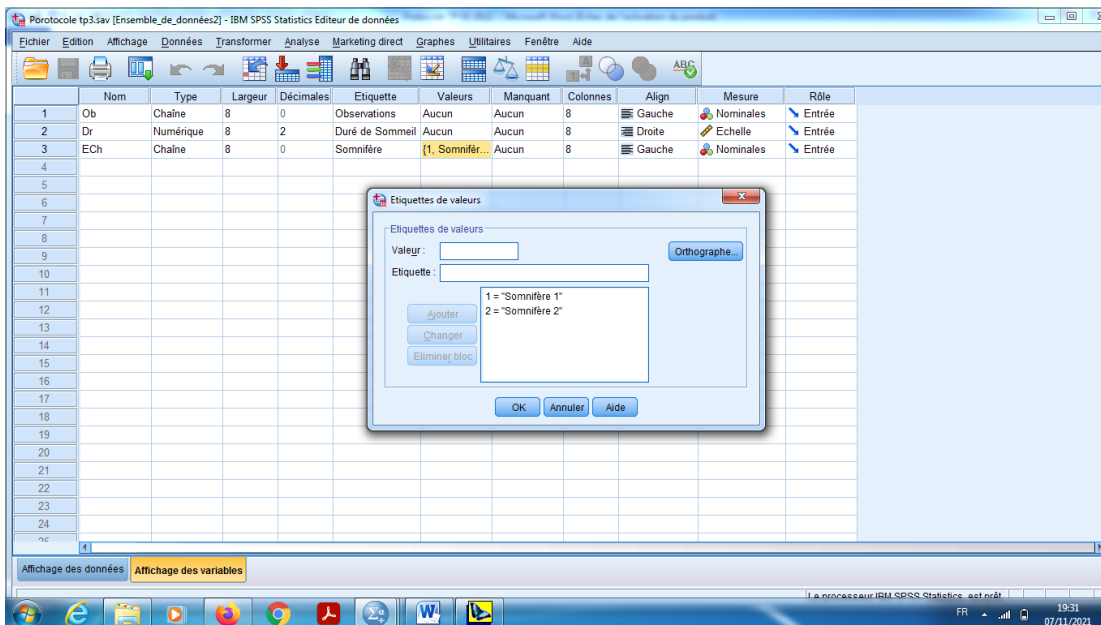
Tout d'abord il faut introduire ces données dans SPSS, on a déjà vu les techniques dans TP 1.

En suivant les étapes suivantes :

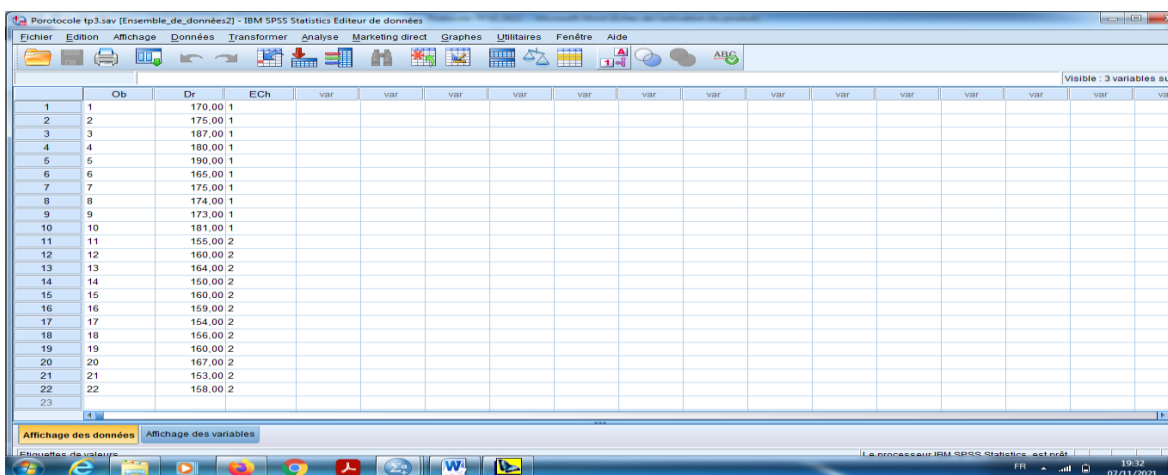
- a) Il faut définir dans la barre en bas « Affichage des variables » : les variables qualitative et quantitative suivantes : observations, la durée de sommeil, et les somnifères.



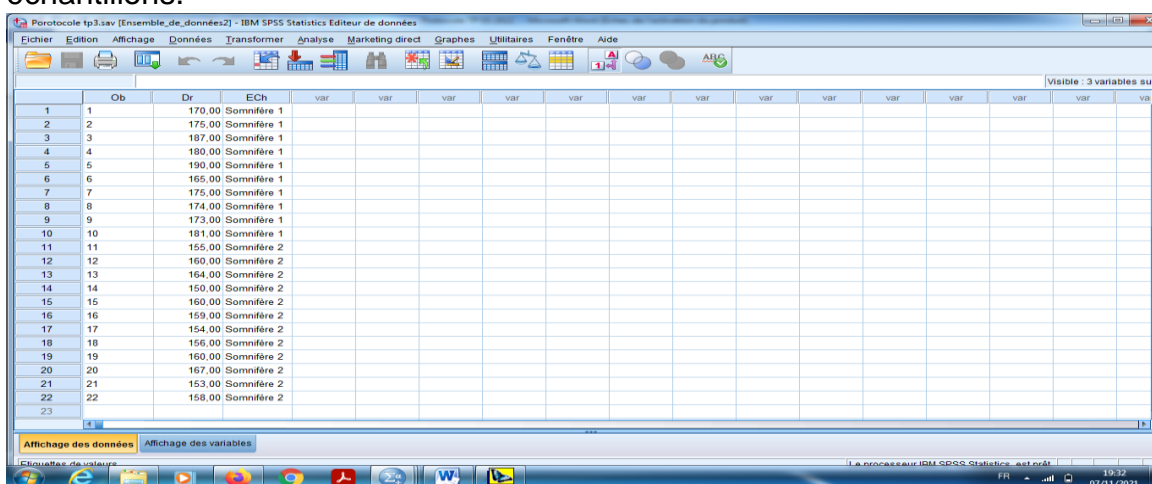
- b) On choisit les modalités pour la variable qualitative qui représente les deux échantillons de somnifères dans « valeurs », on peut prendre comme un exemple (la valeur 1 pour somnifère 1 et la valeur 2 pour somnifère 2).



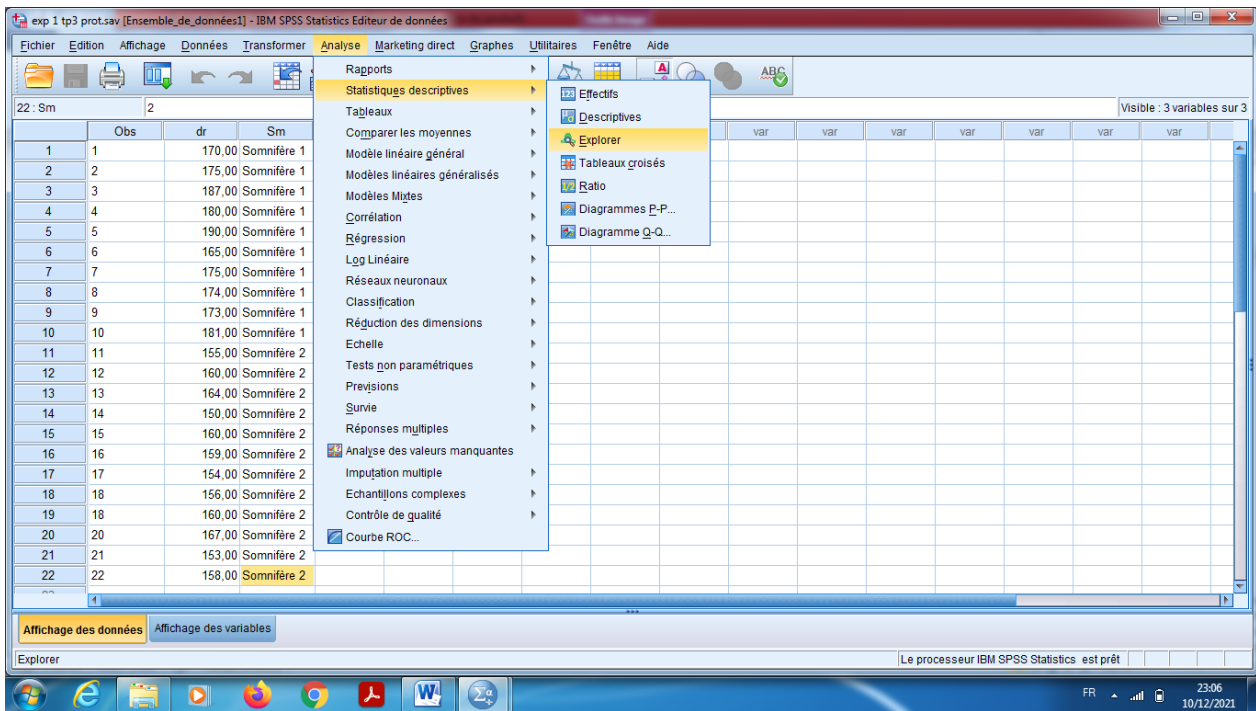
c) On introduit les données dans la barre « Affichage des données ».



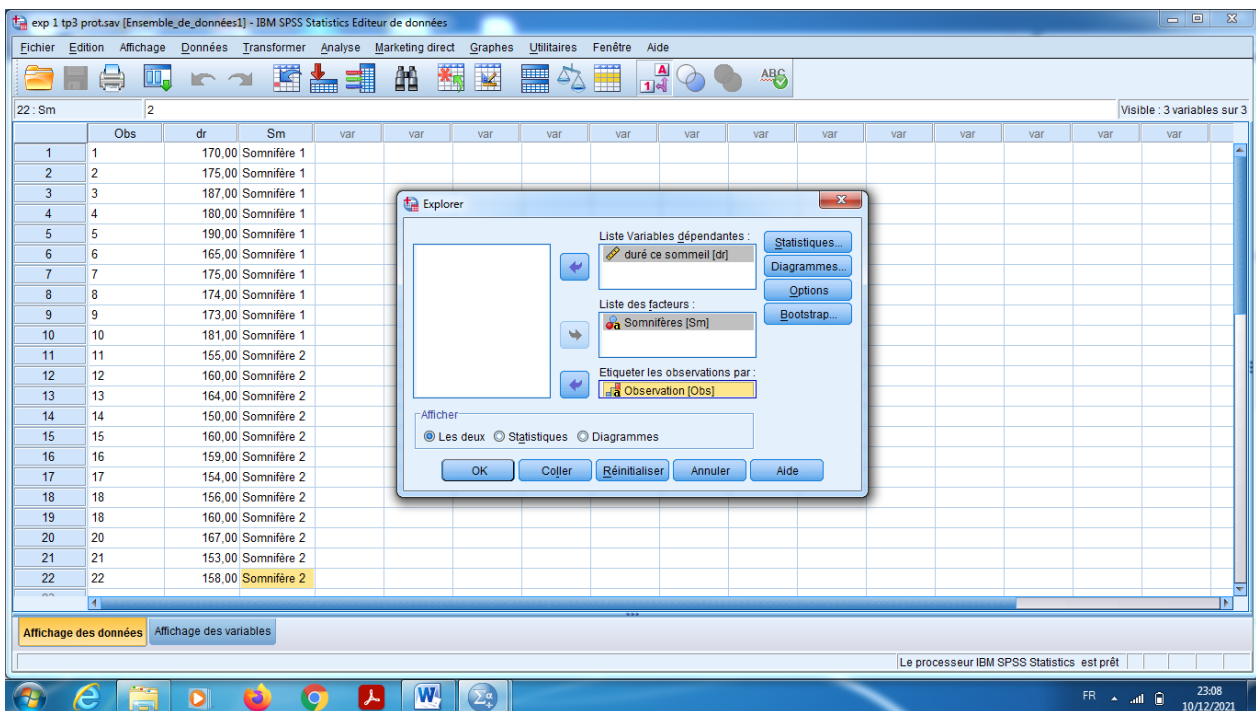
d) En cliquant sur le bouton « Etiquettes des valeurs » pour visualiser le codage des échantillons.



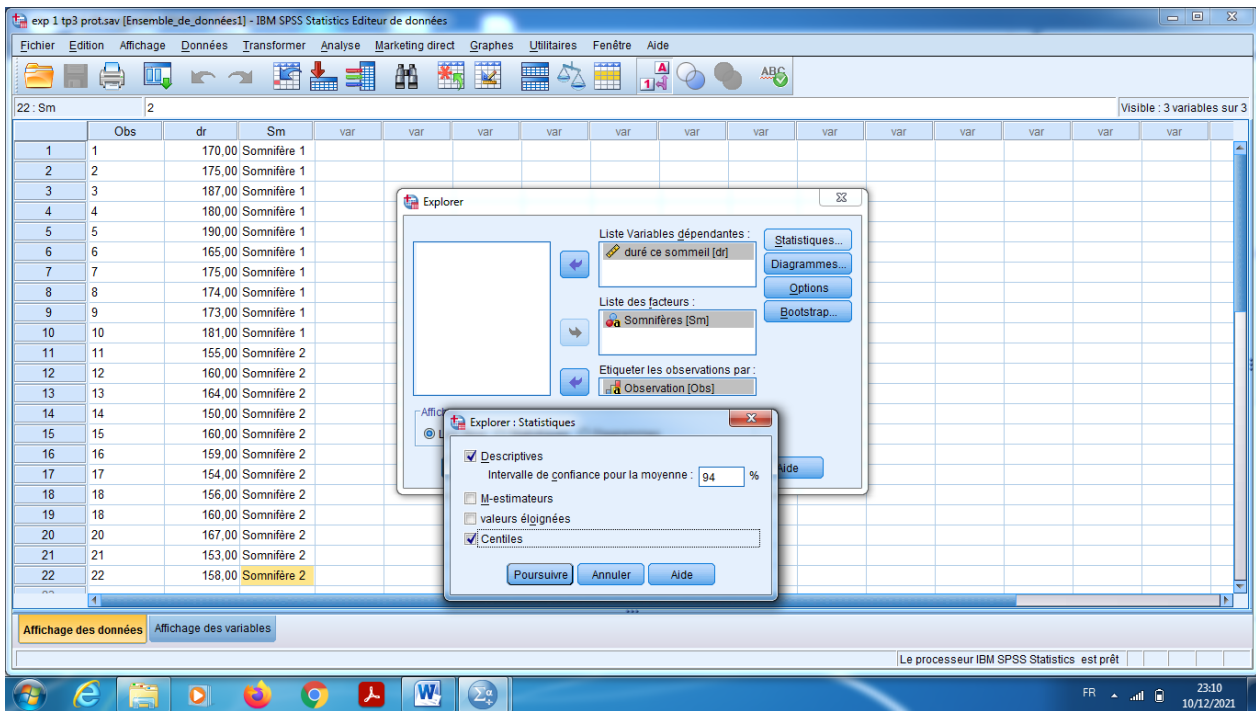
e) En suit, pour obtenir le test de normalité (c'est-à-dire que la variable quantitative à mesuré suit la loi de Gauss), en cliquant sur le bouton « Analyse » qui se trouve dans la barre des outils, et choisir « Statistiques descriptives », et puis « Explorer ».



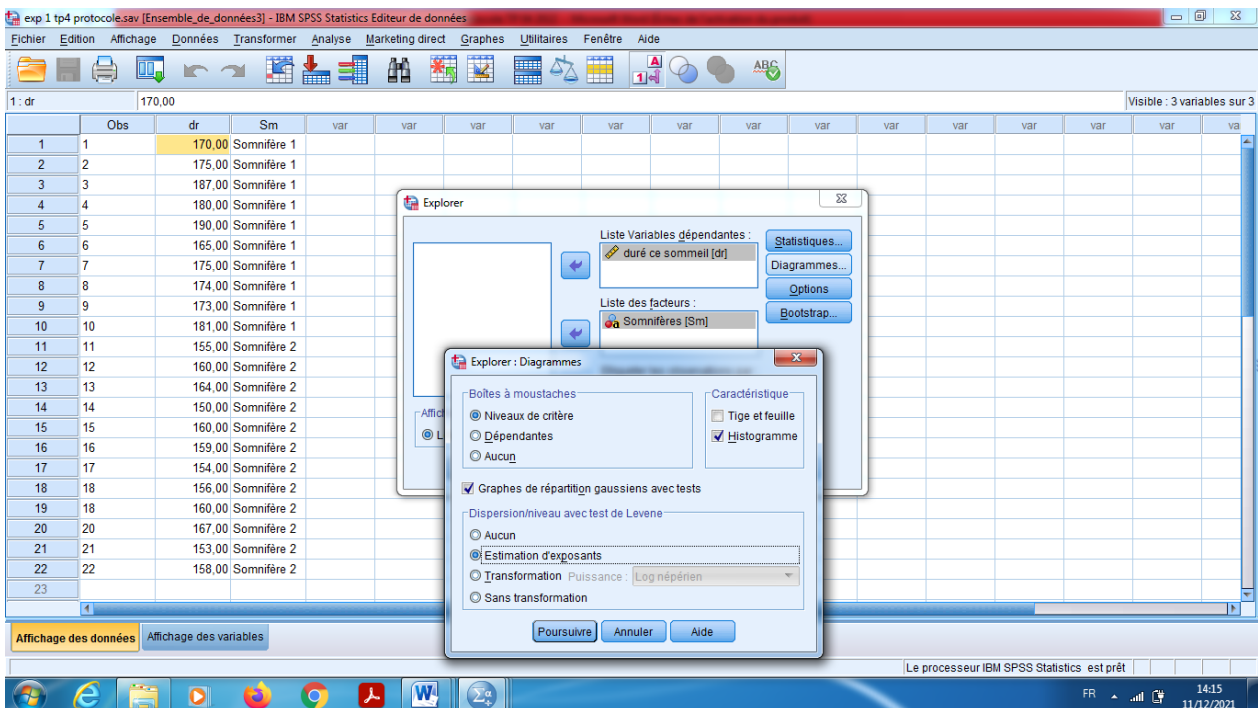
- f) On choisit dans « liste variables dépendantes » la variable quantitative mesurable « Durée de Sommeil ».
- g) On choisit dans la barre « liste des facteurs » la variable qualitative (Echantillons=somnifères)



- 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 répartition gaussiens 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.

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 ».

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

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.

Somnifères		Statistique	Erreur standard	
duré ce sommeil	Somnifère 1	Moyenne	177,0000	
		Intervalle de confiance à 94% pour la moyenne	Borne inférieure 171,8311 Borne supérieure 182,1689	
		Moyenne tronquée à 5%	176,9444	
		Médiane	175,0000	
		Variance	57,778	
		Ecart-type	7,60117	
		Minimum	165,00	
		Maximum	190,00	
		Intervalle	25,00	
		Intervalle interquartile	10,25	
		Asymétrie	,351	,687
		Aplatissement	-,259	1,334
		Somnifère 2	Moyenne	158,0000
			Intervalle de confiance à 94% pour la moyenne	Borne inférieure 155,1268 Borne supérieure 160,8732
Moyenne tronquée à 5%	157,9444			
Médiane	158,5000			
Variance	22,545			
Ecart-type	4,74821			
Minimum	150,00			
Maximum	167,00			
Intervalle	17,00			
Intervalle interquartile	5,75			
Asymétrie	,238		,637	
Aplatissement	-,009		1,232	

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 H0, alors la distribution pour échantillon 1 est gaussienne.

D'autre part dans somnifère 2 (Sig=0,969>0,06) on accepte H0, alors la distribution pour échantillon 2 est gaussienne.

Test d'homogénéité de la variance

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'y a pas des valeurs critiques pour cette variable, (échantillon est homogène par rapport à la variable durée).

The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays the following tables:

Valeurs caractéristiques de Tukey

	duré ce sommeil	Sommeil 1	Sommeil 2				
		173,0000	175,0000	181,0000			
		154,5000	158,5000	160,0000			

Tests de normalité

duré ce sommeil	Sommeil 1	Sommeil 2	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
			Statistique	ddl	Signification	Statistique	ddl	Signification
			,204	10	,200	,964	10	,828
			,170	12	,200 [*]	,977	12	,969

*. Il s'agit d'une borne inférieure de la signification réelle.
a. Correction de signification de Lilliefors

Test d'homogénéité de la variance

duré ce sommeil	Basé sur la moyenne	Statistique de Levene		Signification
		ddl1	ddl2	
		2,408	1 20	,136
	Basé sur la médiane	1,213	1 20	,284
	Basé sur la médiane et avec ddl ajusté	1,213	1 15,105	,288
	Basé sur la moyenne tronquée	2,360	1 20	,140

duré ce sommeil

Histogrammes

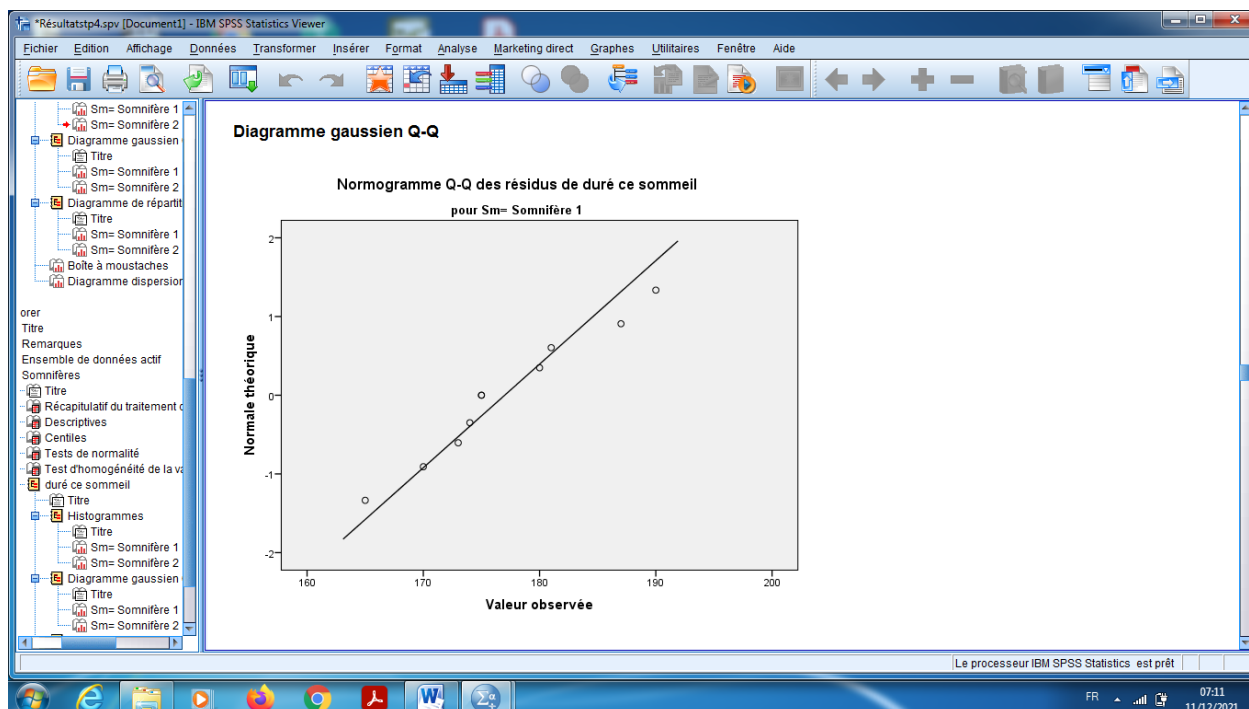
Histogramme

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 types de graphes.

Diagramme Q-Q gaussien

Pour Sommeil 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.