

Protocole de TP 04 :

Validation du test de Student

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

Test d'homogénéité de la variance

Pour validé test de comparaison entre deux échantillons (test de Student), ou test d'ANOVA pour plus que deux échantillons indépendants, 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) ou bien une variable qualitative avec trois modalités et de plus (pour test d'ANOVA).
- b) La distribution soit gaussienne (suit la loi Normale).
- c) L'échantillon est homogène pour la variable quantitative. (toutes les valeurs 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 TP3 et TP 3 bis), 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 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 des variables quantitatives.

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 de 6%, que peut-on dire pour la normalité? Et sur l'homogénéité ?

Remarque:

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 ($n < 30$).

Pour la décision, on utilise souvent la règle suivante

Si Signification (bilatérale) inférieure à 0,05, alors on rejette H0

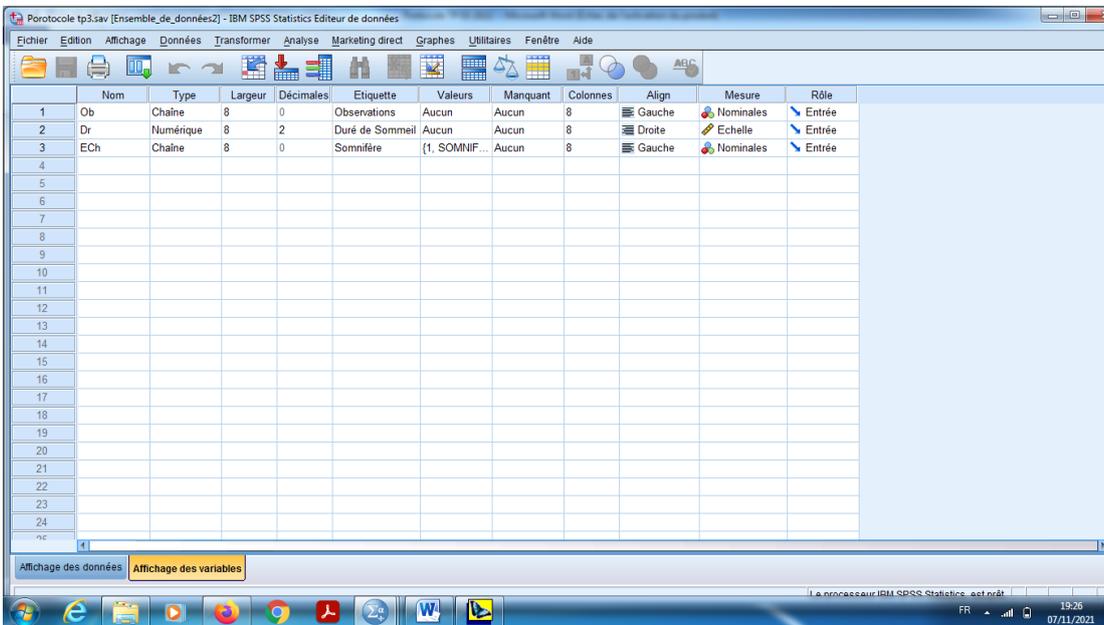
Si Signification (bilatérale) supérieure à 0,05, alors on accepte H0

Solution

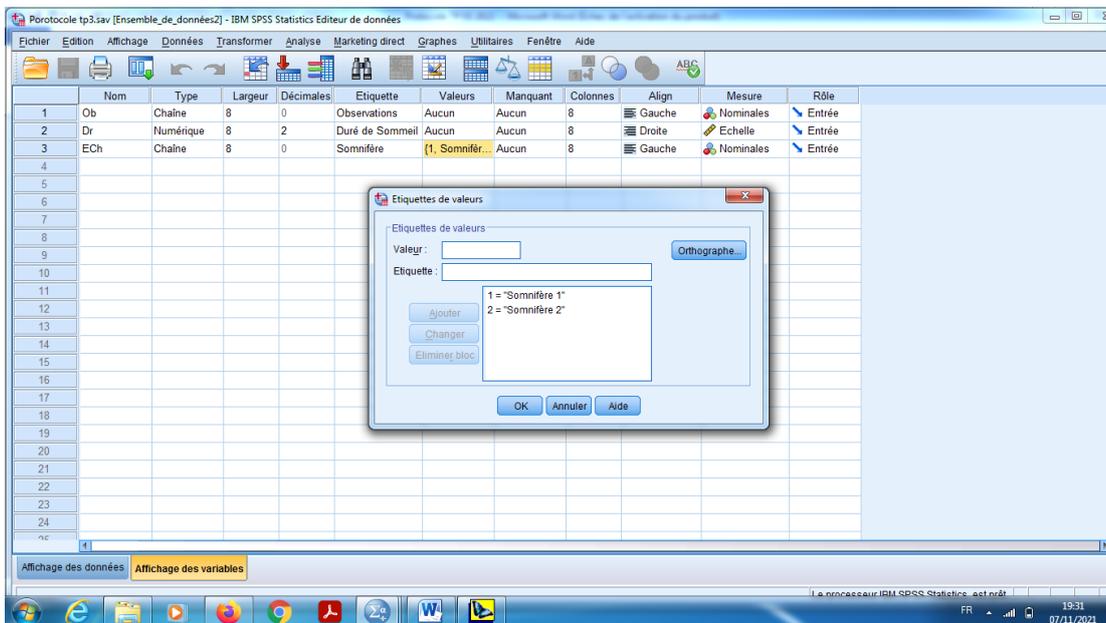
Tout d'abord il faut entrer ces données dans SPSS, on a déjà vu les techniques dans TP 1.

Nous suivant les étapes suivantes :

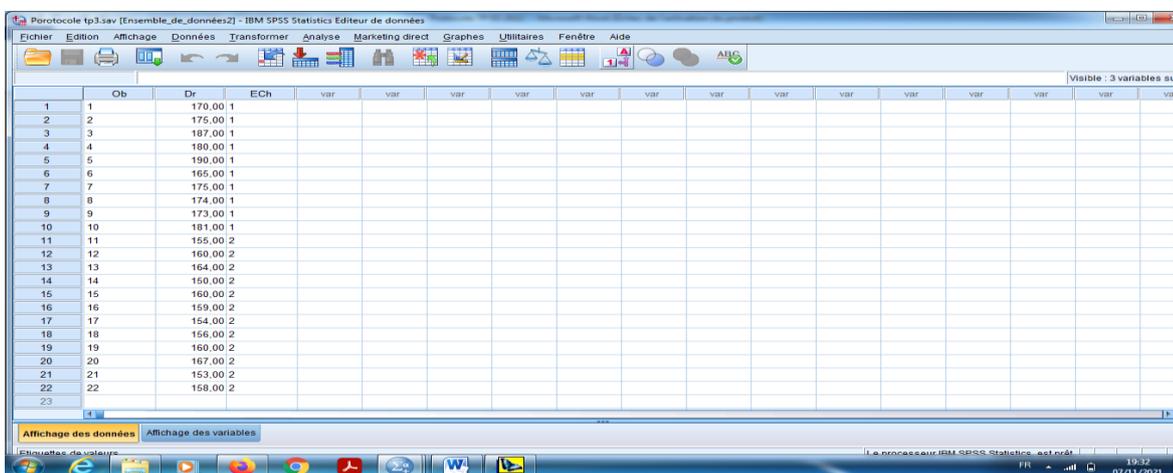
Il faut définir dans la barre en bas « Affichage des variables » : les variables qualitative et quantitative suivantes : observations, la durée de sommeil, et les somnifères.



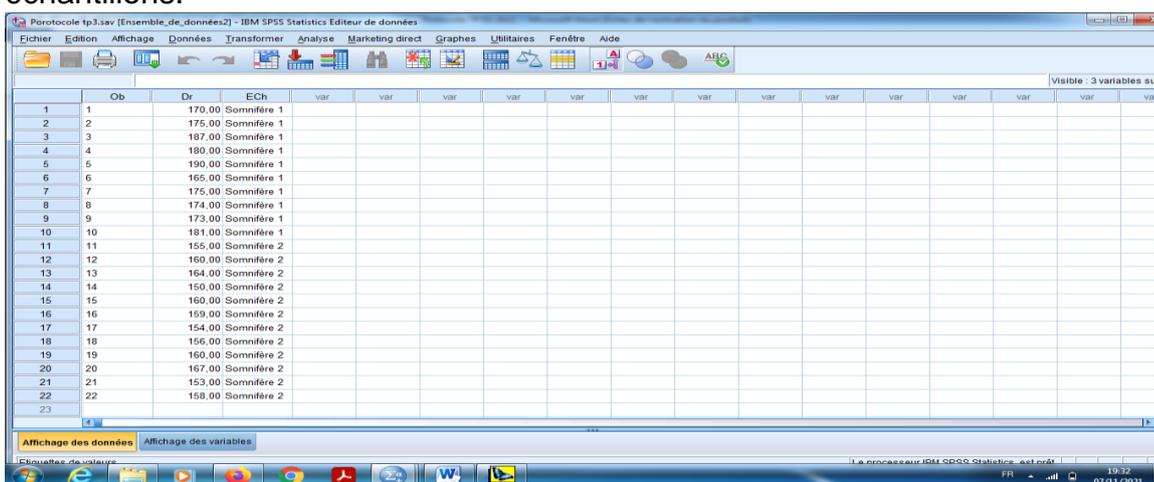
On choisit les modalités pour la variable qualitative qui représente les deux échantillons de somnifères dans « valeurs », on peut prendre comme un exemple (la valeur 1 pour somnifère 1 et la valeur 2 pour somnifère 2).



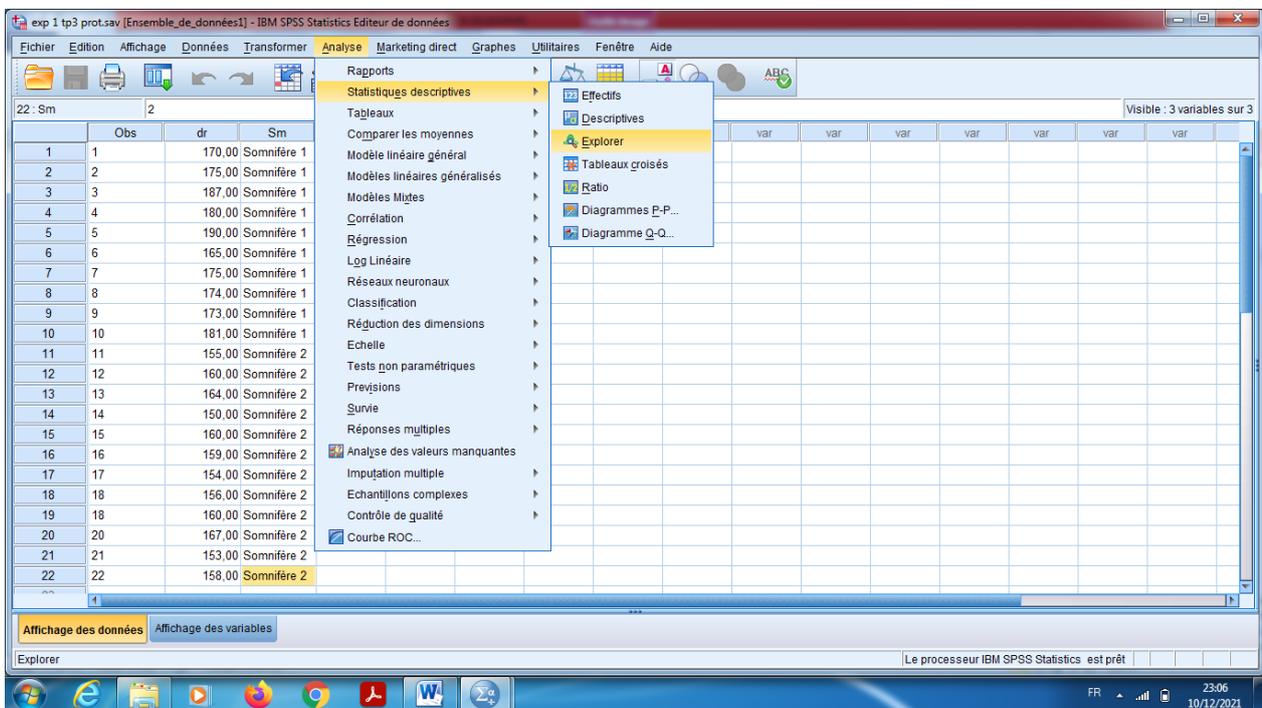
On introduit les données dans la barre « Affichage des données ».



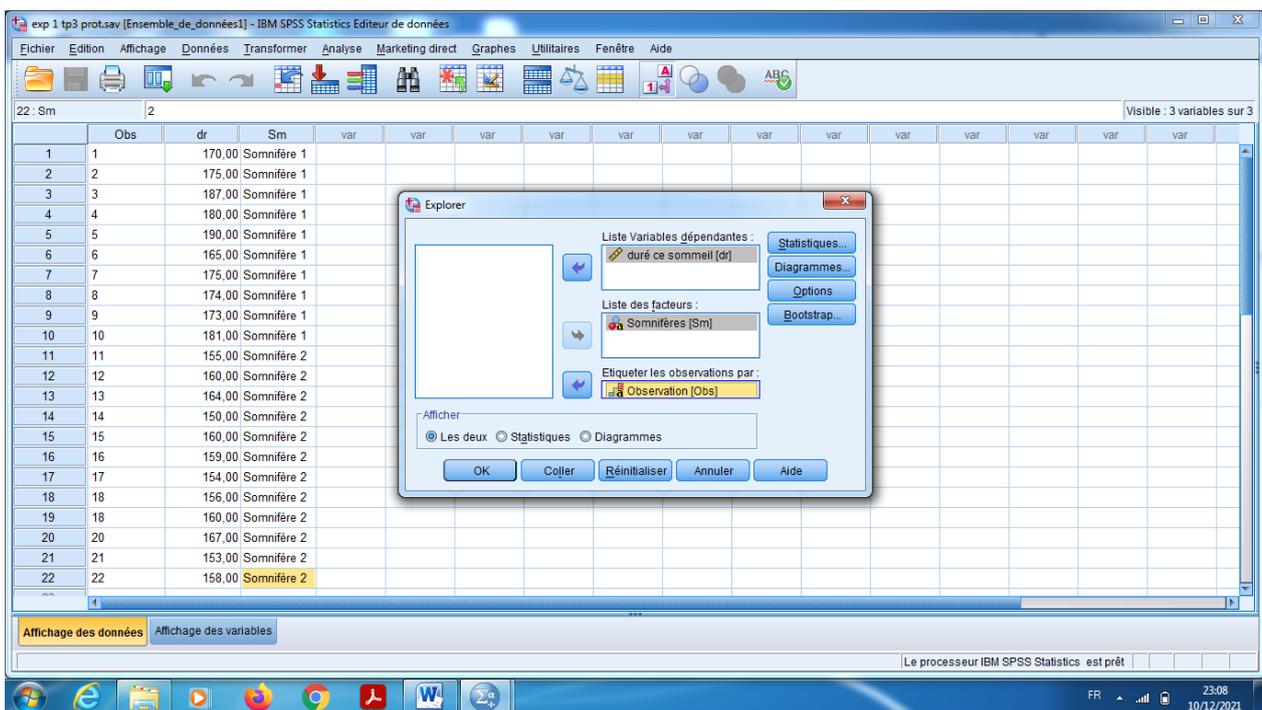
En cliquant sur le bouton « Etiquettes des valeurs » pour visualiser le codage des échantillons.



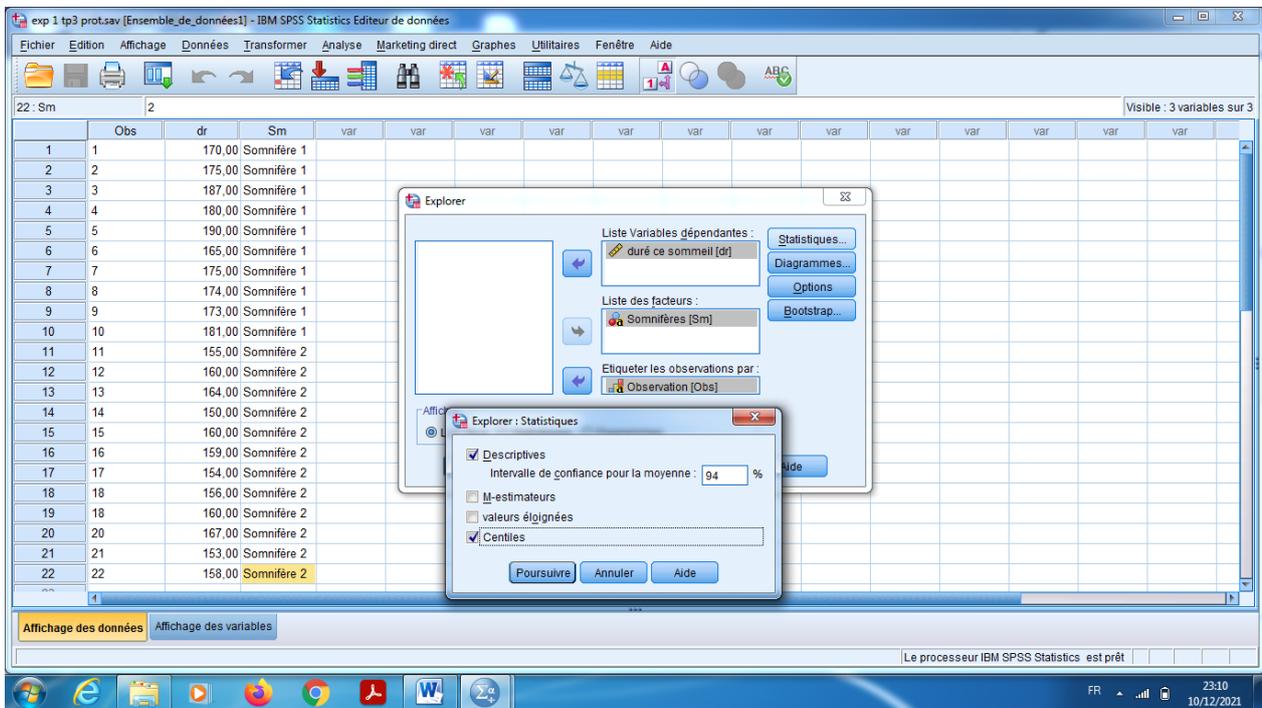
En suit, pour obtenir le test de normalité, en cliquant sur le bouton « Analyse » dans la barre des outils, et choisir « Statistiques descriptives », et puis « Explorer ».



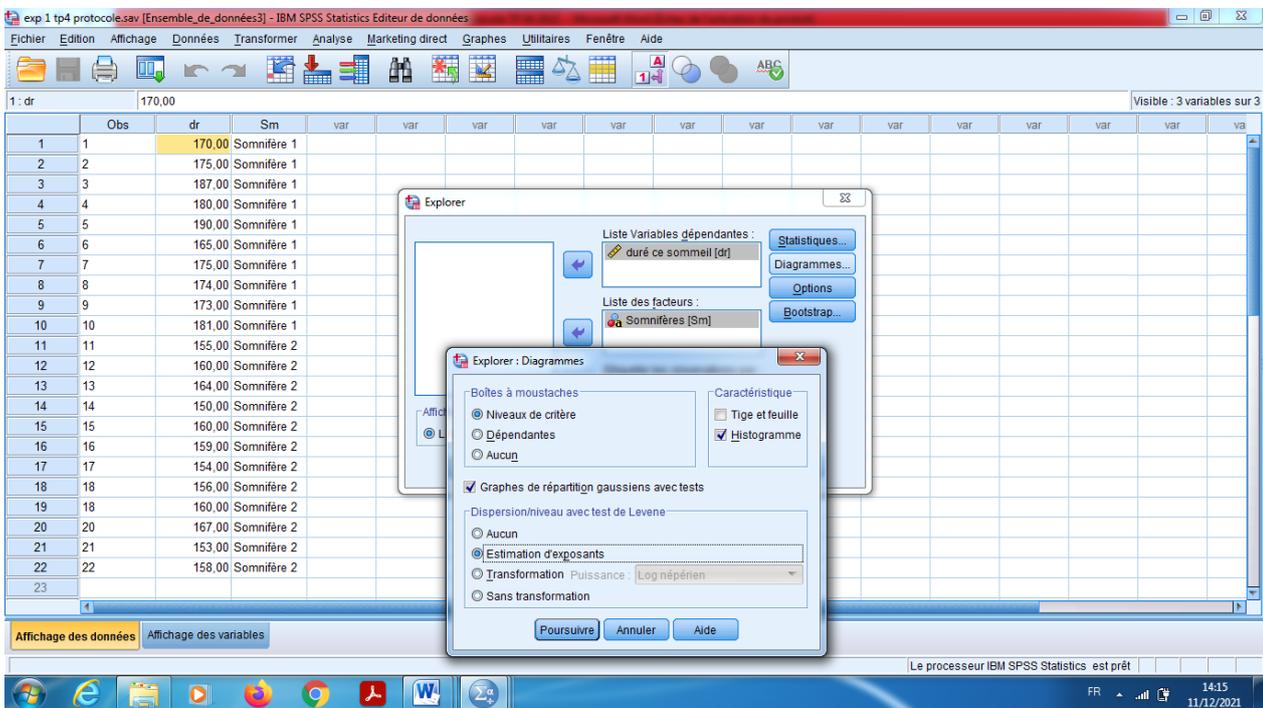
- a- On choisit dans « liste variables dépendantes » la variable quantitative mesurable « Durée de Sommeil ».
- b- On choisit la variable qualitative (Echantillons=somnifères) dans la barre « liste des facteurs ».



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



- f- Sur le bouton « Diagrammes », on garde dans « Boîtes à moustaches » le niveau de critère », 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.
- g- En cliquant sur « poursuivre » .



En fin en cliquant sur « Ok ».

Interprétation des résultats

1) Notre objectif est de valider les conditions comparer les deux types des somnifères (Test d'échantillons indépendants) qui s'appelle test de normalité ainsi 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 la somnifère 1 et 12 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 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.

Pour le coefficient « aplatissement » s'appelle coefficient de Yule de la forme.

$$\delta = \frac{Me - \bar{x}}{\sigma} = \frac{175 - 177}{7,601} = -0,259 < 0, \text{ la série est étalée à gauche (pour Somnifère 1).}$$

Pour le coefficient « aplatissement » (pour Somnifère 2).

$$\delta = \frac{Me - \bar{x}}{\sigma} = \frac{158 - 158,5}{4,748} = -0,009 \sim 0, \text{ la série est presque symétrique « Gaussienne »}$$

The screenshot shows the 'Descriptives' table in IBM SPSS Statistics. The table is organized into two main sections: 'Somnifère 1' and 'Somnifère 2'. Each section lists various statistical measures such as Mean, Standard Deviation, Median, and Skewness. The 'Somnifère 1' section shows a mean of 177.0000 and a standard deviation of 2.40370. The 'Somnifère 2' section shows a mean of 158.0000 and a standard deviation of 1.37069. The table also includes measures of central tendency like the median and mode, and measures of dispersion like the range and interquartile range.

Somnifères		Statistique	Erreur standard
Somnifère 1	Moyenne	177,0000	2,40370
	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

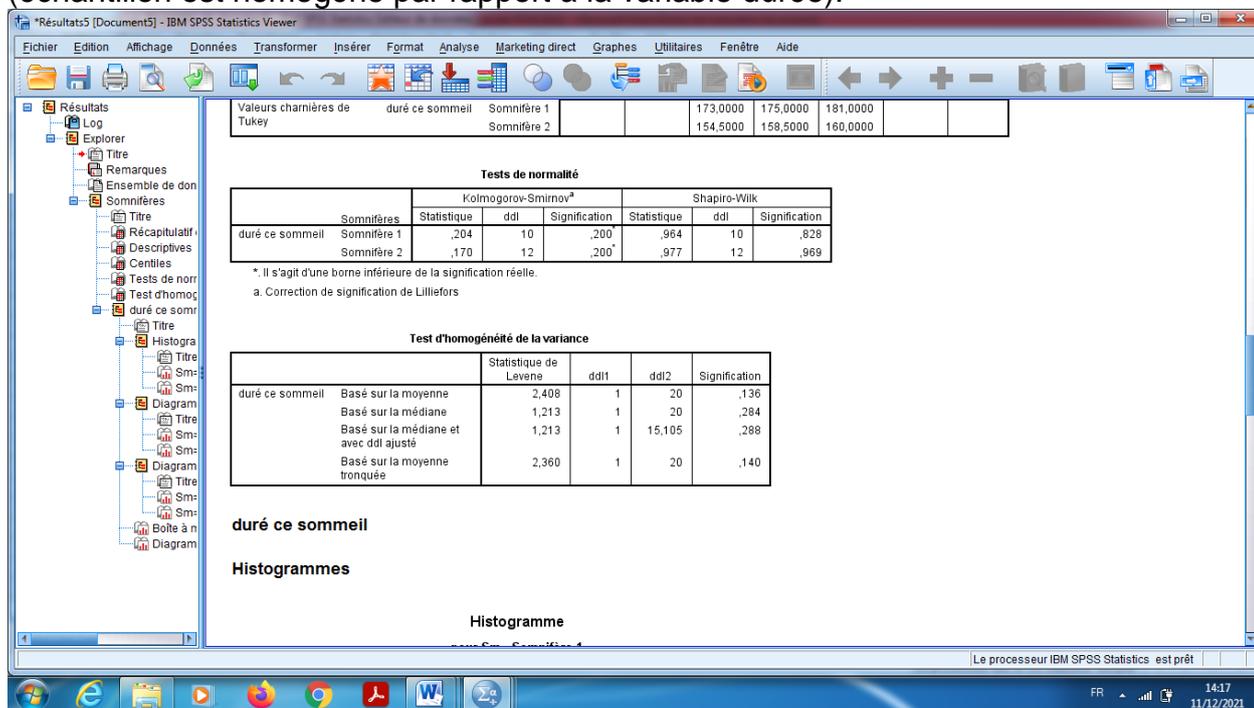
5) **Pour le test de normalité**, il faut choisir le test de Shapiro-Wilk.

On remarque que dans somnifère 1 (Sig=0,828>0,05) on accepte H0, alors la distribution pour échantillon 1 est gaussienne.

D'autre part dans somnifère 2 ($\text{Sig}=0,969>0,05$) on accepte H_0 , 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 ya pas des valeurs critiques pour cette variable, (échantillon est homogène par rapport à la variable durée).

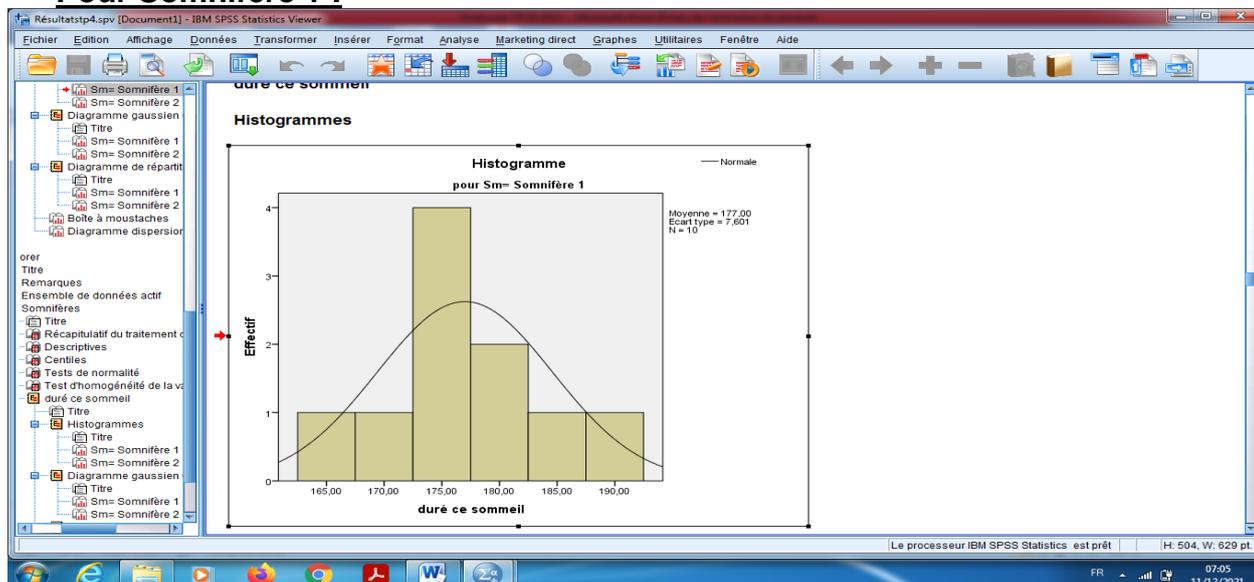


Test de normalité graphiquement

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

a) Histogrammes

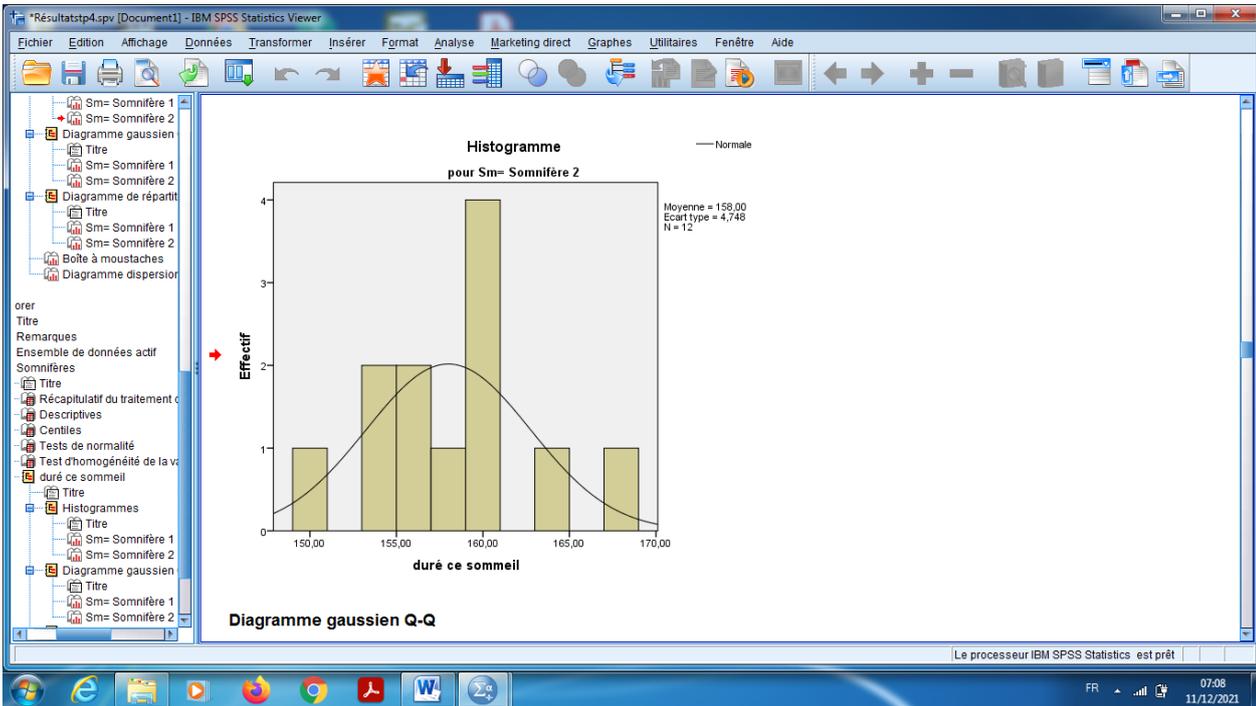
Pour Somnifère 1 :



On remarque ici que l'histogramme en batons est de la même forme que le graphe gaussien.

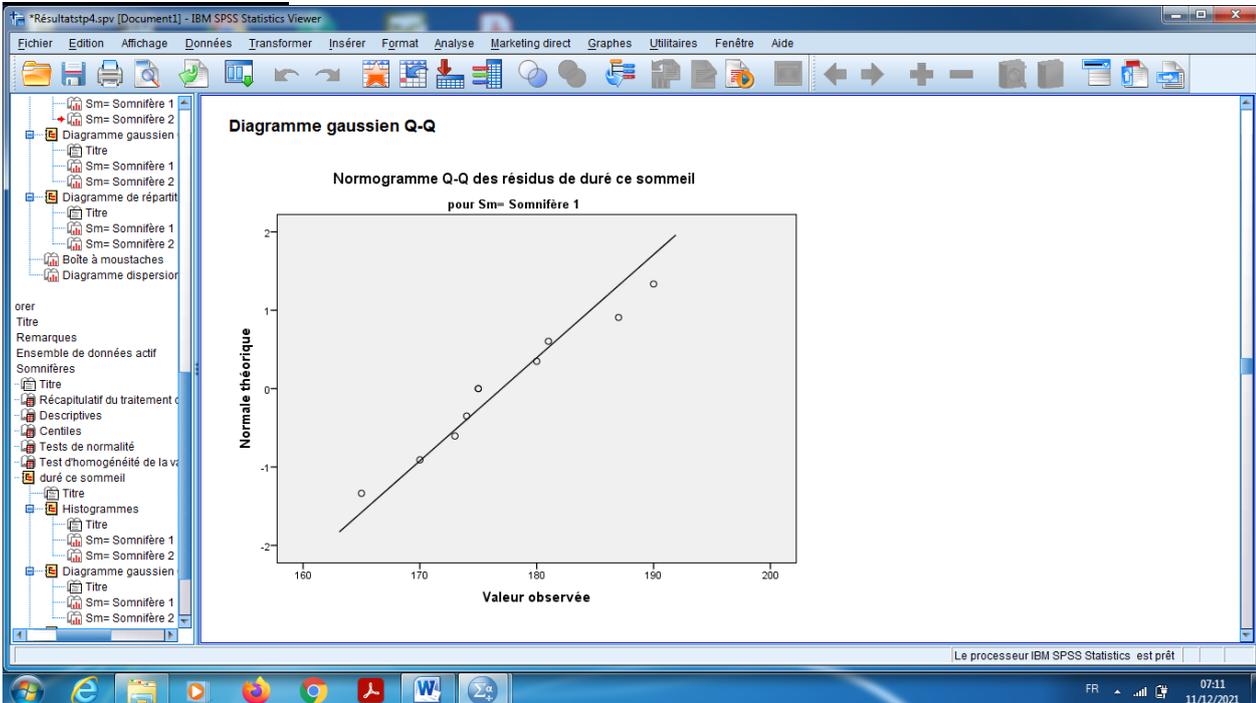
Pour Somnifère 2

L'histogramme en batons est presque de la même forme que le graphe gaussien.



b) 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, alors on peut appliquer le test de comparaison des moyennes.