

# **Méthode de recherche sociale et Techniques de traitement des données**

*Par : Abdeljabar RAFIKI*

*Niveau ; Licences professionnelles*

**Semestre 6**  
**Années universitaire ; 2019/2020**

**Plan du cours**

I- Théorie la statistique uni-variée et sa mise en pratique pas SPSS version 20 FR

II- Théorie statistiques bi-variée et sa mise en pratique par SPSS Version 20 FR

III- Démarche de recherche sociale par l'exemple de diagnostic de niveau de vie et programmes sociaux.

**Objectifs du cours**

L'objectif de ce cours est de mettre à la disposition des étudiants de la licence professionnelle des outils techniques de traitement de données par SPSS dans le domaine de gestion. L'ensemble des techniques, que l'étudiant de niveau licence non spécialiste ait capable d'assimiler sont la statistique uni-variée, bivariées et multi-variée appliquées aux données collectée par voie de questionnaire. Le logiciel que l'on emploiera au cours de ce semestre est le SPSS version 20 Fr.

Le traitement de l'information recueillie par voie d'enquête est une tâche délicate puisqu'il rend compte de plusieurs étapes. Ces derniers commencent par le traitement de l'information brute et se terminent par l'interprétation des résultats. Toutes ces étapes nécessitent de larges connaissances sur l'outil statistique, économétrique, informatique et sur la discipline elle-même objet de l'étude.

Le présent cours s'inscrit dans cette logique. il permettra à l'étudiant d'une part, de manipuler le logiciel statistique de traitement d'enquête «**SPSS** », et d'autre part, de maîtriser la démarche pratique de traitement d'enquête par logiciel. Dans cette perspective, une brève initiation au logiciel s'avère nécessaire suivie du codage des variables, de la saisie des variables et des données sur logiciel, du lancement de l'analyse, du choix d'outil statistique et économétrique de traitement et de l'interprétation des résultats.

L'analyse des données du questionnaire est nécessaire pour l'enrichissement des recherches scientifiques académiques telles que les mémoires de maîtrise ou du master et des thèses de doctorat. Ce cours a pour objectif de présenter séquentiellement les étapes de l'analyse des données recueillie par voie de questionnaire. A des fins de simplification, ce cours sera présenté de façon résumé pour soutenir les apprenants alors que le détail a été présenté et expliqué en classe.

A la fin de ce cours l'étudiant développera des compétences considérables dans les techniques statistiques, les méthodes de traitement de données et l'emploi du logiciel.

## **1- Théorie la statistique uni-variée et sa mise en pratique pas SPSS version 20 FR**

Dans ce chapitre nous allons expliciter les différentes techniques et les analyse statistiques unidimensionnelles. Ces analyses diffèrent selon la nature des variables statistiques « Qualitatives ou quantitatives ». Ces variables sont des caractéristiques ou des propriétés qui changent d'un individu à l'autre.

Toute information collectée par voie de questionnaire ou autres moyens pourra être traité à des fins de confirmation ou d'infirmerie d'une hypothèse de recherche fixée aprioris.

Les variables qualitatives (nominales ou ordinales) soumises au traitement statistique ne se prêtent qu'aux calculs des fréquences et éventuellement le mode. Par contre les variables quantitatives se prêtent aux différents calculs statistiques tels que le minimum, le maximum, l'étendue, les quintiles d'ordre  $\alpha$ , le mode, la médiane, les moyennes, la moyenne tronquée, la variance, l'écart type, l'erreur standard à la moyenne, le Skewness et le Kurtosis.

Pour les représentations graphiques des variables, on rencontre souvent le diagramme en secteur pour les variables qualitatives et l'histogramme pour les variables quantitatives sans oublier le diagramme en bâton pour les variables discrètes.

Plusieurs logiciels de statistique peuvent être utilisés pour effectuer ces traitements statistiques, mais nous nous limitons à SPSS version 20 vu ses potentialités statistiques. Après avoir présenté les procédures des calculs statistiques sous SPSS, nous l'exploitons dans le traitement des enquêtes par voie de questionnaire.

### **A- Principes de fonctionnement de SPSS<sup>1</sup>**

Lorsque l'utilisateur dispose des données sur les variables à étudier statistiquement, il peut exploiter les potentialités de SPSS pour un traitement de

---

<sup>1</sup> Statistical Package for the Social Sciences.

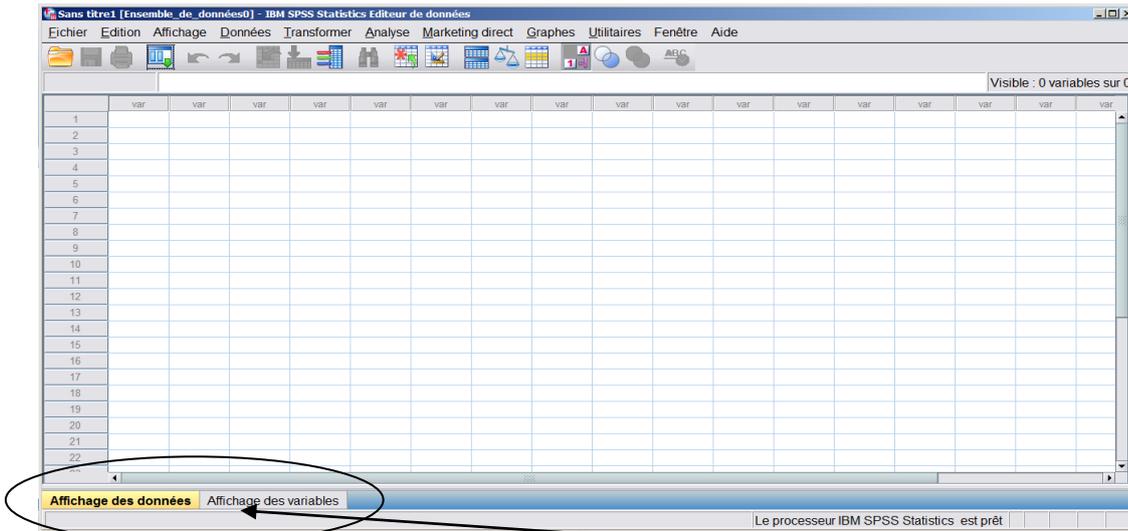
qualité. La tenue de la statistique descriptive par ce logiciel passe par trois étapes :

- La première étape se traduit par la création ou la déclaration des variables statistiques ;
- La deuxième étape concerne la saisie des données sur les variables déclarées et leur pondération en cas de configuration des observations par des poids différents ;
- La troisième étape consiste au lancement de l'analyse statistique souhaitée sans oublier la pondération s'elle y a lieu.

Avant d'explicitier ces étapes, il est nécessaire de présenter une initiation à SPSS pour familiariser le lecteur avec son langage et de rendre sa compréhension banale. Le logiciel SPSS est un logiciel puissant mais agressif, il fonctionne sous un mode interactif et sous un mode programmation de style Visuel Basic. Les résultats des analyses statistiques demandées (output) sont présentés généralement dans des tableaux difficilement déchiffrables et souvent non commentés.

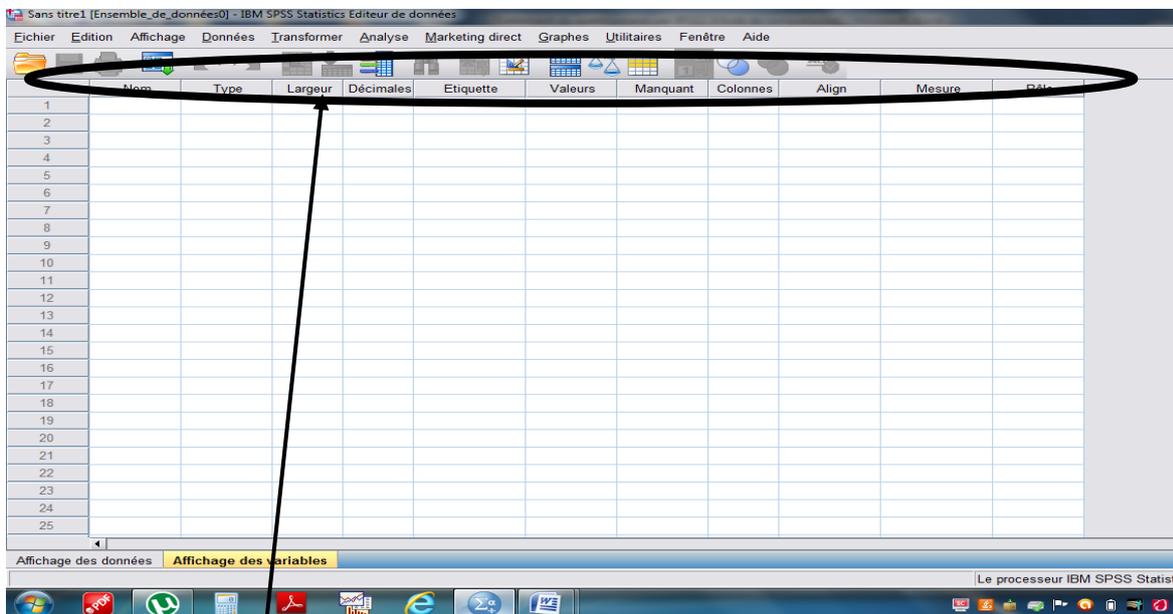
L'ouverture d'une nouvelle session sous SPSS, passe par une série d'étapes successives. Une étape préliminaire du codage des variables doit être réalisée avant de procéder à l'entrée des données sur ordinateur. Cette nouvelle session commence par :

- Dans le sous-menu Programmes du menu Démarrer de Windows, sélectionnez IBM SPSS Statistics 20 puis annulez la fenêtre qui s'ouvre automatiquement avec le lancement de SPSS.
- La fenêtre de l'Editeur de données apparaît, comme l'illustre la Figure suivante :



Dans cette fenêtre figure en bas à gauche **deux onglets** : affichage des données et affichage des variables. En mode «affichage des variables», chaque ligne représente une variable. En mode «affichage des données», chaque colonne représente une variable et chaque ligne représente une observation.

Pour commencer, **dans une première étape**, cliquez d'abord, sur affichage des variables pour en créer les variables à étudier, la fenêtre d'affichage des variables apparaît :



**Les caractéristiques de chaque variable doivent être définies dans les colonnes de S.P.S.S. suivantes :**

**Nom** : Sert à donner un nom à la variable. Il est recommandé de ne pas utiliser de caractères accentués. Il doit bien représenter la variable pour vous permettre de la reconnaître facilement.

**Type** : Sert à décrire la variable et le format de la colonne. Il est recommandé de laisser l'option «numérique» qui est indiquée par défaut pour les variables quantitatives et chaîne pour les variables qualitatives. Pour les décimales, vous pouvez les laisser égales à 2 ou les mettre à 0, selon la situation.

**Décimale** : Indique le nombre de chiffre après la virgule qui apparaîtra dans l'onglet affichage des données pour les variables numériques.

**Etiquette** : Permet de donner une description plus explicite à la variable qui sera affichée dans les différents calculs statistiques au niveau d'output.

**Valeur** : Permet de définir les valeurs que peuvent prendre la variable, ceci ne veut pas dire que le SPSS n'acceptera pas d'autres valeurs lors de la saisie dans l'éditeur de données. La colonne « valeur » prend de l'importance dans la codification des variables qualitatives d'un questionnaire.

**Manquantes** : Permet d'indiquer si la variable peut accepter ou non des valeurs manquantes. L'emploi de la colonne manquante est courant dans le traitement de questionnaire pour indiquer les non-réponses.

La gestion de valeur manquante est de deux types :

**1- Valeur manquante par défaut** : Si aucune valeur n'est saisie pour une variable numérique, la valeur manquante par défaut est représentée par une virgule dans l'Editeur de données.

**2-Valeurs manquantes spécifiées par l'utilisateur** : Si vous avez prévu d'utiliser des codes spécifiques (ex. 9, 99, etc.), il faut les définir dans «Manquantes».

**Colonne** : Permet de terminer la largeur de la colonne. Vous pouvez augmenter ou réduire la taille de la colonne.

**Alignement** : sert à aligner les données à l'intérieur des cellules.

**Echelle** : permet de définir le type de l'échelle (nominale, ordinale, intervalles).  
 Mais, Attention : le type d'échelle n'a pas d'influence sur le type d'analyse qu'on peut effectuer avec SPSS.

**Remarque :**

Les règles de base pour les noms des variables sont les suivantes : Le nom doit commencer par une lettre.

- Le nom de variable ne peut pas se terminer par un point
- La longueur du nom de variable ne doit pas excéder huit caractères dans SPSS 10.
- Les noms de variable ne peuvent pas contenir d'espaces ou de caractères spéciaux (par exemple, !, ?, ' et \*).
- Chaque nom de variable doit être unique. Aucune duplication n'est p.

Il n'y a pas de différence entre majuscule et minuscule des noms des variables. Lorsque la caractéristique de chaque variable est décrite par les éléments ci-dessus (nom ; type ; colonne, manquant ; etc.) on dit qu'on crée (ou déclare) sous SPSS les variables statistiques à étudier.

On crée maintenant deux variables « sexe et salaire des employés ». La variable sexe est qualitative- nominale et la variable salaire est quantitative. Par le même procédé de création ci-dessus, on obtient la fenêtre suivante :

The screenshot shows the 'Edit Data' window in IBM SPSS Statistics. The title bar reads '\*Sans titre1 [Ensemble\_de\_données0] - IBM SPSS Statistics Editeur de données'. The menu bar includes 'Fichier', 'Edition', 'Affichage', 'Données', 'Transformer', 'Analyse', 'Marketing direct', 'Graphes', 'Utilitaires', 'Fenêtre', and 'Aide'. The main area is a table with the following columns: 'Nom', 'Type', 'Largeur', 'Décimales', 'Etiquette', 'Valeurs', 'Manquant', 'Colonnes', 'Align', 'Mesure', and 'Rôle'. Two variables are defined: 'Sexe' (Type: Chaîne, Largeur: 8, Décimales: 0, Etiquette: 'Sexe des emplo...', Valeurs: '(,00, Fémini...', Manquant: '99.00', Colonnes: '8', Align: 'Gauche', Mesure: 'Nominales', Rôle: 'Entrée') and 'Salaire' (Type: Numérique, Largeur: 8, Décimales: 2, Etiquette: 'Salaire des emp...', Valeurs: 'Aucun', Manquant: 'Aucun', Colonnes: '8', Align: 'Droite', Mesure: 'Echelle', Rôle: 'Entrée'). The status bar at the bottom indicates 'Affichage des données' and 'Affichage des variables'.

	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	Sexe	Chaîne	8	0	Sexe des emplo...	(,00, Fémini...	99.00	8	Gauche	Nominales	Entrée
2	Salaire	Numérique	8	2	Salaire des emp...	Aucun	Aucun	8	Droite	Echelle	Entrée
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

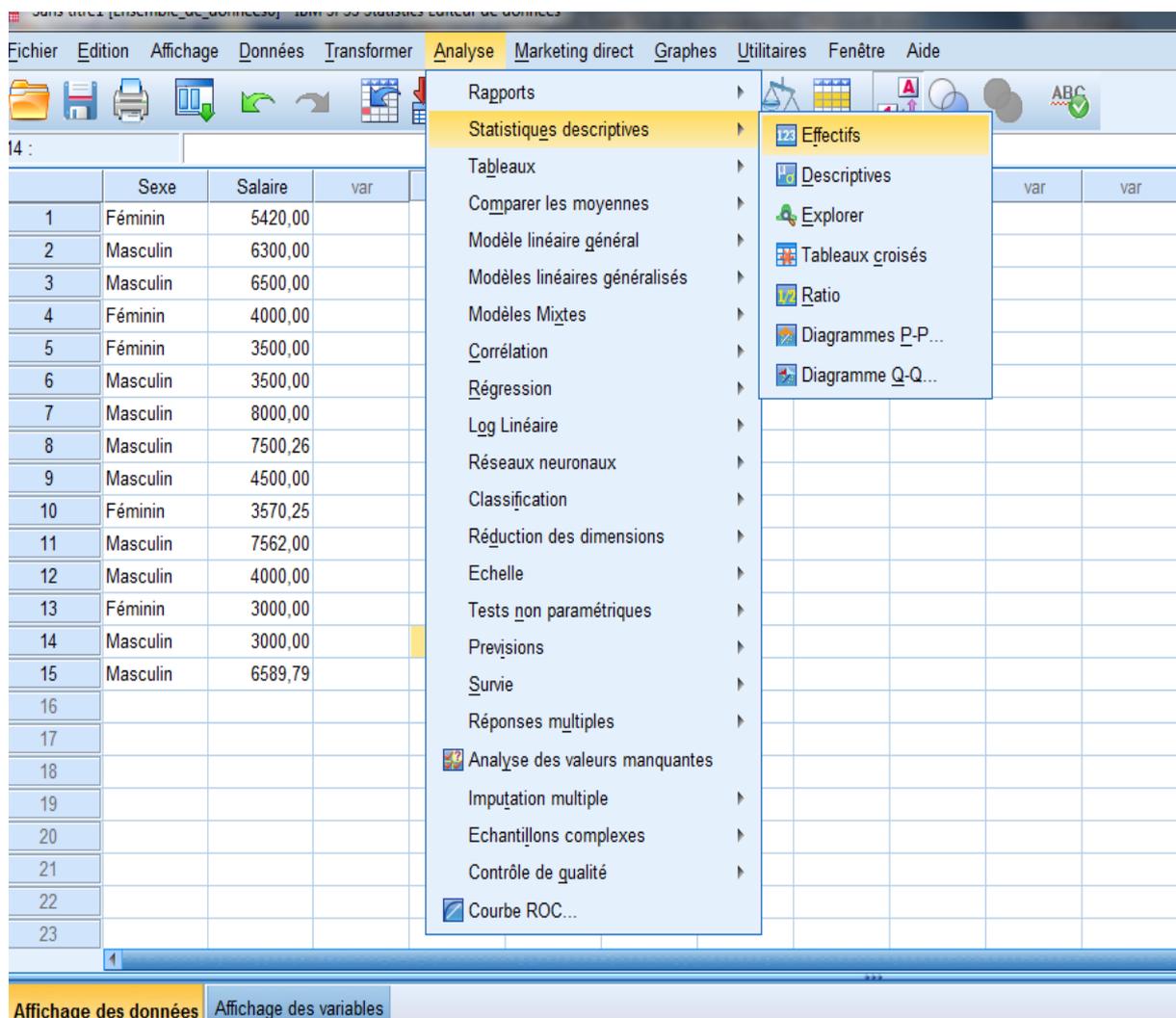
Après la déclaration de toutes les variables, on passe dans **une 2<sup>ème</sup> étape** à la saisie des données proprement dite, elle est accessible à travers un simple clic



Au niveau de chaque étape, on peut toutefois, enregistrer le fichier des travaux sous un nom explicite qui prend l'extension par défaut \*.SAV

Après la création des variables et la saisie des données, on passe à la **3<sup>ème</sup> étape** qui est le lancement de l'analyse statistique suivant la nature des données et l'objectif de l'étude. Pour lancer les calculs de la statistique descriptive, suivez les instructions suivantes :

→ Cliquez sur le menu analyse de la fenêtre d'éditeur de données de SPSS, puis pointez sur Statistique Descriptive et cliquez sur effectifs par exemple.



→ Dans boîte de dialogue qui apparaît à l'écran, déplacez le (s) variables dans la zone de traitement par un appui sur un petit bouton sous forme de flèche (►) ou par un double clic sur la variable.

The screenshot shows the SPSS Data Editor window with a data grid. A dialog box titled 'Effectifs' is open, allowing the user to select variables for frequency analysis. The 'Variable(s) :' list contains 'Sexe des employés' and 'Salaire des employés'. The 'Afficher les tableaux d'effectif' checkbox is checked. Buttons for 'Statistiques...', 'Diagrammes...', 'Format...', and 'Bootstrap...' are visible on the right side of the dialog.

	Sexe	Salaire	var								
1	Féminin	5420,00									
2	Masculin	6300,00									
3	Masculin	6500,00									
4	Féminin	4000,00									
5	Féminin	3500,00									
6	Masculin	3500,00									
7	Masculin	8000,00									
8	Masculin	7500,26									
9	Masculin	4500,00									
10	Féminin	3570,25									
11	Masculin	7562,00									
12	Masculin	4000,00									
13	Féminin	3000,00									
14	Masculin	3000,00									
15	Masculin	6589,79									

→ Sélectionnez les options souhaitées, puis validez par OK.

Les résultats apparaissent dans une nouvelle fenêtre appelée Viewer :

The screenshot shows the SPSS Viewer window displaying the results of the frequency analysis. The main content area shows the following tables:

**Effectifs**

[Ensemble\_de\_données0]

**Statistiques**

Sexe des employés

N	Valide	15
	Manquante	0

**Sexe des employés**

	Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide Féminin	5	33,3	33,3	33,3
Masculin	10	66,7	66,7	100,0
Total	15	100,0	100,0	

Le Viewer correspond, comme son nom l'indique, à la page d'affichage des résultats des calculs statistiques. Cette page constitue un fichier de résultat que l'on appelle Output.

A partir de ce Viewer, on peut utiliser les mêmes fonctions que dans la feuille de données. On peut copier, coller, imprimer, exporter et enregistrer les résultats

affichés. Lors de l'enregistrement, le fichier des résultats prend par défaut l'extension \*.SPO.

## **B- Création des tableaux statistiques et représentation graphique de la variable qualitative sous SPSS**

Soit l'exemple de la taille des villes présenté ci-dessus :

<b>Groupe sanguin</b>	A	B	AB	O	<b>Total</b>
<b>Effectif</b>	30	20	10	5	<b>65</b>

Dans le cas d'une variable qualitative, comme le cas de cet exemple, on ne peut calculer que les fréquences relatives et de représenter graphiquement la variable par des diagrammes en secteur ou en tuyaux d'orgue. Sous SPSS on procède de la manière suivante :

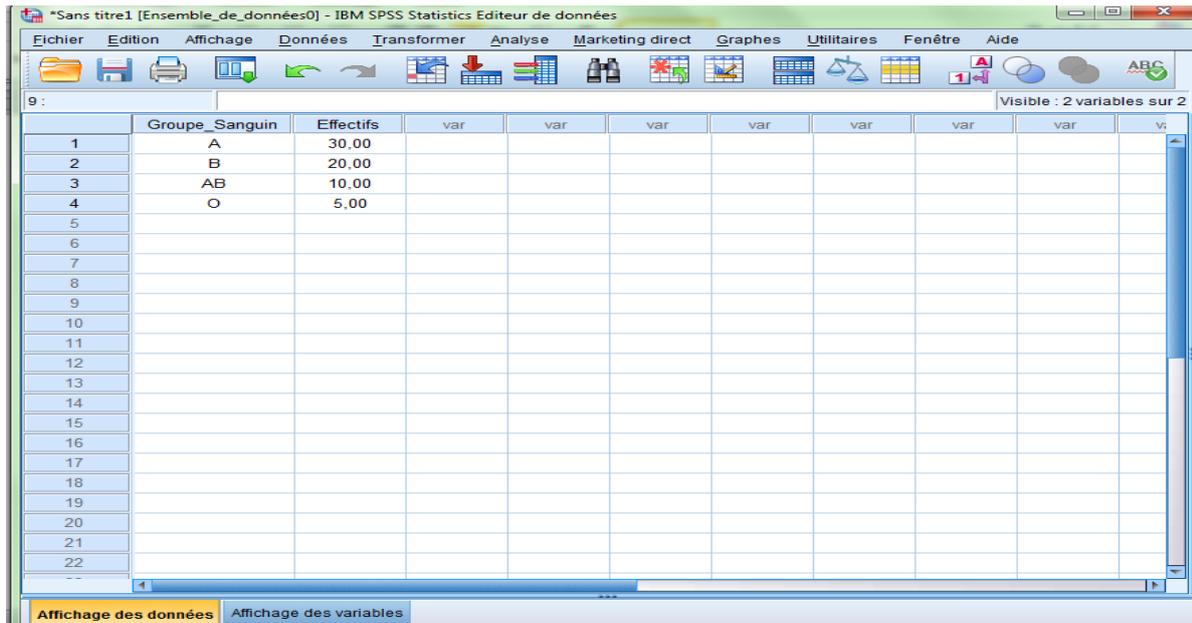
- 1) Création de deux variables l'une relative aux effectifs et l'autre relative au groupe sanguin
- 2) Saisie des données sur les deux variables
- 3) Pondération de la variable groupe sanguin par l'effectif
- 4) Lancement du calcul statistique des fréquences et options graphiques

### **Procédure :**

- Cliquez sur l'onglet affichage des variables ;
- Tapez dans la première ligne le nom de la variable « Groupe\_Sanguin »<sup>2</sup> puis spécifier dans la colonne « type » 'chaîne'.
- Tapez dans la deuxième ligne « effectifs » puis spécifier le type numérique.
- Cliquez dans l'onglet affichage des données puis entrez les données sur les deux variables comme indiqué dans le tableau ci-dessus. Vous obtenez une fenêtre identique à celle-ci :

---

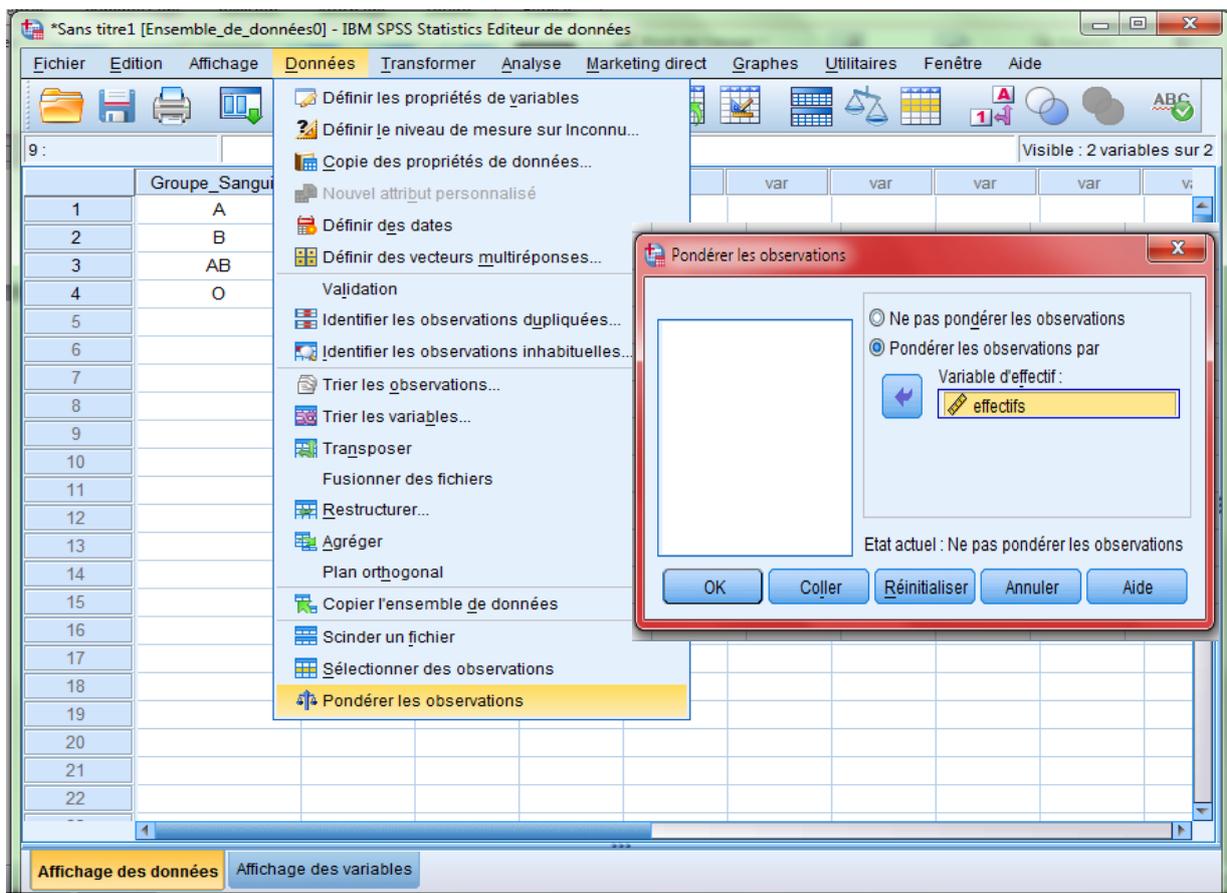
<sup>2</sup> Les noms des variables ne contiennent pas de guillemets



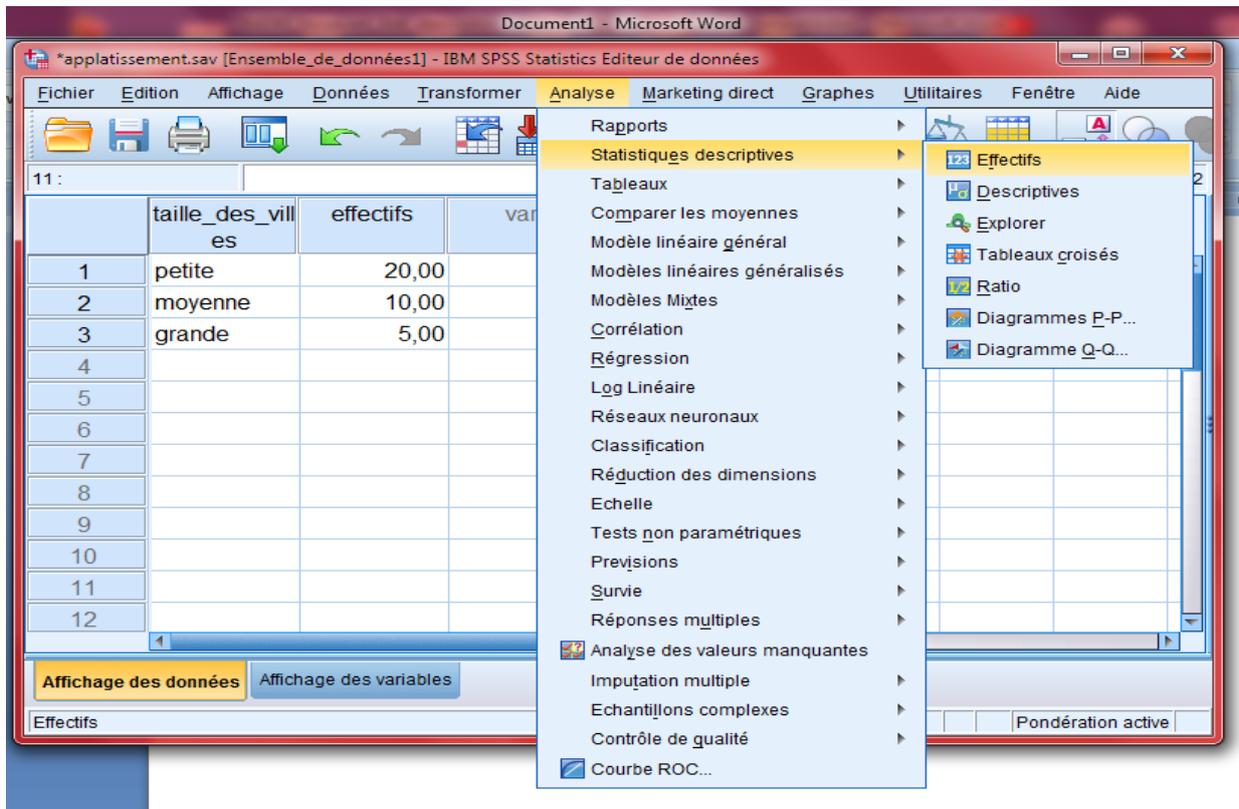
→ Cliquez dans le Menu donnée/Pondérer les observations

→ Cochez l'option pondérer les observations

→ Dans la petite fenêtre qui apparaissent à l'écran (voir figure ci-après), déplacer effectifs puis validez par OK.



→ Cliquez sur le Menu Analyse/Statistique Descriptive/effectifs, comme indiqué sur la figure ci-après.



Ensuite, dans la fenêtre qui apparaît automatiquement à l'écran (identique à celle-ci) :



Dans la section type de diagramme /Sélectionnez Diagramme puis cochez diagramme en secteur ou en bâton/ poursuivre/OK.



Après la validation par ok, les résultats suivants sont affichés dans l'environnement Viewer :

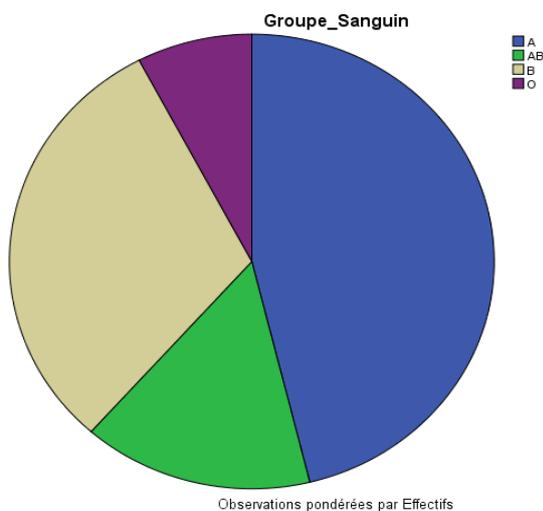
### Statistiques

Groupe\_Sanguin

N	Valide	65
	Manquante	0

### Groupe\_Sanguin

		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	A	30	46,2	46,2	46,2
	AB	10	15,4	15,4	61,5
	B	20	30,8	30,8	92,3
	O	5	7,7	7,7	100,0
	Total	65	100,0	100,0	



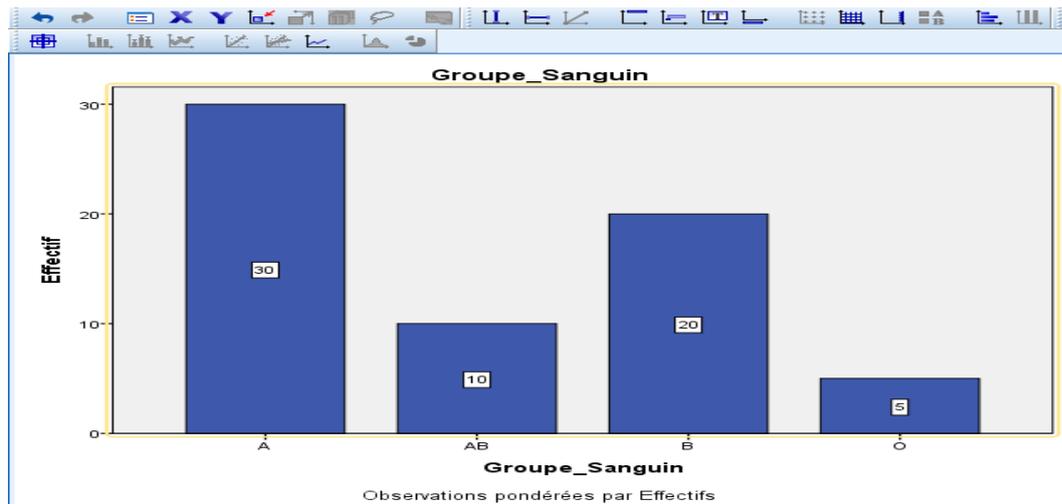
**Lecture des résultats :**

Le premier petit tableau indique une taille d'échantillon de 65 individus sans données manquantes.

Le deuxième tableau affiche le calcul des fréquences en pourcentage et les fréquences cumulées pour chaque modalité. On constate que la colonne du pourcentage est la même que celle du pourcentage valide en raison d'absence des données manquantes.

En fin, la représentation graphique par secteur est affichée en montrant la part de chaque modalité dans le total. Il est toujours possible de modifier l'aspect des graphiques par un double clic sur le graphe puis un double clic à côté.

On recommence la même procédure en demandant cette fois-ci le diagramme en tuyaux d'orgue en cochant diagramme en bâton dans l'option diagramme de SPSS:



### **C-- Création des tableaux statistiques et représentation graphique de la variable quantitative sous SPSS**

L'objectif de ce paragraphe est de calculer les fréquences, et de représenter graphiquement la variable continue. Les graphiques adéquats de la variable continue sont l'histogramme, la courbe des fréquences ajustée à l'histogramme et la courbes cumulative. Pour illustrer ces travaux sous SPSS, on prend l'exemple suivant :

La répartition des exportations annuelles en tonnes de 90 pays de différentes régions du monde est :

Exportations en 10 <sup>3</sup>	Effectifs	Centre des classes
[10-20[	10	15
[20-30[	15	25
[30-40[	25	35
[40-50[	15	45
[50-60[	10	55
[60-70[	10	65
[70-80[	5	75
<b>Total</b>	<b>90</b>	<b>--</b>

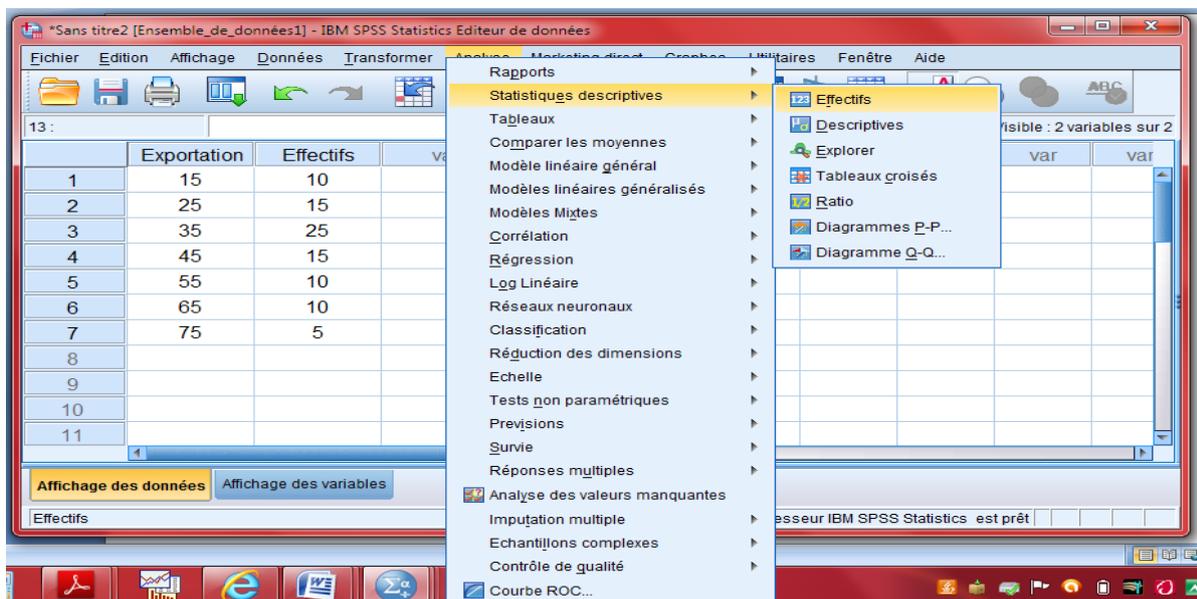
- **Calcul des fréquences par SPSS**

**1<sup>ère</sup> étape** : création de deux variables relatives à l'effectif et à l'exportation (voir dessus), puis la saisie des données sur ces deux variables. Il faut toutefois signaler, que le SPSS n'est pas programmé pour la saisie des classes. Le remède à ce problème est de saisir les centres des classes telles qu'elles sont indiqués dans le tableau statistique ci-dessus.

**2<sup>ème</sup> étape** : Pondération de la variable exportation par effectifs à partir du menu données/Pondérer les observations.

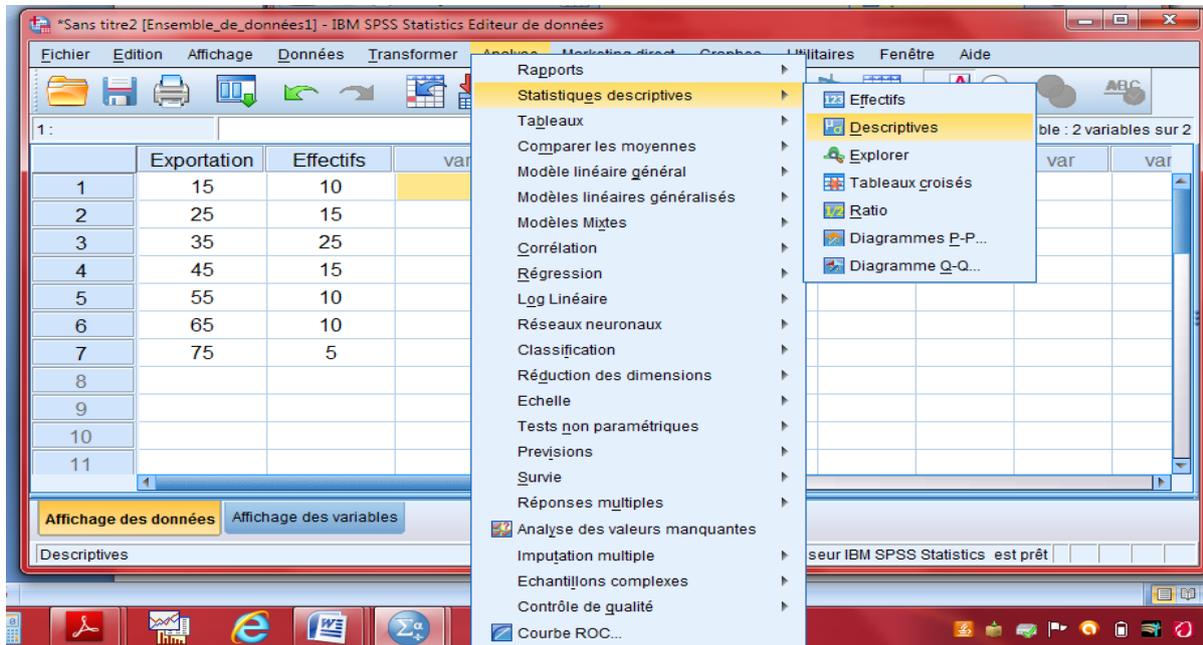
**3<sup>ème</sup> étape** : Lancement de l'analyse statistique. Pour le calcul des fréquences, On peut accéder par deux menus différents :

**1<sup>ère</sup> façon**



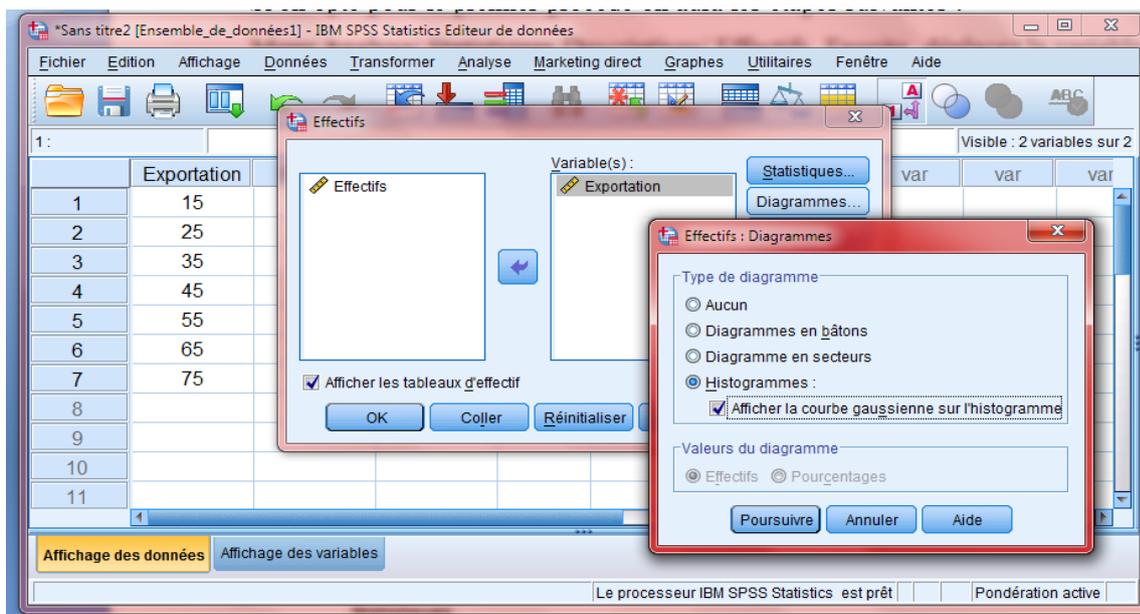
## .2<sup>ème</sup> façon :

La première méthode diffère de la seconde par les graphiques qu'elle génère automatiquement avec le lancement de l'analyse.



Si on opte pour le premier procédé on procède de la manière suivantes :

Menu Analyse/ Statistiques Descriptives/ Effectifs. Ensuite, déplacez la variable exportation dans la zone de variable/ cliquez sur Diagramme/Cochez histogramme et afficher la courbe gaussienne sur l'histogramme/poursuivre/Ok.



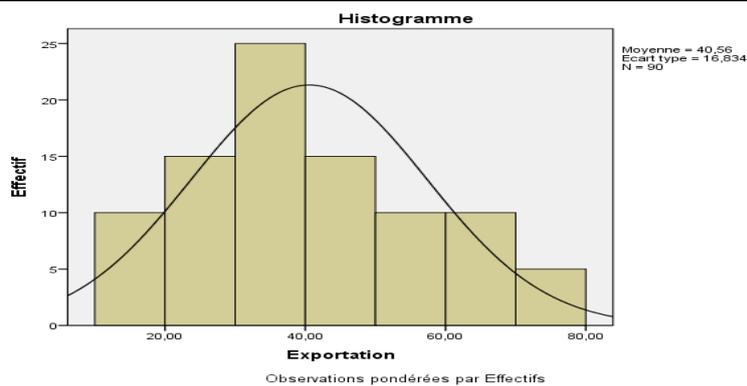
Après validation par OK, on obtient les résultats suivants dans l'éditeur Viewer :

## Exportation

N	Valide	90
	Manquante	0
Minimum		15,00

## Exportation

	Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
15,00	10	11,1	11,1	11,1
25,00	15	16,7	16,7	27,8
35,00	25	27,8	27,8	55,6
45,00	15	16,7	16,7	72,2
55,00	10	11,1	11,1	83,3
65,00	10	11,1	11,1	94,4
75,00	5	5,6	5,6	100,0
Total	90	100,0	100,0	



- Le 1<sup>er</sup> tableau montre une taille d'échantillon de 90 avec absence de données manquantes sur les exportations.
- Le 2<sup>ème</sup> tableau contient les fréquences absolues, les fréquences relatives et les fréquences cumulées croissantes.
- Le graphe montre l'histogramme en termes d'effectifs avec la même amplitude des classes et une courbe de fréquence ajustée normalement à l'histogramme.

**C-1 Paramètres de position et de dispersion****→ Paramètres de position**

Dans ce paragraphe, on présente aux lecteurs d'une part, la démarche du calcul par SPSS, des statistiques suivantes : le mode, la médiane, les quantiles d'ordre  $\alpha$ , la moyenne arithmétique, la moyenne géométrique, quadratique et harmonique ; et d'autre part, la représentation graphique de la distribution statistique par la boîte à moustache (Box-Plot)

Avant de passer au fond du problème, deux remarques suscitent notre intérêt dans ce propos. Premièrement, ces statistiques que l'on calculera ne sont valables que pour les variables quantitatives, autrement dit, les variables qualitatives ne se prêtent pas à ces calculs sauf le mode. Deuxièmes, le SPSS ne distingue pas entre le cas discret et le cas continu que ce soit en présence ou en absence de pondération. Dans le cas continu, quel que soit la nature de l'amplitude, ce sont les centres des classes qui sont considérés comme valeurs prises par la variable à étudier.

**Exemple : L'ancienneté de 200 employés d'une société est réparti ainsi :**

Ancienneté (en années)	Effectifs	Centre des classes
[1-4 [	14	2,5
[4-8[	30	6
[8-12[	35	10
[12-16[	45	14
[16-21[	30	18,5
[21-25[	25	23
[25 et plus	21	27
<b>Total</b>	<b>200</b>	---

1- Calculer la somme, le minimum, le maximum, le mode, la médiane, les quartiles, les déciles, les centiles et la moyenne arithmétique.

2- Représenter graphiquement cette distribution par la boîte à moustache

3- Donner la procédure de calcul, sous SPSS, de la moyenne géométrique, quadratique et harmonique.

**1) Procédure sous SPSS pour le calcul de la somme, du minimum, du maximum, du mode, de la médiane, des quartiles, des déciles, des centiles et la moyenne arithmétique.**

Après avoir saisi les données dans l'éditeur de données de SPSS, suivez les instructions suivantes :

→ Sélectionner le menu données pour pondérer les observations ;

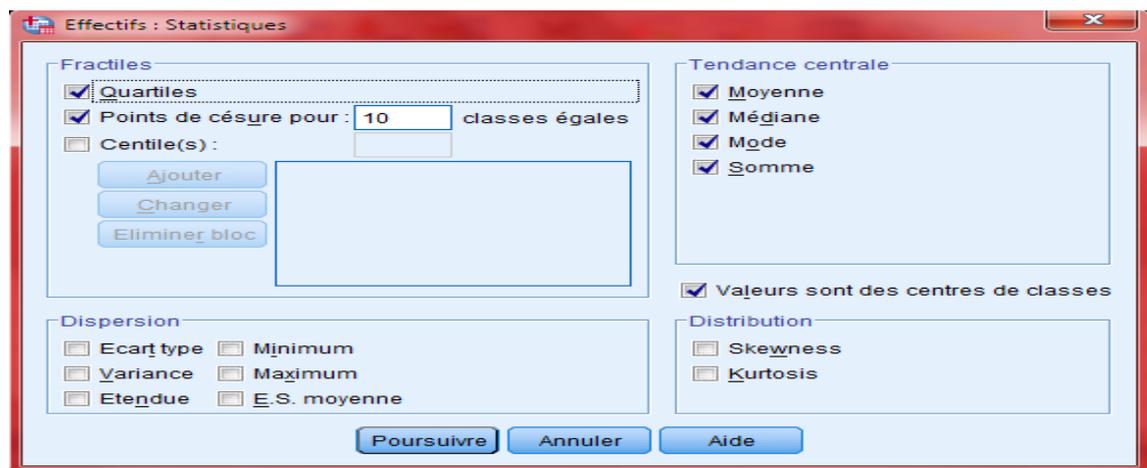
→ Cliquez dans le menu analyse /Statistiques Descriptives/Effectifs

→ Cliquez sur Statistiques/

→ Dans la section « Fractiles » Cochez Quartiles, points de césure pour les classes égales et taper le nombre de classes, les centiles et taper l'ordre du centile puis ajouter.

→ Dans la section tendance centrale, cochez moyenne, médiane, mode et somme puis cochez les valeurs sont les centres de classes pour un caractère continu.

→ Dans la section dispersion, cochez le minimum et le maximum. Cliquez sur poursuivre/ OK.



**Remarque sur certaines options ci-dessus :**

- 1) L'option « Valeurs sont des centres de classes » à un sens précis. Si les valeurs dans les données représentant des centres de classes sélectionnez cette option pour estimer la médiane et les centiles des données originales, non regroupées. Par contre si le caractère est discret, cette option doit être décochée.
- 2) Si le nombre de classe précisé dans l'option «points de césure pour les classes égale » est de 10 par exemple ceci signifie le calcul des 9 déciles. Cette option n'est rien d'autre qu'un quantile d'ordre  $\alpha$ .

**Après la validation de la commande, Le SPSS fournit l'Output suivant :**

**Statistiques**

Anciennete		
N	Valide	200
	Manquante	0
Moyenne		14,4600
Médiane		13,8500 <sup>a</sup>
Mode		14,00
Minimum		2,50
Maximum		27,00
Somme		2892,00
Centiles	10	4,5682 <sup>b</sup>
	20	7,3538
	25	8,5846
	30	9,8154
	40	11,8500
	50	13,8500
	60	16,2200
	70	18,6636
	75	20,3000
	80	21,9364
	90	25,3478

Pour le calcul du mode le SPSS prend le centre de la classe modale, il ne procède pas par la méthode des diagonales. En outre, le calcul des centiles est basé sur les données d'un échantillon aléatoire.

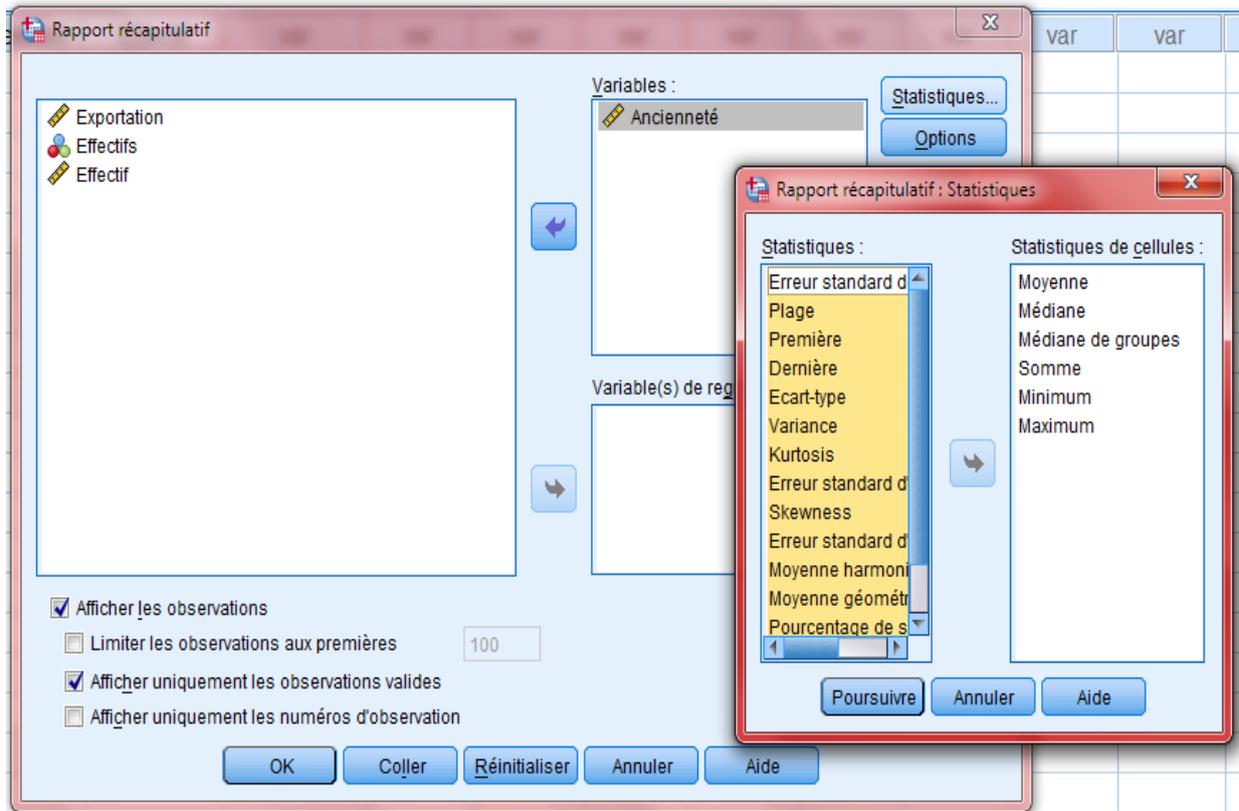
Dans SPSS, il existe d'autres entrées pour effectuer le même calcul, c'est le cas par exemple de ce procédé :

→ Menu Analyse / Rapport /Récapitulatif des observations

→ Déplacez la variable à étudier (ici c'est l'ancienneté) dans la zone de traitement.

*La zone réservée au variable(s) de regroupement doit être vide en absence de variable de regroupement. Il ne faut pas déplacer la variable effectif dans cette zone. A la fin de ce chapitre, on présente un exemple d'illustration de la variable de regroupement*

→ Cliquez sur Statistique /Déplacez les options souhaitées/poursuivre/ OK.



Le listing des résultats est:

	Ancienneté
1	2,50
2	6,00
3	10,00
4	14,00
5	18,50
6	23,00
7	27,00
Médiane	14,0000
Médiane de groupes	13,8500
Total Somme	2892,00
Minimum	2,50
Maximum	27,00

**Remarque :**

Le menu Analyse/Statistiques Descriptives/ Descriptive/ Option / ne fournit que quelques calculs statistiques tels que la moyenne, l'écart type, la variance, la somme, le minimum le maximum, le Skewness et le Kurtosis ; l'écart à la moyenne et l'étendu.

## → Paramètres de dispersion par SPSS

Soit l'exemple suivant :

Longueur en mètre	Effectifs
[18-18,5[	4
[18,5-19[	8
[19-20 [	14
[20-21 [	9
[21-21,5 [	10
[21,5-22,5[	5
<b>Total</b>	<b>50</b>

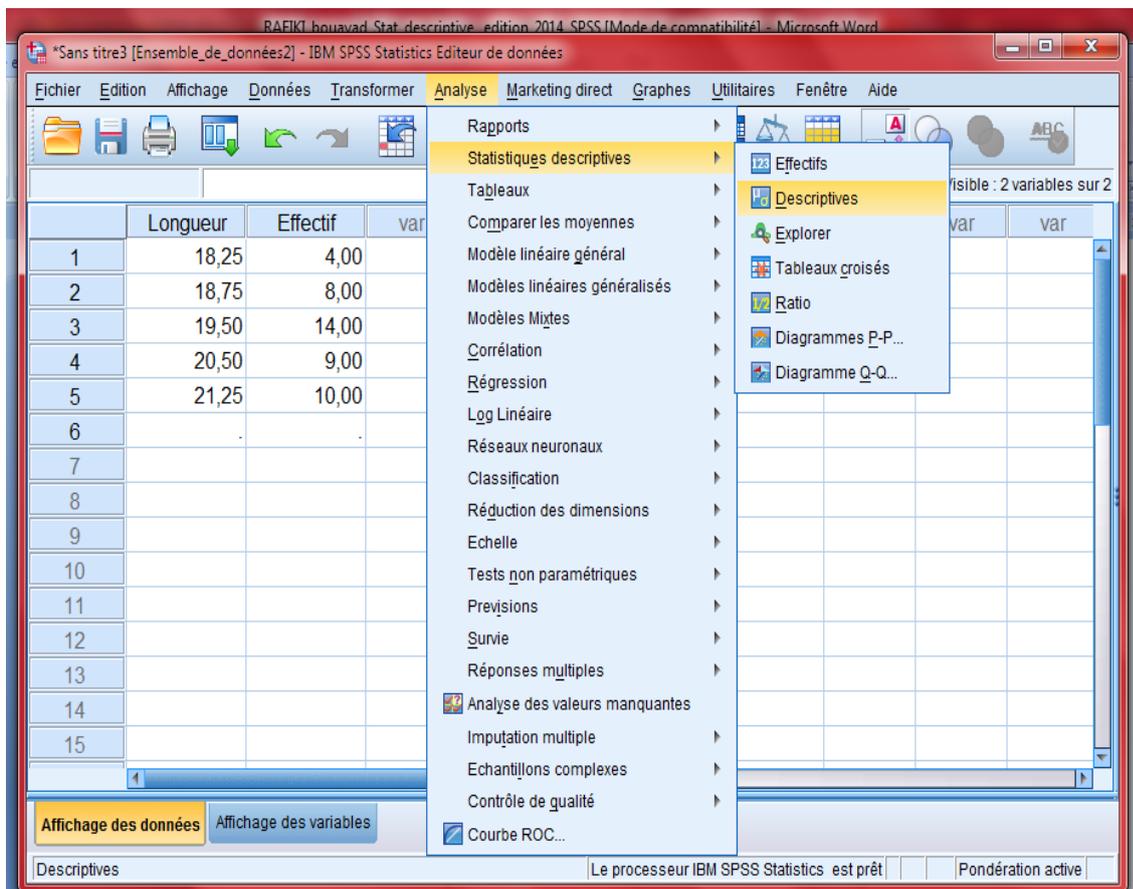
Après avoir saisi ces données sous SPSS, calculer l'étendu, l'écart type, la variance et l'erreur standard.

### Procédure sous SPSS

Il existe plusieurs entrées (procédé) sous SPSS.

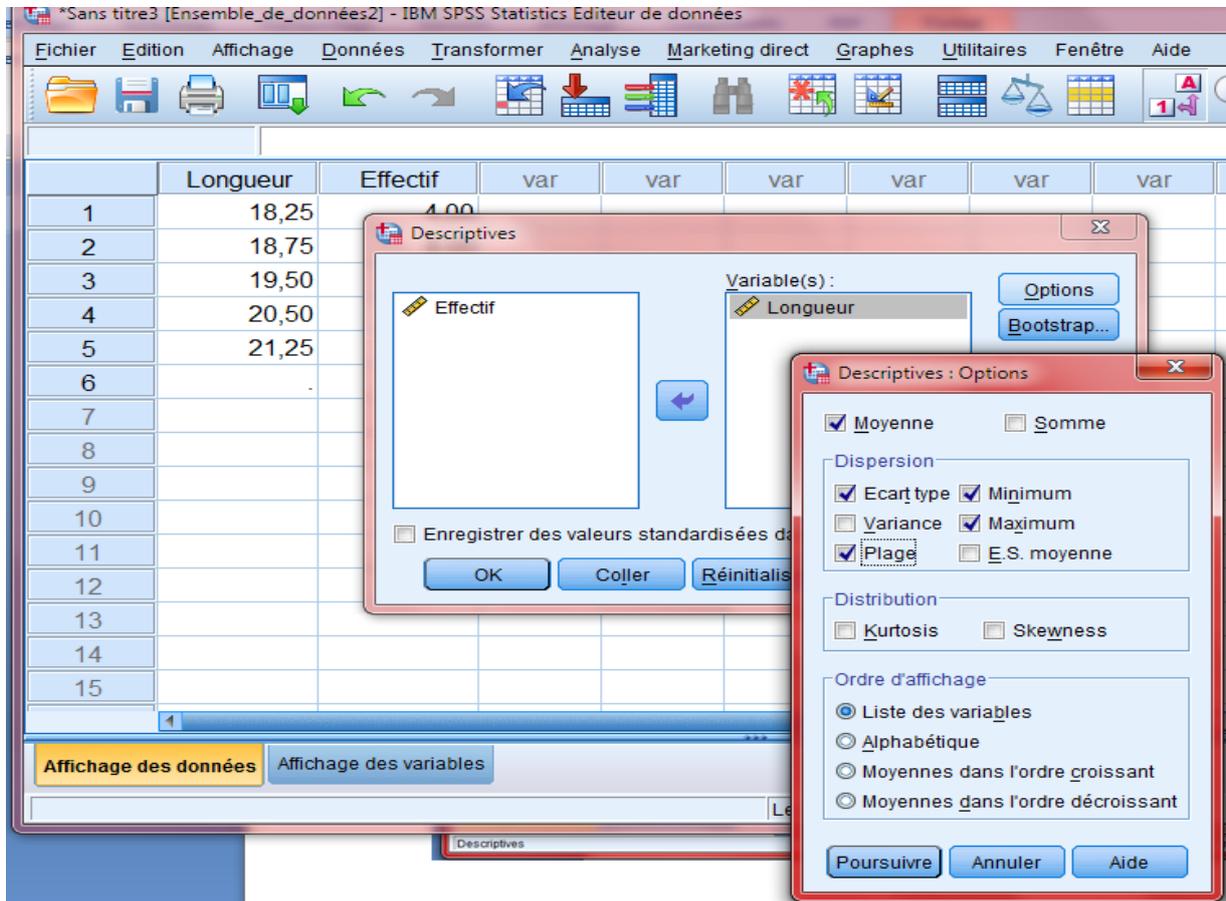
#### 1<sup>ère</sup> entrée :

→Sélectionnez menu analyse/Statistiques Descriptive/Descriptives



→ Sélectionner option et cochez plage (Etendu), variance et écart type/E.S. moyennes/poursuivre/OK

Les autres options ont été déjà explorées dans le chapitre précédent.



Les résultats affichés dans Viewer sont les suivants :

### Statistiques descriptives

	N	Intervalle	Moyenne		Ecart type	Variance
	Statistique	Statistique	Statistique	Erreur std	Statistique	Statistique
Longueur	50	3,75	20,0600	,16406	1,16009	1,346
N valide	50					

N : désigne la taille d'échantillon (ou population)

Intervalle : désigne l'étendu

La moyenne de la série statistique est de 20,06

L'erreur std : désigne l'erreur standard de la moyenne, il s'agit de l'écart type des moyennes

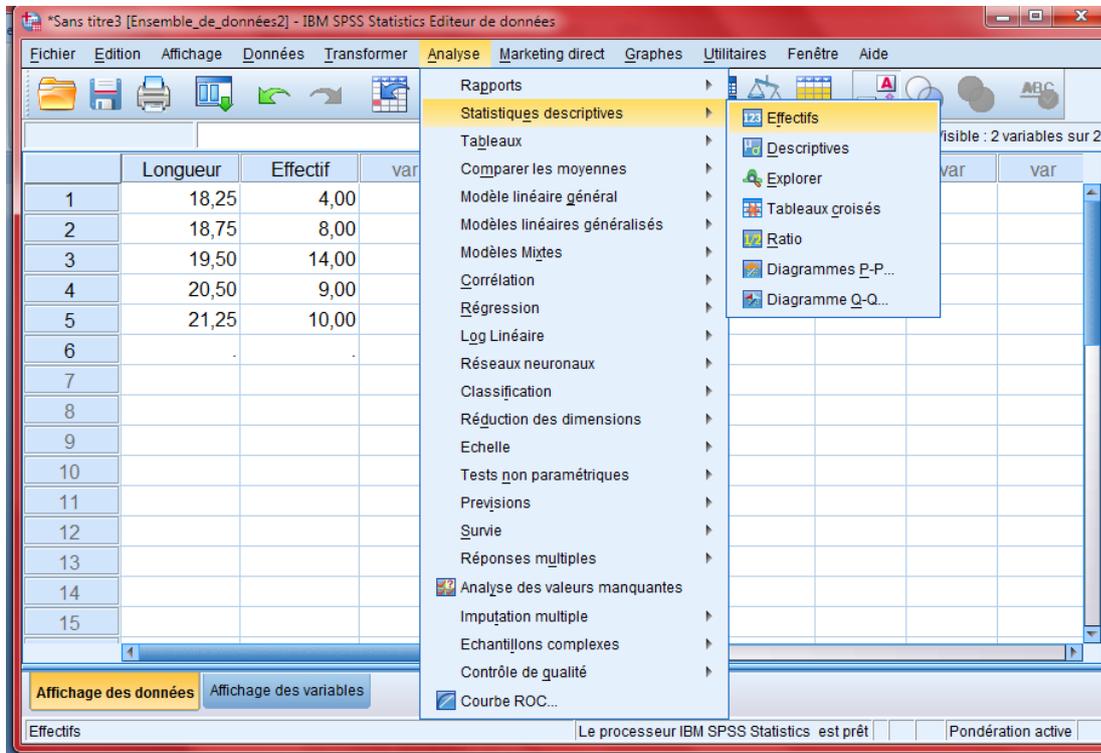
échantillonnales. Il est donné par la formule :

$$\text{Erreur Std} = \frac{S_{\text{Corr}}}{\sqrt{N}} = \frac{1,16009}{\sqrt{50}} = 0,16406$$

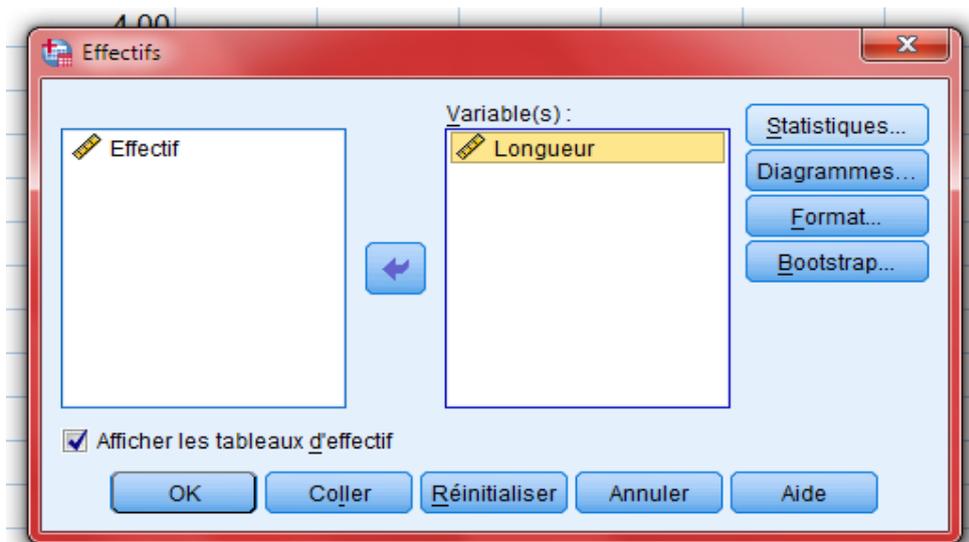
Les statistiques de l'écart type et de la variance sont des estimateurs sans biais de l'écart type et de la variance de la population.

**2<sup>ème</sup> entrée :**

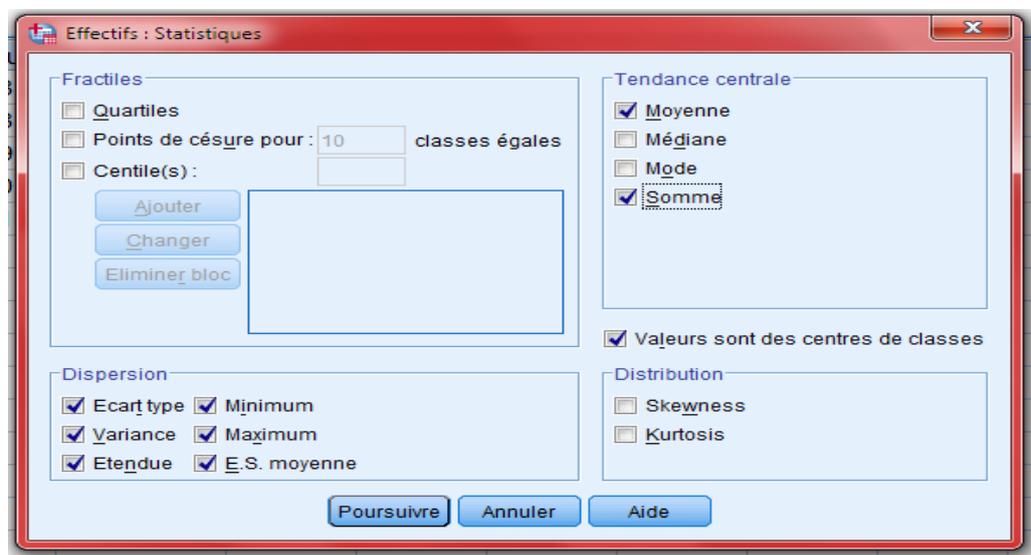
→ Sélectionnez menu analyse/Statistiques Descriptive/Effectifs



Une petite fenêtre de commande apparaît à l'écran :



→ Déplacez la variable longueur vers la zone Variable (s) puis Cliquez sur Statistiques et cochez, variance/écart type/ Variance/Etendu/ Minimum, Maximum /E.S. moyenne /Somme/poursuivre/OK.



Les résultats sont affichés dans Viewer comme suit :

### Statistiques

#### Longueur

N	Valide	50
	Manquante	0
	Erreur std. de la moyenne	,16406
	Ecart-type	1,16009
	Variance	1,346
	Intervalle	3,75

#### 3<sup>ème</sup> Entrée

→Sélectionnez menu analyse/Statistiques Descriptive/Explorer

→Sélectionner Statistiques et cochez Descriptives/poursuivre/OK.

L'Out put produit par SPSS est le suivant :

#### Descriptives

	Statistique	Erreur standard
Moyenne	20,0600	,16406
Intervalle de confiance à 95% pour la moyenne	Borne inférieure	19,7303
	Borne supérieure	20,3897
Moyenne tronquée à 5%	20,0528	
Médiane	19,5000	
Variance	1,346	
Ecart-type	1,16009	
Minimum	18,25	

<b>Maximum</b>	<b>22,00</b>	
<b>Intervalle</b>	<b>3,75</b>	
<b>Intervalle interquartile</b>	<b>1,94</b>	
<b>Asymétrie</b>	<b>,136</b>	<b>,337</b>
<b>Aplatissement</b>	<b>-1,167</b>	<b>,662</b>

Cette entrée génère deux nouvelles statistiques qui ne figurent pas dans les autres entrées, il s'agit de l'intervalle de confiance de la moyenne à 95% et de la moyenne tronquée à 5%.

\* L'intervalle de confiance désigne que la moyenne de la population est encadrée par 19,73 et 20,38 avec une probabilité de 0,95. Cet intervalle est calculé sur la base de la moyenne de l'échantillon qui est de 20,06.

\* la moyenne tronquée n'est pas trop différente de la moyenne courante. Ceci signifie qu'il n'y a pas assez des valeurs aberrantes dans cette série statistique. Car la moyenne tronquée est calculée après avoir éliminé 5% des observations des deux côtés de la distribution supposée ranger par ordre croissant.

Les autres statistiques sont éclairées ci-haut.

#### 4<sup>ème</sup> entrée :

On peut accéder à ces statistiques et bien d'autres statistiques à travers le menu analyse/ rapport/Récapitulatif des observations/ déplacer la variable d'étude dans la zone variables puis déplacer les statistiques souhaitées/poursuivre/ OK.

The screenshot shows the IBM SPSS Statistics interface. The main window displays a data table with the following data:

	Longueur	Effectif	var
1	18,25	4,00	
2	18,75	8,00	
3	19,50	14,00	
4	20,50	9,00	
5	21,25	10,00	
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			

The 'Rapport récapitulatif' dialog box is open, showing the 'Effectif' variable selected in the 'Variables' field. The 'Statistiques' list includes Mean, Median, Mode, Sum, Kurtosis, etc. The 'Statistiques de cellules' list includes Number of observations, Range, Minimum, Maximum, etc. The 'Options' button is visible.

Les résultats produits par SPSS sont les suivants :

**Récapitulatif des observations**

		Longueur
1		18,25
2		18,75
3		19,50
4		20,50
5		21,25
Total	N	45
	Intervalle	3,00
	Minimum	18,25
	Maximum	21,25
	Ecart-type	1,01171
	Variance	1,024
	Erreur standard de la moyenne	,15082
	Premier	18,25
	Dernier	21,25

## II- Théorie statistiques bi-variée et sa mise en pratique par SPSS Version 20 FR

### A- Production du tableau de contingence par SPSS.

#### → Méthode de saisie du tableau de contingence sous SPSS

Supposons que nous disposons d'un tableau à double entrée ventilant le niveau d'instruction par sexe d'un échantillon de 2306 individus. Les deux variables que contient le tableau sont qualitatives nominales.

	Analphabète	Primaire	Secondaire	Professionnel	Supérieur	Total
Féminin	250	50	350	46	400	1096
Masculin	100	370	150	390	200	1210
Total	350	420	500	436	600	2306

#### ◆ Création des variables

Les variables qualitatives sont généralement codifiées par des codes numériques. On attribue des codes allant de 0,1,2 .....n, respectivement aux modalités de X et de Y. Dans notre exemple, la codification se présente ainsi :

Variable « niveau d'instruction »	Codification	Variable « Sexe »	Codification
Analphabète	1	Féminin	1
Primaire	2	Masculin	2
Secondaire	3		
Professionnel	4		
Supérieur	5		

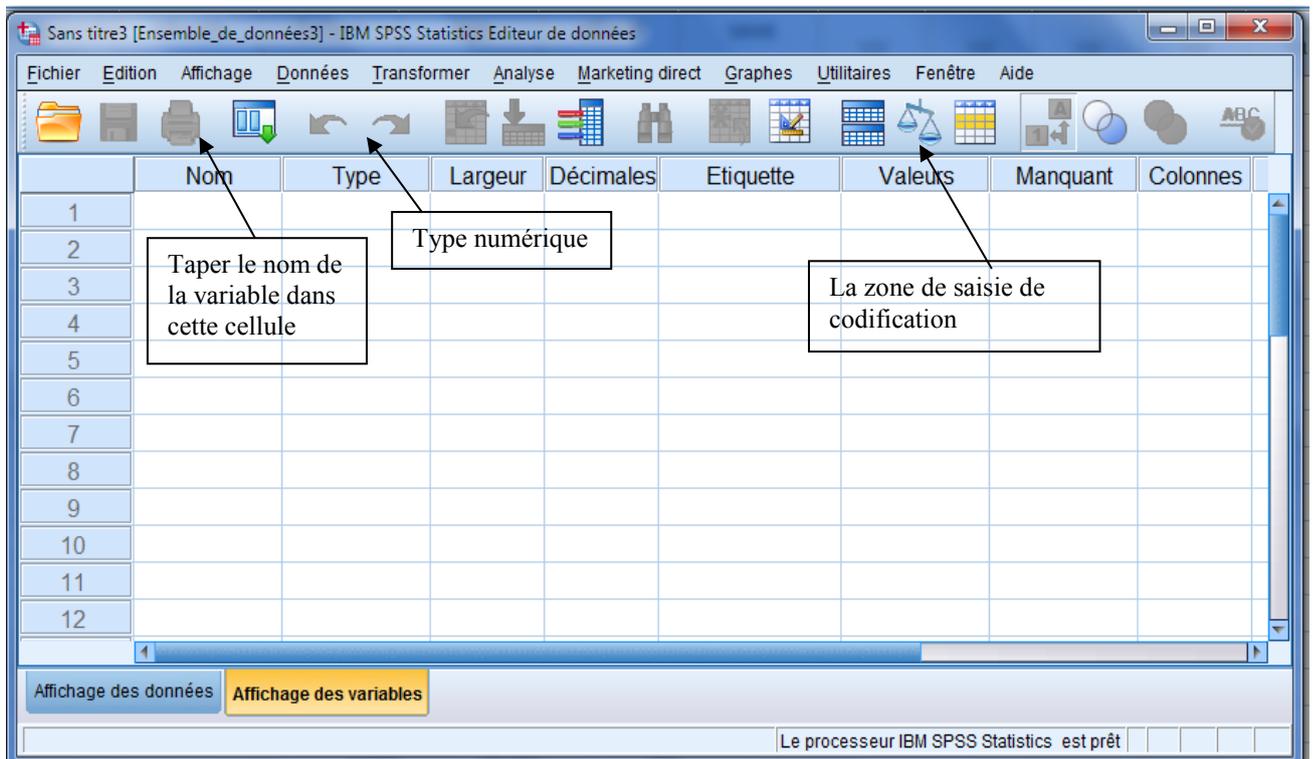
*Ces codes seront déclarés en même temps que les variables dans SPSS.*

- Dans SPSS, le tableau de contingence ci-dessus se produit par la déclaration de trois variables « **niveau d'instruction ; sexe et effectifs** ». la variable « niveau d'instruction » est nommée par exemple « Niveau\_d\_instruction » de type numérique et ses modalités sont codifiées respectivement de 1 à 5. La variable « Sexe » est nommée par son intitulé, de type numérique et codifié de 1 à 2. La dernière variable est l'effectif, elle joue le rôle de pondération donc elle doit être déclarée numérique et n'a pas besoin d'être codifiée. La saisie

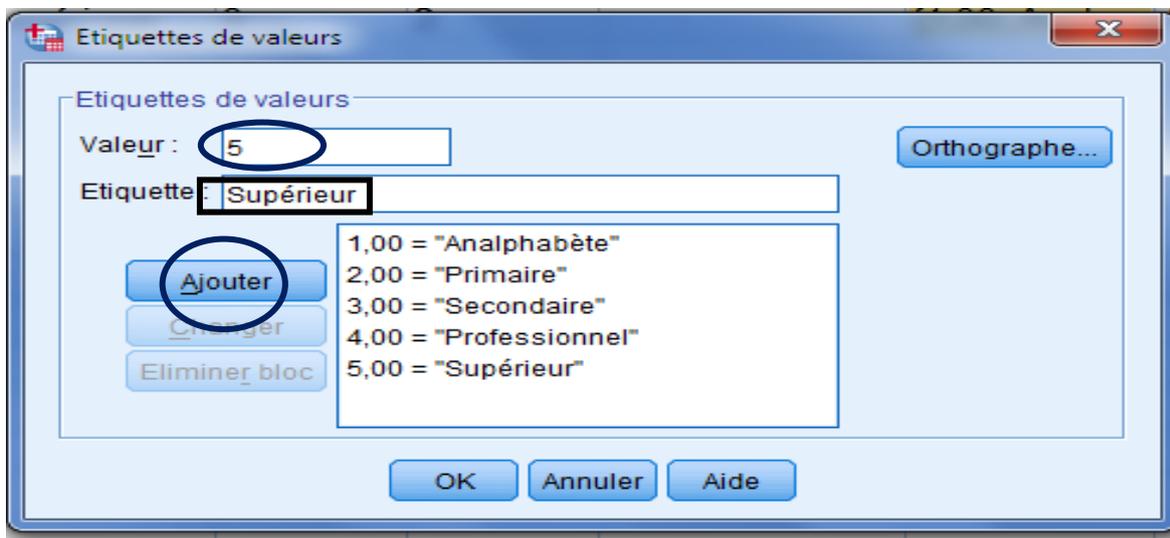
des caractéristiques des variables se fait à travers l'onglet « affichage des variables » de l'éditeur de données de SPSS.

- Pour commencer cliquez sur l'onglet affichage des variables et créer trois variables numériques ainsi que leur codification comme l'illustre la figure suivante :

### Fenêtre de déclaration des variables



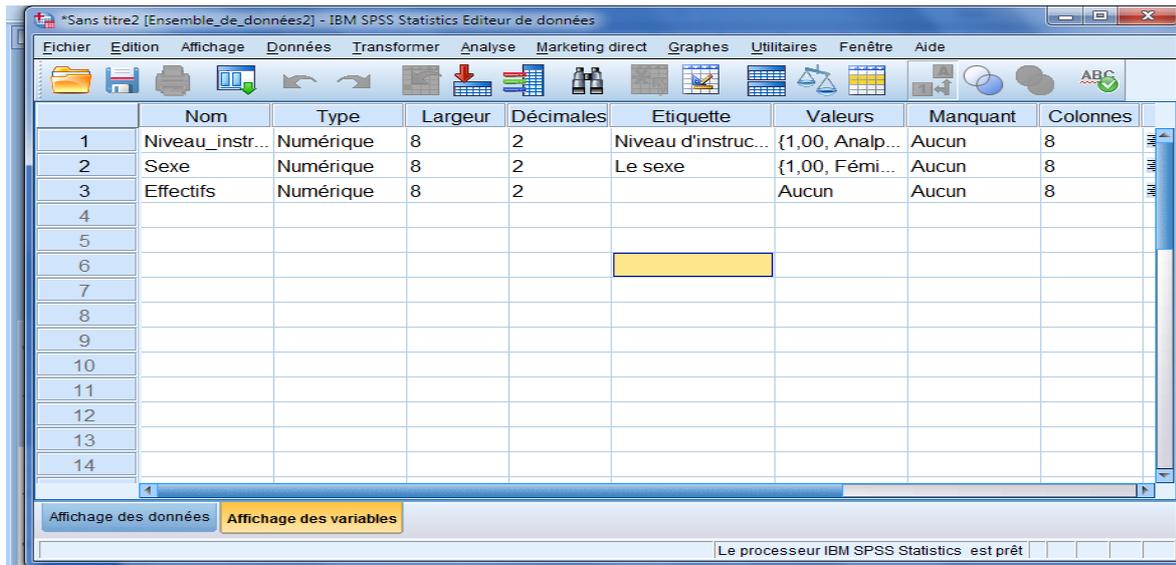
Dès le premier clic sur la zone valeurs, une petite boîte de dialogue nommée « **Etiquettes de valeur** » apparaît à l'écran (voir figure ci-après)



- 1- Taper le code de la modalité dans la zone valeur
- 2- Taper la modalité dans la zone Etiquette.
- 3- Cliquer sur ajouter.
- 4- Valider par OK après avoir terminé la saisie de tous les codes.

Ces manipulations aboutissent à la même fenêtre que celle-ci

### Fenêtre de saisie des variables relatives au niveau d'instruction et sexe



#### ◆ Saisie des données des variables

Basculez maintenant au mode d'affichage des données, les données du tableau croisé initial, doivent être saisies de la même manière telles qu'elles apparaissent dans le tableau suivant :

Niveau_instruction	Sexe	Effectifs
Analphabète	Féminin	100
Primaire	Féminin	370
Secondaire	Féminin	150
Professionnel	Féminin	390
Supérieur	Féminin	200
Analphabète	Masculin	250
Primaire	Masculin	50
Secondaire	Masculin	350
Professionnel	Masculin	46
Supérieur	Masculin	400

En voici la fenêtre de l'affichage de données de SPSS après la saisie de ces données :

## Fenêtre de saisie des données des variables niveau d'instruction et sexe

	Niveau_instruction	Sexe	Effectifs	var	var	var	var	var	var
1	Analphabète	Féminin	250,00						
2	Primaire	Féminin	50,00						
3	Secondaire	Féminin	350,00						
4	Profession...	Féminin	46,00						
5	Supérieur	Féminin	400,00						
6	Analphabète	Masculin	100,00						
7	Primaire	Masculin	370,00						
8	Secondaire	Masculin	150,00						
9	Profession...	Masculin	390,00						
10	Supérieur	Masculin	200,00						
11									
12									

### ◆ Pondération des variables par l'effectif

La dernière étape est la pondération des variables par l'effectif. Dans le menu données/ pondérer les observations/Déplacer la variable effectif dans la zone correspondante/OK.

Pour obtenir le tableau croisé initial, il suffit de procéder par le Menu Analyse/Statistique Descriptives/tableaux croisés/déplacez Sexe par ligne et Niveau\_d\_instruction par colonne/valider par OK.

En voici les résultats explorés par SPSS.

Tableau croisé Le sexe \* Niveau d'instruction

Effectif		Niveau d'instruction					Total
		Analphabète	Primaire	Secondaire	Professionnel	Supérieur	
Le sexe	Féminin	250	50	350	46	400	1096
	Masculin	100	370	150	390	200	1210
Total		350	420	500	436	600	2306

## B- Le lancement des analyses statistiques relatives au tableau croisé

- **Calcul des fréquences conjointes par SPSS:**

Menu Analyse /Statistiques descriptives/Tableau croisé/ déplacement des variables en ligne et en colonne/ Cellules/ cocher total/poursuivre/OK

- **Calcul des fréquences marginales de X par SPSS.**

Menu Analyse /Statistiques descriptives/Tableau croisé/ déplacement des variables en ligne et en colonne/ Cellules/ cocher lignes/poursuivre/OK

- **Calcul des fréquences marginales de Y par SPSS.**

Menu Analyse /Statistiques descriptives/Tableau croisé/ déplacement des variables en ligne et en colonne/ Cellules/ cocher Position /poursuivre/OK

- **Calcul des effectifs théoriques (Attendus)**

Menu Analyse /Statistiques descriptives/Tableau croisé/ déplacement des variables en ligne et en colonne/ Cellules/ cocher attendus /poursuivre/OK

### **C- Mesure d'association entre deux variables qualitatives nominales par le Khi-2**

**Les étapes du test sont :**

#### **Etape 1 : Formulation des hypothèses à tester :**

$H_0$  : X et Y sont indépendantes

$H_1$  : X et Y sont dépendantes

#### **Etape 2 : Calcul des effectifs théoriques de chaque cellule**

$$O_{théoriques} = \frac{n_{i*} * n_{*j}}{n_{**}}$$

**Etapes 3 :** Comparaison entre les effectifs théoriques et observés par le calcul de la statistique de Khi-2

$$KHi-2 = \sum_{i=1}^q \sum_{j=1}^p \frac{(O_{théoriques} - O_{Observées})^2}{O_{théoriques}}$$

#### **Etape 4 : Prise de décision statistique**

Si Khi-2 = 0 ou très proche de zéro alors les deux variables sont indépendantes

Si Khi-2 est loin de zéro alors les deux variables sont dépendantes.

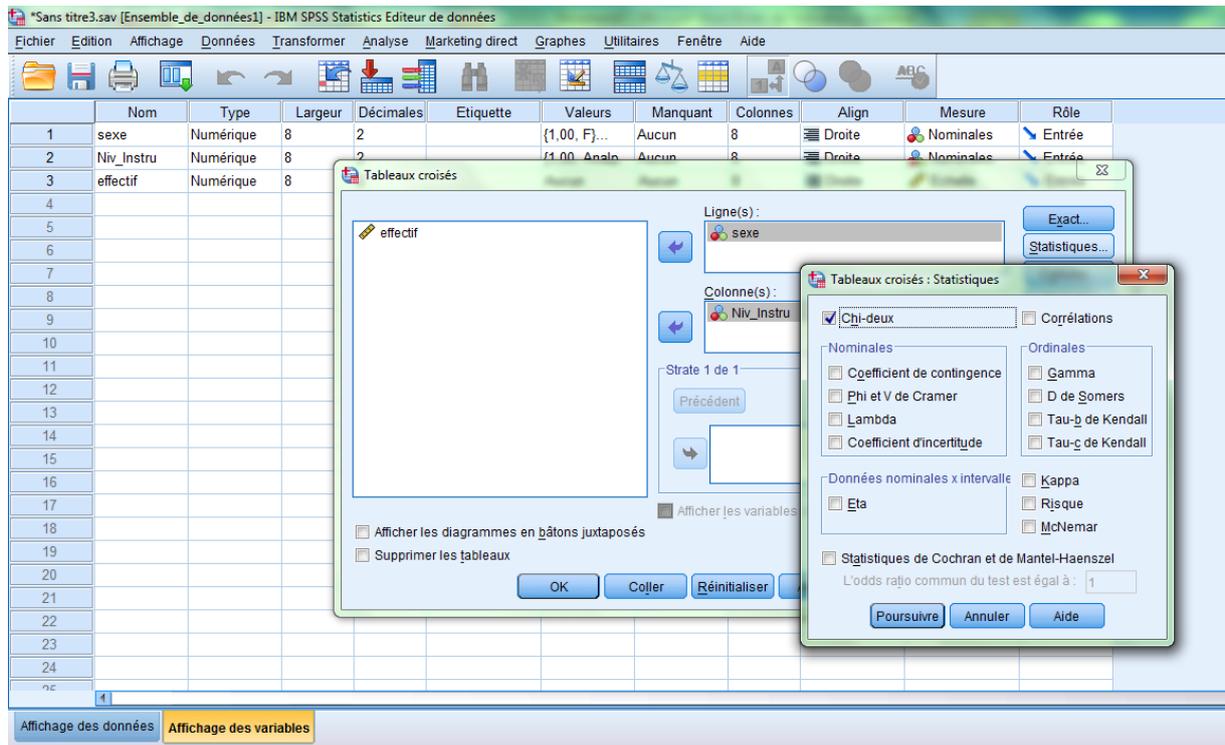
#### **Application par SPSS :**

**On reprend le tableau ci-dessus croisant le sexe avec le niveau d'instruction.**

**Objet :** on va chercher s'il existe une association entre le niveau d'instruction et le sexe ou aucune association ne peut être décelée.

### Procédure sous SPSS.

Menu Analyse/ Statistiques descriptives/Tableau croisé/Statistiques/Cochez Chi-2/poursuivre/Ok



### Le listing des résultats

#### Tests du Khi-deux

	Valeur	ddl	Signification asymptotique (bilatérale)
Khi-deux de Pearson	722,304 <sup>a</sup>	4	,000
Rapport de vraisemblance	797,193	4	,000
Association linéaire par linéaire	2,274	1	,132
Nombre d'observations valides	2306		

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5.  
L'effectif théorique minimum est de 166,35.

La statistique de khi-2 est de 722,304, elle est loin de zéro d'où les deux variables sont liées. En ce sens, la tendance générale dans les données montre

que le sexe féminin est généralement analphabètes ou ont un niveau secondaire.  
Alors que les masculins ont un niveau primaire ou professionnel.