

Traitement et analyse de données avec SPSS

Dr. Ir. Epiphane SODJINOU
Agroéconomiste, Biostatisticien

Contenu

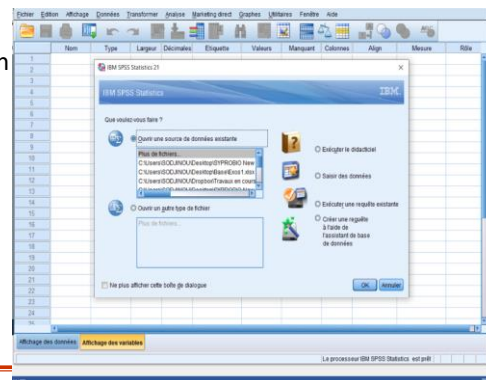
1. Introduction
2. Présentation du logiciel
3. Gestion de bases de données
4. Statistiques descriptives
5. Corrélations dans SPSS
6. Régressions dans SPSS
7. Analyse de la variance dans SPSS
8. Analyses multivariées dans SPSS

1. Introduction

- SPSS = Statistical Package for Social Science
 - Favorise la gestion des données dans un environnement graphique convivial associant menu descriptif et boîtes de dialogue
 - Son utilisation repose largement sur la manipulation de fenêtres et de menus déroulants → Le mode de fonctionnement des menus et la gestion des fenêtres répond aux principes généraux du système d'exploitation Windows.
 - Offre un langage de commande qui permet d'écrire des programmes afin d'optimiser des tâches de production
 - Permet d'effectuer des analyses sur de grandes bases de données

1. Introduction

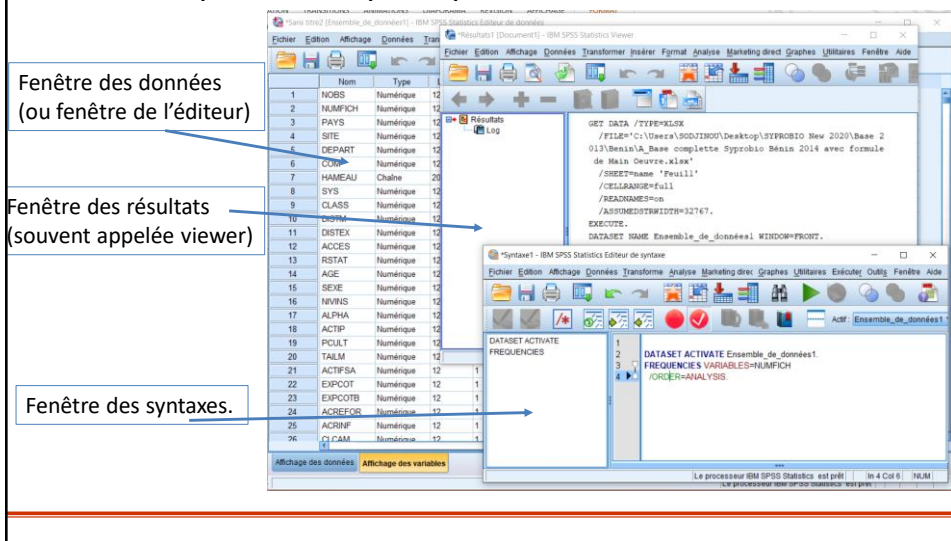
- Pour le démarrage
 - si vous avez un raccourci du logiciel sur le bureau, double-cliqué là-dessus
 - sinon, il il faut aller dans *Démarrer + IBM SPSS Statistics* ou cliquer sur *Démarrer + PASW Statistics 18*, selon la version installée
- La fenêtre de démarrage qui se présente
 - donne la liste des fichiers SPSS récemment ouverts. Pour ouvrir un fichier, il suffit de faire un double-clic sur son nom.
 - permet d'exécuter le didacticiel (aide), d'exécuter ou de créer un requête d'importation de données d'un fichier non SPSS



2. Présentation du logiciel

2.1. Les fenêtres

- SPSS comporte trois principales fenêtres



2. Présentation du logiciel

2.1. Les fenêtres

Fenêtre des données

- Permet d'entrer les données
- Présente deux onglets (en bas à gauche) :
 - affichage des variables
 - affichage des données.
- En cliquant sur affichage des variables, vous aurez l'ensemble de vos variables si vous avez déjà ouvert une base. Vous pouvez aussi ajouter ou supprimer des variables dans cette fenêtre
- Dans l'onglet « **affichage des variables** »
 - Une ligne équivaut à un cas, c'est-à-dire un individu ou une observation
 - Chaque colonne équivaut à une variable ou une information sur le cas

2. Présentation du logiciel

2.1. Les fenêtres

Fenêtre des résultats

- Affiche les résultats des commandes exécutées par SPSS
- Donne les sortis de tableaux, de graphiques, etc.
- S'ouvre automatiquement dès qu'on lance une procédure qui doit afficher des résultats

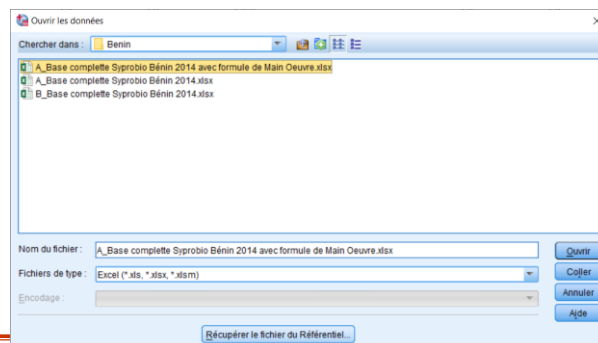
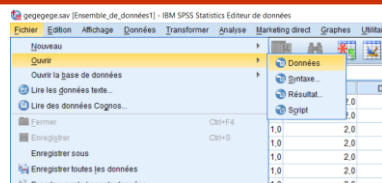
Fenêtre des syntaxes

- Permet de saisir les commandes SPSS
- SPSS permet en effet d'obtenir la syntaxe de toute opération avec le bouton *coller*
- En cliquant sur *coller*, vous obtenez automatique la commande, correspondant à l'analyse en cours, dans la fenêtre syntaxe

3. Gestion de bases de données

3.1. Importation de données dans SPSS

- Fichier > Ouvrir > Données
- Dans la boîte de dialogue qui s'affiche, préciser le type de fichier dans « **Fichiers de type** » puis indiquer l'emplacement du fichier dans « **Chercher dans** »



3. Gestion de bases de données

3.2. Fusion de fichiers

Ajout d'observations

- Il s'agit de fusionner deux fichiers contenant les mêmes nombres de variables des individus différents

Comment faire:

- ouvrez le fichier 1,
- Allez dans **données** → **fusionner les fichiers** → **ajouter des observations**.

Fichier 1

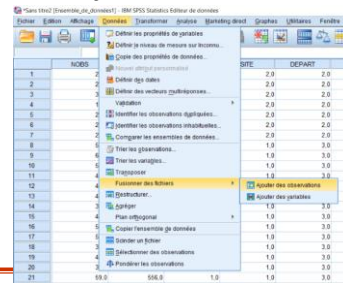
Individus	V1	V2	...	VJ-1	VJ
1					
2					

Fichier 2

Individus	V1	V2	...	VJ-1	VJ
1					
2					
...					
600					

Fusion des deux fichiers

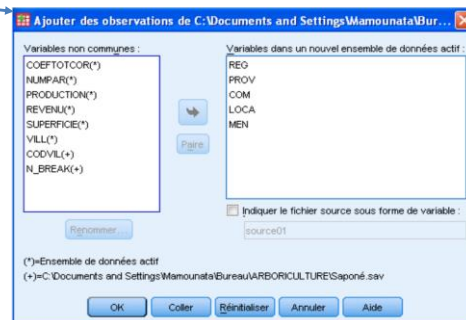
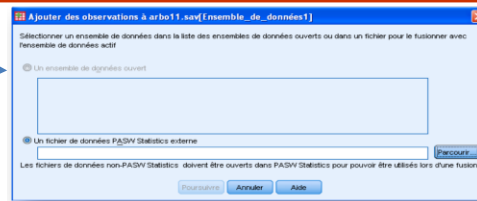
Individus	V1	V2	...	VJ-1	VJ
1					
2					
...					
500					
501					
...					
1 100					



3. Gestion de bases de données

3.2. Fusion de fichiers

- Première fenêtre de l'ajout des observations
- Cliquer sur parcourir
- Puis allez dans le répertoire où se trouve le fichier et sélectionnez-le
- La figure suivante apparaît avec deux grande zone
 - variables non communes** (cette contiendra les variables qui ne figure pas à la foi dans les deux fichiers)
 - variables dans un nouvel ensemble de données actif** (c'est les variables que SPSS vous propose pour la construction du fichier concaténé)
- Cliquez sur ok, et le tour est joué, le fichier



3. Gestion de bases de données

3.2. Fusion de fichiers

Ajout de variables

- Il s'agit de fusionner deux fichiers contenant les mêmes nombres d'observations mais des variables différentes

Fichier 1

individu	D1	D2	DJ-1	DJ
1				
2				
...				
100				

+

Fichier 2

individu	E1	E2	...	EK-1	EK
1					
2					
...					
100					

Fichier fusionner

individu	D1	D2	DJ-1	DJ	E1	E2	...	EK-1	EK
1									
2									
...									
100									

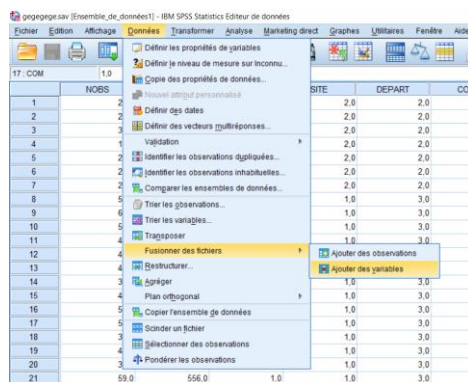
3. Gestion de bases de données

3.2. Fusion de fichiers

Ajout de variables

Comment faire:

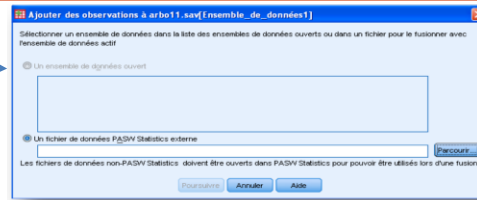
- Trier d'abord les deux fichiers suivant la (ou les) variable(s) clé(s), c'est-à-dire suivant un identifiant
- Ouvrez le fichier 1,
- Cliquez sur **données** → **fusionner les fichiers** → **ajouter des variables**



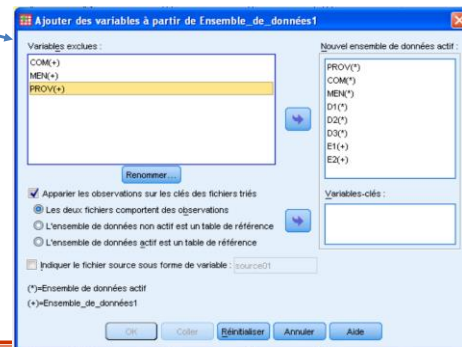
3. Gestion de bases de données

3.2. Fusion de fichiers

- Première fenêtre de l'ajout des observations
- Cliquer sur parcourir
- La nouvelle fenêtre a deux panneaux:



- A gauche les Variables exclues (variables appartenant aux deux fichiers) ;
- à droite l'ensemble des variables du fichier fusionné.
- Apparier les observations sur les clés des fichiers triés



3. Gestion de bases de données

3.2. Fusion de fichiers

Table de consultation

- Table de consultation : nous avons deux fichiers concernant deux enquêtes. Une sur les ménages et l'autre les des individus tiré dans les ménages. Donc nous avons un fichier ménages et un fichier individus. Généralement le fichier individus est largement supérieur (en terme de taille) au fichier ménage car dans un ménage donné on peut tirer plusieurs personnes.
- **Mise en pratique :**
 - On reprend le même processus comme celui dans le cas de l'ajout de variable mais en cochant cette fois l'option « *l'ensemble des données actif est une table de référence* » ou « *l'ensemble des données non actif est une table de référence* » selon le fichier sur lequel on travaille est le fichier ménage ou non.

3. Gestion de bases de données

3.3. Caractéristiques d'une variable

- Dans l'onglet « affichage des variables », les caractéristiques suivantes: **Nom, Type, Largeur, Décimales, Etiquette, Valeur, Manquante, Colonnes, Align, Mesure et Rôle.**

gegegege.sav [Ensemble_de_données1] - IBM SPSS Statistics Editeur de données

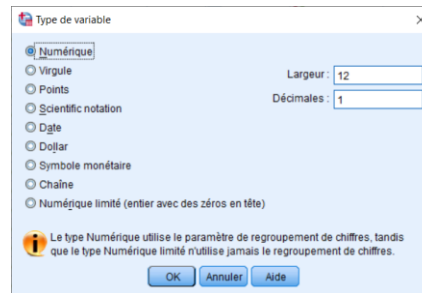
	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	NOBS	Numérique	12	1		Aucun	Aucun	12	Droite	Echelle	Entrée
2	NUMFICH	Numérique	12	1		Aucun	Aucun	12	Droite	Echelle	Entrée

- Nom :**
 - Nom de la variable, c'est-à-dire un abrégé (au maximum en 8 lettres) de la variable
 - Eviter d'y mettre des caractères accentués comme é, à, è, etc.
 - Exemple pour la question : quel est l'âge du chef de ménage ? On peut prendre comme **nom** de la variable : **AGECM**

3. Gestion de bases de données

3.3. Caractéristiques d'une variable

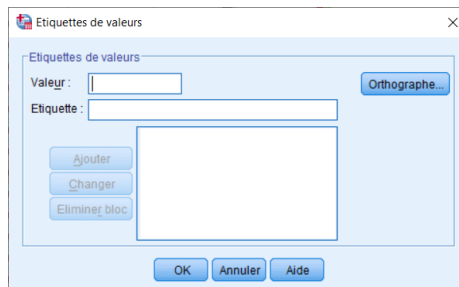
- Type:**
 - spécifie la nature numérique, alphanumérique, format date, etc. de la variable
 - Virgule** et **points** désignent les types de séparateurs de milliers des nombres ;
 - chaîne** est relative aux chaînes de caractère non numériques



3. Gestion de bases de données

3.3. Caractéristiques d'une variable

- La **largeur** de la variable indique
 - le nombre de positions occupées par les valeurs prises par cette variable
- **Décimal**
 - désigne le nombre de décimal à *afficher* après la virgule
- **Etiquette** de variable:
 - Permet d'indiquer la signification réelle des variables
 - Exemple: *Age du chef de ménage.*
- **Valeur** :
 - Permet d'enregistrer les modalités des variables



3. Gestion de bases de données

3.3. Caractéristiques d'une variable

- **Manquant** :
 - permet de donner une valeur pour les individus qui n'ont pas de réponse pour une question donnée.
- **Colonne** :
 - indique la largeur de la colonne de la variable en question
- **Aligner** :
 - indique si les valeurs de la variable doivent être centrées dans les cellules ou aligner à gauche ou à droite. Par défaut SPSS aligne les nombres à droite et les chaînes à gauche.
- **Mesure** :
 - indique l'échelle de mesure de la variable. Il y'a 3 échelles :
 - **nominal** pour les variables discrètes non ordonnées (comme le numéro de la région ou la province) ;
 - **ordinal** pour les variables discrètes ordonnées
 - **échelle** désigne les variables continues.
- **Rôle** :
 - permet de donner le rôle de la variable dans les futures analyses (variable dépendante, indépendantes, discriminante).

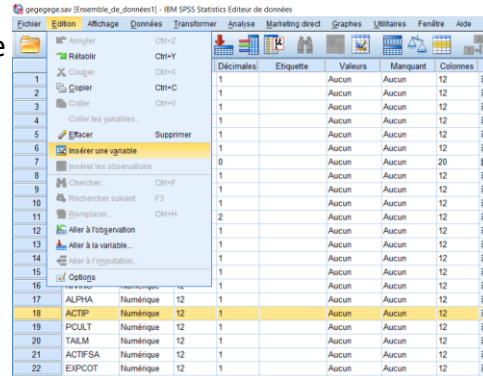
3. Gestion de bases de données

3.4. Insertion et tri

Insertion de variables

- Pour insérer une variable
 - Placez vous dans la fenêtre affichage de variables,
 - Sélectionnez la variable au-dessus de laquelle (ou avant laquelle) vous souhaitez insérer la nouvelle variable

- Cliquez sur **Edition**, puis sur **insérer une variable**: SPSS insère la variable juste en haut de votre sélection



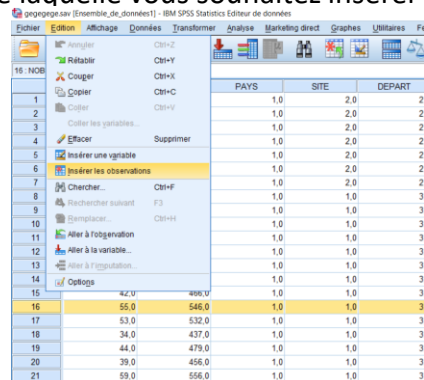
3. Gestion de bases de données

3.4. Insertion et tri

Insertion d'observations

- Pour insérer une observation
 - Placez vous dans affichage de données
 - Cliquez sur la ligne au-dessus de laquelle vous souhaitez insérer la nouvelle observation

- Cliquez sur **Edition**, puis sur **insérer les observations**: SPSS insère l'observation juste en haut de votre sélection

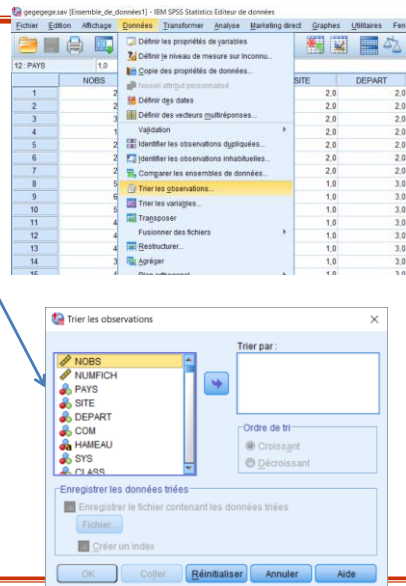


3. Gestion de bases de données

3.4. Insertion et tri

Trier les données selon une ou plusieurs variables

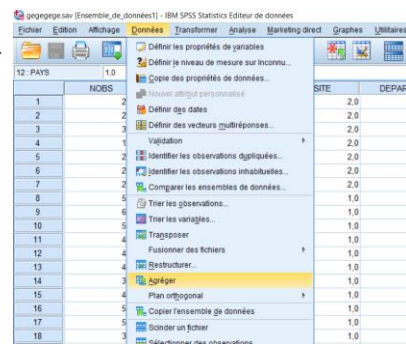
- Pour trier votre fichier selon une ou plusieurs variables, Cliquez sur **Données** → **trier les observations** puis, dans la fenêtre qui s'affiche,
- Sélectionner la (ou les) variable(s) selon laquelle (ou lesquelles) vous voulez trier (par ordre de préférence)
- Envoyez la (ou les) variable(s) sélectionnées dans le panneau **Tri par**
- Choisissez l'**Ordre de tri** : croissant ou décroissant
- Cliquez sur ok



3. Gestion de bases de données

3.5. Agréger des variables

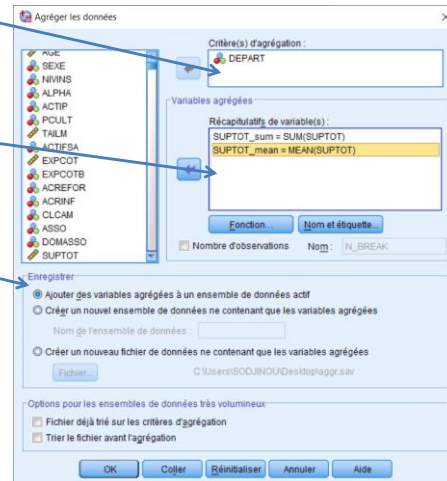
- Supposons qu'on a un fichier contenant des ménages, et d'autres variables telles que la commune du ménage, la superficie agricole du ménage, etc.
- On sait que la superficie agricole totale d'une commune est la somme des superficies agricoles des ménages de la commune en question ; la superficie agricole moyenne par ménage d'une commune est la moyenne de la variable superficie dans la commune en question
- La commande agréger permet de calculer ces nouvelles variables.
- Procédez comme suit :
 - Dans **affichage des données**,
 - Cliquez sur **donnée**
 - puis sur **agréger**



3. Gestion de bases de données

3.5. Agréger des variables

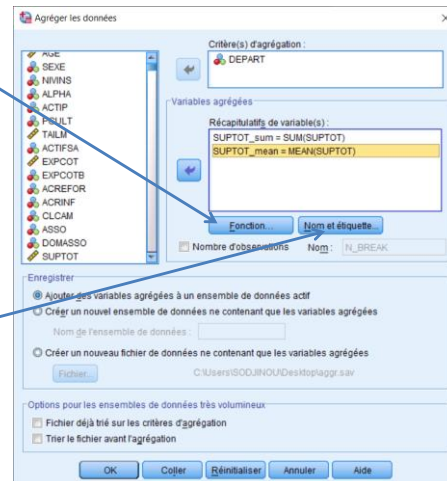
- Envoyez dans le panneau **critère(s) d'agrégation** la variable département
- Envoyez dans le panneau **Récapitulatifs de variable(s)** la variable superficie
- En bas de votre fenêtre, vous avez 3 options :
 - Ajouter les variables agrégées à un ensemble de données actif
 - Créer un nouveau fichier contenant seulement les variables agrégées : dans ce cas le logiciel crée le fichier et l'ouvre automatiquement
 - Créer un nouveau fichier contenant seulement les variables agrégées : dans ce cas le logiciel crée le fichier avec un nom et un emplacement que vous indiquez avec le bouton **Fichier**



3. Gestion de bases de données

3.5. Agréger des variables

- Par défaut SPSS vous donnera **la valeur moyenne**, pour changer cela, cliquez sur l'option **fonction**.
- Dans la fenêtre qui s'affiche, vous avez une panoplie de fonctions. Choisissons ici : somme des valeurs. Faites encore entrer la variable superficie et garder la fonction moyenne.
- Puis changer les noms et étiquettes de vos nouvelles variables avec le bouton **Nom et étiquette** puis faire OK.

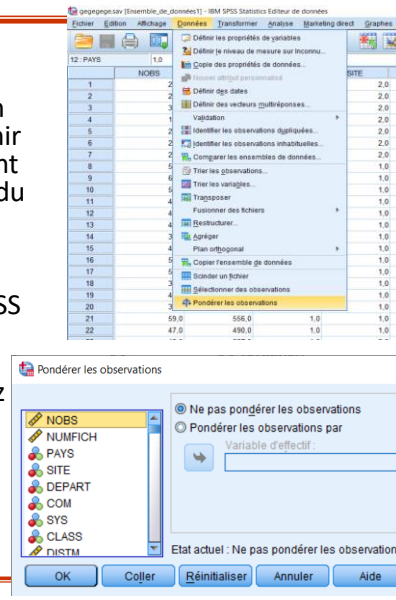


3. Gestion de bases de données

3.6. Pondération, calcul et recodage

Pondération

- Utile lorsqu'on cherche à extrapoler les résultats d'enquête sur toute la population
→ faire une pondération, c'est-à-dire définir un coefficient de pondération ou coefficient d'extrapolation → rattacher chaque individu à valeur de la variable coefficient d'extrapolation
- Avant la sortie des résultats, indiquer à SPSS cette variable d'extrapolation
- Cela se fait de la manière suivante : cliquez sur **donnée**, puis sur **pondérer les observations**
- Dans la fenêtre qui s'affiche, choisissez **pondérer les observations par**.

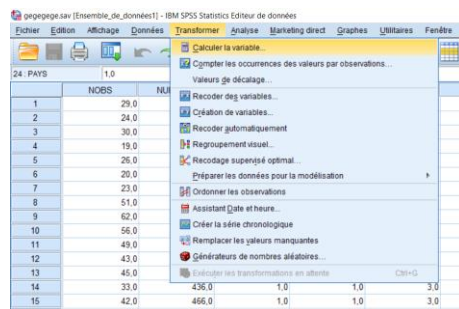


3. Gestion de bases de données

3.6. Pondération, calcul et recodage

Calcul

- La fonction calculer permet de créer une nouvelle variable à l'aide d'anciennes variables.
 - Par exemple pour avoir une variable *rendement à l'hectare*, il suffit de diviser la *production* par la *superficie*
- Quelques options de cette fonction: logarithme, exponentiel, valeur absolue, etc.
- Pour utiliser cette fonction
 - cliquer sur **Transformer** → **calculer la variable**

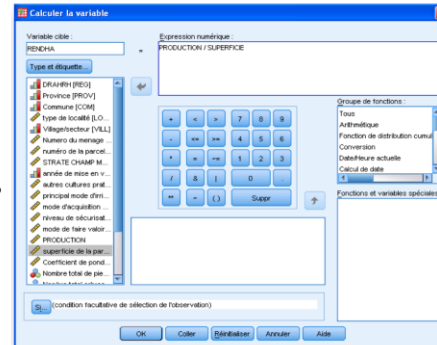


3. Gestion de bases de données

3.6. Pondération, calcul et recodage

Calcul

- Dans le panneau **Variable cible**, le nom de la nouvelle variable (Exemple RENDHA : pour dire *rendement à l'hectare*)
- Juste en bas du panneau **Variable cible**, cliquez sur **Type et étiquette** pour précisez surtout le type de la variable
- Dans le panneau **Expression numérique**, écrivez l'expression nécessaire au calcul de la nouvelle variable (Exemple: PRODUCTION/SUPERFICIE)

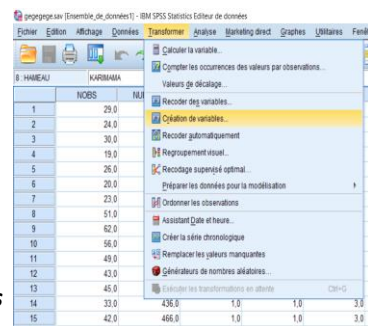


3. Gestion de bases de données

3.6. Pondération, calcul et recodage

Recodage de variable

- Parfois, on a envie de recoder une variable en classe de valeur. Par exemple lorsqu'on a la superficie exploitations, on peut vouloir recoder cette variable en cinq groupes de 1 à 5
 - Superficie < 2 ha → 1
 - 2 ha ≤ superficie < 4 ha → 2
 - 4 ha ≤ superficie < 6 ha → 3
 - 6 ha ≤ superficie < 8 ha → 4
 - superficie > 8 ha → 5
- Pour cela :
 - Cliquez sur **transformer** ; vous aurez deux options de recodage: **recodage des variables** et **création de variables**.
 - Prenez l'option **création de variables** qui permet de préserver la variable source, contrairement à l'option **recodage de variable** qui écrase la variable source

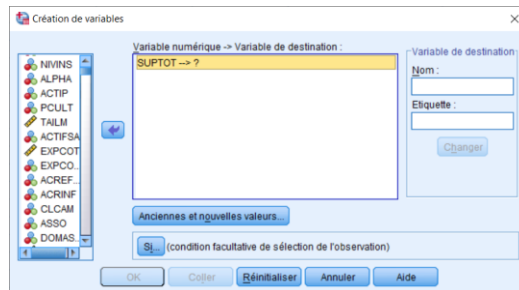


3. Gestion de bases de données

3.6. Pondération, calcul et recodage

Recodage de variable

- Faire passer la variable que voulez recoder dans la fenêtre variable numérique variable
- Dans la zone nom écrivez SUPCLAS; comme étiquette on peut mettre : Superficie par classe
- Maintenant cliquez sur **anciennes et nouvelles valeurs**.

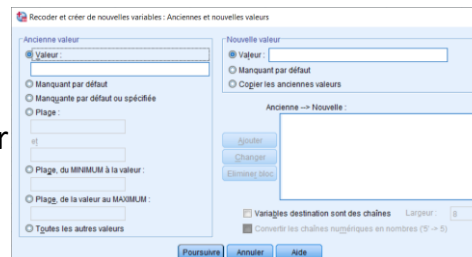


3. Gestion de bases de données

3.6. Pondération, calcul et recodage

Recodage de variable

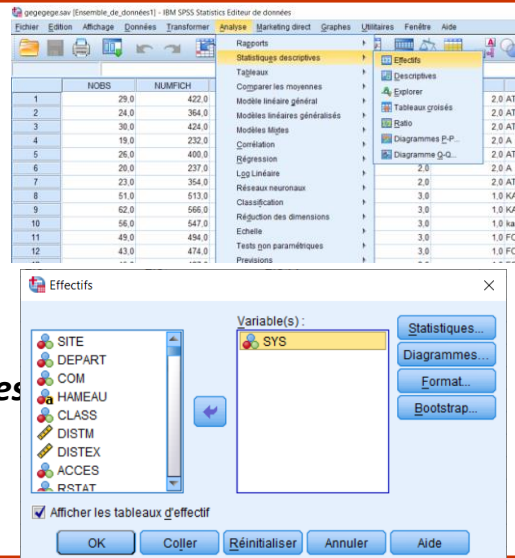
- Dans la nouvelle fenêtre qui s'affiche, vous devez spécifier le mode de recodage.
- Saisissez dans la zone **Intervalle** : 0 à 2; puis tapez 1 dans la zone **Valeur** en haut à droite, puis cliquez sur **Ajouter**. Encore tapez dans **Intervalle** 2 à 4 puis dans **Valeur** mettez 2 puis cliquez sur **Ajouter**, ainsi de suite jusqu'à la fin.
 - N.B.: Ne pensez pas qu'il aura une ambiguïté sur les bornes: tous les 2 seront coder en 1, les 4 en 2.
- Cliquez sur **Poursuivre**, ainsi vous revenez à la précédente fenêtre, cliquez sur remplacer puis sur ok. Et voila votre variable codifiée.



4. Statistiques descriptives

4.1. Fréquences

- Cliquez sur **Analyse**
- Choisissez **statistiques descriptives**
- Choisissez **effectif**
- Choisissez les variables pour lesquelles vous désirez des statistiques.
- Envoyer ces variables dans le panneau **Variables**
- Cliquez sur ok pour terminer

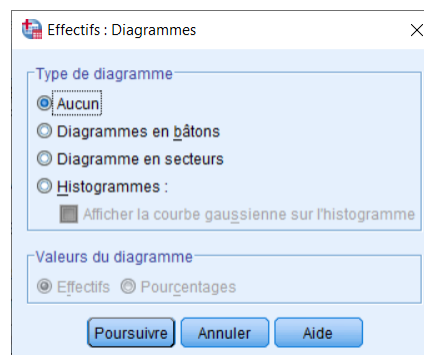


4. Statistiques descriptives

4.1. Fréquences

Fréquences

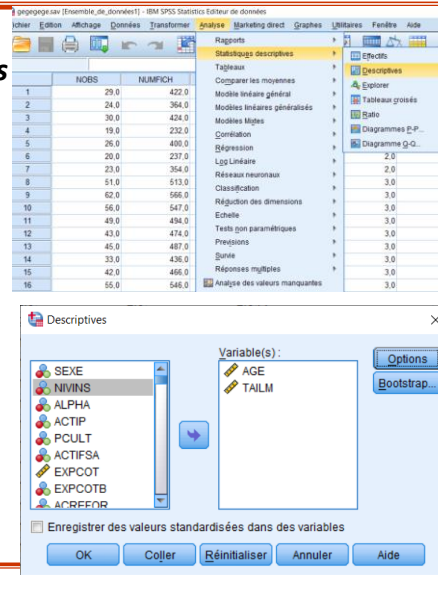
- On peut activer l'option « **Diagrammes** » pour produire le tableau des effectifs avec un graphique correspondant
- Le choix de diagramme se fait comme suit :
 - Pour les données quantitatives classifiées : **Diagramme en bâtons**, (Tranche d'âge, Salaire, ...)
 - Pour les données qualitatives : **Diagramme en secteurs**, (Sexe, Statut matrimonial...)



4. Statistiques descriptives

4.2. Moyennes et autres

- Cliquez sur **Analyse**
- Choisissez **statistiques descriptives**
- Choisissez **Descriptives**
- Choisissez les variables pour lesquelles vous désirez des statistiques.
- Envoyer ces variables dans le panneau **Statistiques**
- Cliquez sur l'onglet **Options** pour mieux préciser les statistiques (moyenne, écart-type, plage, etc.) que vous souhaitez
- Cliquez sur **poursuivre** puis sur **ok** pour terminer

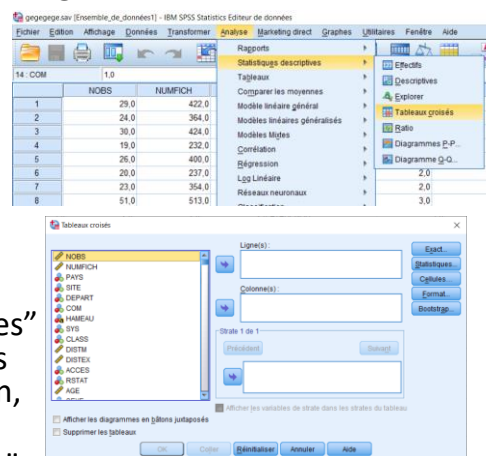


4. Statistiques descriptives

4.3. Tableaux croisés

Croisements entre 2 variables catégorielles

- Cliquez sur **Analyse**
- Choisissez **statistiques descriptives**
- Choisissez **Tableaux croisés**
- Choisissez les deux variables que vous désirez croisées (une dans "Ligne(s)" et une dans "Colonne(s)").
- Cliquez sur le bouton "cellules" et cochez votre souhait sous "pourcentages" (row, column, total)
- Cliquez sur "continue" et "ok"

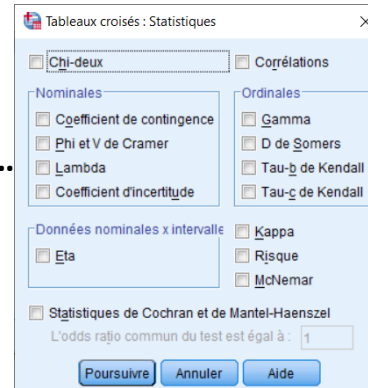


4. Statistiques descriptives

4.3. Tableaux croisés

Croisements entre 2 variables catégorielles

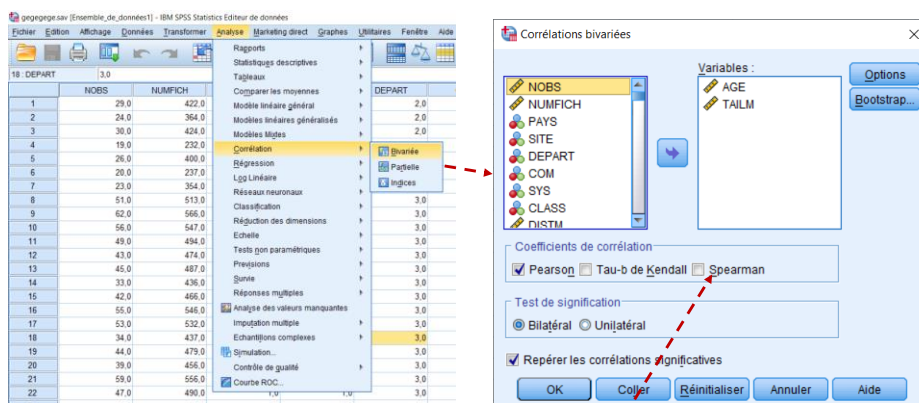
- On peut activer l'onglet « **Afficher les diagrammes en bâtons juxtaposés** » pour obtenir le **graphique** des tableaux croisés.
- Pour obtenir des statistiques pour tests d'hypothèses, il y a lieu d'activer le bouton « **Statistiques...** »
 - On peut faire, par exemple, le test de Khi-deux ou encore calculer le coefficient de contingence



5. Corrélations dans SPSS

5.1. Corrélation de Pearson

- Analyse > Corrélation > Bivariée**

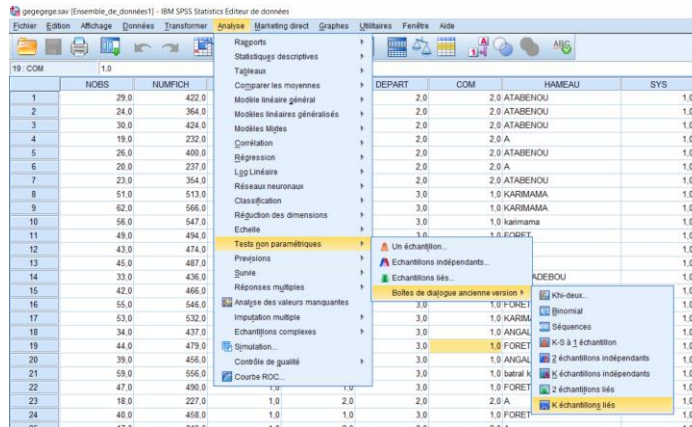


- La même boîte de dialogue permet de calculer aussi le coefficient de corrélation de Spearman

5. Corrélations dans SPSS

5.2. Corrélation de Kendal

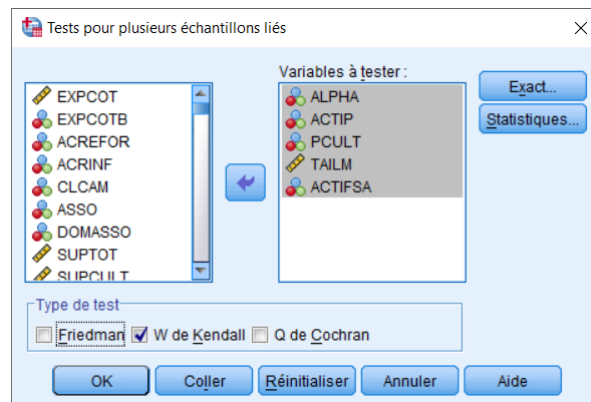
- **Analyse > Tests non paramétriques > Boîtes de dialogue ancienne version > k échantillons liés**



5. Corrélations dans SPSS

5.2. Corrélation de Kendal

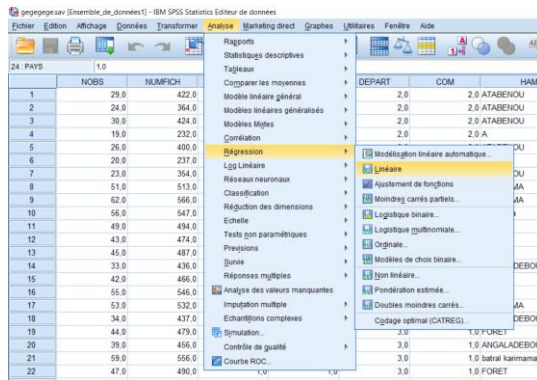
- **Analyse > Tests non paramétriques > Boîtes de dialogue ancienne version > k échantillons liés**



6. Régressions dans SPSS

6.1. Régression linéaire

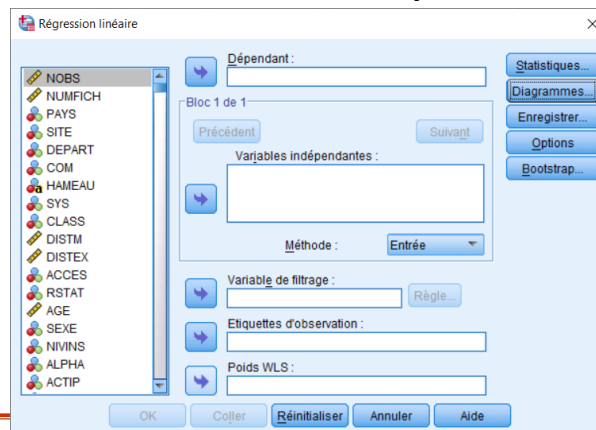
- **Analyse > Régression > Linéaire**



6. Régressions dans SPSS

6.1. Régression linéaire

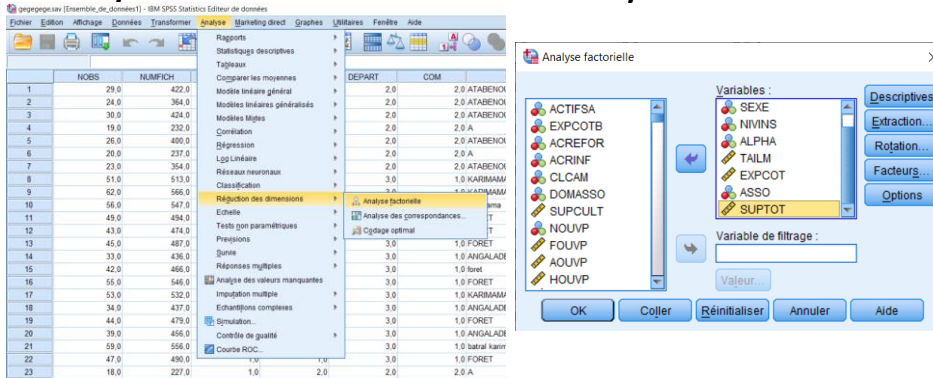
- Dans la boîte de dialogue qui s'affiche, mettre la variable dépendante dans « **Dépendant** » et les variables explicatives dans « **Variables indépendantes** »



7. Analyses multivariées dans SPSS

7.1. Analyse factorielle

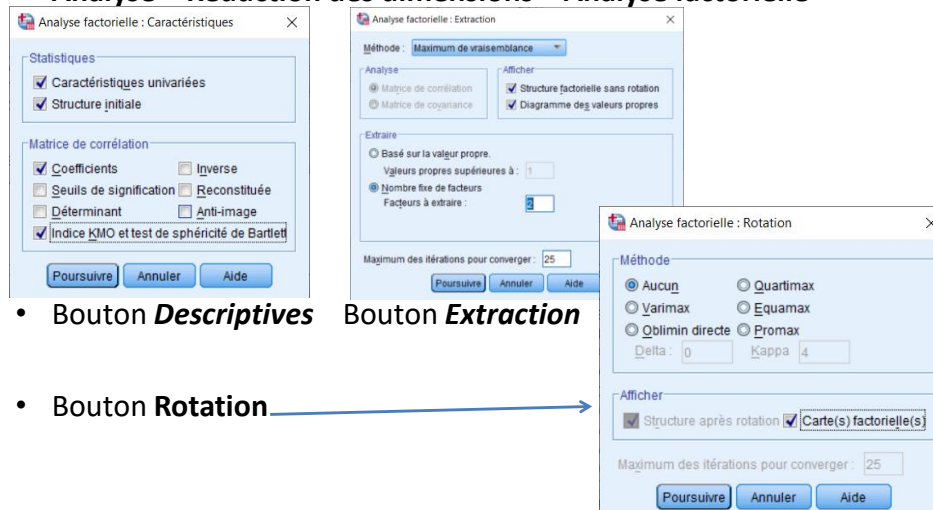
- **Analyse > Réduction des dimensions > Analyse factorielle**



7. Analyses multivariées dans SPSS

7.1. Analyse factorielle

- **Analyse > Réduction des dimensions > Analyse factorielle**

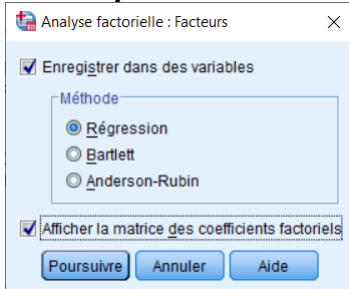


- Bouton **Descriptives**
- Bouton **Extraction**
- Bouton **Rotation**

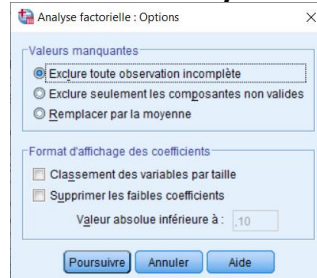
7. Analyses multivariées dans SPSS

7.1. Analyse factorielle

- **Analyse > Réduction des dimensions > Analyse factorielle**



- Bouton **Facteurs**



- Bouton **Options**

7. Analyses multivariées dans SPSS

7.2. Analyse Discriminante

- But de l'analyse discriminante
 - d'étudier les relations entre une variable qualitative et un ensemble de variables explicatives quantitatives
- Objectifs principaux assignés à l'analyse discriminante :
 - Déterminer les variables explicatives les plus discriminantes vis à vis des classes déterminées
 - Déterminer à quel groupe appartient un individu à partir de ses caractéristiques
 - Mais surtout à **valider une classification** ou à **faire un choix entre plusieurs classifications pour savoir laquelle est la plus pertinente**. L'analyse discriminante intervient donc a posteriori d'une classification

7. Analyses multivariées dans SPSS

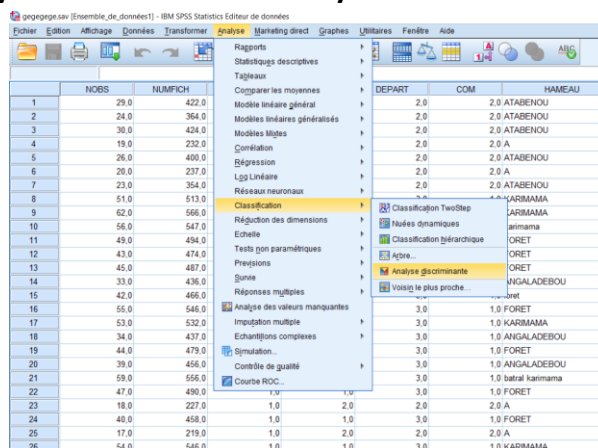
7.2. Analyse Discriminante

- Conditions à remplir :
 - Les variables explicatives doivent être métriques
 - Elles ne doivent pas être trop corrélées entre elles. Cela se vérifie par l'observation des corrélations entre les variables. Si c'est le cas, on peut passer par une analyse factorielle qui permet de réduire les données à quelques axes. Ces axes sont, par propriété, non corrélés entre eux
- Une analyse discriminante se déroule en 3 étapes:
 - On vérifie l'existence de différences entre les groupes
 - On valide l'étude
 - On vérifie le pouvoir discriminant des axes
 - On juge la qualité de la représentation du modèle

7. Analyses multivariées dans SPSS

7.2. Analyse Discriminante

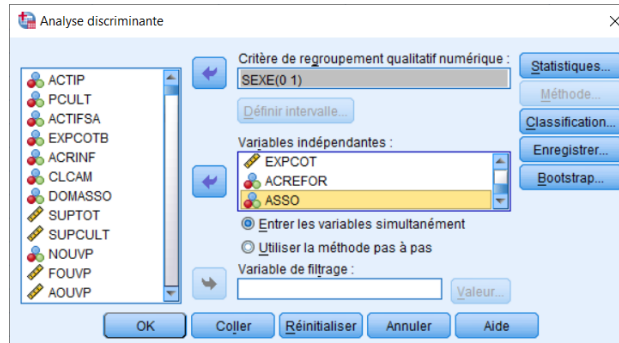
- Analyse > Classification > Analyse Discriminante



7. Analyses multivariées dans SPSS

7.2. Analyse Discriminante

- **Analyse > Classification > Analyse Discriminante**

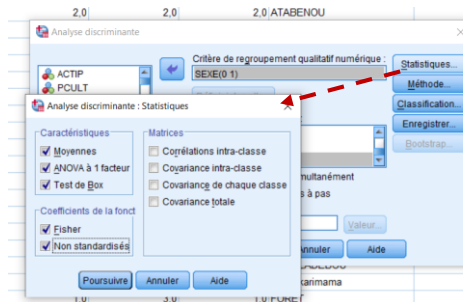


- Dans «**les critère de regroupement**», il faut indiquer la variable à expliquer en la sélectionnant dans la partie de droite puis en cliquant sur la flèche qui pointe vers la droite. SPSS demande alors de définir l'intervalle, c'est-à-dire les différentes modalités que la variable peut prendre.

7. Analyses multivariées dans SPSS

7.2. Analyse Discriminante

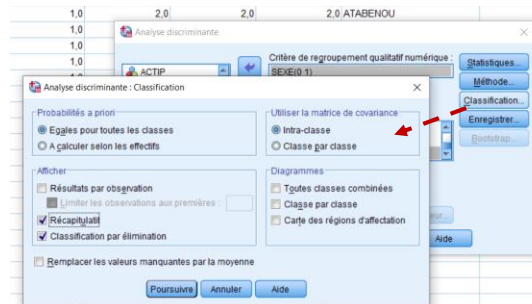
- **Analyse > Classification > Analyse Discriminante**
- Dans «**les variables explicatives** » (i.e. *Independents*), il faut indiquer les variables métriques que l'on souhaite intégrer à l'analyse. Il est important de choisir « **la méthode pas à pas** »
- Trois options s'offrent : « Statistiques... », « Méthodes... » et « Classification... ». On ne touchera pas aux différentes options de « Méthodes... »
- **Statistiques**
 - Cochez «**Moyennes**», «**ANOVA à 1 facteur**» et «**Test de Box** », « Fisher » et « Non Standardisés »



7. Analyses multivariées dans SPSS

7.2. Analyse Discriminante

- **Analyse > Classification > Analyse Discriminante**
- **Classification...**
 - Cochez «Récapitulatif» (option qui permet l’affichage de la matrice de confusion) et «Classification par élimination» (Leave-one-out classification)



8. Autres

8. Exportation des résultats

Exportation vers Excel ou Word

- À partir du fichier “outputs” dans lequel se trouve les résultats à exporter
- Cliquez sur **“Fichier”** puis sélectionnez **“Exporter”**
- Choisir le type de fichier (word, excel, etc.) et le nom du fichier ainsi que le repertoire où vous désirez l’exporter sur votre ordinateur.
- Vous pouvez aussi choisir si vous désirez exporter le document en entier ou seulement une partie (ex. seulement le graphique)
- Cliquez sur **“ok”**.

