

Analyse de données quantitatives (LPOLS1221)
Séance 2 : Data Management & Statistiques descriptives

1. Transformations et créations de variables

Nom de la procédure (Menu Transform)	Fonction
Compute Variable	Créer/transformer des variables sur base d'opérations arithmétiques
Count Values within cases	Compte les valeurs apparaissant au sein des différentes variables
Recode into different variables	Recodage spécifique de valeurs tout en créant une/des variables
Recode into same variables	Recodage spécifique de valeurs au sein de la variable d'origine /\

2. Gestion des données

Nom de la procédure (Menu Data)	Fonction
Define variables Properties	Paramétrage des variables → Permet d'obtenir un code de syntaxe
Select Cases	Sélectionne un sous-échantillon sur base de condition spécifique
Split Cases	Permet d'appliquer des analyses sur plusieurs sous-échantillons spécifiques.
(Menu Utilities)	
Define variable set	Crée un groupe de variable spécifique
Use variable set	Permet de n'afficher que les variables présentes dans le groupe sélectionné

3. Analyse descriptive

Nom de la procédure (Menu Analyze/Descriptive Statistics)	Fonction
Frequencies	Fonction adaptée à l'analyse descriptive univariée de variables nominales ou ordinales. Si vous voulez utiliser cette procédure afin de produire des statistiques sur des variables de type numérique, veillez à décocher l'option « <i>Display frequency tables</i> » et à sélectionner les statistiques voulues dans l'onglet <i>statistics</i> . Cependant, vous verrez que d'autres procédures sont mieux adaptées à l'analyse des variables de type quantitatif. Possibilité d'obtenir des graphiques univariés
Descriptives	Permet d'obtenir des statistiques descriptives univariées rudimentaires. Plutôt adapté à l'analyse des variables quantitatives.
(Menu Analyze/Compare means)	
Means	Procédure idéale pour produire très rapidement des statistiques de base (moyenne, écart-type par défaut) sur des variables quantitatives. La procédure permet également de séparer l'information en fonction de groupes particuliers. Pour cela, il faudra renseigner une variable d'intérêt (celle dont on veut avoir la moyenne) en « <i>Dependent list</i> » et une variable de regroupement en « <i>Independent list</i> ». L'onglet « options » permet de sélectionner des statistiques supplémentaires ou d'en enlever. Parmi elles : moyenne (mean), médiane (median), nombre de cas (Number of Cases), écart-type (Standard deviation), minimum, maximum, etc. Permet donc d'obtenir des statistiques descriptives bivariées.

(Menu Graphs/Legacy Dialogs)

Bar	Graphique utile pour la représentation de variables qualitatives. L'option « Bar clustered » permet une représentation bivariable.
Line/Pie	Permet d'autres représentations de variables qualitatives. Leur interprétation est cependant souvent moins aisée que celle du diagramme en barres.
Histogram	Graphique utile pour la représentation univariée de variables quantitatives.
Box-Plot	Graphique utile pour la représentation conjointe d'une variable qualitative et quantitative
Scatter/dot	Graphique utile pour la représentation bivariable de variables quantitatives

Exercices

Utilisez la base de données EVS pour réaliser ces exercices.

1. Créez un jeu de variables, comprenant les variables suivantes et utilisez le jeu de variable ainsi que les nouvelles variables qui seront créées : les variables

Year: survey year
Country: country code
v46: don't like as neighbours: people with criminal record
v47: don't like as neighbours: people of different race
v48: don't like as neighbours: left wing extremists
v49: don't like as neighbours: heavy drinkers
v50: don't like as neighbours: right wing extremists
v51: don't like as neighbours: large families
v52: don't like as neighbours: emotionally unstable people
v53: don't like as neighbours: muslims
v54: don't like as neighbours: immigrants/foreign workers
v55: don't like as neighbours: people with AIDS
v56: don't like as neighbours: drug addicts
v57: don't like as neighbours: homosexuals
v58: don't like as neighbours: jews
v59: don't like as neighbours: gypsies
v60: don't like as neighbours: christians
V264_LR: which political appeals to you most left/right scale
V302: Sex respondent
V303: Respondent Year of birth
V336_r: education level (recoded)
V355_r: educational level father (recoded)

2. Sélectionnez uniquement les belges
3. Créez une variable « Age » à partir de la variable Year et V303. Paramétrez cette variable. Faites une analyse univariée de cette variable et commentez.
4. Créez une variable « Intolérance » qui va compter le nombre de critères des voisins qui irrite le répondant. Faites une analyse univariée de cette variable et commentez.
5. Créez une variable « couleur politique » à partir de la variable v264_LR: les scores 1 à 4 représenteront la gauche, les score 5 et 6 le centre et les scores allant de 7 à 10, la droite. Faites une analyse univariée de cette variable et commentez.
6. Peut-on observer des différences de moyenne de la variable Intolérance en fonction du fait de voter à gauche, au centre ou à droite ? Réalisez une analyse bivariée sous forme de tableau, représentez-la graphiquement et commentez.
7. Réalisez la même analyse mais pour les ressortissants de Grèce. Est-ce que le phénomène observé est similaire ?
8. Créez un graphique pertinent qui permet de visualiser la relation entre le niveau d'éducation des répondants de Belgique et le niveau d'éducation de leur père. Que peut-on en dire ?
9. Observe-t-on des différences d'âge en fonction de l'orientation politique des répondants ? Produisez un tableau, représentez la relation graphiquement, commentez.
10. Trouvez un moyen pour produire deux fois le graphique réalisé au point 9, un pour les hommes et un pour les femmes, en effectuant une et une seule procédure graphique.