

Analyse statistique de données qualitatives et quantitatives en sciences sociales : TP

**IMPORTATION, PRÉPARATION DES DONNÉES,
RAPPORTS ET MISE EN FORME**



Objectifs de la séance

1. Présenter un exemple de thématique qui pourrait être traitée dans le cadre du travail que l'on vous demande de réaliser à la fin de ce cours.
2. Aborder les aspects techniques qui précèdent la mise en œuvre des méthodes d'analyses vues au cours théorique : la phase de préparation des données dans le logiciel SAS.

Exemple de thématique

- **Variables d'intérêt** : l'importance que l'individu accorde au fait de vivre dans un pays démocratique (IMPLVDM) et la manière dont il évalue le fonctionnement de la démocratie dans son pays (DMCNTOV).
- **Explicatives socio-démographiques** :
 - La région du pays (REGION) – On se limite à un seul pays, dans ce cas, la Belgique.
 - Le genre du répondant (GNDR)
 - Son âge (AGEA)
 - Son niveau d'éducation (EISCED)
 - Son statut occupationnel (MNACTIC)
 - Son niveau d'aisance – financière – (HINCFEL)
- **Explicatives « thématiques »** :
 - Son intérêt pour la politique (POLINTR)
 - Le fait qu'il appartienne ou non à une minorité ethnique (BLGETMG)
 - Sa position sur l'échelle gauche-droite (LRSCALE)
 - La place qu'il estime occuper dans la société (PLINSOC)

Critères pour le choix des variables

INTÉRÊT « SOCIOLOGIQUE » !!!

ADÉQUATION PAR RAPPORT AUX MÉTHODES

- Choisir une ou deux variable(s) d'intérêt (relatives à une même thématique) présentant suffisamment de niveaux (par exemple : échelle de Likert de 0 à 10).
- Travailler sur un seul pays (au choix) – et faire éventuellement des comparaisons inter-régionales
- Prendre un nombre suffisant de variables explicatives (pour la sélection des modèles)
- Ne pas sélectionner uniquement des explicatives socio-démographiques (intérêt sociologique limité) mais cibler aussi des variables liées à la thématique susceptibles d'apporter un éclairage intéressant.
- Avoir au moins une variable continue (l'âge dans l'ESS est l'une des seules disponibles)

Phase de préparation des données: Exercice à réaliser

1. Télécharger la base de données ESS2012 depuis le site de l'ESS et l'importer dans SAS
2. Le nouveau data set doit contenir uniquement :
 - Les réponses des 1869 Belges
 - Les variables DMCNTOV (variable d'intérêt), REGION, EISCED, HINCFEL, AGEA (explicatives socio-démo), POLINTR, PLINSOC, BLGETMG (explicatives thématiques)
3. Assigner si nécessaire des labels aux variables et des formats aux valeurs afin d'aider à l'interprétation des sorties.
4. Réaliser un premier rapport en format pdf reprenant la table des fréquences des 8 variables retenues avec des titres explicites et certains critères de mise en forme (style, format, pas de date, etc.).
5. Créer de nouvelles variables (versions recodées des variables initiales) et définir leurs labels.
6. Réaliser un second rapport en format pdf présentant les tables de contingence pour chaque croisement 2 à 2 entre les variables explicatives (recodées ou formatées) et la variable réponse (binarisée). Interpréter les premières tendances.
7. Créer un nouveau data set SAS reprenant les fréquences absolues de la variable réponse en fonction du niveau d'éducation et l'imprimer dans la fenêtre de sortie.

Instructions « Data step »

```
DATA output-SAS-data-set;
  LENGTH new-variable(s) $ length;
  SET input-SAS-data-set;
  WHERE where-expression;          /* Sélection des observations */
  KEEP variable-list ;            /* Sélection des variables */
  IF expression THEN statement ;
  ELSE IF expression THEN statement ; /* Création des nouvelles variables */
  LABEL variable = 'label'
        variable = 'label'
        variable = 'label';
  FORMAT variable(s) format;
RUN;
```

Procédures

```
PROC CONTENTS DATA = SAS-data-set <varnum>;  
    /* varnum = option de procédure pour lister les variables dans l'ordre où elles apparaissent dans le  
    data set plutôt que par ordre alphabétique (par défaut) */  
RUN;
```

```
PROC PRINT DATA = SAS-data-set;  
    VAR variable(s);  
RUN;
```

```
PROC FORMAT;                                     /*Définition de formats (! Avant l'assignation dans le code !) */  
    VALUE format-name range1 = 'label'  
        range2 = 'label'  
        .... ;  
    VALUE format-name range1 = 'label'          /* NB : Le nom du format commence par un $ pour les valeurs caractères*/  
        range2 = 'label'  
        .... ;  
RUN;
```

Procédures (suite)

```
PROC FREQ DATA = SAS-data-set < option(s) > ;  
    TABLES variable(s) </option(s)>;           /*one-way frequency table */  
RUN;
```

```
PROC FREQ DATA = SAS-data-set < option(s) >;  
    TABLES variable1*variable2 </option(s)>;   /* two way frequency table */  
RUN;           rows      columns
```

PROC FREQ Options :

- NLEVELS (affiche un résumé du nombre de niveaux pour chaque variable listée dans TABLES)
- ORDER = DATA | FORMATTED | FREQ | INTERNAL

TABLES Statement Options :

- NOPERCENT, NOROW, NOCOL
- OUT = SAS-data-set

Instructions globales

- Assigner une librairie :

LIBNAME libref 'chemin du folder'; /*libref : max 8 caractères, commence par une lettre ou un underscore*/

Ensuite, pour stocker les datasets dans la librairie permanente : les appeler par « libref.filename »

- Diriger les sorties vers une fichier pdf

ODS pdf file = 'chemin\nom_du_fichier.pdf' < style = journal > ;

SAS code to generate a report(s)

ODS pdf close;

- **TITLE***n* 'text';
- **OPTIONS** nodate nonumber center ... ;

!!! Caractéristiques des instructions globales : elles restent d'application jusqu'à ce qu'elles soient modifiées, supprimées ou que l'on mette fin à la session SAS.

Structure du programme SAS

ESS6e02_1_F1.sas

- LIBNAME LIBRARY '*chemin vers le dossier destiné à stocker les formats*';
- LIBNAME OUT '*chemin vers le dossier destiné à stocker la base de donnée*';
- PROC FORMAT LIBRARY=LIBRARY ; /*Définition de formats stockés dans la librairie permanente LIBRARY */
 VALUE format-name range1 = 'label'
 range2 = 'label'
 ... ;
RUN;
- DATA OUT. ESS6e02_1_F1;
 INFILE '*chemin vers le fichier txt*\ESS6e02_1_F1.txt';
 INPUT variables-names <\$> column_start-column_stop;
 FORMAT variable(s) format; /*Assignation des formats de manière permanente*/
 LABEL variable = 'label'
 variable = 'label'
 ... ;
 IF variable = value THEN variable = value ; /*Modification des valeurs pour les NSP REFUS MISSING*/
RUN;