



IUT STID, 1^{ère} année

Découverte de logiciels statistiques

Feuille 4 : SPSS

Vous devez me faire **un compte rendu de ce TP**. Vous le rédigerez sous Word et me l'enverrez au format PDF ou DOC avec la feuille de calcul SPSS à l'adresse vivien.rossi.stid@free.fr au plus tard une semaine après la dernière séance passée sur ce TP.

Ce compte-rendu sera évalué et participera à la moyenne du module, **il est donc à réaliser seul**.

1 Quelques fonctions utiles

- Il est possible, lorsque l'on dispose d'un fichier complet, de ne traiter que certaines des observations : pour cela, on fait appel au menu *Données* → *Sélection des observations*. Une boîte de dialogue s'ouvre, on a alors la possibilité de sélectionner les observations selon :
 - une condition logique (dépendant par exemple d'une variable, ex : sélectionner les individus dont la taille est supérieur à 1,50 m)
 - un échantillonnage aléatoire (les individus sont sélectionnés de manière aléatoire)
 - un intervalle de temps ou d'observation (on sélectionne, par exemple, les individus numérotés de 1 à 50)

Les individus non sélectionnés voient alors leurs numéros barrés et les traitements s'effectuent seulement sur les individus sélectionnés. Enfin, dans le cas où les individus ont été sélectionnés selon une condition logique, SPSS crée une variable filtre qui prend la valeur 1 pour les individus sélectionnés et la valeur 0 pour les individus non sélectionnés. Cette variable filtre reste en permanence dans le fichier et il est alors possible de s'en servir à d'autres fins.

- Il est également possible de faire en sorte que chaque ligne (ie : chaque observation) corresponde à plusieurs individus. Pour cela, on entre, en bout de ligne l'effectif correspondant à l'observation (dans une variable appelée, par exemple, effect) et on fait appel au menu *Données* → *Pondérer les observations*. Après avoir entré la variable de pondération effect dans la boîte de dialogue correspondante, chaque ligne sera pondérée de la valeur de la variable effect.

2 Statistique descriptive : tableaux croisés

Les données avec lesquelles nous allons travailler concernent la répartition des universités françaises en ce qui concerne leurs moyens d'existence. Ces données sont stockées dans un fichier Microsoft Access nommé *TP4SPSS.mdb*. Pour importer ce fichier sous SPSS, on effectue les opérations suivantes :

1. *Fichier* → *Capturer une base de donnée* → *Nouvelle requête* → ouverture d'une boîte de dialogue
2. Dans la boîte de dialogue, choisir *MS Access Data Base* avant de cliquer sur *Suivant*
3. On appuie alors sur le bouton *Parcourir* de la nouvelle fenêtre et on recherche le fichier *ACCESS* à ouvrir *TP4SPSS.mdb*
4. SPSS va alors commencer l'importation de la base : dans la première boîte de dialogue, SPSS demande quelles tables l'utilisateur veut importer. On double-clique sur la (les) table(s) qui nous intéresse.
5. Sélectionner l'unique table présente : " Universités françaises ". A gauche apparaissent les champs (ou variables) que contient la table
6. On peut supprimer certains de ces champs (par exemple, le champ " Numéro ") ou modifier leur ordre puis on passe à la boîte de dialogue suivante.
7. L'étape suivante ne nous concerne pas, on appuie donc sur *Suivant*. SPSS ne peut enregistrer que des noms de variables restreints : il faut donc lui redonner des noms de variables convenables pour chaque champ sélectionné de la base de donnée. SPSS conservera le nom initial du champ dans la base de donnée comme description de la variable.
8. La fenêtre suivante ne nous concerne pas. On finit donc la manipulation en appuyant sur *TERMINER*. Une nouvelle base de donnée apparaît que l'on enregistre.

2.1 Création de classes

Le budget par étudiant est donné en francs convertissez le en euros en créant une nouvelle variable budget par étudiant en euros (*Transformer* → *Calculer*).

En fonction de la répartition des budgets par étudiants des différentes universités, on décide de créer 4 types d'universités : les universités ayant un très gros budget (plus de 1500 euros par étudiant), les universités ayant un gros budget (entre 1050 euros et 1500 euros par étudiant), les universités ayant un budget moyen (entre 750 euros et 1050 euros par étudiant) et les universités ayant un budget faible (moins de 750 euros par étudiant). Pour créer les classes on s'y prend de la façon suivante :

1. Créer une nouvelle variable « Richesse » qui prendra les valeurs 0 si l'université a un budget faible, 1 si elle a un budget moyen, 2 si elle a un gros budget et 3 si elle a un très gros budget (Entrer ces données dans *Valeurs* au niveau de la saisie des variables).
2. Utiliser la fonction *Transformer* → *Calculer*. Dans la boîte de dialogue qui s'ouvre, cliquer sur le bouton *SI* et entrez la condition « budget/etud < 5000 », cliquer sur *Poursuivre*. Entrer « Richesse » dans la variable de destination et la valeur 0 (correspondant à la condition choisie) ; faire OK.
3. Renouveler l'opération 3 fois pour les autres conditions

2.2 Inférence sur les classes

On veut ensuite afficher le tableau croisé donnant la répartition des universités selon leur type et leur richesse :

1. utiliser la fonction *Analyse* → *Statistique Descriptives* → *Tableaux croisés*. En ligne, entrer la richesse et en colonne, le type de l'université. Demander également le diagramme. Pour information, dans le menu *Cellule*, cocher aussi les options *Pourcentage lignes* et *Pourcentage colonnes*
2. Le diagramme peut être modifié de manière à être plus parlant : le sélectionner, appuyer sur *Options de diagramme* et choisir échelle 100%. On obtient un diagramme dit de fréquences de la variable « Type » (colonne) conditionnellement à la variable « Richesse » (ligne)
3. Enfin, supprimer les universités polyvalentes en double-cliquant sur sa légende et en entrant « Universités polyvalentes » dans la case *omettre*
4. Afficher le diagramme de fréquences de la variable « Richesse » en fonction de la variable « Type »
5. Commenter les résultats

Rq : On peut aussi obtenir, sur ce tableau croisé, le coefficient de corrélation (appelé ici, V de Cramer) ainsi que les effectifs théoriques d'indépendance qui ont permis de le calculer : dans le menu *Analyse* → *Statistiques descriptives* → *Tableaux croisés* choisir les statistiques « Phi et V de Cramer » et dans le menu *Cellules*, cocher les effectifs théoriques et les effectifs observés.

3 Applications

3.1 Guerre 14-18

Nous travaillons ici avec des données issues d'un fichier Excel. Importons tout d'abord ce fichier. . . Pour ce faire, on peut utiliser une procédure similaire à celle que nous avons utilisée pour importer un fichier Access. Cependant, il existe une méthode plus simple (bien que moins personnalisée) que nous allons utiliser. Faire *Fichier* → *Ouvrir* → *Données*. Dans « Type », choisir « Excel » et chercher le fichier désiré (ici, il est appelé TP4SPSS.xls) et faire *Ouvrir*. Une boîte de dialogue propose une plage de données (qui est la plage de la feuille de calcul comportant des valeurs) ainsi que de donner comme nom aux variables les valeurs la première ligne (option que nous choisissons).

Le fichier ouvert par SPSS contient les données concernant 30 individus passés en conseil de guerre en 1914. C'est donc un échantillon de l'ensemble des individus passés en conseil de guerre en 1914. On y lit la profession de l'individu, rangées en 4 catégories (i1 : Marchand / Cultivateur. . . , i2 : Journalier / Domestique. . . , i3 : Tailleur de pierre / Scieur de bois / Terrassier. . . i4 : Acrobate / Auteur. . .), le type d'insoumission pour lequel il était jugé, rangé en 4 catégories (j1 : Antimilitariste / antipatriotique. . . , j2 : Refus de partir, j3 : Alcoolique, j4 : Oubli / Négligence. . .) et enfin, la peine (en nombre de mois).

1. Enregistrer le fichier.
2. Etudier les liens entre la profession, le type d'insoumission et la peine.

Indications : On pourra répartir les peines en trois classes : 0 - pas de prison, 1 - peine légère, 2 - peine lourde (>12 mois).

3.2 Peine de mort

Voici un tableau concernant 4764 homicides jugés en Floride de 1973 à 1979. Ces données ont été publiées, pour la première fois, dans le New-York Times du 11 mars 1979 et maintes

fois utilisées par les statisticiens. La question posée par le journaliste était la suivante : Qui, des meurtriers noirs et blancs, est le plus souvent condamné à mort ?

Meurtrier	Victime	Peine de mort	Autre peine
Blanc	Blanc	72	2074
Blanc	Noir	0	111
Noir	Blanc	48	239
Noir	Noir	11	2209

1. Après avoir correctement identifié la population et les variables, entrer ces données sous SPSS (on utilisera, pour cela, la fonction *Pondérer les variables*).
2. Quelle réponse donneriez vous à la question des journalistes ? (justifier)