

ASI (L2) : TP6 Fichiers et graphiques sous *Excel*

Objectifs du TP :

Savoir lire des fichiers *.DBF* et importer des fichiers-texte; savoir faire des tris à plat et des tris croisés; enfin, savoir faire des graphiques de courbes et d'histogrammes avec *Excel*.

Rappel :

Pour chaque dossier traité (*ELF*, *VINS*, *PBIO*) ne pas oublier de faire lire le descriptif correspondant avec *Notepad* à savoir *ELF_DSC.TXT*, *VINS_DSC.TXT* et *PBIO_DSC.TXT* afin de savoir quelles sont les unités, quels sont les codes.

1. Importer des fichiers avec *Excel*

1.1 Importation de fichiers *Dbase*

Comme indiqué en cours, une variable statistique correspond à une colonne de valeurs. Pour garantir l'homogénéité de typage de la colonne (toutes les données doivent avoir le même type), le format standard proposé par *Excel* avec les fichiers *.XLS* n'est pas suffisant. Il vaut mieux utiliser le format non propriétaire *.DBF* défini dans les années 80 par le logiciel *Dbase*.

Pour ouvrir un fichier *.DBF* il suffit de faire **Fichier / Ouvrir** et d'utiliser les boîtes de dialogue affichées pour sélectionner le fichier. Pour enregistrer au format *Dbase* il suffit de faire **Enregistrer Sous** et de sélectionner *Dbase3* mais attention : cela ne sauvegarde que l'onglet courant, les colonnes doivent avoir un nom d'au plus huit lettres majuscules non accentuées (un seul mot), les données doivent être contigues et de même type (ne faire aucune manipulation de sauvegarde).

On pourra lire les fichiers ELF.DBF et VINS.DBF (après les avoir recopiés en local sur le disque D:). Faire tout de suite après et avant toute autre manipulation **Format / Mise en forme automatique**.

1.2 Importation de fichiers-texte

Les colonnes d'un fichier-texte peuvent être délimitées par un caractère ou être formatées ("cadrées") à l'aide de caractères espace : *Excel* parle alors de fichier de données "en largeur fixe". Leur lecture se fait à l'aide du menu **Fichier / Ouvrir**. L'assistant d'import est censé être "évident". Faire les manipulations et commenter au passage les options.

Faire un essai avec le fichier ELF.TXT.

2. Comptages, tris à plat et tris croisés

On fera toutes les manipulations avec le fichier ELF.DBF.

Montrer comment on utilise le menu

Données / Rapport de Tableau Croisé Dynamique.

Commenter les étapes de l'assistant :

- Etape 1 : Liste ou base de données

(car on utilise un fichier .DBF)

- Etape 2 : Base_de_données

(car la base est automatiquement définie quand on utilise un fichier .DBF)

- Etape 3 : Localisation = "Feuille existante" (pour l'instant)

Détailler la gestion de la disposition : indiquer ce qu'est **LIGNE**, **COLONNE**, **DONNEES**. Effectuer le tri à plat du sexe en ligne avec les effectifs absolus : mettre **NUM** comme **Données** et **SEXE** en **Ligne**.

Recommencer avec **SEXE** en **Colonne**.

Effectuer les calculs suivants :

- le tri à plat de **SEXE** en ligne avec les effectifs relatifs,
(cliquer deux fois sur **Nombre de NUM** apparu en données, commenter la rubrique "*Synthèse par*", cliquer sur **Options** et choisir dans la liste déroulante nommée *Afficher les données* la rubrique **% du total**).
- le tri à plat des niveaux d'étude,
- le tri croisé sexe / niveaux d'étude en effectifs absolus,
- le tri croisé sexe / niveaux d'étude en effectifs relatifs (% ligne),
- le tri croisé sexe / niveaux d'étude en effectifs relatifs (% colonne),
- le tri croisé sexe / niveaux d'étude en effectifs relatifs (% total).
- le comptage de l'âge avec les effectifs relatifs,
- le calcul dit *multi-critères* de la moyenne de l'âge par sexe.

3. Graphiques de courbes sous *Excel*

On fera toutes les manipulations avec le fichier **VINS.DBF**.

Montrer comment tracer un nuage de points pour tracer la colonne **UK** (une seule série) : on sélectionne la colonne, on clique sur le bouton de l'assistant graphique (on peut aussi utiliser **Insertion / Graphique**).

Descendre sur **Nuages de points**, cliquer de suite sur **Terminer**.

Déplacer le graphique, redimensionner le graphique avec les poignées sur les bords, montrer ce qu'affiche un clic-droit sur le graphique et en particulier commenter ce qu'on peut faire avec **Emplacement**, **Options du graphique** et **Type de graphique**.

Faire un deuxième graphique de **UK** avec des points reliés par des segments de droite.

Montrer comment tracer un nuage de points pour tracer la colonne **UK** en fonction de la colonne **RFA** (deux séries) : sélectionner **RFA** puis à l'aide de la touche **CTRL** sélectionner **RFA** de façon à avoir deux colonnes sélectionnées non contigues. Faire un graphique en nuage de points avec points et lignes.

Comme ce n'est pas très "beau", insérer un onglet, y recopier les colonnes **RFA** et **UK** (dans cet ordre), utiliser le menu **Données / Trier** de façon à trier les deux colonnes ensemble par **RFA** croissant. Montrer que le graphique est alors plus "correct". Indiquer qu'il y a croissance.

Faire les graphiques suivants dans ce nouvel onglet :

- avec des point seuls,
- avec des points reliés par des segments de droite,
- avec des points seuls et la droite de régression linéaire,
- avec des points seuls et la droite de régression linéaire et avec l'équation et le R^2 .

Pour la droite de régression, l'équation et le R^2 , il faut mettre la souris sur un point de la courbe, faire un clic-droit et valider **Ajouter une courbe de tendance**. Choisir le type "Linéaire" puis cliquer sur l'onglet **Options** avant de cocher "équation" et "coefficient de détermination".

Insister sur le problème du choix des axes : "qui est X ? ", " qui est Y ?" en fonction d'une causalité éventuelle. Par exemple pour tracer "C=consommation d'essence" et "D=distance parcourue", le seul tracé "raisonnable" est $C=f(D)$.

Montrer comment calculer ρ , a et b sous *Excel* :

$$\begin{aligned}\rho &= \text{COEFFICIENT.CORRELATION}(x;y), \\ a &= \text{INDEX(DROITEREG}(y;x;"VRAI");2), \\ b &= \text{INDEX(DROITEREG}(y;x;"VRAI");1).\end{aligned}$$

4. Histogrammes de fréquences avec *Excel*

On fera toutes les manipulations avec le fichier `ELF.DBF`.

Montrer comment tracer un histogramme des effectifs à partir des tris à plat pour tracer les fréquences des modalités de la variable **SEXE**. Faire écrire le texte **Hommes** et le texte **Femmes** de façon à pouvoir sélectionner

Hommes	Femmes
35	64

Faire remarquer que le tracé des histogrammes est le même pour les effectifs absolus et pour les effectifs relatifs (mais que les valeurs sur les échelles changent). La conclusion est qu'il vaut mieux utiliser les effectifs relatifs pour pouvoir comparer les graphiques avec les pourcentages.

Tracer aussi l'histogramme des niveaux d'étude avec les labels en toutes lettres, à savoir

NR	Prim	Bepc	Bac	Sup
3	6	30	21	39

Faire interpréter avec une phrase chacun des histogrammes, comme par exemple : (pour le sexe) *il y a en gros deux fois plus de femmes que d'hommes, ce qui est compréhensible vu le sujet de l'étude* ; (pour l'âge) *il y a beaucoup de gens avec un niveau supérieur puis avec le niveau bepc.*

Faire tracer les histogrammes à partir des deux tris croisés **Sexe / Etude** puis **Etude / Sexe** sans se préoccuper des labels.

Faire remarquer que seul l'histogramme **Sexe / Etude** montre qu'il n'y a pas le même profil pour les hommes que pour les femmes alors que l'histogramme **Etude / Sexe** ne fait que reproduire la disproportion 1/3 - 2/3 entre les hommes et les femmes.

5. Tracés complémentaires (s'il reste du temps)

5.1 Histogrammes probabilistes

Faire copier la feuille BINOMIALE.XLS de K:\Stat_ad.

Leur faire tracer l'histogramme de p_i en fonction des x_i .

Montrer que le graphique est dynamique : si on change p alors le graphique change.

Faire interpréter la position du mode en fonction de p (graphique symétrique pour $p = 0.5$; mode à gauche pour $p < 0.5$; mode à droite pour $p > 0.5$).

Peut-on mettre une courbe d'interpolation linéaire, polynomiale, exponentielle ?

En particulier, pour la régression linéaire, quelle est la valeur de ρ^2 ? Est-ce une bonne droite de régression ?

5.2 Tracés d'histogrammes multiples

Prendre le dossier PBI0. Combien faudrait-il faire de tris à plat ? de tris croisés ? Lesquels choisir ?

Essayer de donner à chaque fois qu'on fait le tri à plat d'une variable le tableau récapitulatif avec l'affichage "intelligent" vu en cours (une QL par ligne, les plus fortes modalités de la QL par ordre décroissant), les lignes triées par mode.

5.3 Tracés de courbes multiples

Reprendre le dossier VINS. Combien de courbes y aurait-il à faire ? lesquelles retenir ?

5.4 Tracés techniques

Peut-on réaliser des tracés en "tige et feuille" (*stem and leaf* en anglais) avec *Excel* ? Essayer avec la variable AGE du dossier ELF.

Peut-on réaliser des tracés en "boîtes à moustaches" (*boxplot* ou *box and whisker diagram* en anglais) avec *Excel* ? Essayer avec toutes les variables du dossier VINS.