

Logiciel EPIDATA en 10 pages

Niveau 1 avec Entry et Analysis 2.2

Bernard BRANGER – Réseau « Sécurité Naissance – Naître ensemble » des Pays de la Loire
1, allée Baco – 44000 NANTES. Tél 02 40 48 55 81 –
Courriel : bernard.branger@naître-ensemble-ploire.org
Juillet 2010

- A quoi sert EPIDATA ?

Saisir des questionnaires (EPIDATA ENTRY), et exploiter les données (EPIDATA ANALYSIS).

- Est-ce mieux que d'autres logiciels ?

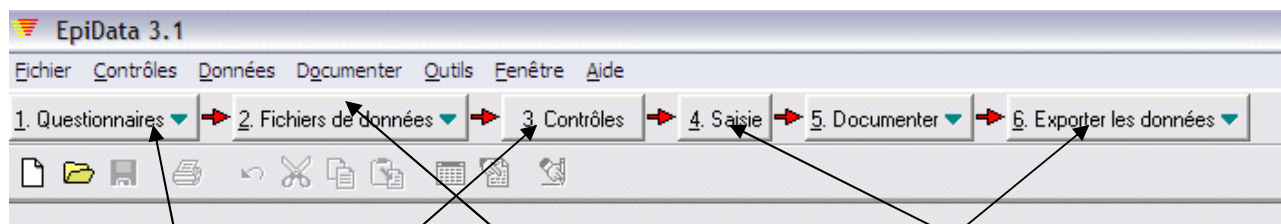
EPIDATA va remplacer EPI-INFO 6.0 qui reste sous DOS et n'évolue plus. EPIDATA semble plus facile à utiliser que EPI-INFO Windows, logiciel lourd qui génère des fichiers *.mdb. **EPIDATA est libre de droit** et peut-être utilisé gratuitement. La méthode présentée ici correspond à un niveau 1 de prise en main. Deux autres guides spécifiques pour l'analyse et le contrôle à la saisie sont disponibles.

- Télécharger et installer EPIDATA sous Windows XP ou Windows 7

- Télécharger : par Internet, ouvrir <http://www.epidata.dk>. Cliquer sur download puis effectuer deux manœuvres :
 - dans la partie *Epidata Analysis* 2.1 (en anglais), cliquez sur [Setup\(exe\) \(1.9 Mb\)](#). Le fichier à télécharger est *setupepidatastat.exe*. Cliquer sur Enregistrer et le sauvegarder (sous C : , ou sous « Mes documents », ou dans « Téléchargements », ou sur le bureau, par exemple).
 - dans la partie *Epidata Entry Français*, cliquez [FR_setup](#) (1.1 Mo). Le fichier à télécharger est *setup_epidata_fr.exe*. Cliquer sur Enregistrer (idem précédemment).
- Installer dans votre ordinateur : se mettre à l'emplacement du téléchargement, cliquer d'abord sur *setup_epidata_fr.exe* pour décompresser et installer *Epidata Entry*. Accepter « noms des champs : automatique » +++. Puis cliquer ensuite sur *setupepidatastat.exe* pour le décompresser et installer *Epidata Analysis*.

Accepter les chemins proposés C:\Program Files\Epidata sous Windows XP ou C:\Programme(x86)\Epidata sous Win 7. C'est plus commode, sinon recommencer avec un choix d'installation « Custom » pour choisir vos propres chemins d'installation. Accepter les icônes sur le Bureau.

Attention ! Dans Vista et Windows 7, la sauvegarde des fichiers *.rec, *.qes, *.chk, *.htm et *.pgm se trouve quelquefois dans la fenêtre « Fichier de compatibilité ». Pour les voir : cliquer sur *Ordinateur, C : , Programmes (ou Programmes Files, Epidata, Samples* : en haut d'une colonne cliquer sur « Fichier de compatibilité » pour voir les fichiers (et les copier par exemple).

EPIDATA Entry (pour la saisie) 3.1 → Premier écran : 5 à 6 boutons principaux sont à noter.

- Faire un questionnaire
- Créer un fichier de données
- Faire des contrôles à la saisie (calculs automatiques, bornes...)
- Saisir les données
- Exporter des données dans d'autres formats, si nécessaire (dBase, Excel, SPSS).

Faire un questionnaire

On distingue en épidémiologie :

- *les variables qualitatives* à coder avec des lettres ou des chiffres ou en type oui/non
- *les variables quantitatives* qui peuvent prendre toutes les valeurs possibles avec ou sans décimales. Attention, le système par défaut est décimal ; tous les autres systèmes tels que la semaine et les jours (heptagésimal), l'année et le mois (duodécimal), les heures et les minutes (sexagésimal) doivent être convertis en décimales. Exemple : 28 SA + 3 jours = 28.42 semaines ; ou 1 h 30 = 1.5 heures ; ou 2 ans 9 mois = 2.75 années.
- *les variables de dates* généralement avec les jours, le mois et l'année
- *les variables de texte* en clair

Une variable a deux composantes :

- **un nom** de 3 à 10 lettres sans espace, sans accent +++, plutôt en EN MAJUSCULE (pour éviter les accents), ne commençant pas par un chiffre, sans caractères particuliers. Voir détail en annexe.
- **un champ** qui en définit des caractéristiques :
 - # ou ## ou ##.## pour les variables contenant des chiffres (ayant valeur qualitative codée ou quantitative)
 - <Y> pour les variables de type oui/non (Yes/No)
 - _ ou __ ou ___ (trait sous le 8 du clavier) pour les qualitatives à 1,2 ou 3 caractères avec des lettres
 - <dd/mm/yyyy> pour les variables de date (d = day = jour, m = month = mois, y = year = année)
 - _____ pour du texte (réponses en clair).

Exemple à faire

```

<IDNUM >                FICHE #####
HOPITAL ____ SERVICE __ UNITE __ SEXE #
DATENAISS <dd/mm/yyyy>  ADMIUNIT <dd/mm/yyyy>          SORTIUNIT <dd/mm/yyyy>
AGE ###.#              ENDOSCOP <Y>          AMBULA #
DUREE ###             ASA #                  COMMENT _____
  
```

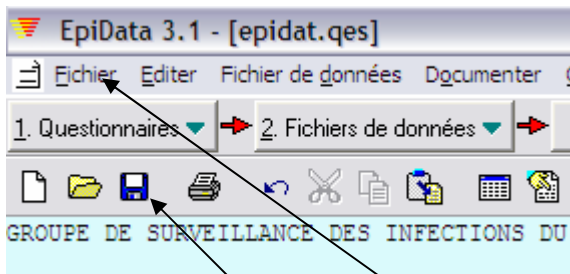
- **<IDNUM>** : nombre qui s'incrémentera à chaque enregistrement (à partir de 1 pour la première fiche)
- **SEXE #** : 1 pour Homme et 2 pour Femme, chiffres mais pas pour en faire une moyenne !
- **DATENAISS <dd/mm/yyyy>** : jour, mois, année

- **AGE ###.#** : âge calculée par différence entre date de naissance et date d'admission avec 3 chiffres et une décimale. L'intervalle de décimale est toujours un point.
- **ENDOSCOPIE** : Y pour Yes, N pour No ou rien (champ vide) ➔ à éviter +++, préférer 1 pour « Oui » et 2 pour « Non ».
- **AMBULA #** : codage avec un chiffre avec 1 pour oui et 2 pour non, 9 pour ne sait pas (éviter le zéro)
- **ASA #** : chiffre de 1 à 5 (on pourra en faire une moyenne)
- **COMMENT** : peut recueillir du texte libre (mettre en majuscule pour éviter les accents)

NB : toujours une espace entre la dernière lettre de la variable et le signe du champ +++

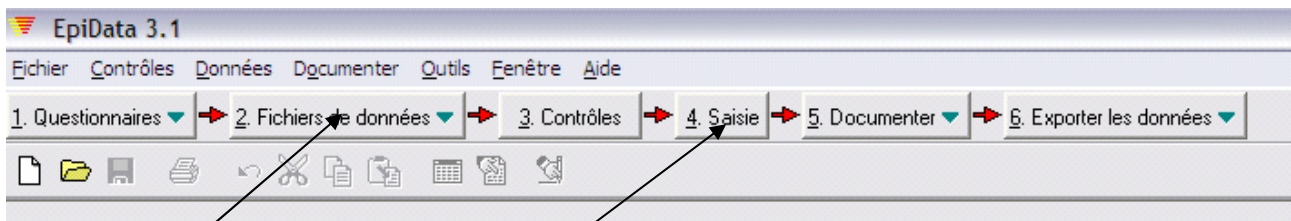
Il est fortement conseillé pour la version de EPIDATA actuelle (3.1) de mettre le maximum de variables codées avec des # et donc de coder en chiffres (et non pas avec des lettres ni avec Yes et No), même pour des variables qualitatives +++

Sauver le questionnaire



Cliquer sur l'icone Sauver ou Fichier – Enregistrer sous. Donner un nom évocateur : il sera en suffixe de type *.QES. Noter le répertoire (ou chemin).

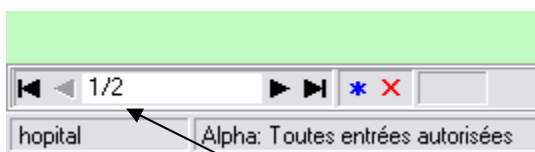
Créer le fichier de données



Cliquer sur **2. Fichier de données**. Donner le même nom que les données évocateur : il sera en suffixe de type *.REC. Noter le répertoire (ou chemin). Cette partie est à faire à la création du fichier, mais pas pour les saisies ultérieures.

Saisir les premières fiches

- Cliquer sur **4. Saisie**
- Sélectionner le fichier de type *.REC (connaître son chemin..) : la première fiche s'affiche, sinon la première non saisie en cas de saisie précédente.

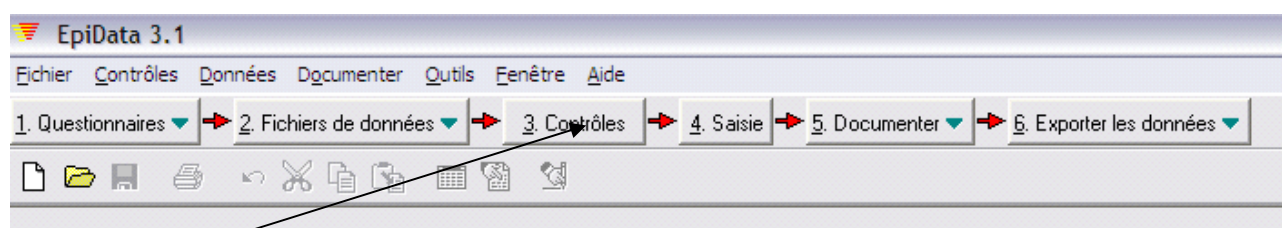


- Vous êtes ici sur la fiche 1 dont le numéro est noté en bas à gauche

- Remplissez au fur et à mesure : les dates peuvent être entrées en 01022007 ou même 0102 qui deviendra automatiquement 01/02//2007. Pour les nombres avec décimales, ne pas mettre le point. Les champs remplis font passer le curseur directement sur le champ suivant, sinon faire ↵¹.
- Faire ↵ sur les champs calculés (voir infra).
- On peut revenir sur les champs, et sur les fiches. Un champ Yes/No ne supporte que les lettres Y et N.
- Pour chercher un champ donné : se mettre sur le champ et faire *ctrl-F* ou cliquer sur *Fenêtre* et *Aller à*. Une fenêtre d'affiche : remplir le champ de la valeur cherchée et faire ↵ → le premier enregistrement correspondant s'affiche. Faire F3 pour chercher l'enregistrement suivant.

Etablir des contrôles de saisie

Les contrôles de saisie permettent de fixer des codages de remplissage (surtout d'il y a plusieurs personnes qui saisissent) ou faire des calculs automatiques ou pour organiser la saisie. Ne pas multiplier les contrôles et ne pas perdre de temps à « peaufiner ». Réserver cette partie pour les calculs automatiques d'âge, de durée de séjour ou d'âge gestationnel. Exemple : l'âge peut être calculé à la saisie par différence entre la date d'admission et la date de naissance.



Ouvrir **3. Contrôles**. Ouvrir le fichier *.REC à contrôler. Par référence au fichier réalisé page 2, la variable AGE à calculer nécessiterait les opérations suivantes : cliquer sur AGE, une fenêtre s'ouvre, cliquer sur Editer, écrire ou compléter alors pour avoir **l'âge en jours** :

```
AGE
AGE = admiunit - datenaiss
END
```

Sinon, pour **l'âge en années** (avec, en moyenne, 365.25 jours par an) :

```
AGE
AGE = (admiunit - datenaiss)/365.25
END
```

Sauver et passer à une autre variable. On peut déterminer des bornes (RANGE) comme 15-45, ou des valeurs imposées (LEGAL) (1, 2 pour le sexe). Les contrôles sont sauveés dans un fichier de type *.chk.

Pour faire un changement de page, pour les questionnaires longs et pour ne pas se perdre dans les dédales des variables, mettre dans la variable du haut de page (par exemple SEXE)

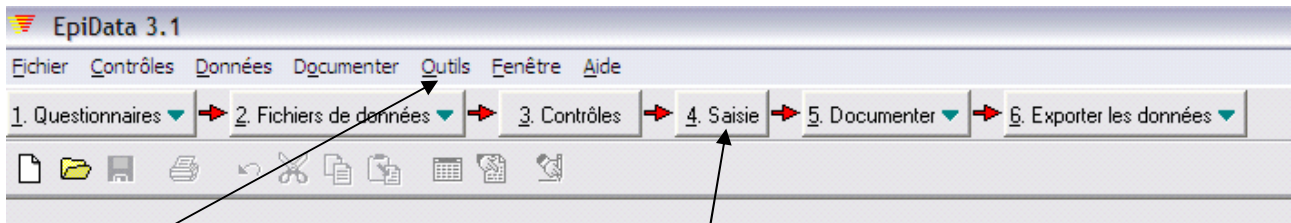
```
SEXE
topofscreen 2
end
```

→ **Pour consulter la liste des codes à la saisie, faire F9 sur le champ de la variable et choisir soit avec la souris, soit avec les flèches ↑↓ et faire ↵.**

¹ ↵ est la touche "Entrée" du clavier, grosse touche située à droite du clavier alphabétique ou du clavier numérique.

Revenir sur le questionnaire

Il est fréquent d'être obligé de revenir sur le premier questionnaire déjà fait : revenir sur 1. Questionnaires, le changer (en supprimant des variables, en en ajoutant ou en modifiant), et sauver.



Cliquer sur *Outils* puis sur *Modifier .REC avec .QES modifié* → le nouveau fichier REC sera modifié dans sa structure (mais pas dans le contenu des variables). Ouvrir *Saisie* de nouveau, mais attention, les fiches déjà remplies doivent être modifiées avec les nouvelles modalités.

Penser à sauver vos données régulièrement

Les données "stratégiques" sont sauveés dans un fichier *.rec. Sauvez régulièrement sur une clé USB ou un autre ordinateur (*copier et coller*). Les fichiers *.qes et *.chk doivent être sauveés une fois pour toutes s'ils ne sont pas modifiés.

Pour charger des fichiers déjà réalisés sur un autre support

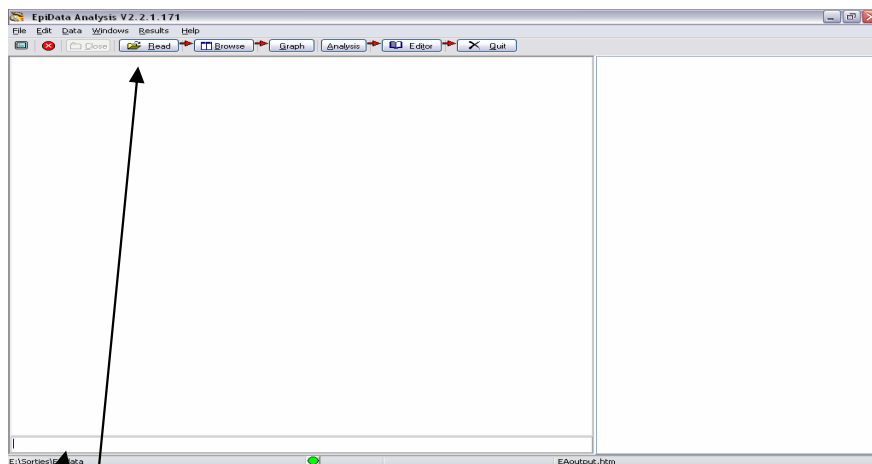
Ouvrir la clé USB ou la disquette. Sélectionner les fichiers en *.qes, en *.rec ou en *.chk. Faire «*copier*». Puis se mettre dans *C:\Program files\epidata\samples* et faire «*coller*». Le chemin désormais pour ouvrir et sauver les données est donc *C:\Program files\epidata\samples\l fichier.rec*

- EPIDATA : exploitation de données avec Analysis 2.2



EpiData Analysis.Ink

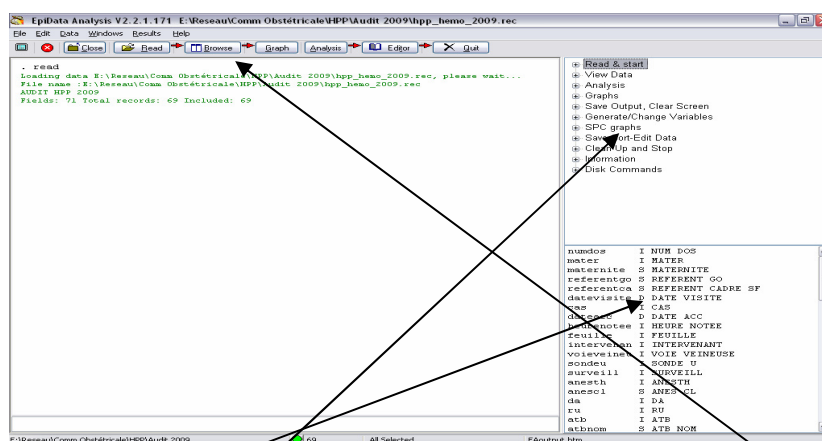
Un document de 17 pages est plus détaillé → s'y référer, sinon ici c'est plus rapide... Cliquer sur l'icone sur le bureau pour ouvrir le fichier.



Ouvrir un fichier

cliquer sur le bouton **Read Data**: choisir un chemin à la recherche de fichiers *.REC. comme *C:\Program Files\Epidata\Samples* ou *C:\Epi6fr* si des fichiers *.rec ont déjà été créés dans Epi 6.0.

A signaler la ligne de commande pour écrire ou saisir en cliquant.



- **Fenêtre des commandes** : faire F2
- **Fenêtre des variables** : faire F3 → variable de type *S* = *String* = codage qualitatif ou de type *I* = *Integer* = quantitatif.
- Cliquer sur l'icone Tableur ou écrire *Browse* dans la ligne de commande **ou F6 ou cliquer sur Browse Data** pour voir le fichier comme un tableur. Revenir en cliquant sur Close. Voir la liste des commandes en cliquant sur **Help**.

Les commandes de base dans la fenêtre des commandes

Les conseils donnés ici concernent l'écriture dans la fenêtre des commandes. Pour ceux qui aiment cliquer, on obtient les mêmes résultats en cliquant sur Analysis, choisir *Frequency* ou *Means* ou *Tables*, puis cliquer sur ce que l'on désire. De toute manière, le logiciel écrit sur les résultats les lignes de commande correspondant. Penser quand même qu'en appuyant sur la touche ↑ une ou plusieurs fois on retrouve les commandes précédentes.

NB : *varqual* est une variable qualitative (sexe par exemple), *varquant* est une variable quantitative (age par exemple).

→ Pour décrire

- **freq varqual** pour la fréquence des données codées (qualitatives) ; **freq *** trie toutes les variables qualitatives. Pour ajouter les pourcentages en colonne, faire (après un espace) **/c**, et pour avoir les pourcentages **cumulés** **/cum**, et pour avoir aussi les manquants **/m**
- **means varquant** pour les données quantitatives seules : donne, entre autres, le nombre d'observations, la moyenne (mean), la médiane, le minimum, le maximum, l'écart-type (Dev Stand)
- **describe** pour décrire toutes les variables quantitatives. Voir F1 pour d'autres modalités.

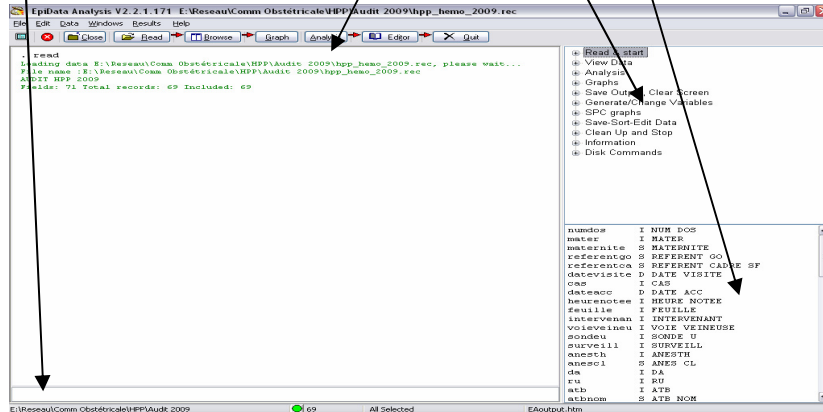
→ Pour comparer

- **means varquant varqual**. A signaler que EPIDATA demande que la *varqual* soit de type numérique (c'est-à-dire # dans Epidata Entry !) → c'est pour cela que les variables doivent être plutôt de type #. **means varquant varqual /t** pour avoir le test statistique (test de F ou test de t pour deux).
- **tables varqual1 varqual2** pour croiser deux données qualitatives → pour les pourcentages en lignes **/r** (pour row, et suite à un espace), ou pour les pourcentages en colonnes **/c** (pour

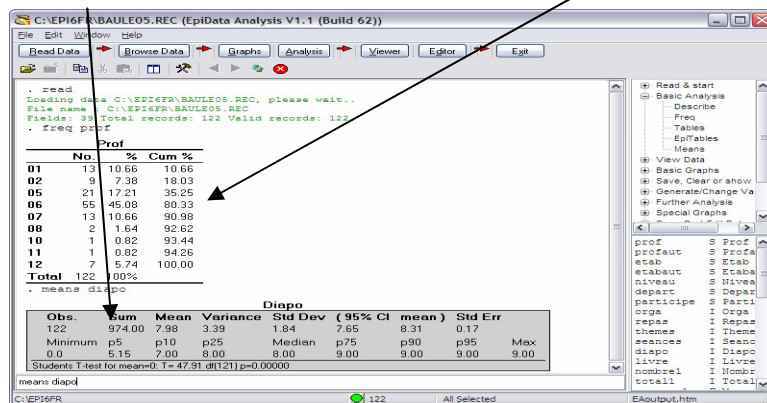
column), pour les tests statistiques /t (c'est le χ^2), pour le test de Fisher avec tableau 2x2 /ex, /rx pour le risque relatif, /o pour l'odds ratio (attention à l'ordre des codages).

Procédure :

- écrire les commandes ou les cliquer avec le bouton **Analysis**
- Si vous préférez écrire : écrire le nom de la variable/ou cliquer sur une variable de la liste des variables, et faire ↵ après les deux termes écrits ou cliqués

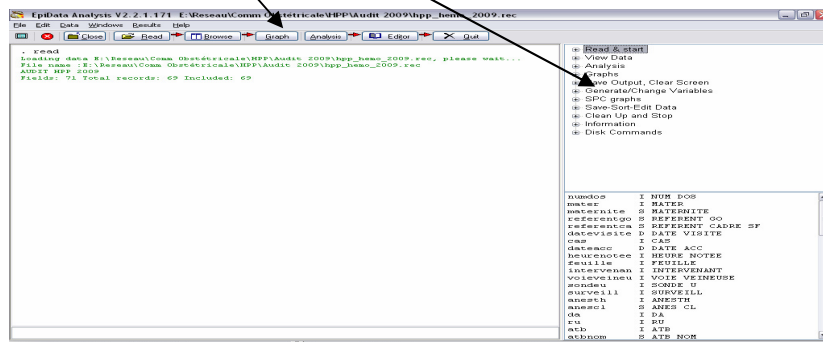


- lire les résultats dans la fenêtre des résultats après par exemple freq varqual /c /cum (ici appelée prof de 01 à 12) et means varquant (ici appelée diapo, note de 0 à 9)



Les graphes

Cinq types de graphes sont utilisables : barres, histogramme, boxplot, scatter (diagramme de points) et lignes. Exemple de sortie avec bar repas. Procédure : écrire dans la fenêtre de commande bar varqual (ici « repas ») ou cliquer sur **Graphs** ou sur le bouton **Graphs**



Les graphes n'ont qu'une valeur indicative : il faudra les reprendre, en les simplifiant, dans un document de résultats (sous Word par exemple avec le module Microsoft Graph plutôt que sous Excel). Voir aussi les boxplots par exemple. Voir document Analysis / Niveau 2.

Pour imprimer les sorties

Pour imprimer les sorties déjà effectuées : cliquer sur **File** et *Print* (cocher *Tout*)

Pour sauver les sorties, les visualiser et les imprimer

Les sorties sont sauvées automatiquement à la fermeture dans un fichier *EAoutput.htm* situé dans le répertoire par défaut. Si on veut donner un nom plus explicite, avant d'exploiter, cliquer sur *Save*, *Clear*, *Show output* ou écrire dans la fenêtre des commandes `Logopen " fichier.htm"`. Un fichier de nom *fichier.htm* sera sauvé dans un répertoire par défaut. Fermer les sorties avec `Logclose`. On peut rouvrir ultérieurement le *fichier.htm* qui s'ouvrira dans Internet Explorer (ou équivalent), ou clic droit, ouvrir avec... Word par exemple.

Pour visualiser les données comme dans un tableur

- Sans modifier les données : `browse var1 var2` ou cliquer sur le bouton **Browse**. Les données peuvent être ordonnées, et copiées (clic droit sur une colonne)

- Pour modifier quelques données sans passer par la saisie : la commande `update` ou `update var1` permet de modifier des valeurs. Les données sont modifiées dans le fichier d'origine si on fait `savedata` à la fin.

Changer, créer des variables

→ Il est possible de regrouper des codages ensemble (variables qualitatives) ou de déterminer un seuil dans les variables qualitatives. Exemple à écrire sur la ligne de commande pour transformer une variable qualitative codée en 1, 2 et 3 (`var1`) en une autre variable (ici quantitative) (`varquant`) en deux classes 1 et 2 (le codage 1 de `var1` devenant 1 dans `varquant`, et 2 et 3 devenant 2 :

```
define varquant # ↵
if (var1 = 1) then varquant=1 else varquant=2 ↵
```

→ Autre exemple avec l'âge et fixer un seuil à 30 ans ; le nom de la variable créée est libre :

```
define old30 _ ↵      if age<30 then old30=1 ↵      if age>= 30 then old30=2 ↵
```

Ceci aboutit au même résultat que : `if age<30 then old30=1 else old30=2` ↵

→ Attention : `age<30` → les âges non codés (manquants) sont inclus. Si doute, écrire :

```
if (age<30) and (age>0) then old30=1 ↵
```

→ Pour exclure un codage, utiliser les deux caractères "<" et ">" : tous les sexes codés « pas 1 » seront 2

```
if sexe <> 1 then sexe =2 ↵
```

→ Pour sélectionner des observations : `select age>30` ↵ OU `select sexe=1` ↵.

Pour enlever la sélection, `select` ↵ (y penser +++)

→ Les codages sont bruts pour les données quantitatives, avec des " " pour les variables qualitatives codées avec le `_`. Exemple : `define sx # ↵ if sexe="M" then sx=1 ↵`

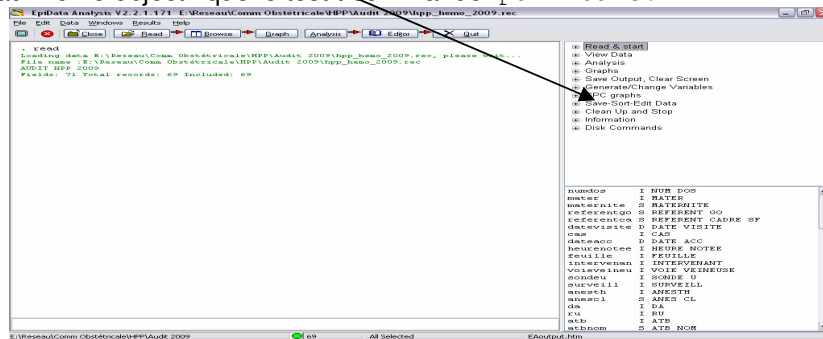
→ Essayer la commande **recode** pour faire des classes à partir de variable quantitative. Exemple : tous les 10 (ans pour l'âge par exemple : `define classe # ↵ recode varquant to classe by 10 ↵ freq classe`

→ Les commandes `freq` et `tab` écrivent les codes tels qu'ils ont été rentrés dans EPIDATA ENTRY : soit avec des « 1 » et des « 2 » par exemple, soit avec « Homme » et « Femme » par exemple si avec la commande F9 un comment legal avait défini les « 1 » et les « 2 ». Si on veut attribuer des codes à un libellé plus explicite, cela s'appelle un « label de variable ». La commande est la suivante : `labelvalue sexeU/1="Homme"U/2="Femme"` (U étant un espace).

Autres analyses

Voir à *Further analysis* dans la boîte des commandes

- `regress varquant1 varquant2` : Régression linéaire de Y en X
- `correlate varquant1 varquant2` : Coefficient de corrélation de Pearson
- `kwallis varquant varqual` : Test de Kruskal-Wallis pour comparer des moyennes entre elles (test non paramétrique au même objectif que le test t commandé par `means`).



Sauver un programme d'analyse – Le lancer ensuite

- `Savepgm "fichier.pgm"` : sauve les commandes réalisées dans un *fichier.pgm*
- `Run fichier.pgm` : lance l'analyse automatiquement.

Pour faire un programme : leçons niveaux 2. **Exit** pour quitter

Annexe

Importer les données d'un fichier Excel

→ Deux solutions :

- enregistrer, dans Excel 2003 (mais n'existe pas dans Excel 2007 ni 2010), le fichier Excel au format *.dbf (dBase III par exemple) : faire *Fichier/Enregistrer sous/Type de fichier* : *.dbf, puis dans EPIDATA Entry : "Données", "Importer" : dBase : la conversion est simple avec le même nom en *.rec. Ouvrir dans EPIDATA Analysis.

- enregistrer, dans Excel 2007 ou 2010, au format *.csv (avec un « ; ») ou au format *.txt (avec « tab » comme séparateur). Dans EPIDATA Analysis, ouvrir le fichier en format « csv » ou « txt » dans EPIDATA (en prenant "all" comme type de fichiers) : le fichier est lu directement. Faire « savedata » pour créer un fichier équivalent en « *.rec »

→ Attention : les noms des variables doivent être simples : pas plus de 10 lettres, sans espace, sans accent, sans signes particuliers... La première ligne des données (la deuxième ligne sous Excel) doit être "propre" : tout rempli, en particulier si on veut garder la forme "numérique" des nombres ou la forme "Date"; sinon, par défaut ce sera une variable "Texte". Il ne doit y avoir que des nombres dans les colonnes pour les nombres (pour faire des moyennes...).

Epidata Saisie : noms des variables (noms de champs)

→ Le type des noms des champs peut être modifié a priori : *Fichier/Options/Créer le fichier de données/* : le faire avant de démarrer le questionnaire (et peut être changé après).

Dans tous les cas, les voyelles avec accent (é,à,è,ë) et les caractères comme « % » ou « -« ou « / » sont éliminés des noms de champs. Ne pas commencer par un chiffre (Epidata placerait alors la lettre « N » devant le chiffre).

Option 1 : *Fichier/Options/Créer le fichier de données/* **Nom de champ = 1er mot**

En cas de noms de champs identiques (comme « date »), Epidata placera en fin de champ un chiffre pour les distinguer comme date, puis date1, puis date2.... Les espaces sont supprimés.

- Questionnaire

```
Date naissance <dd/mm/yyyy>           Date enquête <dd/mm/yyyy>
Datentrée <dd/mm/yyyy>                 Déclenchement #
```

- Champs dans le fichier de données (*cliquer dans Documenter/Structure du fichier*)

Num.	Nom	Libellé	Type	Taille	Contrôles
1	date	naissance	Date (dmy)	10	
2	date1	enquête	Date (dmy)	10	
3	datentre		Date (dmy)	10	
4	dclenchem	nt	Numérique	1	

- Dans Analysis

Name	Type	Length	Decimals	Label
DATE	Date	10	0	date naissance
DATE1	Date	10	0	date1 enquête
DATENTRE	Date	10	0	datentre
DCLENCHEM	Integer	1	0	dclenchement

Le nom « DATE » plante EPIDATA, car c'est une commande : à éviter donc +++

Option 2 : *Fichier/Options/Créer le fichier de données/* **Nom de champ = automatique (10 premières lettres)**

Option à privilégier car les noms de champs sont mieux repérés. Ne pas forcément donner un nom de 10 lettres (une ou deux peuvent suffire comme « sx » pour le sexe). Les espaces sont supprimés.

- Questionnaire

```
Date naissance <dd/mm/yyyy>           Date enquête <dd/mm/yyyy>
Datentrée <dd/mm/yyyy>
Déclenchement #
```

- Champs dans le fichier de données (*cliquer dans Documenter/Structure du fichier*)

Num.	Nom	Libellé	Type	Taille	Contrôles
1	datenaissa	Date naissance	Date (dmy)	10	
2	dateenqute	Date enquête	Date (dmy)	10	
3	datentre	Datentrée	Date (dmy)	10	
4	dclenchem	Déclenchement	Numérique	1	

- Dans Analysis

Name	Type	Length	Decimals	Label	Value label	Missing
DATENAISSA	Date	10	0	date naissance		
DATEENQUTE	Date	10	0	date enquête		
DATENTRE	Date	10	0	datentrée		
DCLENCHEM	Integer	1	0	déclenchement		