

Logiciel EPIDATA en 10 pages

Niveau 1 avec Entry et Analysis 2.2 – Version 2016

Bernard BRANGER – Association Epi-Σûre
11 bis, rue Gabriel Luneau – 44000 NANTES
Tél 06 32 70 33 80 – Mail branger44@gmail.com

- A quoi sert EPIDATA ?

Saisir des questionnaires (EPIDATA ENTRY), et exploiter les données (EPIDATA ANALYSIS) sous Windows. La saisie sous Mac ou Linux est possible mais pas encore l'analyse. On peut importer ou exporter des fichiers sous d'autres formats (Excel, dBase...).

- Est-ce mieux que d'autres logiciels ?

EPIDATA est un logiciel tout-en-un qui permet la saisie et l'analyse de données. Pour la saisie, c'est une alternative majeure à la saisie sous EXCEL. Pour l'analyse, il permet de faire des analyses univariées, mais ne prétend pas remplacer les poids lourds comme SPSS, SAS, STATA ou « R ».

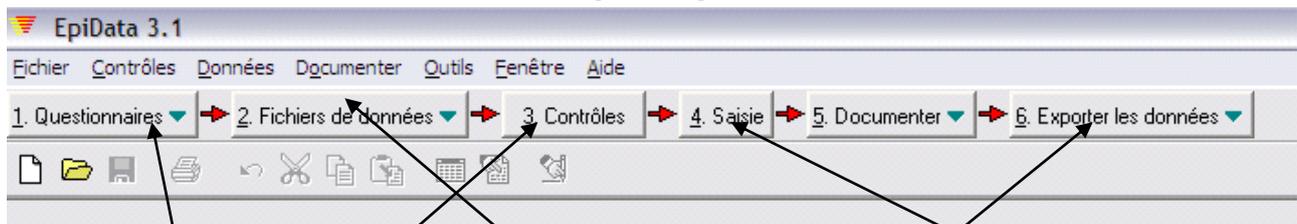
EPIDATA est libre de droit et peut-être utilisé gratuitement. La méthode présentée ici correspond à un niveau 1 de prise en main. Deux autres guides spécifiques pour l'analyse et le contrôle à la saisie sont disponibles.

- Télécharger EPIDATA sous Windows XP ou Vista ou Windows 7 ou 8 ou 10

- Créer un dossier "EPIDATA" sur le bureau de Windows
- Télécharger : par Internet, ouvrir <http://www.epidata.dk>. Cliquer sur `download` puis effectuer deux manoeuvres (**ne pas utiliser Manager 1.4.x, ni Entry Client 1.4.x**)
 - dans la partie *Epidata Entry Français*, cliquez **FR setup** (1.1 Mo). Le fichier à télécharger est *setup_epidata_fr.exe*. Cliquer sur Enregistrer dans le dossier EPIDATA du bureau. Le document « Course Materials in French » est fait par B. Branger
 - dans la partie *Epidata Analysis 2.2* (en anglais), cliquez sur **Setup(exe) (2.9 Mb)**. Le fichier à télécharger est *setup.epidataanalysis.exe*. Cliquer sur Enregistrer dans le dossier EPIDATA du bureau
- Développer ces deux fichiers dans votre ordinateur : se mettre à l'emplacement du téléchargement EPIDATA sur le bureau, 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* → choisir Custom : accepter le chemin proposé C:\Program Files\Epidata sous Windows XP ou C:\Programme(x86)\Epidata sous Win 7, et prévoyez les sorties des analyses dans EPIDATA du bureau de Windows.

- EPIDATA Entry 3.1 (pour la saisie)

➔ 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**

La saisie d'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) à *déconseiller* +++
 - _ ou __ ou ___ (trait sous le 8 du clavier) pour les variables de texte (lettres et chiffres)
 - <dd/mm/yyyy> pour les variables de date (d = day = jour, m = month = mois, y = year = année)
 -

Exemple à faire

```

FICHE <IDNUM >
HOPITAL ___ SERVICE ___ UNITE ___ SEXE #
DATENAISS <dd/mm/yyyy> ADMIUNIT <dd/mm/yyyy> SORTIUNIT <dd/mm/yyyy>
POIDS ###.# TAILLE #.## IMC ##.##
AGE ###.# ENDOSCOPI <Y> AMBULA #
DUREE ### ASA # COMMENT _____
    
```

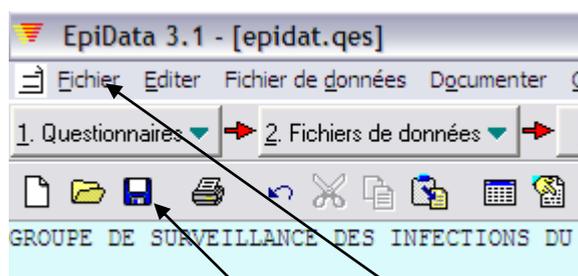
- **<IDNUM>** : nombre qui s'incrémentera à chaque enregistrement
- **HOPITAL, SERVICE ou UNITE** avec des lettres ou des chiffres

- **SEXE #** : 1 pour Homme et 2 pour Femme, chiffres pour en faire des % plus tard.
- **DATENAISS** <dd/mm/yyyy> ainsi que ADMITUNIT et SORTUNIT : jour, mois, année
- **POIDS ###** pour les 100 kg et plus; **TAILLE #.##** en mètres avec deux décimales, **IMC ##.##** avec deux décimales (par calcul à partir des deux précédents)
- **AGE ###.#** : âge calculé par différence entre date de naissance et date d'admission avec 2 ou 3 chiffres (pour les centenaires éventuellement) et une décimale. L'intervalle de décimale est un point.
- **ENDOSCOPIE** : Y pour Yes, N pour No ou rien (champ vide)
- **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 +++

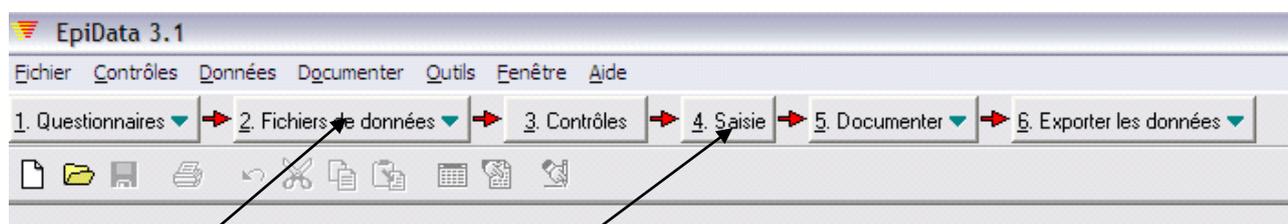
Il est fortement conseillé pour la version de EPIDATA actuelle (3.1) de mettre le maximum de variables codées avec des # et donc de coder en chiffres (et non pas avec des lettres ni avec Yes et No), même pour des variables qualitatives +++

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

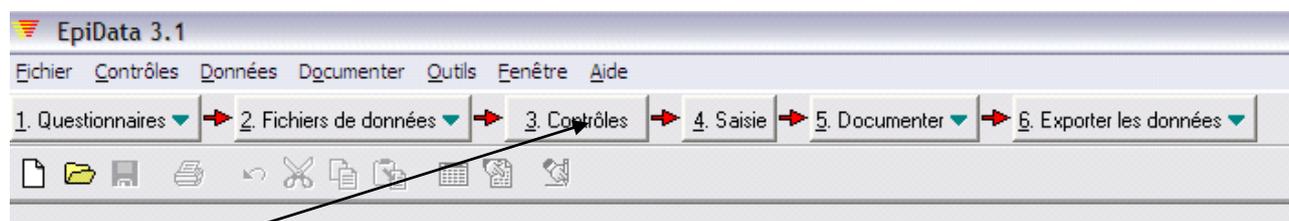
- Cliquer sur **4. Saisie**
- Sélectionner le fichier de type *.REC (connaître son chemin..) : la première fiche s'affiche, sinon la première non saisie en cas de saisie précédente.



- 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 entrée 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 sauvées 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)

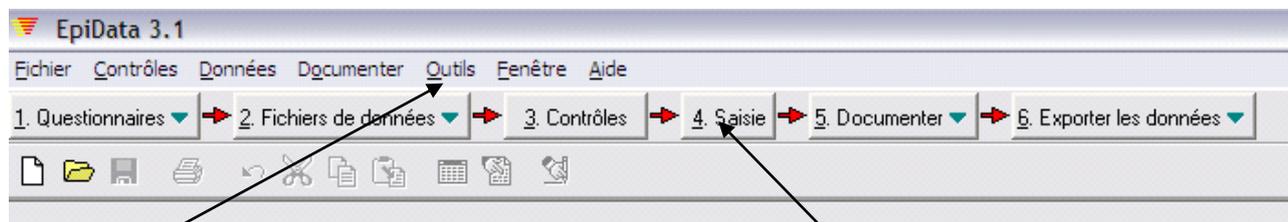
```
SEXE
topofscreen 2
end
```

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

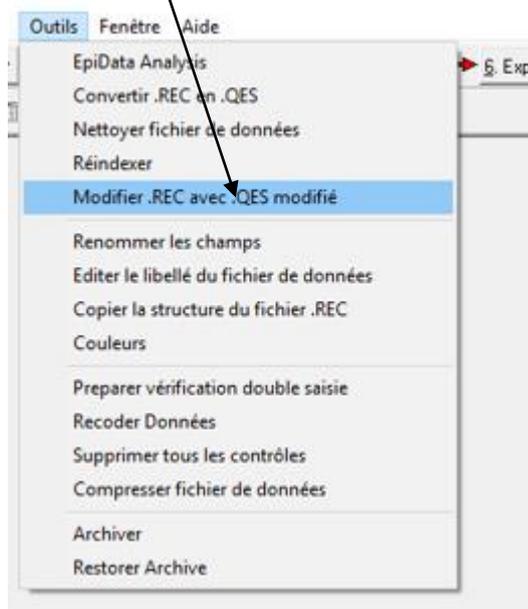
➔ Pour consulter la liste des codes à la saisie, faire F9 sur le champ de la variable et choisir soit avec la souris, soit avec les flèches ↑↓ et faire ↵.

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

Pour charger des fichiers déjà réalisés sur un autre support

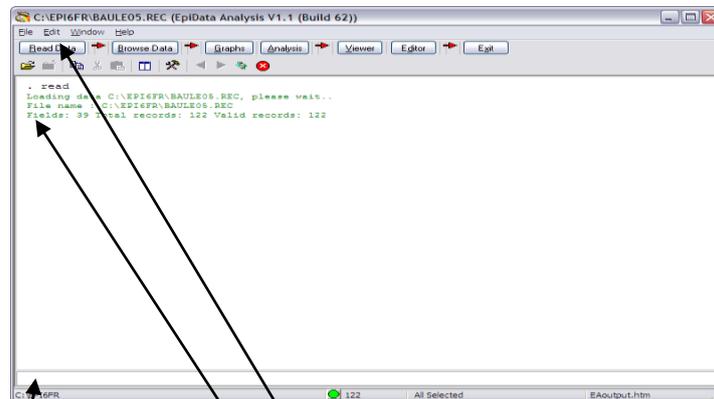
Ouvrir la clé USB ou le disque externe. Sélectionner les fichiers en *.qes, en *.rec ou en *.chk. Faire «copier». Puis se mettre dans *C:\Program files\epidata\samples* et faire «coller». Le chemin désormais pour ouvrir et sauver les données est donc *C:\Program files\epidata\samples\fichier.rec*

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 : saisie de données Cliquer sur le bureau pour ouvrir le fichier.



- EPIDATA : exploitation de données avec Analysis 2.2

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

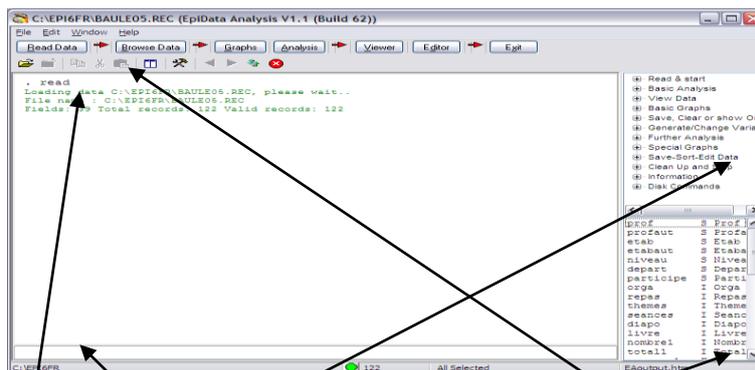


Ouvrir un fichier

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.

La fenêtre des résultats affiche les caractéristiques du fichier.

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'icône Tableau ou écrire *Browse* dans la ligne de commande **ou F6 ou cliquer sur **Browse Data**** pour voir le fichier comme un tableau. 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 (sinon taper `set echo=on`). 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 (âge par exemple).

➔ Pour décrire

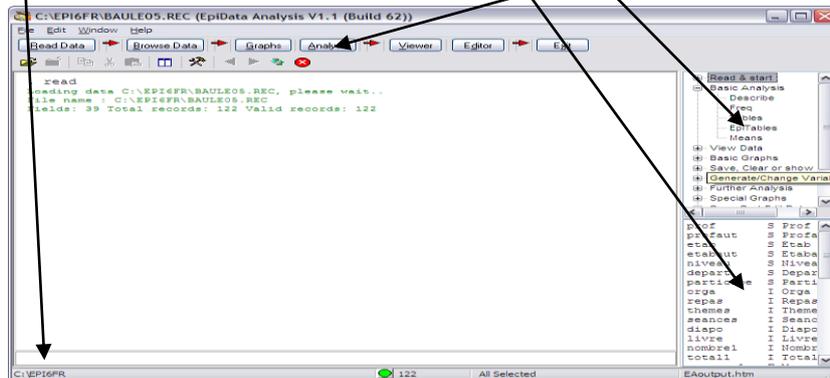
- **freq varqual** pour la fréquence des données codées (qualitatives) ; **freq *** trie toutes les variables qualitatives. Pour ajouter les pourcentages en colonne, faire (après un espace) **/c**, et pour avoir les pourcentages **cumulés /cum**, et pour avoir aussi les manquants **/m**
- **means varquant** 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

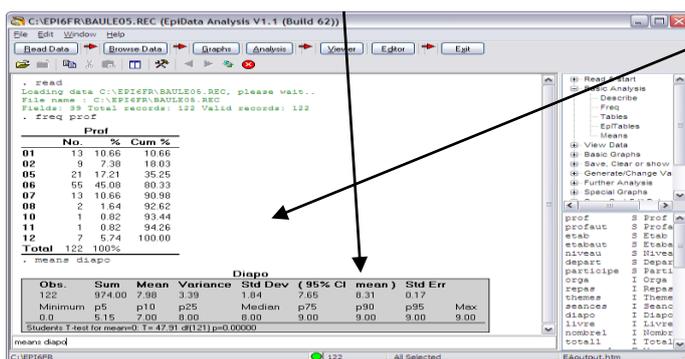
- **means varquant varqual** pour comparer une variable quantitative selon deux ou plus de deux modalités. A signaler que EPIDATA demande que la **varqual** soit de type numérique (c'est-à-dire # ou ## ou ### dans Epidata Entry). **means varquant varqual /t** pour avoir le test statistique (test de F ou test de t pour deux moyennes).
- **tables varqual1 varqual2** pour croiser deux données qualitatives (la première en colonne, la deuxième en ligne) ➔ 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**, **/rr** pour le risque relatif , **/o** pour l'odds ratio. Pour avoir les codages dans l'ordre 1,2,...., ajouter **/sa**, sinon d'autres modalités permettent d'inverser l'ordre 2,1... soit en colonnes (**/scd**), soit en lignes (**/srd**).

Procédure :

- écrire les commandes ou les cliquer dans + **Basic Analysis** ou cliquer sur le bouton **Analysis**
- écrire le nom de la variable ou cliquer sur une variable de la liste des variables
- faire ↵ après les deux termes écrits ou cliqués

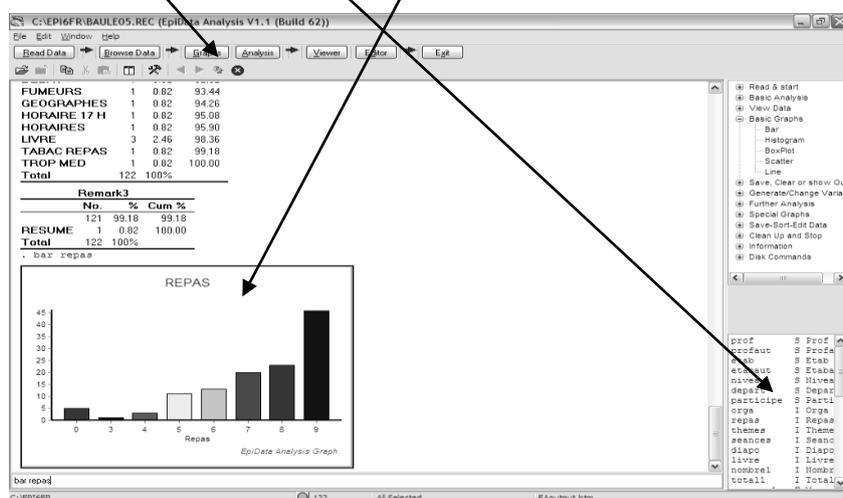


- lire les résultats dans la fenêtre des résultats après par exemple **freq varqual /c /cum** (ici appelée **prof** de 01 à 12) et **means varquant** (ici appelée **diapo**, note de 0 à 9)



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 **Basic graphs** ou sur le bouton **Graphs**



Les graphes n'ont qu'une valeur indicative : il faudra les refaire, 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. Un texte est disponible pour travailler des graphes présentables.

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.

La liste des commandes est aussi sauvée dans le même répertoire sous forme de **temp.pgm** à ouvrir avec Word ou le bloc-notes. **EAoutput** et **temp** s'incrémentent jusqu'à 5 versions (de **EAouput.htm** à **EAoutpt4.htm** (une nouvelle élimine la dernière)

Pour visualiser les données comme dans un tableur et les modifier si besoin

- Sans modifier les données : `browse var1 var2` ou cliquer sur le bouton **Browse** ou F6. 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 une ou des valeurs. Attention les données sont modifiées dans le fichier d'origine.

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 +++)

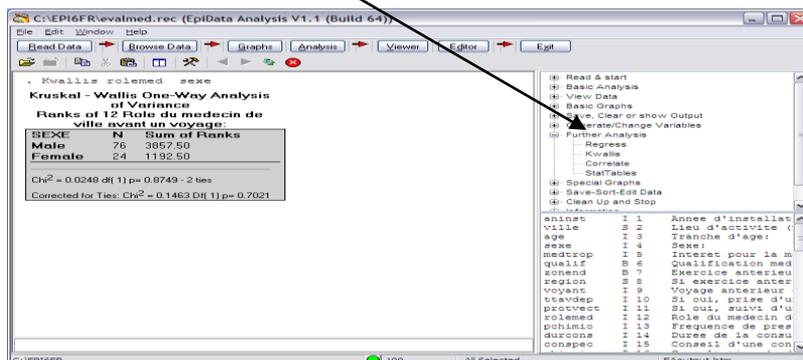
→ Les codages sont bruts pour les données quantitatives, avec des " " pour les variables qualitatives codées avec le _. Exemple : define sx # if sexe="M" then sx=1

→ 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

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*, ou trouver *temp.pgm* et le copier.

- *Run fichier.pgm* : lance l'analyse automatiquement.

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

Annexe

Importer un fichier Excel

➔ Deux solutions :

- enregistrer, dans Excel 2003 et avant, 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.

- l'export au format *.dbf n'est plus possible à partir de Excel 2007 (2010 et 2013) : deux solutions

- ✓ Enregistrer, dans Excel, au format *.csv; ouvrir le fichier dans EPIDATA (en prenant "all" comme type de fichiers).
- ✓ Copier tout le fichier Excel à la souris, revenir à Analysis et écrire la commande `read /cb` (cb pour clipboard) : ça marche pour les petits fichiers « simples » et « propres »

➔ Attention : les noms des variables doivent être simples : pas plus de 10 lettres, sans espace, sans accent, sans signes particuliers... (sinon EPIDATA écrit v1, v2.....) 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...).

Notes

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

