

# EPIDATA : Analysis 2

## Niveau 2 pour 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](mailto:bernard.branger@naître-ensemble-ploire.org)  
Janvier 2014

### - Télécharger et installer EPIDATA

#### sous Windows XP ou Vista ou Windows 7 ou 8

- Télécharger : par Internet, ouvrir <http://www.epidata.dk>. Cliquer sur download puis effectuer deux manœuvres :
  - dans la partie **Epidata Analysis**, cliquez sur [Setup\(exe\) \(2.9 Mb\)](#). Le fichier à télécharger est *setup-epidataanalysis.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 sur [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 Vista ou C:\Programmes\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.

**Note** : Le texte du guide est en arial 10.

**Les indications entourées sont des boutons à cocher.**

Les indications en courrier new 10 sont du texte à écrire dans les lignes de commande (les minuscules et majuscules sont équivalentes).

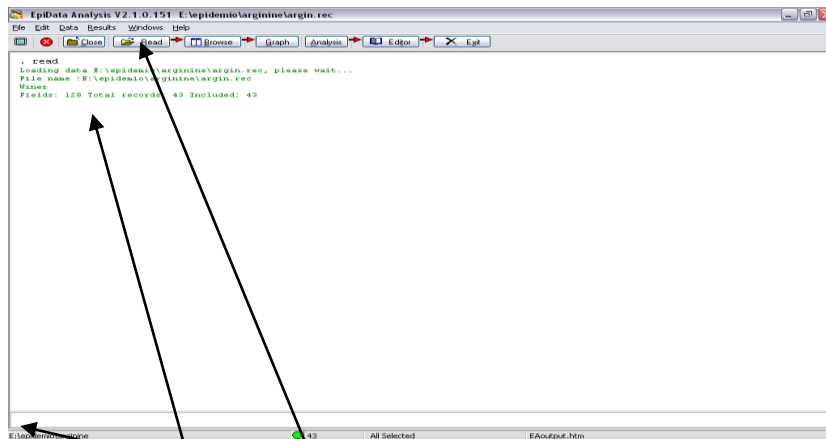
Les indications en arial 10 italique sont du texte de commandes de menus à cliquer. Les sorties logicielles sont en Time New Romans 12. *NB* : varqual est une variable qualitative (sexe par exemple), varquant est une variable quantitative (age par exemple). **Aide : appuyer sur F1 sur le clavier pour comprendre et s'améliorer**

**Attention !** Dans Vista et Windows 7 ou 8, la sauvegarde des fichiers \*.rec, \*.ges, \*.chk, \*.htm et \*.pgm se trouve quelquefois dans une fenêtre « Fichier de compatibilité ». Pour les voir dans le répertoire *Samples* par exemple : 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).

Pour les pro, pour supprimer cette possibilité : clic droit sur *EpiData.exe* et *EpiDataStat.exe*, Propriétés, **Compatibilité**, cocher  Exécuter ce programme en mode de compatibilité pour... Prendre Windows XP (Service Pack 3) (si c'est votre version de XP).

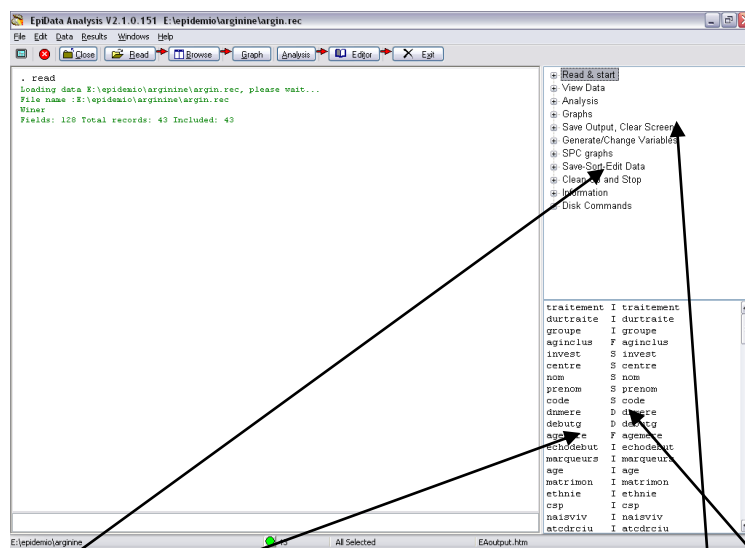
## - EPIDATA Analysis : principes

Cliquer sur l'icône **Epidata Analysis** sur le bureau pour ouvrir le fichier : il s'agit ici de fichiers \*.rec créés dans EPI6 ou EPIDATA. Pour lire des fichiers d'autres formats, voir en fin de document.



- **Ouvrir un fichier** : cliquer sur le bouton **Read** : choisir un chemin à la recherche de fichiers \*.REC. comme *C:\Program Files\Epidata\Samples* ou *C:\Epi6\fr* si des fichiers \*.rec ont déjà été créés dans Epi 6.0. La fenêtre des résultats affiche les caractéristiques du fichier écrits en vert. A noter que toutes les sorties sont dirigées vers un répertoire par défaut (qui dépend de l'installation que vous avez effectuée). A signaler la ligne de commande pour écrire ou saisir en cliquant → très important : utiliser les flèches ↑ ou ↓ pour reprendre une commande déjà effectuée. On peut faire *Read* suivi de ↵ (touche entrée)

## - Aspect de l'écran



- **Fenêtre des commandes** : faire F2
- **Fenêtre des variables** : faire F3 → variable de type *S = String* = codage qualitatif ou de type *I = Integer* = ou *F = Float* = *quantitatif*, ou *D = Date*
- Cliquer sur **Browse** ou écrire *Browse* dans la ligne de commande ou F6 pour voir les données disposées en **tableur**. On peut écrire *update* avec ou sans *variables* (revenir en cliquant sur **Close**) : on peut modifier les erreurs qui seront corrigées sur le fichier initial (attention donc).
- Voir la liste des commandes en cliquant sur **Help** ou F1.

## - Pour imprimer les résultats

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

## - Pour sauver les résultats

Les résultats sont sauvés 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 output*, *Clear screen* ou écrire dans la fenêtre des commandes `logopen "fichier.htm" /close` (ou `/append` pour écrire sur un fichier déjà créé). Un fichier de nom `fichier.htm` est créé ; terminer par `logclose` pour quitter ce fichier de résultats.

## - Bien différencier les types de variables

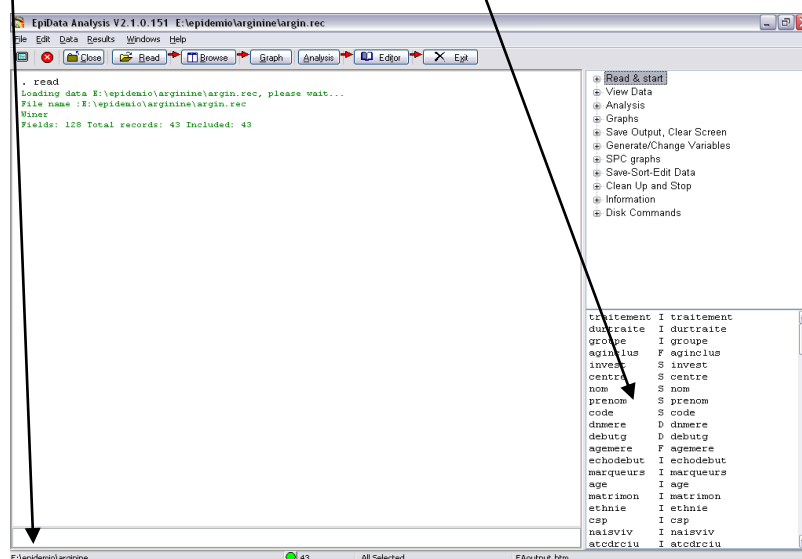
- **Variable qualitative** avec des catégories ou des codes
  - ✓ Description d'une variable avec des pourcentages : exemple `freq sexe`
  - ✓ Croisement de deux variables avec des tableaux : exemple `tables sexe departement`
- **Variable quantitative** avec un format numérique (de type `##.##` ou `###` qui peut prendre toutes les valeurs).
  - ✓ Description d'une variable avec des indices comme la moyenne, la médiane, l'écart-type, les minimum et maximum... : exemple `means age` ou `means glycemie`
  - ✓ Comparaison de deux moyennes (c'est-à-dire comparaison d'une variable quantitative selon une variable qualitative) avec le test de t et le test de Kruskal-Wallis : exemple `means age sexe` ou `means age glycemie` OU `means age periode` OU `kwallis age sexe`
  - ✓ Comparaison de deux variables quantitatives avec le coefficient de corrélation ou la droite de régression : exemple `regress age TA_systolique` OU `correlate age poids`
- **Variable de date**
  - ✓ Calculs de durées par des différences ; extractions de l'année, du mois, de la semaine...

## Les commandes de base dans la fenêtre des commandes

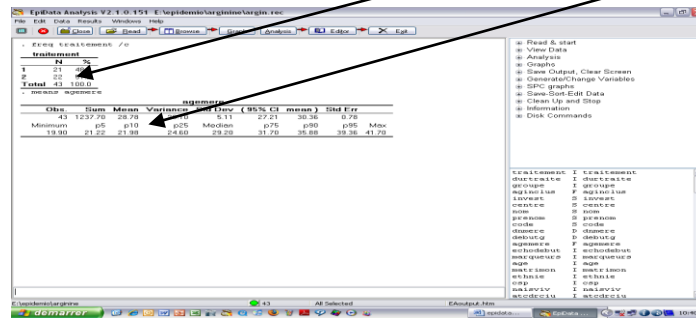
Les conseils donnés ici concernent l'écriture dans la fenêtre des commandes. Pour ceux qui aiment cliquer, on obtient les mêmes résultats en cliquant sur **Analysis**, choisir *Frequency* ou *Means* ou *Tables*, puis cliquer sur ce que l'on désire. De toute manière, le logiciel écrit sur les résultats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

### 1. Procédure si l'on préfère écrire dans la fenêtre des commandes

- écrire les commandes
- écrire le nom de la variable ou cliquer sur une variable de la liste des variables
- faire `↵` après les deux ou trois termes écrits ou cliqués



- lire les résultats dans la fenêtre des résultats après par exemple `freq traitement /c` ou `means agemere`



## 2. Données qualitatives

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) .

### FREQ ou freq (en majuscules ou minuscules)

La commande `freq varqual` permet de visualiser la fréquence des variables qualitatives (codées ou en classes). Les options sont les suivantes (écrire `freqUvarU/options`) → le caractère U est un espace)

- `freq *` Affiche la fréquence de toutes les variables du fichier. A faire au début de l'analyse pour vérifier les effectifs et les erreurs de saisie.
- `/c` Montre les effectifs des classes et les pourcentages
- `/m` Inclue les observations manquantes (codées avec un point)
- `/nm` Exclue les observations manquantes (par défaut)
- `/cum` Montre les pourcentages cumulatifs de haut en bas dans le tableau
- `/ci` Montre l'intervalle de confiance des pourcentages (formule binomiale)
- `/D1 /D0` Spécifie le nombre de décimales des pourcentages (défaut 1 décimale, ce qui est suffisant)

Pour les noms des variables dans les réponses

- `/v` seulement les codages
- `/v1` codages et labels (en cas de F9 à la saisie)
- `/vn` seulement les labels

### TABLES ou TAB

La commande `tables varqual1 varqual2` (on peut ajouter `var3`, voir plus bas) permet de croiser deux variables qualitatives. Le tableau présenté est de type *tables colonne ligne* ou encore *tables maladie exposition* dans les enquêtes de cohorte (avec les pourcentages en lignes) ou cas-témoins. Les options sont les suivantes (écrire `tablesUvar1Uvar2U/options`) → le caractère U est un espace)

- `/f` Montre deux tableaux successifs pour chaque variable. Equivaut à `freq var1 var2`
- `/m` Avec les manquants codés avec "."
- `/r` Pourcentages en lignes → préférer de manière générale +++**
- `/c` Pourcentages en colonnes
- `/tp` Pourcentages sur le total de l'effectif de la population

- `/OA` Analyse d'une épidémie avec le taux d'attaque (ou taux d'incidence pour *attack rate* = AR) pour les exposés et pour les non-exposés et risque relatif ; /CI Intervalle de confiance ; /OBS : montre les effectifs à droite du tableau des taux d'attaque. /CC Odds ratio en cas d'enquêtes cas-témoins
- `/D0 /D1 /D2` Nombre de décimales dans les pourcentages. Par défaut = 1 décimale (suffisant)
- `/t` Test du  $\chi^2$  de Pearson
- `/ex` Test exact de Fisher
- `/rr` Risk Ratio = risque relatif (noter toujours les pourcentages en lignes) pour les enquêtes de cohorte. *Supprime la commande EPITABLES de la version 1.*
- `/o` Odds Ratio pour les enquêtes cas-témoins. Par défaut, le codage le plus élevé est mis en première ligne de l'exposition et en première colonne de la maladie. On peut inverser cet ordre avec /SA. L'Odds Ratio est calculé comme dans Epi6 MLE
- `/nt` En cas de `tables var1 var2 var3` présente les tableaux par la var3 ; avec /RR ou /O ; présente les RR et OR bruts et de Mantel-Haenszel ; /NC : sans les tableaux de var3
- `/i` Détails d'explications des tests
- `/v` Montre les valeurs de codages seuls (1 par exemple)
- `/vl` Montre les valeurs de codage et les labels (1 Homme par exemple) (si le F9 a été utilisé lors de la saisie ou avec la commande `labelvalue`, voir infra, )
- `/vn` Montre le nom des labels seuls (Homme par exemple)

**Les commandes tables d'utilisations plus rares**

- `/q` Ne présente pas le résultat du tableau (à quoi cela sert ?)
- `/gam` Test de  $\gamma$  de Goodmann & Kruskals
- `/W=variable` Utilise la variable comme poids
- `/Sxxx` Ordre des codages des variables dans les tableaux où x est remplacé par :  
R = (row) ligne, C = colonne, A = codage ascendant, D = codage descendant, T = total, L = label . Les commandes suivantes sont possibles :  
`/SA` , `/SD`, `/SLA`, `/SLD`, `/SRAT`, `/SCAT`, `/SRDT`, `/SCDT`, `/SRD=x`,  
`/SCD=x`, `/SRA=x`, `/SCA=x` (e.g. `/sca=2`)

→ `tables traitement mat /r /t /ex`

Trait						
Mat	1	%	2	%	Total	%
1	2	{66.7}	1	{33.3}	3	{100.0}
2	19	{47.5}	21	{52.5}	40	{100.0}
<b>Total</b>	21	{48.8}	22	{51.2}	43	{100.0}

Expected value < 5 in 2 cells ( 50 percent). Chi<sup>2</sup> may not be valid.  
Chi<sup>2</sup>= 0.41 Df(1) p = 0.5218  
Fisher exact: p= 0.6069

Dans cet exemple, le  $\chi^2$  vaut 0.41 et p=0.52 soit non significatif : mais il n'est pas valide. Il faut alors prendre le Fisher qui vaut 0.61 (toujours non significatif).

→ tables allmere allpere /r /t

ALLMERE								
ALLPERE	1	%	2	%	3	%	Total	%
1	44	{46.8}	48	{51.1}	2	{2.1}	94	{100.0}
2	35	{22.4}	117	{75.0}	4	{2.6}	156	{100.0}
3	18	{32.7}	33	{60.0}	4	{7.3}	55	{100.0}
<b>Total</b>	97	{31.8}	198	{64.9}	10	{3.3}	305	{100.0}

Expected value < 5 in 2 cells ( 22 percent). Chi<sup>2</sup> may not be valid.  
Chi<sup>2</sup>= 19.70 Df(4) p = 0.0006

Dans cet exemple, le  $\chi^2$  n'est pas valide et on ne peut conclure, même si c'est très significatif ( $p < 0.001$ ). Le test de Fisher n'est pas possible en dehors des codages à deux classes. Il faut alors regrouper des codages (le « 3 » devient « 2 » selon une plausibilité clinique ou biologique plutôt que statistique).

```
. if allmere=3 then allmere=2
. if allpere=3 then allpere=2
. tables allmere allpere /r /t
```

ALLMERE						
ALLPERE	1	%	2	%	Total	%
1	44	{46.8}	50	{53.2}	94	{100.0}
2	53	{25.1}	158	{74.9}	211	{100.0}
<b>Total</b>	97	{31.8}	208	{68.2}	305	{100.0}

Chi<sup>2</sup>= 14.11 Df(1) p = 0.0002

Le regroupement permet de dire que 46.8 % est différent de 25.1 % avec  $p < 0.001$ . L'utilisation du Fisher n'est pas utile car les effectifs sont suffisants.

→ tables arret allaite\_m /R /T /EX /RR /O : présente en plus les risques relatifs et les odds ratios avec les intervalles de confiance .

Dans l'exemple, `allaite_m` représente le fait d'avoir été allaité par sa propre mère (0 = non et 1 = oui) : c'est le facteur d'exposition (exposure). `Arret` représente l'arrêt précoce de l'allaitement maternel le premier mois (1=oui et 2=non) : c'est la variable de jugement ou d'intérêt (outcome).

ARRET						
ALLAITE_M	1	%	2	%	Total	%
0	32	{15.0}	181	{85.0}	213	{100.0}
1	5	{5.0}	96	{95.0}	101	{100.0}
<b>Total</b>	37	{11.8}	277	{88.2}	314	{100.0}

Outcome: ARRET=1 Exposure: ALLAITE\_M=0  
 Chi<sup>2</sup>= 6.688 p= 0.010  
 Fisher exact: p= 0.0086  
 Crude OR = 3.394 (95% CI: 1.20 - 10.37)  
 MLE: OR= 3.384 (Exact 95% CI: 1.25 - 11.49)  
 p (MLE OR >= 3.38 if population OR = 1.0)= 0.0058  
 RR= 3.035 (95% CI: 1.22 - 7.56)

La statistique de comparaison de deux taux (15.0 % et 5.0 %) est calculée avec le  $\chi^2$  et le test de Fisher, tous les deux significatifs ( $p=0.01$ ). Puis l'odds ratio est calculé (« Crude OR » = OR brut) ici de 3.4 (une seule décimale peut suffire). Son intervalle de confiance à 95 % est calculé : 1.2 – 10.4 → son intervalle ne comprend pas 1, il est significativement différent de 1. La mention *MLE* est un autre calcul de l'OR, qui est voisin ; préférer le Crude OR.

Un risque relatif est calculé avec RR qui vaut 3.03 avec un intervalle de confiance de 95 % de 1.2 – 7.6 qui est aussi significativement différent de 1 : c'est le rapport de 32/213 sur 5/101 ou encore (aux arrondis près) de 15.0 sur 5.0 → d'où l'intérêt des pourcentages en ligne et de la présentation du tableau dans ce sens.

### STATTABLES

La commande `stattables var1 var2` (et éventuellement `var3`) montre un tableau résumé du dénombrement des couples de valeurs `var1 var2 var3`, et les exporte dans un nouveau fichier avec une variable nouvelle de dénombrement (`count`)

```
/close          Permet de fermer le fichier en cours et d'ouvrir un nouveau fichier avec les valeurs de
                dénombrement des couples de valeur

/save="filename" Sauve le nouveau fichier

/m             Avec les manquants

/mean /SD /max /min /p5 /p10 /p25 /p50 /p75 /p90 /p95 /p99 /sum /des (min median max)
/iqr (p25 p75) /idr (p10 p90) /isr (p5 p95)/mv.
```

Exemple : `Stattab sex agegrp /Mean="weight height" /Des="weight,height"`

### 3. Variables quantitatives

- **DESCRIBE** : `describe *` pour décrire toutes les variables quantitatives avec moyenne, médianes, percentiles divers.... `describe age` pour décrire la variable `age` ou `describe age poids taille` pour décrire les trois variables. Préférer `means` en général (mais un par un). Commande non maîtrisée car se bloque si une seule variable vide !
- **MEANS** : `means agemere`

AGEMERE									
	Obs.	Sum	Mean	Variance	Std Dev	( 95% CI	mean )	Std Err	
	308	9247.0	30.02	22.28	4.72	29.49	30.55	0.27	
Minimum		p5	p10	p25	Median	p75	p90	p95	Max
	17.00	21.45	25.00	27.00	30.00	33.00	36.00	39.00	42.00

Tout n'est pas à communiquer : la moyenne (`mean`) est le plus commode avec l'écart-type (= `Std Dev` = déviation standard), mais la médiane (`median`) peut être intéressante surtout si elle est différente de la moyenne : on peut aussi ajouter le minimum, le maximum et l'étendue (différence entre les deux). Mais plus on est précis (beaucoup d'informations), moins on est « percutant ». A signaler  $p$  = percentile et  $p5$  = 5<sup>ème</sup> percentile.

- `means varquant /t` : le « Students's test » compare la moyenne à 0 : rarement utile, en tout cas pas ici pour l'âge de mères, mais pour des différences par exemple dans l'hypothèse de comparaison à 0 (séries appariées). Elle peut servir à comparer avec une valeur théorique : créer d'abord la différence de chaque observation avec cette valeur ; comparer cette différence à 0.
- Avec une donnée qualitative : `means varquant varqual` ou `means varquant /by=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 #. Ci-dessous `means agemere arret` (équivalent à `means agemere /by=arret`)

AGEMERE									
ARRET	Obs.	Sum	Mean	Variance	Std Dev	( 95% CI	mean )	Std Err	
1	37	1056.00	28.54	44.92	6.70	26.31	30.78	1.10	
2	271	8191.0	30.23	19.00	4.36	29.70	30.75	0.26	

ARRET	Minimum	p5	p10	p25	Median	p75	p90	p95	Max
1	17.00	17.90	19.60	23.00	29.00	33.50	38.40	40.10	41.00
2	19.00	23.00	25.00	27.00	30.00	33.00	36.00	38.40	42.00

Le premier tableau décrit les moyennes, médianes... pour chaque valeur dans le premier tableau : voir à `means` plus haut. Le second tableau est le résultat du test de comparaison de deux moyennes : la valeur est à lire sous « p value » : ici  $p < 0.05$ , les deux moyennes diffèrent (28.5 ans et 30.2 ans).

➤ `means agemere arret /T.` Lire le test de l'ANOVA ici ( $p < 0.05$  donc significatif)

et non pas là : ce « p » là veut dire que les deux variances sont différentes (44.92 et 19.00)

Source	SS	df	MS	F	p Value
Between	92.38	1	92.38	4.19	0.0415
Within	6748.46	306	22.05		
Total	6840.84	307	22.28		

Bartlett's test for homogeneity of variance

Chi<sup>2</sup> = 14.412 df(1) p = 0.000

Si le fichier comporte un faible échantillon (moins de 30) ou que la variable n'a pas une distribution « normale » (par exemple si la moyenne est très différente de la médiane), ou si le test de Bartlett est significatif, la méthode à utiliser est la suivante : `kwallis varquant varqual` pour le test de Kruskal-Wallis (test non paramétrique au même objectif que le test t commandé par `means`) (dans *Further analysis*, voir *infra*).

Ici `kwallis agemere arret`. Lire le test ici : non significatif ( $p = 0.1330$ )

**Kruskal - Wallis One-Way Analysis of Variance  
Ranks of AGEMERE**

ARRET	N	Sum of Ranks
1	37	4955.00
2	271	42631.00

Chi<sup>2</sup> = 2.2459 df( 1) p = 0.1340 - 22 ties

Corrected for Ties: Chi<sup>2</sup> = 2.2570 Df( 1) p = 0.1330

➤ **CORRELATE**

Compare deux variables quantitatives différentes à l'aide du coefficient de corrélation r de Pearson r va de -1 à + 1. Le test du r à 0 est le même que celui avec le test de la pente avec `correlate`.

Exemple : lien entre l'âge des mères (`agemere`) et le poids de naissance (`pn`)

```
correlate agemere pn
          pn          agemere
pn          1.000
agemere    0.027          1.000
Total N = 113 Included: N = 110
```

r = 0.027 (faible, mais le test statistique n'est pas mis, c'est avec `regress`)

➤ **REGRESS**

Compare deux variables quantitatives différentes

`regress Y X` avec formule de la régression linéaire  $Y = A + B * X$  où B est la pente et A l'intercept

Exemple : `pn` (poids de naissance) par rapport à l'âge de la mère → `regress pn agemere`

Le test de la pente à 0 (zéro) est ici ( $p = 0.78 \Rightarrow$  non significatif  $> 0.05$ )



La formule de la droite est la suivante :

$$Y(\text{ici pn}) = 2192.40 + 4.07 * X(\text{ici âge de la mère})$$

```

regress pn agemere
Source Sum Sq      Mean Sq      df      Number of obs=110
Model      41223.55      41223.55      1      F(1,108)      0.08
Residual  54787881.87    507295.20     108    Prob > F      0.78
Total     54829105.42    503019.32     109    R-squared     0.00
                                Root MSE     712.25

Variable      Beta      LCL      UCL      SE      t      P>|t|
agemere      4.07      -24.22  32.35    14.27    0.29    0.78
Intercept    2192.40    1308.31  3076.49  446.06  4.92    0.00
Total N = 113  Included: N = 110

```

### ➤ COURBE DE SURVIE (lifetable)

Partie réservée aux utilisateurs de courbe de survie en cancérologie par exemple. Il faut deux types de variables :

- Variable de temps en jours, semaines ou mois ; généralement obtenue par une différence de dates : la date de début ou d'inclusion et la date de fin de suivi dite date de point (DP) ou date des dernières nouvelles (DDN)
- Variable « événement » codée souvent en 0 pour vivant (DP) ou perdu de vue (DDN), et 1 pour décès (DDN)

Le résultat est une suite de probabilité de survie selon le temps et une courbe en escalier dite de « Kaplan –Meier ». Les commandes sont de type : `lifetable duree evenement` OU `lifetable duree evenement /by=groupe /t` pour comparer deux courbes de survie. Voir F1 pour toutes les options possibles.

➔ Ce type d'analyse peut concerner tout événement dans le temps pour lequel toutes les données ne sont pas disponibles à un temps donné. L'événement n'est pas forcément un décès.

## Les graphes (voir complément page 17)

### Les graphes les plus courants

- histogramme (**h**istogram) pour les variables quantitatives : `histogram agemere`
- en barres (**b**ar) pour les variables qualitatives (en vertical ou en barres) : `bar profession`
- en camembert (**p**ie = tarte) pour les variables qualitatives : `pie profession`
- en points (**s**catter) pour des couples de valeurs quantitatives : `scatter poidsnaissance agemere`
- en boîtes (**b**oxplot) pour les variables quantitatives : `boxplot agemere` avec la médiane, le 1<sup>er</sup> et le 3<sup>ème</sup> quartiles ; les lignes en haut en bas sont établies par rapport à la médiane ( $\pm 1.5$  fois l'écart inter-quartile).
- en ligne (**l**ine) pour les données dans le temps

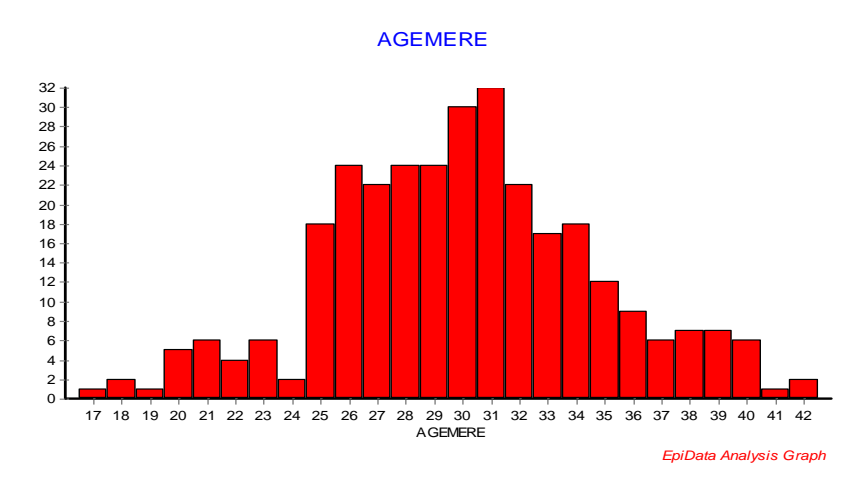
### Commandes supplémentaires

```

/by=varname  Graphes selon une autre variable (scatter, line, box, bar, dotplot)
/edit       Permet de modifier le graphe ➔ richesse des menus (un peu compliqué)
/q         N'affiche pas le graphe
/save=" "   Sauve le graphe avec le nom et le suffixe en *.bmp, ou *.wmf. Par défaut, c'est en *.png

```

➔ `histogram agemere`



Pour les variables quantitatives, la commande `histogram` peut regrouper en classes avec la mention `/width=valeur`. Sinon, il faut mettre la variable en classes, avec la commande `recode` (en précisant la longueur de chaque classe ici 5 ans), d'abord dans une variable qualitative et faire `bar classe`. Recoder en classes les variables quantitatives : âge des mères, pour faire des classes de 5 ans :

```
define old5 ##
recode agemere to ol5 by 5
freq old5 /c /cum
```

	No.	%	Cum %	
		6	1.9	1.9
<b>15 - 19</b>	4	1.2	3.2	
<b>20 - 24</b>	23	7.3	10.5	
<b>25 - 29</b>	112	35.7	46.2	
<b>30 - 34</b>	119	37.9	84.1	
<b>35 - 39</b>	41	13.1	97.1	
<b>40 - 44</b>	9	2.9	100.0	
<b>Total</b>	314	100%		

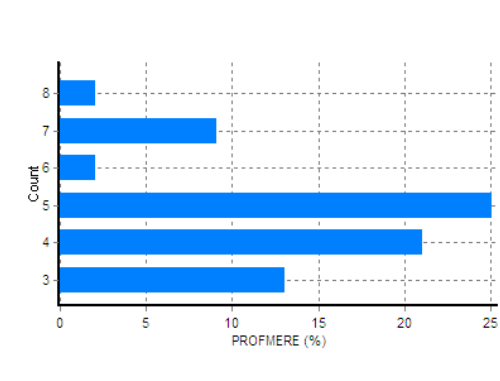
```
bar old5 /edit
```

Pour supprimer le titre : cliquer sur **Titre** et mettre vide le texte à gauche ; idem pour **Footnote** (enlève *Epidata Analysis Graph*). Pour en mettre un autre titre : Chart, Title

Pour barres horizontales : *Edit all, Data* : cliquer sur **Change** et choisir un nouveau graphe

Pour modifier les axes : *Chart, Axis....*

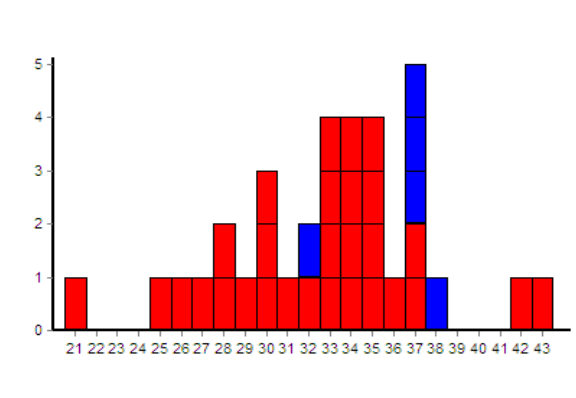
Pour mettre les mêmes couleurs : *Series, Data source* : cliquer sur la couleur pour changer



**EPICURVE**

```
epicurve temps critere [/by=group] [/options]
```

Cette commande montre une courbe épidémique par jour par exemple, avec une autre variable (/by=var)



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

## Divers

### Modifier l'ordre de sortie des fiches : SORT

sort variable1 [variable2] [variable3 ...] : ordonne le fichier selon un nouvel ordre base sur la variable ;  
pour list ou update ou browse. sort ↵ : reviens à l'ordre du fichier primitif

### Changer, créer des variables → ne change pas le fichier original de données +++

#### DEFINE et IF.... THEN..... ELSE .....

→ 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 # ↵1 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 ↵ où else dirige le codage vers toutes les autres valeurs. Attention : age<30 → strictement inférieur à 30 (si on veut garder 30 : mettre age=<30). Si doute pour les manquants, 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 ↵

→ 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 ↵
```

→ A noter que les variables manquantes sont, par défaut, dans EPIDATA sous la forme de « . ». Par exemple, pour remplacer les variables manquantes par un codage « 9 »

La variable est numérique (en #) : if var=. then var=9

La variable est du texte (en \_\_\_\_\_) : if var=. then var= "9"

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

→ Tous les calculs sont possibles. Exemple pour l'IMC avec la taille en mètres :

```
define IMC ##.# ↵ IMC = poids/(taille*taille) ↵ means IMC
```

### Manipuler les variables de date

Les dates à utiliser sont généralement de type « européen » <dd/mm/yyyy> ou « d » est le jour (pour « day »), « m » le mois et « y » l'année en 4 chiffres (pour « year »).

→ Différence entre deux dates comme entre `datejour` et `datenaiss` pour l'âge. Exemple : définir une variable d'âge (avec une décimale pour que ça marche), puis faire `means age` ou `describe age` ou `freq age` ou `bar age` ou `histogramm age`

- Différence en jours : `define agej ##.# ↵ agej=(datejour-datenaiss) ↵`
- Différence en semaines : `define ages ###.# ↵ ages=(datejour-datenaiss)/7 ↵`
- Différence en mois : `define agem ###.# ↵ agem=(datejour-datenaiss)/30.4 ↵`
- Différence en années : `define agey ###.## ↵ agey=(datejour-datenaiss)/365.25 ↵`  
ou `define agean ## ↵ agean=round(int((datejour-datenaiss)/365.25)) ↵`

Cette dernière commande permet d'obtenir un nombre arrondi, sans décimale (sinon, ça ne marche pas directement : préférer les nombres avec décimales).

→ Extraire le jour du mois, le jour de la semaine, le mois ou l'année

```
define day ## ↵ day=day(datejour) ↵ freq day ↵
define jour # ↵ jour=dayofweek(datejour) ↵ freq jour ↵
define mois ## ↵ mois=month(datejour) ↵ freq mois ↵
define an ##### ↵ an=year(datejour) ↵ freq an ↵
```

### Sauver un fichier de données modifiées ou avec des nouvelles variables

```
savedata "fichier.rec" [liste de variables] [/replace]
```

OU `savedata "fichier.rec" [/replace]` pour toutes les variables

Sauve le fichier avec toutes les variables créées, avec la sélection en cours (enlever ou non les `select`). Le fichier ne doit pas exister préalablement. La liste des variables après le nom du fichier sont seules gardées, sinon toutes les variables.


`/replace`      Idem mais écrase un fichier pré-existant

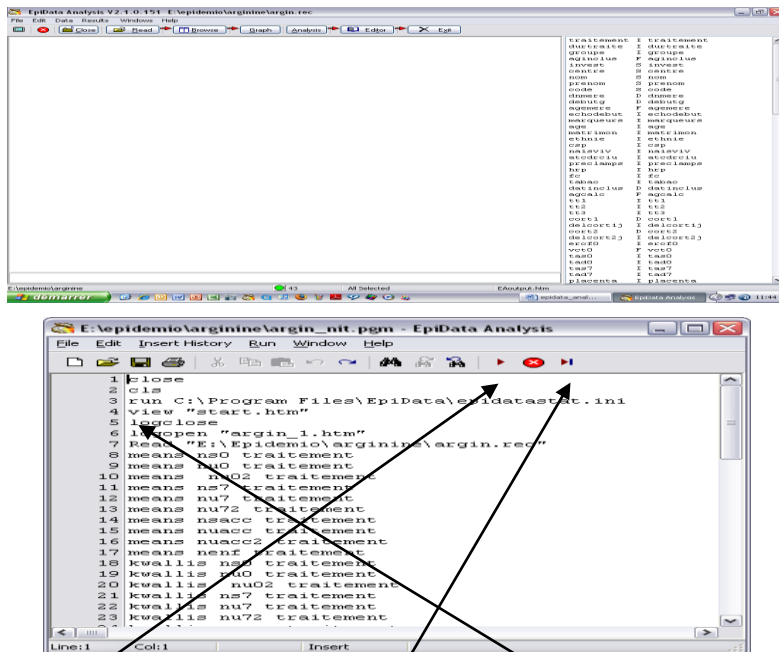
### Sauver un programme d'analyse – Le lancer ensuite

- A la fin d'une série de commandes : `savepgm "fichier.pgm"` permet de sauver les commandes réalisées dans un `fichier.pgm` (sauvé par défaut dans `temp.pgm`).

Exemple de commandes successives sauvées dans un fichier :

```
Read "C:\Program Files\EpiData\samples\echec_pr_314.rec"
freq sexe
means agemere
means agemere sexe
define old5 _____
recode agemere to old5 by 5
freq old5
savepgm "arret1.pgm"
```

- Pour affiner un programme, il faut cliquer sur **Editor** et l'ouvrir dans cette nouvelle fenêtre avec  dans le répertoire correspondant.



Ce programme peut être modifié, complété, tronqué. Le sauve en cliquant sur et le lancer, soit totalement, en cliquant sur , soit partiellement en cliquant sur (après sélection de la partie à relancer). On peut aussi lancer un programme avec un fichier.pgm.

De la sorte, toutes les variables créées une fois peuvent être retrouvées facilement pour les re-créeer ou en garder la trace. Cette manœuvre peut également servir à traiter un fichier plusieurs fois de suite avec des commandes `select` différentes.

## Importer et exporter des fichiers dans d'autres formats que les « \*.rec »<sup>2</sup> : EPIDATA Entry

### 1. Importer et lire des fichiers d'Excel

- 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 directement 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 »

Fichier texte avec ; comme séparateur

```
AGE; SEXE; GROUPE
12;1;2
13;2;2
15;1;1
18;2;1
```

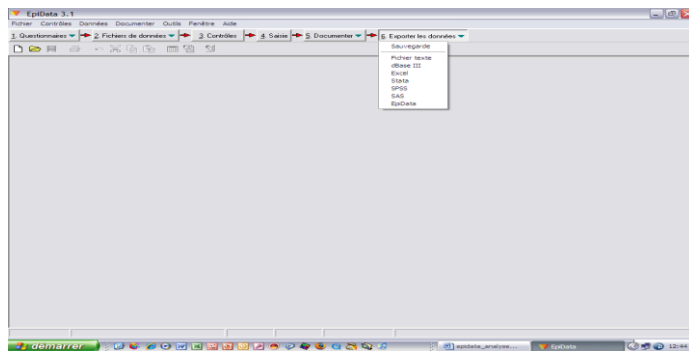
Fichier texte avec espace ou tabulation comme séparateur

```
AGE SEX GROUPE
12 1 2
13 1 5
14 2 3
15 1 5
```

### 2. Exporter des fichiers sous un autre format

Dans EPIDATA Entry : clique sur **Données** et **Exporter**, ou sur et choisir le type d'export. Suivre les instructions. On peut se limiter en exportant des variables choisies. En format texte, on peut choisir le séparateur (onglet Options, le plus commode est le « ; »).

<sup>2</sup> Pour voir les extensions des fichiers sous Windows (comme \*.rec, \*.dbf, \*.txt), se mettre dans *Explorateur, Outils, Options des dossiers, Onglet Affichage*, décocher « Masquer les extensions des fichiers dont le type est connu ».



## Commandes d'état "set" et équivalents (faire F1)

Choix de quelques-unes :

Option	Default Value	Comments or function
BROWSER FONT SIZE	10	Font size in browser and update
DISPLAY COMMAND PROMPT	<b>ON</b>	Show command prompt
DISPLAY MAINMENU	<b>ON</b>	Show main menu
DISPLAY TOOLBAR	<b>ON</b>	Show toolbar
DISPLAY WORKTOOLBAR	<b>ON</b>	Show work toolbar
EDITOR FONT SIZE	10	Font size in editor
EDITOR PRINT INFO	<b>ON</b>	When printing from editor add footer
GRAPH CLIPBOARD	<b>ON</b>	Copy graphs to clipboard
GRAPH FILENAME FOLDER	OFF	Add complete path to filename of saved graph files
GRAPH FILENAME SHOW	OFF	Show the filename in output
GRAPH FONT SIZE	10	Graph text font size
GRAPH FOOTNOTE	EPIDATA ANALYSIS GRAPH	Which footnote, can be empty: "
GRAPH SAVETYPE	PNG	Alternatives to png format are wmf and bmp
GRAPH SIZE X	600	width of graphs in pixels
GRAPH SIZE Y	300	height of graphs in pixels
GRAPH SYMBOL	DEFAULT	The type of symbol used, see <a href="#">Graph options</a>
HISTORY COMMAND PGM	OFF	Add commands from pgm files to history. When off only the name of the pgm is shown
HISTORY COMMENT	<b>ON</b>	Add comments when found in pgm files to history
HISTORY NAME	TEMP	Name of file to use as autosave, when closing the programme. The previous three files will be saved, e.g. as temp3.pgm temp2.pgm temp1.pgm
LANGUAGE	EN	Language file abbreviation
OUTPUT FOLDER	..folder name...	Logfiles and graphs are saved here
<b>OUTPUT NAME</b>	<b>EAOUTPUT.HTM</b>	<b>Name of output files</b>
OUTPUT OPEN	<b>ON</b>	Open logfile automatically
PRINT PREVIEW CM	<b>ON</b>	Use cm for measures when printing
RANDOM SEED	9	Use as seed for random number generation
RANDOM SIMULATIONS	500	Number of simulations
READ DELETED	OFF	Read records marked for deletion
RECODE INTERVAL	TEXT-	Marker to use in value labels with recode
SAVEDATA FIRSTWORD	<b>ON</b>	Use filetype first word. When off Epi6 automatic naming will be used
SHOW COMMAND	<b>ON</b>	show commands after execution
START PAGE	START.HTM	File to show when the programme starts (F8)
VIEWER FONT NAME	VERDANA,COURIER	Must at least contain one proportional and one fixed type font
VIEWER FONT SIZE	10	Font size for output window
WINDOW FONT SIZE	10	Font size for other windows (prompt, variable, history, commands)

## Quelques compléments

- Commande `cls` ou **F12** : pour effacer les résultats et repartir de zéro.
- Commande `clh` pour effacer les commandes précédentes.
- Commande `variables` : liste des variables
- Voir toutes les fonctions et opérateurs en fin de menu après *Help*
- Touches F....

?	Command Reference	F1
	Commands	F2
	Variables	F3
	Command Prompt	F4
📄	Open Program Editor	F5
📄	Browse Data	F6
	History	F7
<hr/>		
	Screen Font	
📄	Default Position	
	Save Current Position	
<hr/>		
✓	Process Toolbar	
	Data Browser	

## Commandes supplémentaires

- Pour transformer une variable quantitative en classes (variable qualitative) avec un codage en # (de 1 à ...) et un label . Deux manières :

```
define classe # → le nombre de dièses soit correspondre au nombre de chiffres de la variable d'origine
recode age to classe by 10
freq classe
```

OU

```
recode age to classe lo-24=1 25-29=2 30-34=3 35-hi=4
freq classe
```

- Pour attribuer un label au nom de la variable

Permet dans les sorties d'avoir un nom plus explicite.

```
label age agemere
```

means age → affichera agemere comme variable

- Pour attribuer un label à un codage de variable, ou un label différent s'il existe

```
labelvalue sexe /1="Garçon" /2="Fille"
```

```
freq sexe → affiche les valeurs du label
```

```
freq sexe /v → n'affiche que les codages de départ 1 et 2
```

```
freq sexe /v1 → affiche les codages 1 et 2 et les labels
```

- Pour obtenir par défaut les % ou les tests statistiques

Ouvrir le fichier `\Epidata\epidatastat.ini` dans l'Explorateur de Windows ou dans Analysis (bouton Edit)

Ajouter le texte suivant :

```
set option freq = "/c /cum" → affiche les % en colonnes par défaut et les % cumulés (avec en plus /m : affiche les manquants
```

```
set option means = "/t" → affiche le "p" test de t ou le p d'ANOVA par défaut
```

```
set option tables = "/r /t /m /rr /ex" → affiche les % en lignes dans les tableaux, le test du  $\chi^2$  et du Fisher, les manquants et le risque relatif par défaut.
```

- Sauver le fichier avec de nouvelles variables créées

```
Savedata "nom et emplacement du fichier en *.rec"
```

Si l'opération est re-itérée, ajouter `/replace`

Attention, cette opération crée un fichier de contrôle en `*.chk` qui peut gêner l'ouverture de ce fichier ensuite.

Faire avant de quitter `erase "nom du fichier en *.chk"`



- Quelques aides pour les graphes → préférer, si on en a l'habitude, les graphes de Word ou de Open Writer

- Si on n'a pas l'habitude, préférer les boîtes de dialogues en cliquant sur **Graph**. Noter qu'en cochant « Editer » à l'onglet n°2 (Graph/Axis), le graphe peut être modifier secondairement (mais çà demande du temps..)

- Dans la barre des commandes pour faire des graphes

= Avant de se lancer, préciser :

set graph footnote="" pour enlever le sous-titre "Epidata Analysis"

set graph colour ="012" pour avoir les trois premières couleurs des barres en rouge, bleu et noir

Ou même set graph colour = "0123456789" pour les avoir en 9 couleurs différentes. Par exemple : 6 = bleu et 9=bleu clair

= Puis écrire : histogram ou boxplot ou bar ou pie (c'est notre camembert..) pour les principaux.

= On peut ajouter une multitude de sous-commandes à suivre sur la même ligne :

/ti="" → enlève le titre du graphe qui ne sert à rien le plus souvent

/xtext ="Titre de l'axe en x"

/ytext ="Titre de l'axe en y"

/PCT pour les axes des Y en %

/width=5 pour la largeur des classes (ou/xinc=5, ou /bins=5 pour le nombre de classes) avec histogram

/by=classe pour faire des barres ou des lignes par codage

/edit pour ouvrir une boîte de dialogue (Graph Form) dans laquelle on peu faire beaucoup de choses (un peu compliquées quand même) pour... faire des graphes simples et dépouillés...

*Exemple de graphe en histogramme*

histogram age /by=sexe /edit /ti="" /xtext="Age en années" /ytext="%" /PCT /width=5 /legend

Dans la fenêtre "Graph Form" : bouton **Edit All**

Pour modifier la légende

Chart - Series – Changer les noms avec Title

Legend  Visible

Position : bottom

Format : changer Color – Frame – Pattern

Pour afficher les valeurs en haut des barres (pour chaque codage)

Series – Marks Style :  Visible Style (Percent)

Format : changer Color – Frame – Pattern

General : Format #0 %

*Exemple de graphe en camembert*

pie prof /ti="" /edit

Dans la fenêtre "Graph Form" : bouton **Edit All**

Pour avoir en face de chaque secteur, le codage et le % :

Series – Marks – Style :  Visible  Multiline – Style (Label et Percent)

Format : changer Color – Frame – Pattern

General – Format #0 %

*Exemple de graphes en barre (en vertical; pour horizontal, voir Chart – Series – Change)*

bar prof /ti="" /edit /xtext="CSP de la mère" /ytext="%" /PCT

Dans la fenêtre "Graph Form" : bouton **Edit All**

Pour avoir les % en haut de chaque barre (appelés "étiquettes de données" dans Word ou Excel

Series – Marks Style :  Visible Style (Percent)

Format : changer Color – Frame – Pattern

General : Format #0 %

*Exemple de graphe de points avec une droite de régression*

regress vary1 varx

→ en déduire l'équation de la droite de régression  $vary2 = A * X + B$  (après define vary2 ###.###)

scatter varx vary1 vary2 /edit

Page d'accueil du module d'édition : chart/series. Cocher et surbriller vary2, cliquer sur **Change** et cliquer sur « Droite ».

Une droite de régression apparaît.

- Pour copier le graphe dans Word ou Powerpoint

Dans la fenêtre "Graph Form" : bouton **Copy** et ctrl-V dans Word.