

*INTRODUCTION AU
LOGICIEL*

Epi Info

Version 3.5

David Moreau

Décembre 2010

INTRODUCTION AU LOGICIEL

EPI INFO V3.5

Décembre 2010

David MOREAU

Statisticien Epidémiologiste

Paris, France

Email : david.moreau@epivf.fr

Site Internet : www.epivf.fr

SOMMAIRE

| | |
|---|-----------|
| INTRODUCTION | 3 |
| I. Pour LANCER EPI INFO | 4 |
| II. CREATION D'UN QUESTIONNAIRE AVEC CREER QUESTIONNAIRE | 6 |
| Questionnaires et masques de saisie dans Epi Info v3.5 | 6 |
| Ouverture d'un nouveau questionnaire avec Créer Questionnaire | 7 |
| Création du masque de saisie d'un questionnaire | 9 |
| Les types de champs | 10 |
| Contrôle de la saisie d'un questionnaire | 12 |
| Sauvegarde d'un questionnaire | 21 |
| III. SAISIE de DONNEES AVEC le MODULE SAISIR | 22 |
| Pour lancer le module SAISIR, la fenêtre du module SAISIR | 22 |
| La saisie des données | 24 |
| Les boutons de commande du module SAISIR | 26 |
| IV. ANALYSE des DONNEES AVEC le MODULE ANALYSE | 29 |
| Pour lancer le module ANALYSE, l'écran du module ANALYSE | 29 |
| Les commandes principales du module ANALYSE | 30 |
| Les graphiques | 50 |
| V. Le MODULE STATCALC | 57 |
| Pour lancer le module STATCALC | 57 |
| Calcul du nombre de sujets nécessaire et de la puissance | 58 |
| INSTALLATION d'EPI INFO | 63 |
| Adresses d'Epi Info | 64 |
| ANNEXES | 65 |
| Les Types de variable | 65 |
| L'Analyse des données | 66 |
| Le calcul des statistiques descriptives | 67 |
| Le calcul des risques pour deux variables binaires | 68 |
| Le calcul du Chi-2 pour les variables en catégories | 70 |
| Fluctuation d'échantillonnage et Intervalle de Confiance | 71 |
| Types d'erreur des tests d'hypothèse, Puissance d'un test statistique | 72 |
| Exercices sur la Puissance | 73 |



Epi Info™ est une série de programmes pour Windows 95, 98, NT, 2000 et XP destinée à être utilisée par les professionnels de la santé. Il permet de manipuler des données épidémiologiques sous forme de questionnaire et de réaliser des analyses statistiques. Avec **Epi Info™** et un ordinateur personnel, les médecins, les épidémiologistes, et les autres personnels de santé peuvent rapidement développer un questionnaire, contrôler et personnaliser la saisie des données, et analyser ces données.

Le logiciel **Epi Info™** est produit par la “Division of Surveillance and Epidemiology, Epidemiology Program Office, Centers for Disease Control et Prevention (CDC)”, à Atlanta, aux Etats-Unis, et par l’Organisation Mondiale de la Santé, à Genève, en Suisse.

Epi Info™ est un logiciel gratuit, sans copyright, donc il peut être copié et distribué librement.

Le but de ce document est de vous montrer l’utilisation de base d’**Epi Info™** version 3.5 pour traiter des questionnaires, de manière pratique.

Il y a trois modules principaux dans Epi Info :

- **Créer Questionnaire**, pour créer un questionnaire et pour contrôler la saisie des données;
- **Saisir des Données**, pour la saisie des données;
- **Analyse des Données**, pour analyser les données en calculant des fréquences, des tableaux croisés, des moyennes, des statistiques d’association, en réalisant des graphiques, et aussi des modèles statistiques plus élaborés comme la régression logistique et l’analyse de survie.

I. Pour LANCER EPI INFO

- ☞ Cliquez avec la souris sur le bouton **Démarrer**, en bas à gauche de l'écran de Windows;
- ☞ ouvrez le menu **Programmes**, et choisissez **Epi Info** ;
- ☞ lancez alors le logiciel en cliquant sur **Epi Info**.

Ecran d'accueil d'Epi Info v3.5



Dans l'écran d'accueil d'Epi Info, vous avez de larges boutons pour lancer les modules principaux du logiciel, et pour quitter le programme (bouton **Quitter**).

Sur la gauche, se trouvent les trois modules que nous allons voir dans ce cours **Créer Questionnaire**, **Saisir les Données** et **Analyse Des Données**. Sur la droite se trouvent trois autres boutons très utiles:

- **Créer Les Cartes**, un programme de création de cartes GIS en y insérant des données d'études épidémiologiques;
- **Créer Les Rapports**, un traitement de texte permettant à l'utilisateur de combiner les sorties des modules **Analyse** et **Saisir**, et toute donnée provenant d'une base Access ou SQL, et de les présenter sous forme de rapport;
- **Site Web d'Epi Info**, pour aller directement à la page d'accueil Epi Info du CDC sur Internet.

En haut de la fenêtre, on trouve des menus déroulant:

- **PROGRAMMES**: pour accéder aux modules principaux du logiciel Epi Info;

- **EDITION**: pour modifier l'image de fond ou les boutons de l'écran d'accueil d'Epi Info;
- **PARAMETRES**: pour modifier la langue du logiciel, le chemin d'accès aux fichiers de données... ;
- **UTILITAIRES**: une série de programmes utiles (Statcalc, Data compare, Table to view, Visualize Data...);
- **AIDE**: pour accéder au Manuel complet d'Epi Info, avec des exercices et des présentations interactives (en anglais).

Pour sélectionner un choix du menu, déplacez le curseur avec la souris et appuyez sur le bouton de gauche (cliquez) pour ouvrir un sous-menu, puis allez sur l'option désirée et cliquez à nouveau.

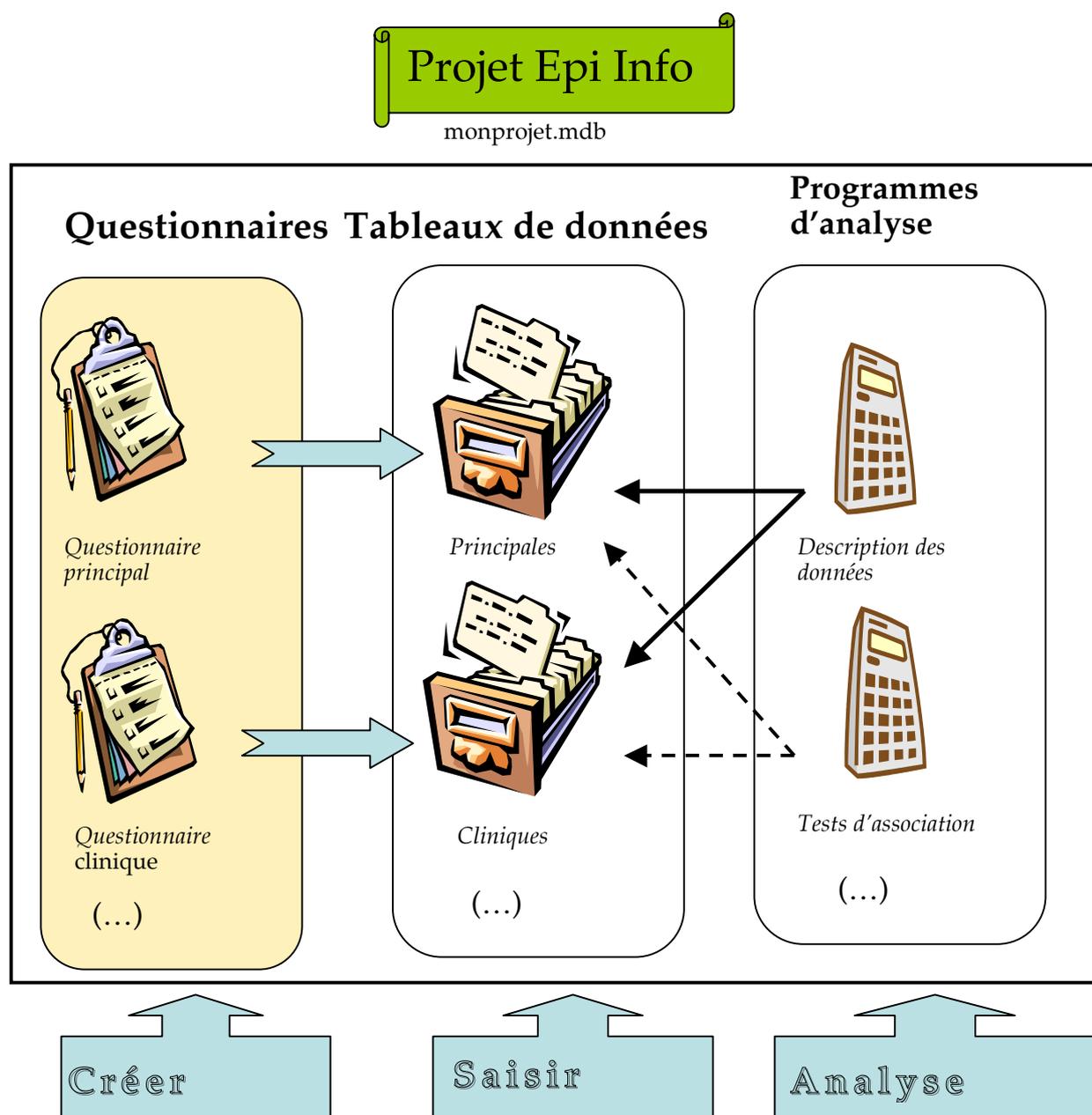
↩ A tout moment, en cliquant sur le petit **X** dans le coin supérieur droit d'une fenêtre, ou en choisissant **Quitter** depuis le menu **Programmes** ou **Fichiers** on quitte le programme en cours.

II. CREATION D'UN QUESTIONNAIRE AVEC LE MODULE CREER QUESTIONNAIRE

Questionnaires et masques de saisie dans Epi Info v3.5

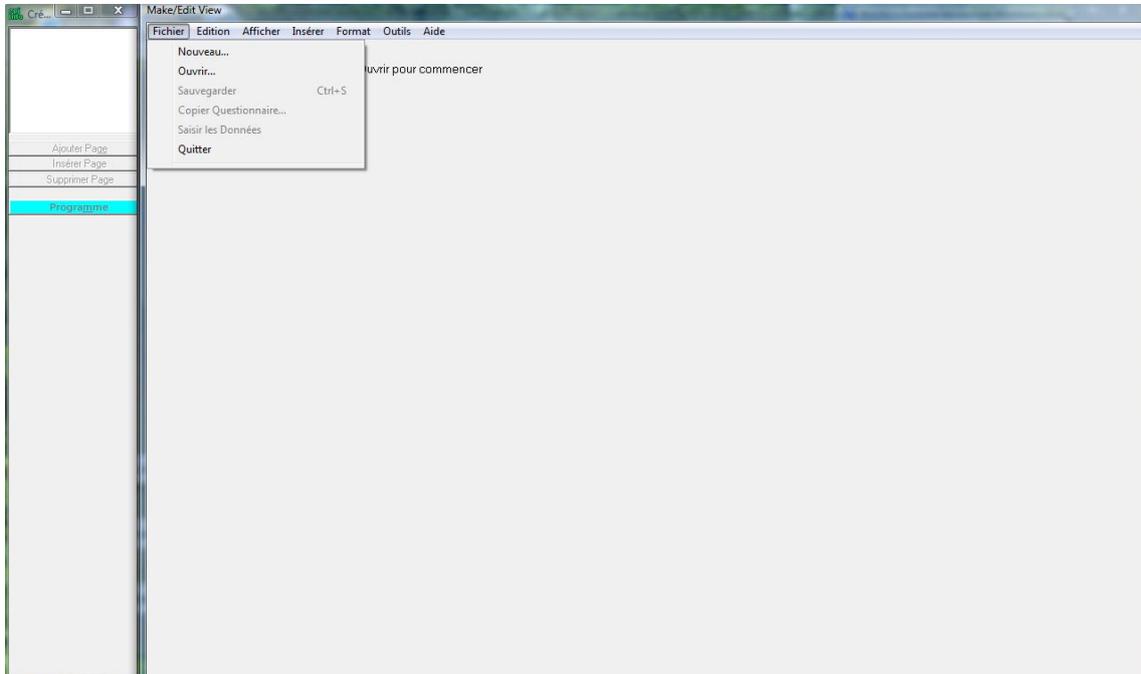
Les questionnaires et les masques de saisie sont appelés des "tables" dans Epi Info v3. Avec le module **Créer Questionnaire** vous créez un "projet" ("project" en anglais) et dans ce "projet", vous créez un ou plusieurs "masques de saisie" (les questionnaires proprement dit). Un projet est un fichier au format MDB (Microsoft DataBase) contenant les questionnaires de votre enquête, les tables de données, et les analyses statistiques que vous appliquez aux données.

Un questionnaire peut avoir autant de pages et de questions que vous voulez.

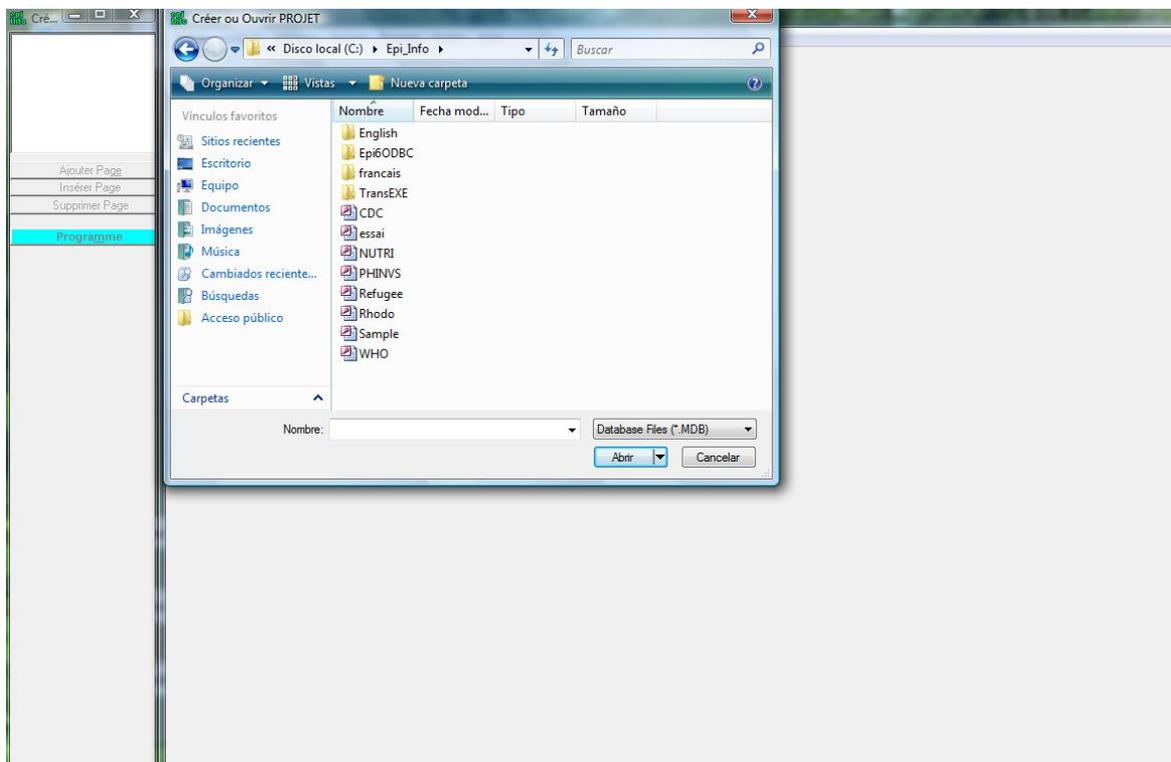


Ouverture d'un nouveau questionnaire avec Créer Questionnaire

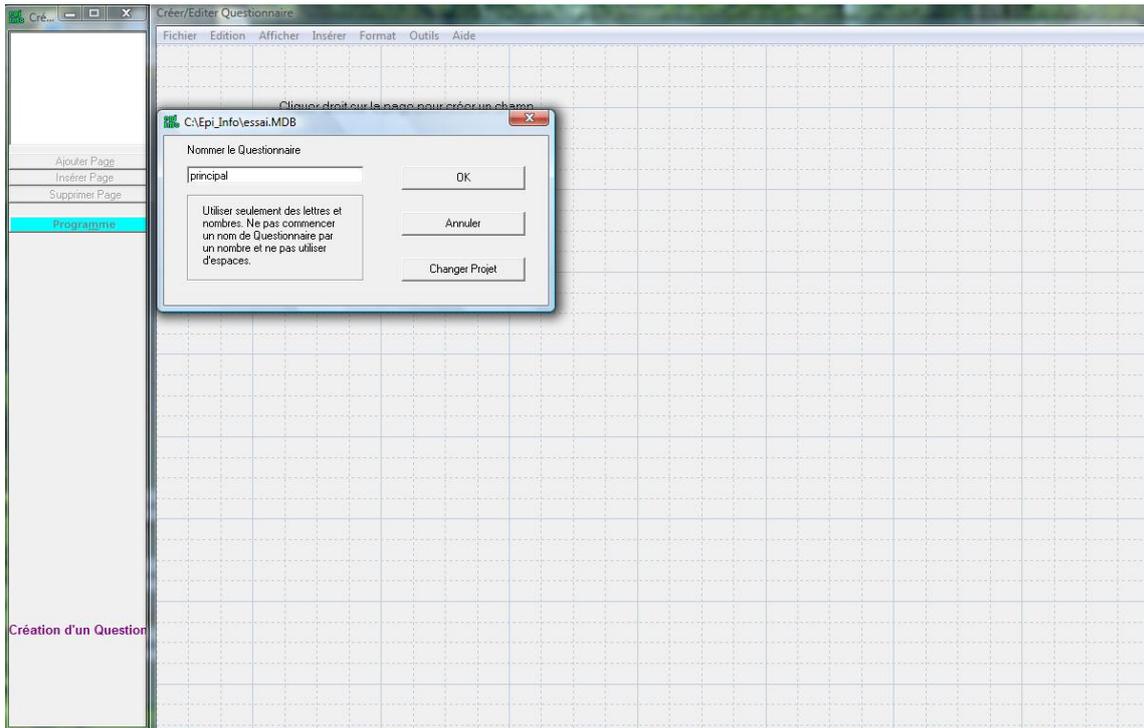
- ✎ Pour lancer le module Créer Questionnaire, cliquez sur le bouton **Créer Questionnaire** de la fenêtre d'accueil Epi Info.
- ✎ Pour ouvrir un nouveau questionnaire (masque de saisie), depuis le menu Fichier, choisissez **Nouveau...**



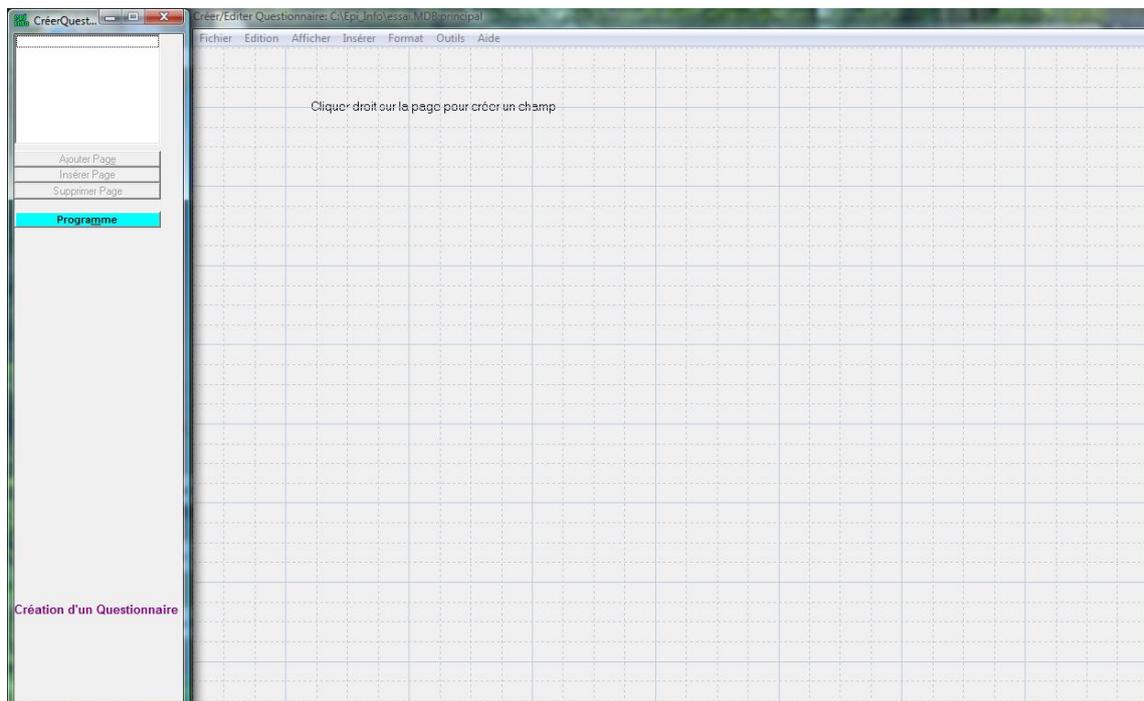
- ✎ La boîte de dialogue "Créer ou Ouvrir PROJET" apparaît; donnez un nom à votre projet (par exemple le nom de votre enquête), puis cliquez sur le bouton **Ouvrir**



⇒ Donnez un nom pour le nouveau questionnaire, et cliquez sur le bouton **OK**.



⇒ Une nouveau questionnaire vide s'affiche avec une grille.



Création du masque de saisie d'un questionnaire

Quand on développe un masque de saisie pour l'utiliser ensuite dans le module **Saisir Les Données**, quelques conventions simples sont nécessaires pour indiquer au programme où créer des champs de saisie de données et quel type de données accepter à ces endroits. Contrairement à de nombreux logiciels de saisie, Epi Info crée des noms de champ automatiquement à partir du questionnaire original. Au niveau d'un champ de saisie, il cherche du texte pour servir de base au nom. Les premiers caractères du texte, hors ponctuation, deviennent le nom du champ. Dans le module **Analyse**, ces mêmes noms de champ de saisie sont appelés des noms de variable, et servent de nom de référence à la question correspondante.

☞ Voyez l'exemple de questionnaire épidémiologique page suivante.

Pour créer un champ dans une vue, vous devez :

- ☞ placez le curseur sur le grille à l'endroit où vous voulez que le champ apparaisse, cliquez ensuite sur le bouton droit de la souris; une boîte de dialogue s'affiche (voir ci-dessous);
- ☞ ❶ tapez le texte de la question ou un titre;
- ☞ ❷ choisissez le type de champ (cf. page 10) et éventuellement sa taille;
- ☞ ❸ modifiez le nom du champ si vous voulez.
- ☞ Vous pouvez aussi sélectionner des caractéristiques particulières pour contrôler la saisie (cf. page 12).
- ☞ Finalement, cliquez sur le bouton OK.

Boîte de dialogue pour créer un champ de saisie dans une vue

Définition du Champ

Question ou Invite
Avez-vous déjà eu des sifflements dans la poitrine ?

Police pour l'Invite

Champ ou Variable
Type: Oui/Non

Police

Nom du Champ
Double Cliquer dans l'invite pour charger
sifflements

Créer
Grille Questionnaire Relié

Répéter le dernier Intervalle
 Requis
 Lecture Seule
 Son
 Conserver la taille de l'image

Tables de Code
Valeurs Légales
Codes
Commentaire Légale

OK Annuler Aide

Les Types de champ

| | |
|-------------------|--|
| Texte | Tout caractère imprimable peut être saisi dans ce champ |
| Etiquette / Titre | Pour ajouter un titre dans le questionnaire |
| Texte[majuscule] | Champ caractères majuscules, il est semblable au champ texte, mais la saisie sera convertie en majuscule |
| Multi ligne | Pour insérer un commentaire |
| Nombre | Champ numérique, seuls des nombres ou des espaces seront acceptés |
| Numéro Téléphone | Numéro de téléphone sous forme ###-####, ou ###-###-#### |
| Date | Champ date au format américain ou européen, les dates seront contrôlées à la saisie |
| Temps | Heures, minutes secondes |
| Date/Heure | Date plus heure |
| Case à cocher | Pour créer des cases à cocher |
| Oui/Non | Dans un champ " Oui/Non ", seuls Y, N et <ESPACE> ou <Entrée> sont permis |
| Option | Pour choisir entre des options exclusives |
| BoutonCommande | Pour créer un bouton dans le questionnaire |
| Image | Pour insérer une image |
| Miroir | Définit un champ devant apparaître sur plus d'une page |
| Grille | Pour créer une table avec des colonnes et des lignes |
| Relier | Pour relier une vue avec une autre vue |

Module Créer Questionnaire affichant un masque de saisie

The screenshot shows a software window titled "Créer/Editer Questionnaire: C:\Users\David\Documents\Boulot\Cours Epi\enquete.MDB:principal" with a menu bar (Fichier, Edition, Afficher, Insérer, Format, Outils, Aide) and a page indicator (Page:1). The main area displays a questionnaire form on a grid background.

Questionnaire ISAAC

Identification

Numéro d'identification

Date de l'entrevue

Age (en année):

Sexe

Symptômes respiratoires

Avez vous déjà eu des sifflements dans la poitrine ?

Durant les 12 derniers mois, combien de crises de sifflements avez-vous eu ?

Jamais

1 à 6 fois

Plus de 6 fois

Avez vous eu de l'asthme ?

Création d'un champ avec des choix exclusifs (type "option") :

Dans notre exemple de questionnaire d'une enquête respiratoire, la question au sujet des crises de sifflements a trois choix exclusifs.

☞ Dans la fenêtre "**Définition de champ**", après avoir écrit la question, sélectionnez le type "**option**", et cliquez sur OK.

⇒ Une nouvelle boîte de dialogue apparaît alors:

Paramétrage pour Option Variable

Select Text on right or left. Enter the Number of choices and press 'ENTER' or 'RETURN'

Texte Activé Droite Gauche

Nombre des Choix

Jamais

1 à 6 fois

Plus de 6 fois

OK

Annuler

Cases d'Option sont pour de choix mutuelles exclusives. Si plus d'un choix peuvent être créés, utiliser les cases à cocher

☞ Entrez le nombre de choix (3) et appuyez sur la touche **Entrée** ;

☞ écrivez les différents choix dans la liste qui apparaît, en appuyant sur **Entrée** après chaque choix;

☞ à la fin, cliquez sur le bouton **OK**.

Contrôle de la saisie d'un questionnaire

Le contrôle de la saisie facilite cette dernière, et permet d'éviter un certain nombre de fautes de frappe. Il définit les valeurs légales, les sauts conditionnels entre des questions. Remarquez que le contrôle de la saisie par le programme est très utile mais n'est pas obligatoire.

Saisie obligatoire

Pour éviter d'avoir des valeurs manquantes, dans la fenêtre "Définition de champ", pour les questions dont vous voulez toujours avoir une réponse, cochez l'option **Requis**.

- ☞ Cliquez deux fois sur le champ "Numéro d'Identification";
- ☞ dans la boîte de dialogue "Définition de champ", cochez l'option **Requis** ;

Définition du Champ

Question ou Invite
Numéro d'identification

Police pour l'Invite

Champ ou Variable
Type: Nombre:
Modèle: ###
Police

Nom du Champ
Double Cliquer dans l'invite pour changer
Identification

Créer
Grille Questionnaire Relié

Répéter le dernier Intervalle
 Requis
 Lecture Seule
 Son
 Conserver la taille de l'image

Tables de Code
Valeurs Légales
Codes
Commentaire Légale

OK Annuler Aide

- ☞ Pour finir, cliquez sur le bouton **OK**.

Saisie Répétée

Avec l'option **Répéter le Dernier**, au cours de la saisie, le champ sera automatiquement rempli avec la dernière valeur saisie par l'utilisateur. Ceci est très pratique si vous avez une série de questionnaires avec les mêmes réponses pour une question donnée.

- ☞ Cliquez deux fois sur le champ "Date de l'entrevue";
- ☞ dans la boîte de dialogue "Définition de champ", cochez l'option **Repeat Last**;

The image shows a screenshot of the "Définition du Champ" (Field Definition) dialog box in Epi Info. The "Question ou Invite" (Question or Prompt) field contains "Date de l'entrevue". The "Type" (Type) is set to "Date" and the "Modèle" (Model) is "DD-MM-YYYY". The "Nom du Champ" (Field Name) is "Datedelentrevue". The "Répéter le dernier" (Repeat Last) checkbox is checked and circled in red. Other options like "Requis" (Required), "Lecture Seule" (Read Only), "Son" (Sound), and "Conserver la taille de l'image" (Keep image size) are unchecked. The "Créer" (Create) section has "Grille" (Grid) and "Questionnaire Relié" (Linked Questionnaire) buttons. The "Tables de Code" (Code Tables) section has "Valeurs Légales" (Legal Values), "Codes", and "Commentaire Légale" (Legal Comment) buttons. "OK", "Annuler" (Cancel), and "Aide" (Help) buttons are at the bottom.

- ☞ Pour finir, cliquez sur le bouton **OK**.

Intervalle de valeurs numériques

Vous pouvez définir une valeur minimale et maximale pour un champ de saisie numérique particulier. Les valeurs entrées en dehors de ces limites seront rejetées au cours de la saisie.

Pour définir un intervalle de valeurs, dans la fenêtre "Définition de champ" du champ numérique, cochez l'option **Intervalle**, et saisissez les valeurs inférieure et supérieure.

- ☞ Cliquez deux fois sur le champ "Age";
- ☞ dans la boîte de dialogue "Définition de champ", cochez l'option **Intervalle**;
- ☞ saisissez les valeurs inférieure et supérieure dans les champs qui apparaissent, en appuyant sur la touche **Entrée** après chacun;

The screenshot shows the "Définition du Champ" dialog box. The "Question ou Invite" field contains "Age (en années)". The "Type" is set to "Nombre" and the "Modèle" is "##". The "Intervalle" checkbox is checked and circled in red. The "Inferieur" field contains "10" and the "Supérieur" field contains "60". The "Nom du Champ" is "Ageenannées".

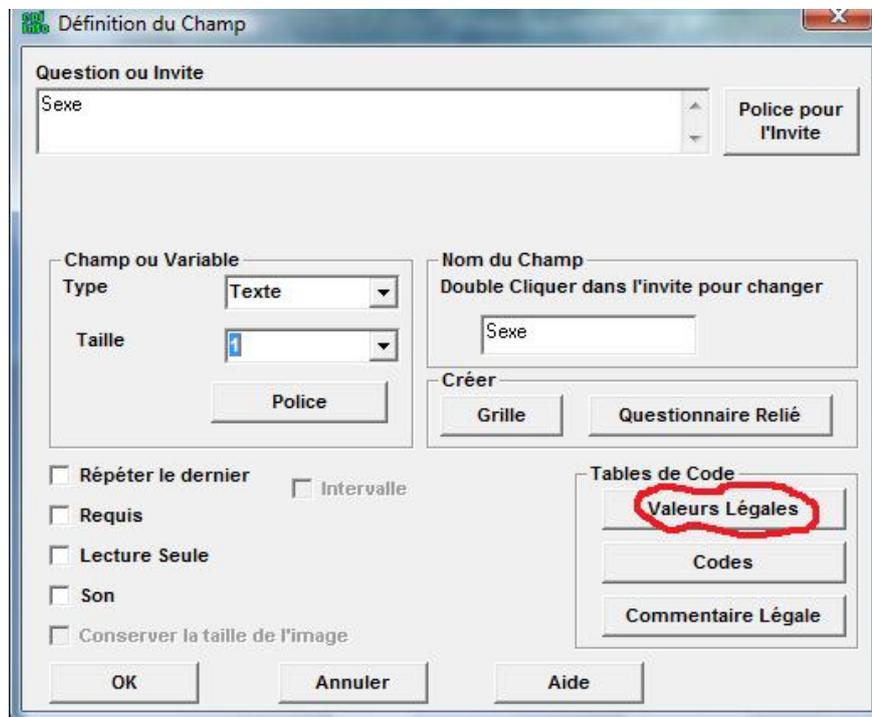
- ☞ Pour finir, cliquez sur le bouton **OK**.

Valeurs légales

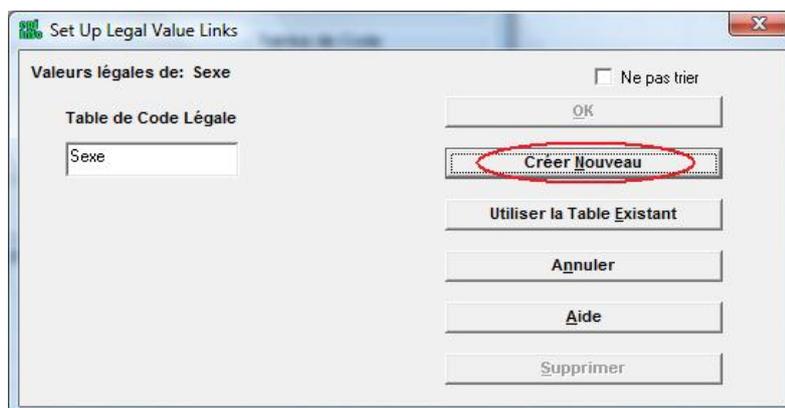
Les valeurs légales sont celles qui sont acceptées à la saisie dans un champ donné. Tout autre valeur sera rejetée.

☞ Cliquez deux fois sur le champ "Sexe";

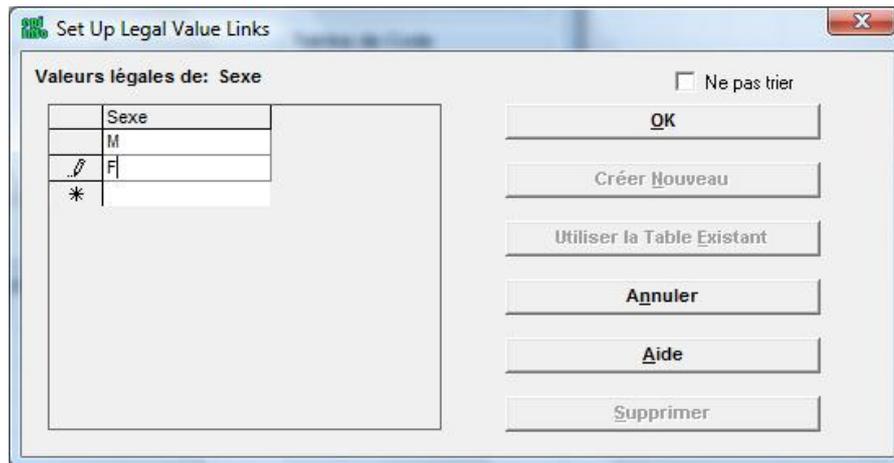
☞ dans la boîte de dialogue "Définition de champ", cliquez sur le bouton **Valeurs légales**;



☞ dans la nouvelle boîte de dialogue qui apparaît, cliquez sur le bouton **Créer Nouveau** ;



☞ saisissez alors les valeurs appropriées dans la liste qui apparaît, en appuyant sur **Entrée** entre chaque.



☞ pour finir, cliquez sur le bouton **OK**, vous retournez sur la boîte de dialogue "Définition de champ", cliquez alors à nouveau sur le bouton **OK**.

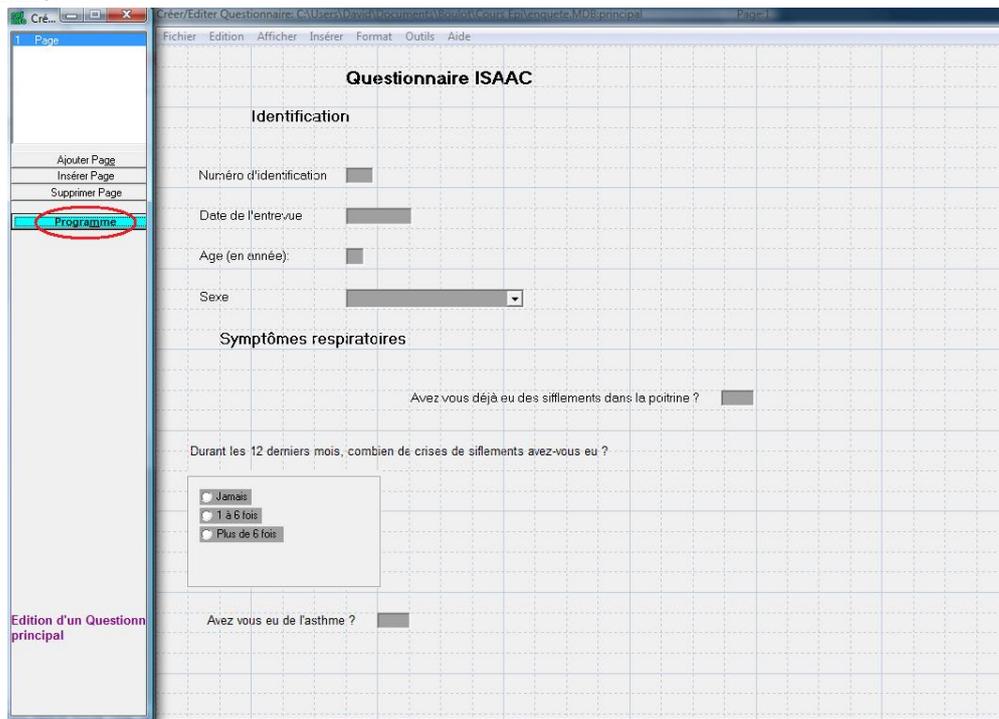
Vous avez ainsi défini de nouvelles valeurs légales.

Saut Conditionnel

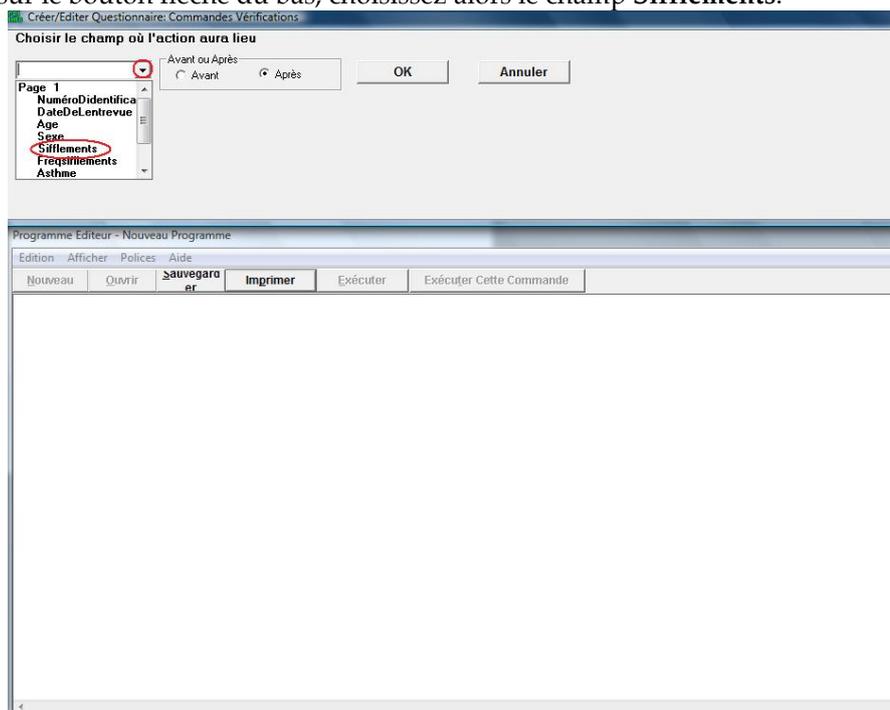
Souvent, les questionnaires contiennent des sections qui doivent être sautées, suivant la réponse donnée à une question précédente. Vous pouvez définir plusieurs valeurs avec différents sauts conditionnels pour un même champ de saisie.

Dans le questionnaire respiratoire, si la réponse à la question sur les sifflements est Non, vous pouvez sauter la question sur la fréquence des crises.

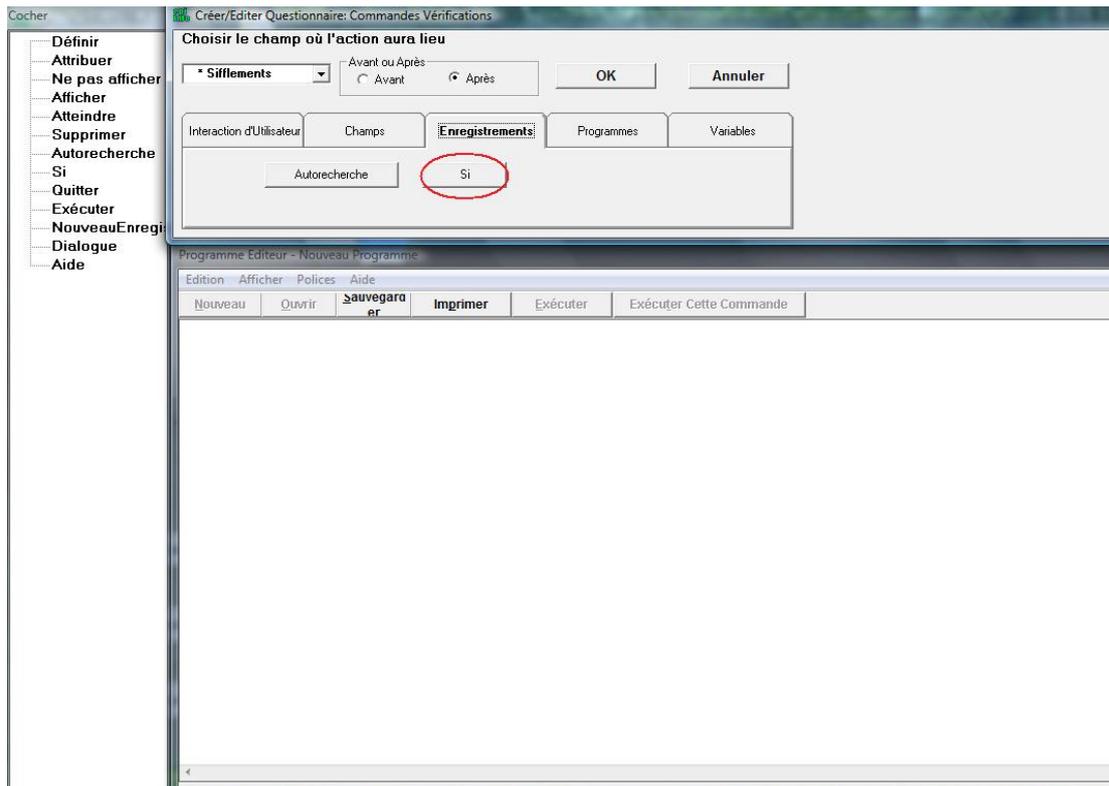
Sur la gauche de la fenêtre "Créer Questionnaire" affichant le questionnaire, cliquez sur le bouton **Programme**.



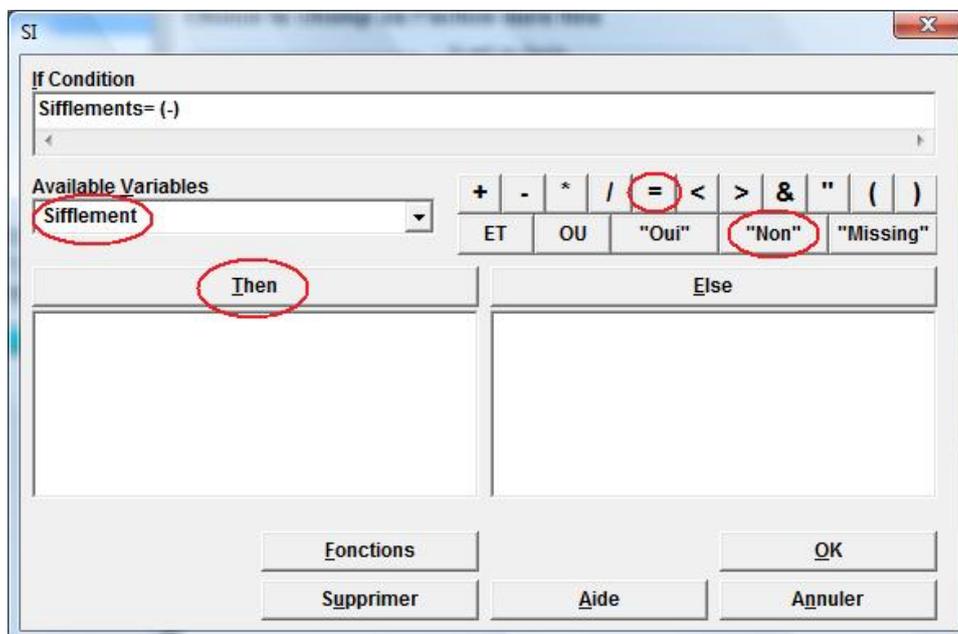
Dans la fenêtre de dialogue "Commandes Vérifications" qui surgit, faites apparaître la liste des champs en cliquant sur le bouton flèche du bas, choisissez alors le champ **Sifflements**.



☞ Dans la rubrique "Enregistrements", choisissez la commande **Si** en cliquant dessus.



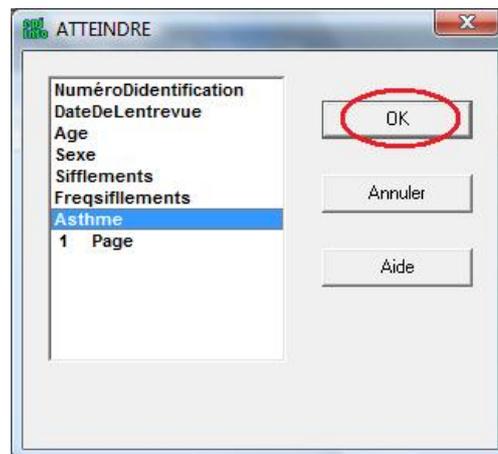
☞ Dans le champ "If Condition" de la boîte de dialogue "SI", sélectionnez à nouveau, depuis la liste déroulante "Available Variables", le champ **Sifflements** et avec les boutons de droite sélectionnez **=** et **"Non"**; cliquez enfin sur le bouton **Then** .



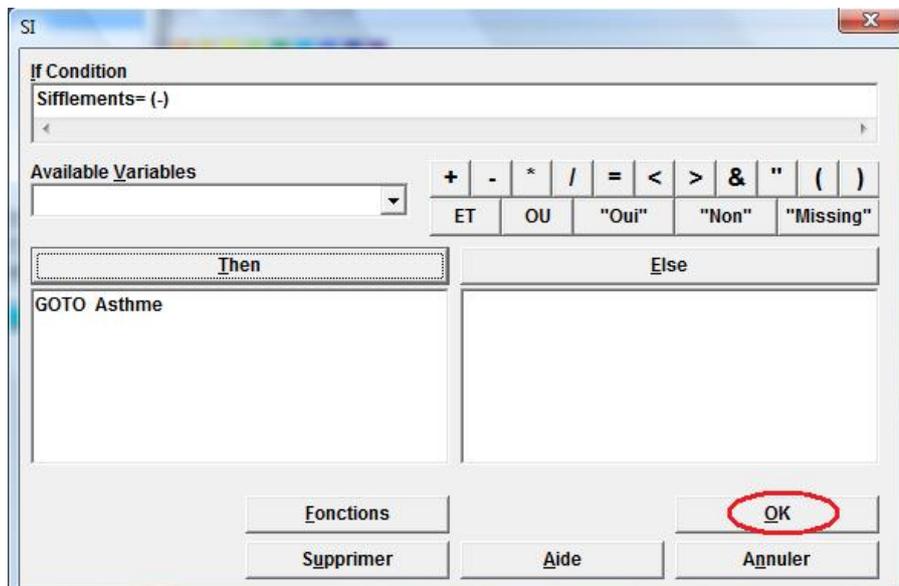
☞ Maintenant dans la fenêtre de dialogue "Commandes Vérifications", dans la rubrique **Fields**, cliquez sur le bouton **Atteindre**.



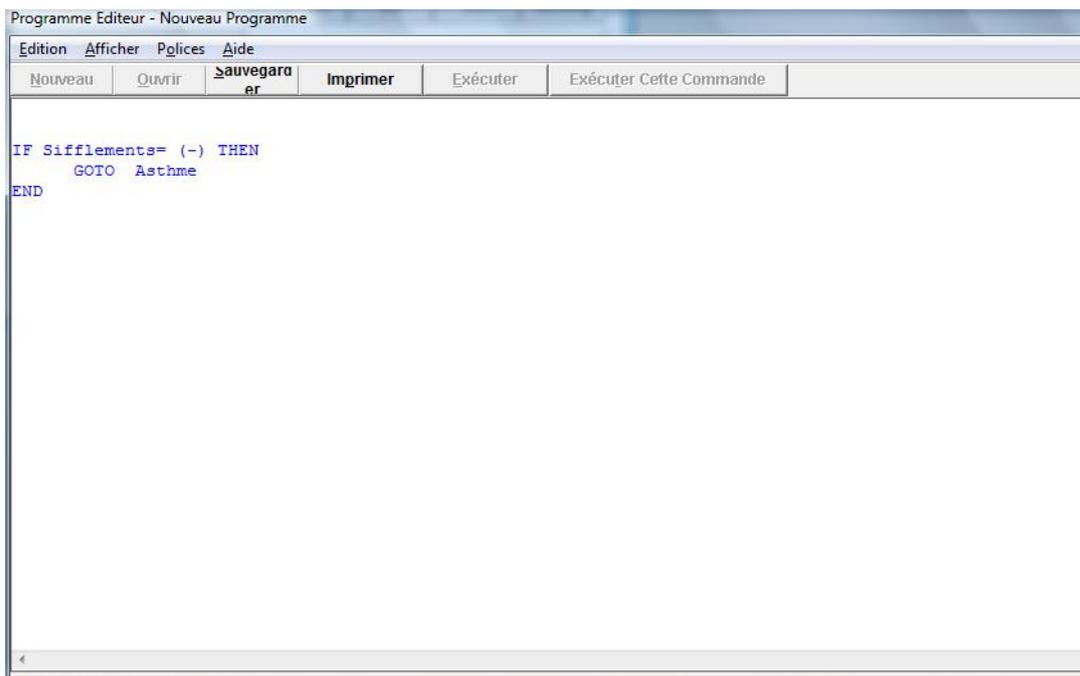
☞ Dans la fenêtre de dialogue "ATTEINDRE", sélectionnez le champ **Asthme**, et cliquez sur le bouton **OK**.



☞ Dans la fenêtre de dialogue "SI" sur laquelle vous revenez, cliquez à nouveau sur le bouton **OK**.



☞ Dans la fenêtre "Programme Editeur", vous devez voir les commandes du saut conditionnel qui apparaissent:



The screenshot shows a window titled "Programme Editeur - Nouveau Programme". The menu bar includes "Edition", "Afficher", "Polices", and "Aide". The toolbar contains buttons for "Nouveau", "Ouvrir", "Sauvegarder", "Imprimer", "Exécuter", and "Exécuter Cette Commande". The main text area contains the following code:

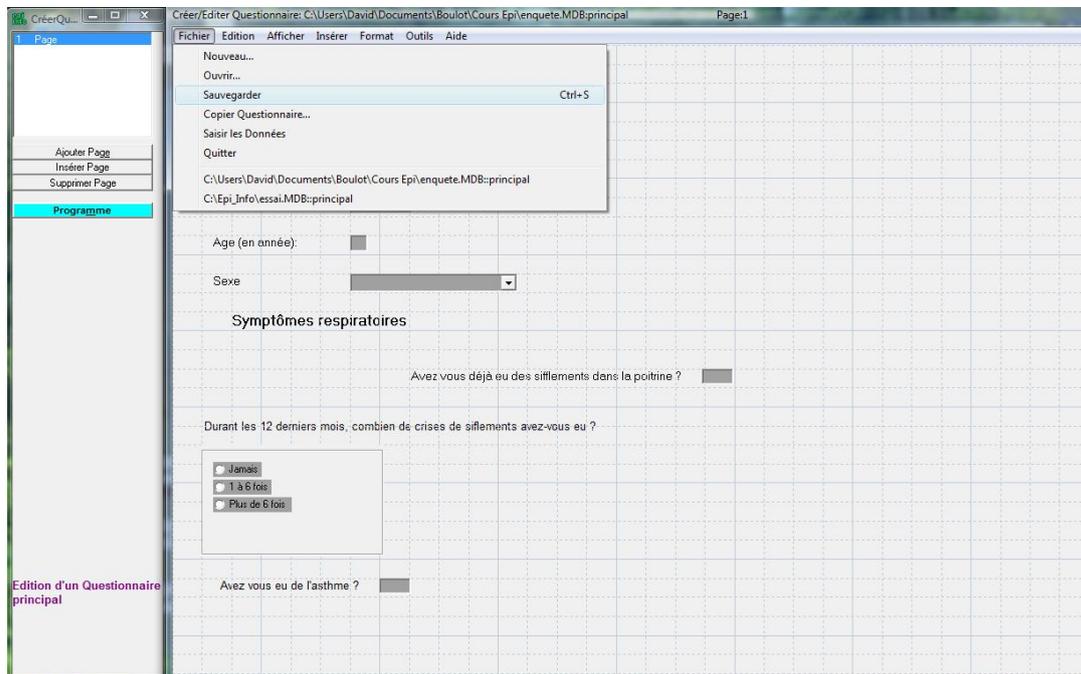
```
IF Sifflements= (-) THEN
  GOTO Asthme
END
```

☞ Pour finir, dans la fenêtre de dialogue "Commandes Vérifications", cliquez sur le bouton **OK**.

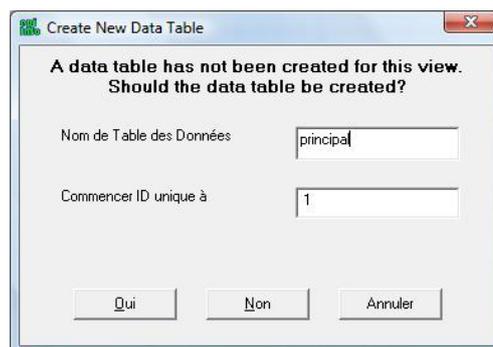
Vous venez de définir un saut conditionnel du champ "Sifflements" au champ "Asthme".

Sauvegarde du masque de saisie du questionnaire

- ✎ Quand vous avez fini d'écrire le masque de saisie du questionnaire, pour enregistrer le fichier, dans le menu **Fichier**, choisissez **Sauvegarder**.



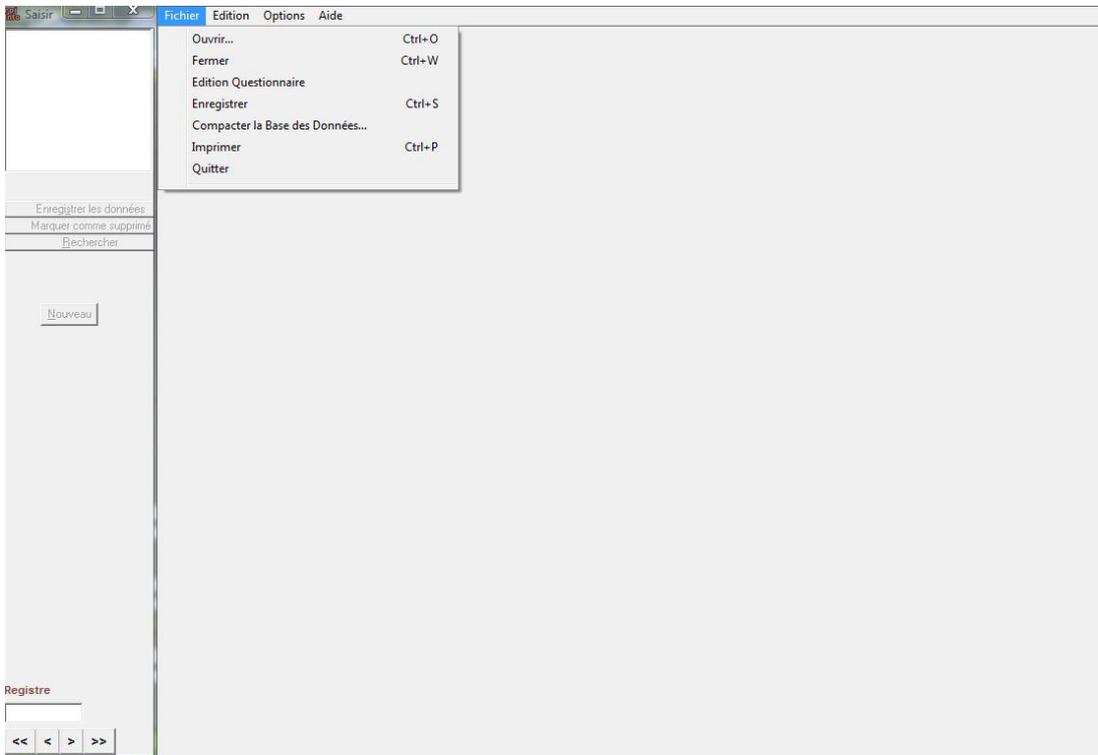
- ✎ En cliquant sur le petit X dans le coin supérieur droit de la fenêtre, ou en choisissant **Quitter** dans le menu **Fichiers**, on quitte le module "Créer Questionnaire".
- ✎ Lorsque l'on quitte le module "Créer Questionnaire" la première fois, une nouvelle boîte de dialogue apparaît avec le nom de la table et le premier numéro d'identification séquentiel des enregistrements, attribué automatiquement par le programme; cliquez juste sur **Oui**.



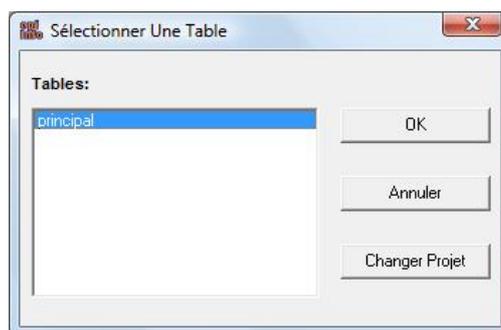
III. SAISIE de DONNEES AVEC le MODULE SAISIR

Pour lancer le module SAISIR

- ☞ Depuis l'écran d'accueil d'Epi Info, cliquez sur le bouton **Saisir Les données**;
- ☞ Pour ouvrir une vue (un questionnaire), dans le menu Fichiers, choisissez **Ouvrir...**, et sélectionnez le nom du projet où se trouve la vue;



- ☞ ensuite, dans la boîte de dialogue "Sélectionner une Table", sélectionnez le nom de la vue, et cliquez sur **OK**.



- ⇒ La vue apparaîtra à l'écran, prête pour l'ajout ou l'édition de nouveaux enregistrements (voir page suivante).

La fenêtre du module SAISIR

The screenshot shows a software window titled "Saisir" with a menu bar (Fichier, Edition, Options, Aide) and a toolbar (Enregistrer les données, Marquer comme supprimé, Rechercher, Nouveau). The main area displays the "Questionnaire ISAAC" form. The "Identification" section includes fields for "Numéro d'identification", "Date de l'entrevue", "Age (en année)", and "Sexe". The "Symptômes respiratoires" section includes a checkbox for "Avez vous déjà eu des sifflements dans la poitrine ?" and a radio button group for "Durant les 12 derniers mois, combien de crises de sifflements avez-vous eu ?" with options "Jamais", "1 à 6 fois", and "Plus de 6 fois". Below this is another checkbox for "Avez vous eu de l'asthme ?". At the bottom left, a "Registre" section shows a list of records with a cursor on record "8" and navigation buttons (<<, <, >, >>). The text "Doit Saisir" is visible at the bottom of the form area.

Au bas de la fenêtre du masque de saisie, les informations de contrôle apparaissent pour le champ contenant le curseur (voir au-dessus, pour le champ "Numéro d'Identification", "Doit Saisir" apparaît).

A côté de la fenêtre du masque de saisie, la fenêtre de gauche indique en haut, le numéro de la page courante, et en dessous des boutons de commande (voir page 26).

En bas de cette fenêtre, vous pouvez voir le numéro de l'enregistrement courant, et au-dessous, des boutons pour se déplacer d'un enregistrement à l'autre :

⏪ : aller au premier enregistrement

⏩ : aller au dernier enregistrement

⏴ : aller à l'enregistrement précédent

⏵ : aller à l'enregistrement suivant

La saisie des données

Saisir des données veut dire entrer les réponses appropriées dans les espaces prévus à cet effet sur l'écran. Le curseur se déplacera automatiquement d'un espace à l'autre.

Chaque espace de saisie est appelé un champ, et chaque exemplaire du questionnaire qui est saisi est appelé un enregistrement. Les enregistrements sont sauvegardés automatiquement dans le projet.

- ✓ En appuyant sur la touche **Entrée** dans un champ de saisie vide, sans taper de données, on crée une valeur manquante.
- ✓ Quand un champ de saisie est rempli, le curseur se déplace automatiquement vers le champ suivant.
- ✓ Les erreurs sont signalées par une boîte de dialogue qui apparaît, et elles peuvent donc être corrigées tout de suite.

☞ Exemple, si vous saisissez une valeur en dehors des limites prédéfinies dans le champ Age, un message apparaîtra vous rappelant l'intervalle de valeurs autorisé.

The screenshot displays the 'Questionnaire ISAAC' software interface. The main window has a menu bar with 'Fichiers', 'Edit', 'Options', and 'Aide'. Below the menu is a toolbar with buttons for 'Sauver les données', 'Marquer l'effacé', 'Chercher', and 'Nouveau'. The main area is divided into sections: 'Identification' with fields for 'Numéro d'identification' (0001), 'Date de l'entrevue' (20-10-2003), 'Age (en années)', and 'Sexe'; and 'Sympt' with a question 'Avez-vous déjà eu des sifflements d...' and radio buttons for 'Jamais', '1 à 6 fois', and 'Plus de 6 fois'. A dialog box titled 'Enter' is open, displaying the error message: 'La valeur n'est pas dans la plage spécifiée. Minimum range:10 Maximum range:60' and an 'OK' button. At the bottom left, there is an 'Enreg.' section with a '1' in a box and a 'Nouveau' button, and navigation buttons '<<', '<', '>', and '>>'. The page number 'Page:10 - 60' is visible at the bottom right.

- ✓ Dans les champs de saisie avec des valeurs légales prédéfinies avec le module Make View, un menu déroulant permet d'accéder avec la souris à ces valeurs.
- ☞ Dans le champ Sexe de notre exemple de questionnaire, en cliquant sur le bouton flèche du bas, apparaissent les deux valeurs prédéfinies "F" et "M".

- ✓ Une fois que le dernier champ de saisie du questionnaire est rempli, le programme va à l'enregistrement vide suivant.
- ✓ A tout moment, vous pouvez aller sur un enregistrement déjà saisi pour le modifier.

☞ Pour quitter le module "Saisir des données" à tout moment, cliquez sur le petit X dans le coin supérieur droit de la fenêtre, ou choisissez **Quitter** dans le menu **Fichiers**.

Les boutons de commande du module SAISIR

Nouveau Nouvel enregistrement

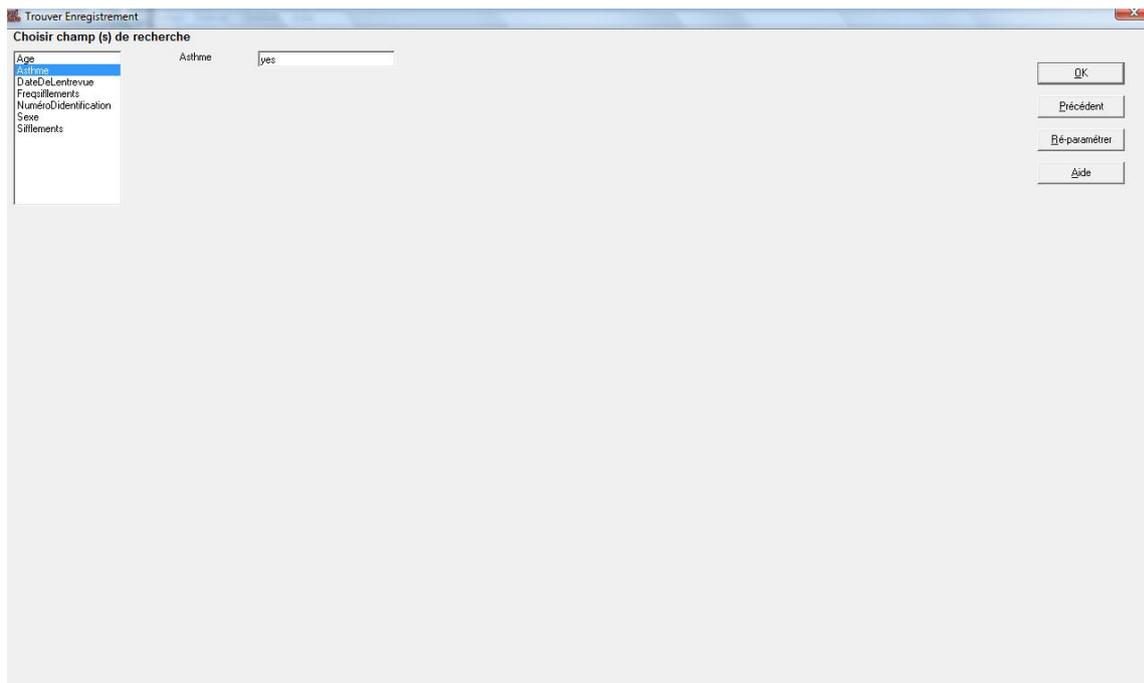
Pour vous déplacer sur le nouvel enregistrement suivant disponible, cliquez sur le bouton **Nouveau**.

Rechercher Recherche d'enregistrements selon des critères définis

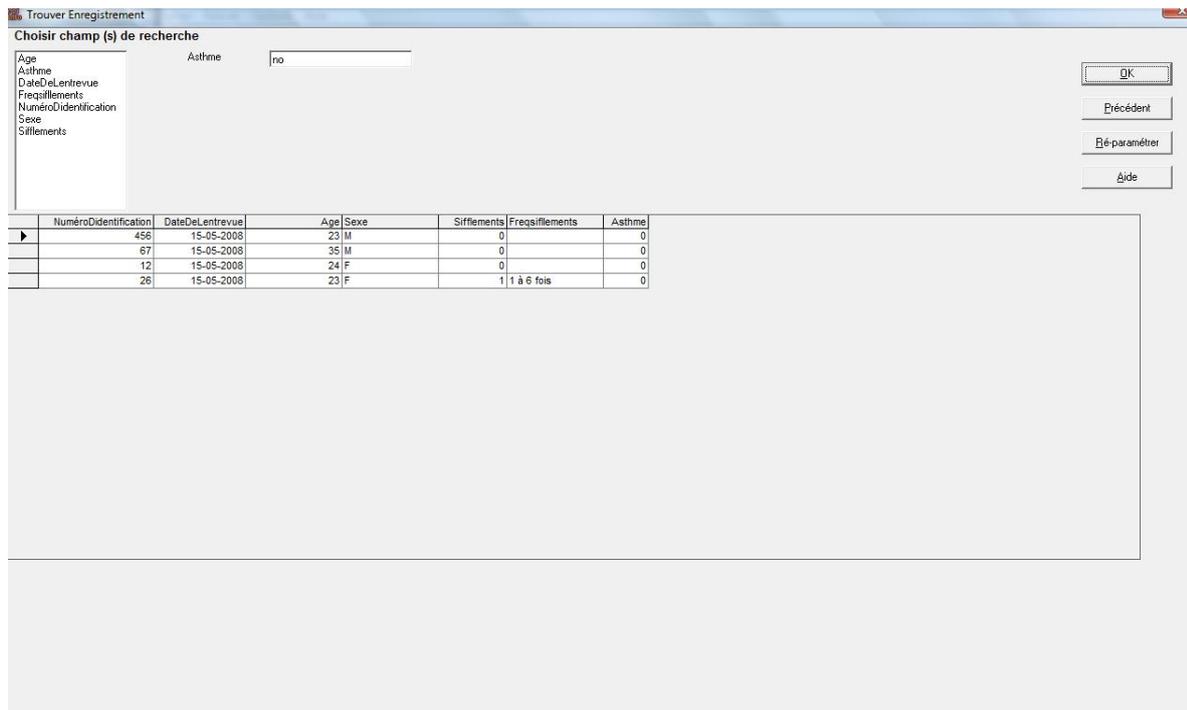
Le module SAISIR vous permet de retrouver des enregistrements selon des critères que vous définissez.

Pour retrouver les enregistrements des asthmatiques dans notre questionnaire respiratoire:

- ☞ cliquez sur le bouton **Rechercher**;
- ☞ dans la boîte de dialogue "Chercher enregistrement", sélectionnez **Asthme** depuis la liste des champs appelée "Sélectionnez champ(s) de";
- ☞ dans le champ de saisie qui apparaît, écrivez "Yes", puis cliquez sur **OK**.



⇒ Les enregistrements qui correspondent au critère de recherche défini, apparaissent alors dans la boîte de dialogue “Trouver Enregistrement” sous forme d’un tableau.



☞ Pour sélectionner un enregistrement, double cliquez sur sa ligne, dans la colonne la plus à gauche. Vous vous retrouverez alors dans la fenêtre masque de saisie affichant l’enregistrement sélectionné.

Si vous cliquez sur le bouton **Rechercher** à nouveau, vous retournez sur la liste des enregistrements sélectionnés. Dans cette boîte de dialogue “Trouver enregistrement”, en cliquant sur le bouton **Ré-paramétrer**, vous réinitialisez les critères de recherche.

Marquer comme supprimé

Suppression d'enregistrements

Ce bouton marque l'enregistrement courant comme supprimé. La mention "Enregistrement Supprimé" apparaîtra au-dessus du numéro de l'enregistrement en bas de la fenêtre Saisir. Un tel enregistrement ne sera pas pris en compte dans le module ANALYSE.

En cliquant sur le bouton **Annuler suppression** (qui remplace le bouton "Marquer comme supprimé"), un enregistrement préalablement supprimé sera à nouveau actif.

Ecran SAISIR affichant un enregistrement supprimé

Questionnaire ISAAC

Identification

Numéro d'identification

Date de l'entrevue

Age (en année):

Sexe

Symptômes respiratoires

Avez vous déjà eu des sifflements dans la poitrine ?

Durant les 12 derniers mois, combien de crises de sifflements avez-vous eu ?

Jamais
 1 à 6 fois
 Plus de 6 fois

Avez vous eu de l'asthme ?

Enregistrement Supprimé
Registre
5 de 8
<< < > >>

Doit Saisir

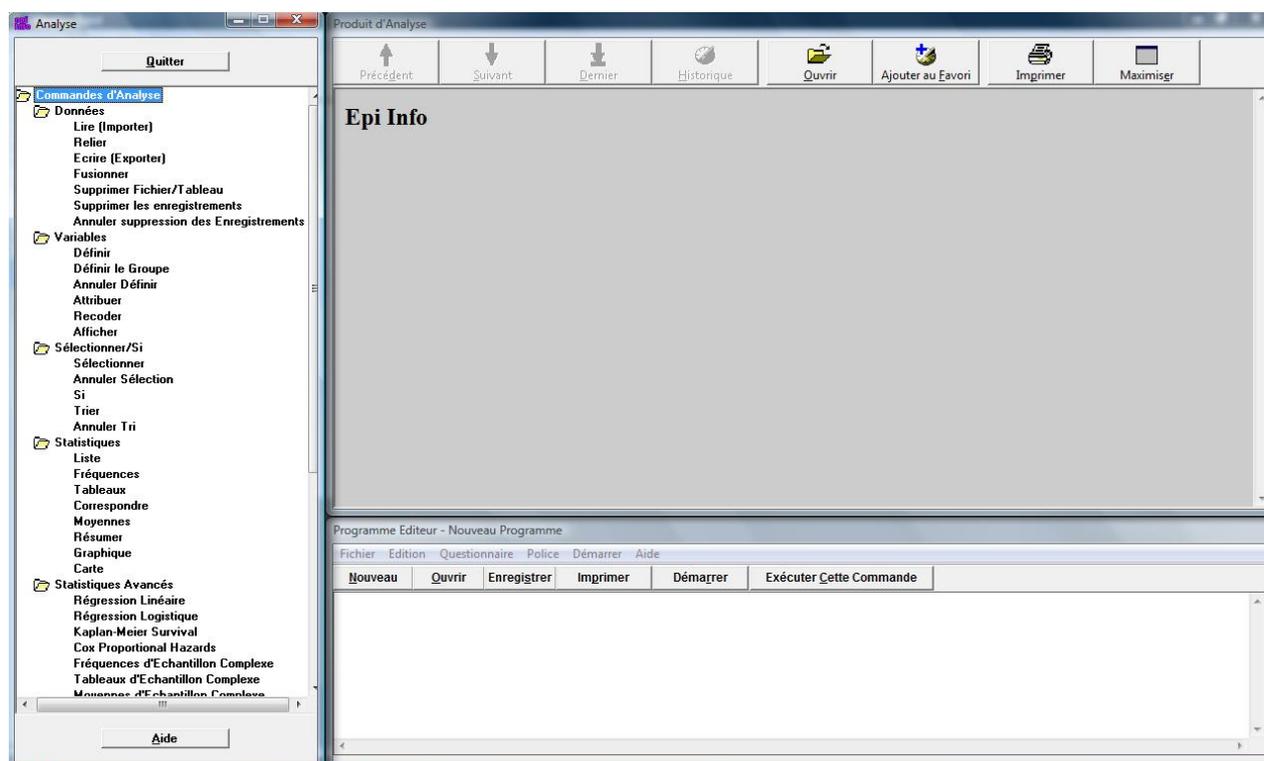
IV. ANALYSE des DONNEES AVEC le MODULE ANALYSE

Le module **ANALYSE Des Données** permet de visualiser les données, de calculer des fréquences, des tableaux croisés, des statistiques, de tester des modèles et de produire différents types de graphiques. Il s'utilise en entrant des commandes. Ces commandes peuvent être tapées une par une à partir du clavier, ou sélectionnées avec la souris, à partir d'une liste.

Pour lancer le module ANALYSE Des DONNEES

☞ Cliquez sur le bouton **Analyse Des Données** dans la fenêtre d'accueil d'Epi Info.

L'écran du module ANALYSE Des DONNEES



L'écran du module ANALYSE Des DONNEES comporte trois fenêtres:

la fenêtre "Analyse" sur la gauche, montrant toutes les commandes disponibles;

la fenêtre "Produit d'Analyse" où les résultats des commandes utilisées vont apparaître;

la fenêtre "Programme Editeur" où les commandes que vous sélectionnez vont s'écrire.

Pour sélectionner des commandes, cliquer dessus dans la fenêtre de gauche, une boîte de dialogue apparaîtra pour compléter la commande qui s'écrira dans l'écran en dessous.

☞ Pour quitter le module ANALYSE à tout moment, cliquez sur le bouton **Quitter** en haut à gauche.

Les commandes principales du module ANALYSE Des DONNEES

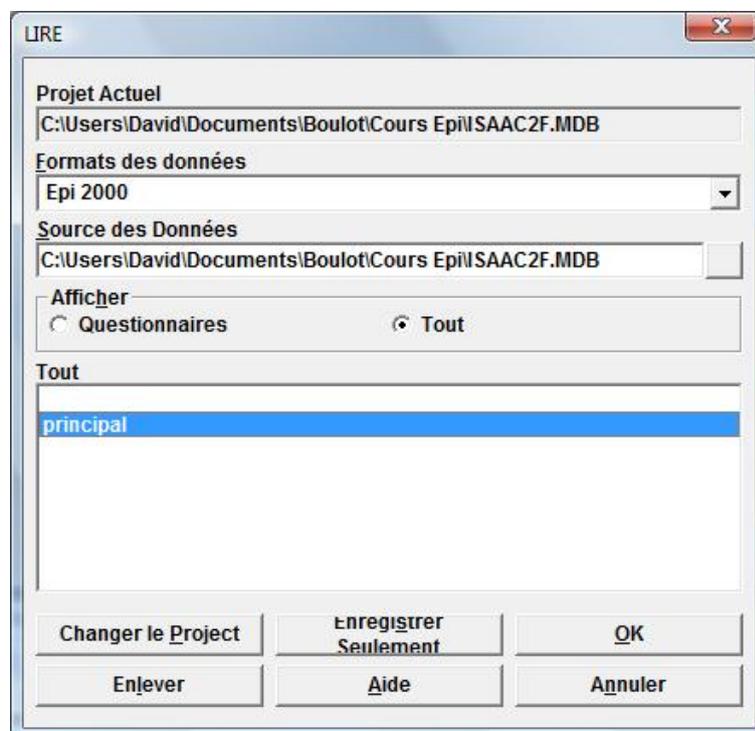
Pour apprendre à utiliser les commandes principales du module d'analyse, nous allons prendre comme exemple un extrait (incomplet, donc les résultats sont fictifs) d'une enquête réalisée chez des enfants scolarisés en France, pour étudier la relation entre des symptômes respiratoires et allergiques et la pollution atmosphérique des salles de classe de leur école. Ces données sont dans le fichier "Isaac2f.mdb" qui peut être téléchargé en allant sur mon site internet dédié à Epi Info à l'adresse : <http://www.epivf.fr>.

LIRE

Tout d'abord, vous avez besoin de sélectionner la base de données sur laquelle vous allez travailler, avec la commande LIRE.

Pour ouvrir le projet appelé "Isaac2f.mdb":

- ☞ Cliquez sur la commande **Lire (Importer)** depuis la fenêtre de gauche, dans le groupe des commandes **■ Données**;
- ☞ dans la boîte de dialogue LIRE, cliquez sur le bouton **Changer le Project** et sélectionnez le projet (ici "Isaac2f.mdb"), dont le nom doit alors apparaître dans le champ appelé **Source des Données**;
- ☞ puis choisissez le format **Epi 2000** depuis le menu déroulant appelé "**Formats des données**";
- ☞ sélectionnez "**Tout**" dans le champ appelé "**Afficher**";
- ☞ choisissez une vue dans la liste en dessous (ici "principal"), et cliquez sur **OK**.



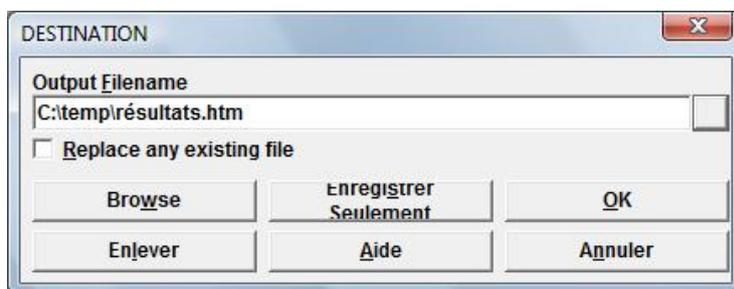
DESTINATION des résultats, FERMER fichier résultats

La commande **Destination** suivie du nom d'un fichier de votre choix, envoie tous les résultats vers ce fichier (de type HTML) et vers l'écran simultanément. Ainsi vous pourrez relire tous les résultats de vos analyses avec n'importe quel éditeur HTML (comme Internet Explorer ou FireFox).

La commande **Fermer** permet de refermer le fichier résultats après l'avoir rempli.

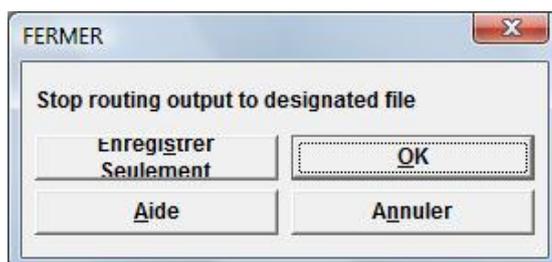
Pour envoyer tous les résultats vers un fichier appelé "resultats.htm", sur le répertoire "c:\temp\" :

- ☞ cliquez sur la commande **Destination** qui se trouve dans la fenêtre de gauche, dans le groupe des commandes ■ Sortie;
- ☞ dans la boîte de dialogue DESTINATION, dans le champ appelé "**Output Filename**", cliquez sur le bouton à droite ;
- ☞ dans la fenêtre de navigation qui apparaît, sélectionner le répertoire dans lequel se trouvera le fichier résultat "c:\temp\", puis tapez le nom du fichier "**résultats.htm**", et cliquez sur le bouton ouvrir.
- ☞ De retour sur la boîte de dialogue DESTINATION, cliquez sur le bouton **OK**.



Pour refermer ensuite le fichier et envoyer les sorties vers l'écran seulement :

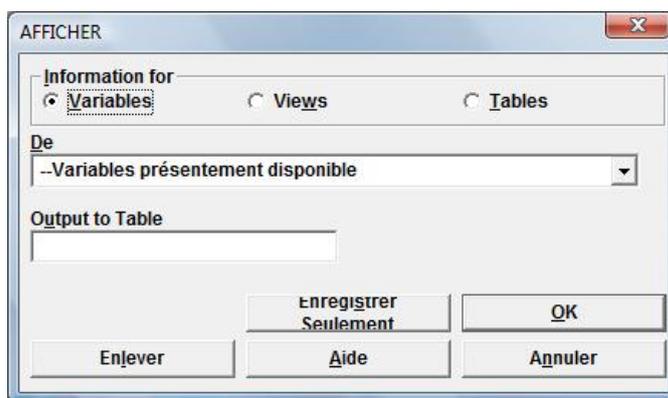
- ☞ cliquez sur la commande **Fermer** qui se trouve dans la fenêtre de gauche, dans le groupe des commandes ■ Sortie;
- ☞ dans la fenêtre de dialogue FERMER qui apparaît, cliquez juste sur le bouton **OK**.



AFFICHER VARIABLES

Une fois que vous avez sélectionné une table de données, la commande **Afficher variables**, avec l'option "Variables actuellement disponibles", donne le nom de la table, et la liste des variables avec leur nom et leur type.

- ☞ Cliquez sur la commande **Afficher** depuis la fenêtre de gauche "Analyse", dans le groupe des commandes **Variables**;
- ☞ dans la boîte de dialogue AFFICHER, sélectionnez "Variables présentement disponibles" dans le menu déroulant appelé "De";
- ☞ cliquez sur **OK**.



⇒ Résultat de la commande **Afficher variables**:

Epi Info

Questionnaire Actuelle: C:\Users\David\Documents\Boulot\Cours Epi\ISAAC2F.MDB:principal
Compte de l'Enregistrement: 1366 Date: 19/10/2010 10:59:53

DISPLAY DBVARIABLES

| Variable | Tableau | Type du champ | Format/Valeur | Info Spéciale | Invite |
|----------------|-----------|---------------|---------------|---------------|----------------|
| age | principal | Nombre | | | age |
| crisessiff | principal | Texte | | | crisessiff |
| dateentrevue | principal | Date | | | dateentrevue |
| eczema | principal | OuiNon | | | eczema |
| identification | principal | Nombre | | | identification |
| PM2_5 | principal | Nombre | | | PM2_5 |
| poids | principal | Nombre | | | poids |
| sexe | principal | Texte | | | sexe |
| sifflements | principal | OuiNon | | | sifflements |
| taille | principal | Nombre | | | taille |
| Language | Defined | Texte | FRANCAIS | Predefined | |

Le résultat se présente sous forme d'un tableau.

Dans la première colonne apparaissent les noms des variables du questionnaire, dans la deuxième le nom du masque de saisie (ici principal) ; puis le type de champ (le type de la variable) ; puis deux colonnes avec des informations complémentaires qui sont vides ici ; enfin apparaît l'invite, soit le texte de la question correspondant à la variable.

Le questionnaire principal du projet "Isaac2f.mdb", ressemble à celui créé dans la partie II "Création d'un questionnaire" de ce document (cf p 10).

On retrouve le nom des variables de la partie identification :

Identification : numéro d'identification ;

Dateentrevue : date de l'entrevue ;

Age : âge des enfants en année ;

Sexe : sexe des enfants.

Le nom des variables de la partie symptômes respiratoires :

Sifflements : avoir des sifflements dans la poitrine ;

crisessiff : la fréquence des sifflements ;

eczema : avoir de l'eczéma.

Trois variables apparaissent en plus par rapport à l'exemple de questionnaire page 10 :

Poids : poids des enfants en kg ;

Taille : taille des enfants en cm ;

PM2_5 : concentration de microparticules de poids moléculaire 2,5 en $\mu\text{g}/\text{m}^3$ dans l'air des salles de classe (c'est une mesure de pollution de l'air).

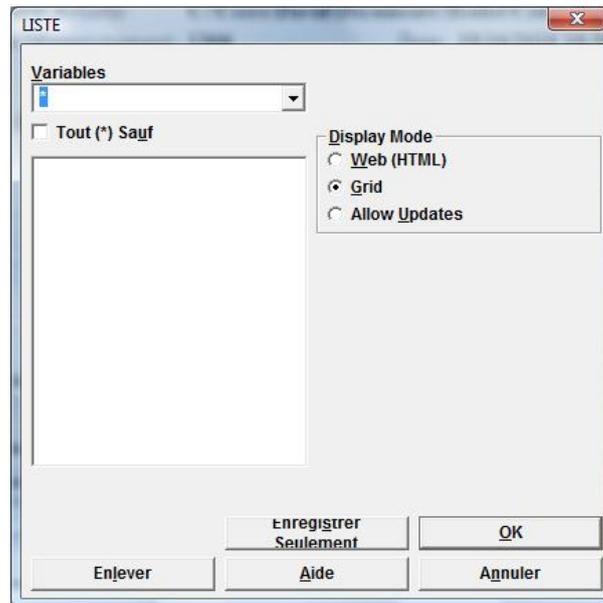
LISTE

La première étape dans une étude, c'est d'examiner les données visuellement, pour avoir une vue d'ensemble et décider qu'elles sont les analyses appropriées à effectuer. Un tableau avec les observations en ligne et les variables en colonne est utile pour cela. La commande **Liste** permet d'obtenir un tel tableau à partir des enregistrements du fichier de données. La commande **Liste** suivie d'un ou plusieurs noms de variable n'affiche le tableau que de ces variables.

☞ Cliquez sur la commande **Liste** depuis la fenêtre "Analyse", dans le groupe des commandes ■ Statistiques;

☞ dans la boîte de dialogue LISTE, l'option **Grid** est sélectionnée dans le champ "Display Mode";

☞ cliquez sur **OK**.



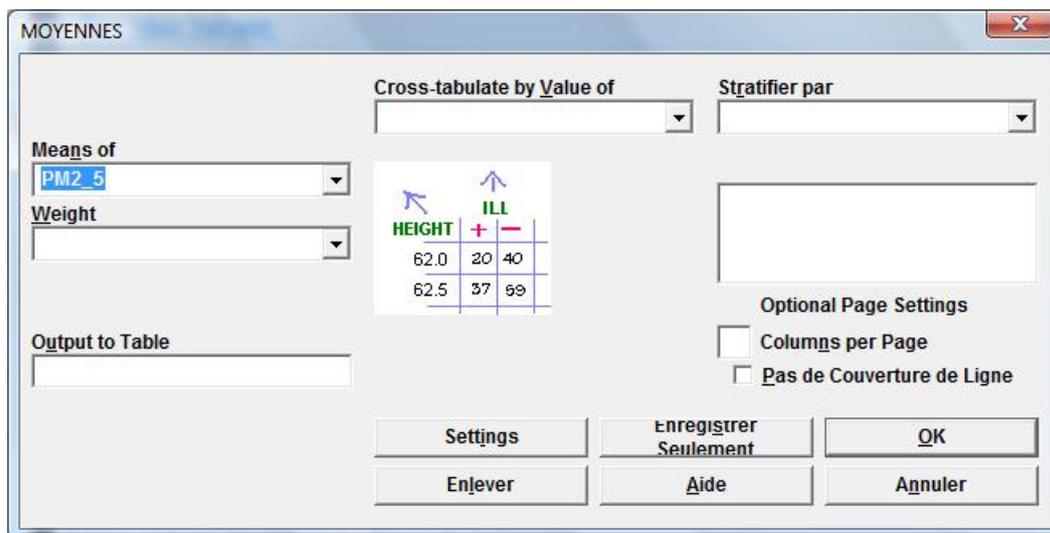
⇒ Le résultat apparaît dans une nouvelle fenêtre:

| identificati | dateentrev | age | sexe | sifflements | crisessiff | eczema | poids | taille | PM2_5 |
|--------------|------------|-----|------|-------------|------------|---------|---------|---------|-------|
| 100001 | 26/10/1999 | 10 | M | Missing | Missing | No | 41 | 1,54 | 23 |
| 100002 | 21/06/1999 | 9 | M | Missing | Missing | No | 37 | 1,43 | 31 |
| 100003 | 26/10/1999 | 10 | M | Missing | Missing | No | 40 | 1,51 | 23 |
| 100004 | 21/06/1999 | 10 | M | Missing | Missing | No | 50 | 1,43 | 31 |
| 100005 | 26/10/1999 | 10 | F | Missing | Missing | No | 48 | 1,5 | 23 |
| 100006 | 26/10/1999 | 11 | M | Missing | Missing | No | 34 | 1,51 | 23 |
| 100007 | 26/10/1999 | 11 | F | Missing | Missing | No | 34 | 1,55 | 23 |
| 100042 | 11/06/1999 | 10 | M | Missing | Missing | No | 45 | 1,44 | 13,9 |
| 100044 | 11/06/1999 | 12 | M | Missing | Missing | No | 34 | 1,43 | 13,9 |
| 100046 | 11/06/1999 | 10 | F | Missing | Missing | No | 30 | 1,45 | 13,9 |
| 100066 | 11/06/1999 | 10 | M | Missing | Missing | Missing | 30 | 1,39 | 11,2 |
| 100068 | 11/06/1999 | 10 | F | Missing | Missing | No | 30 | 1,38 | 11,2 |
| 100073 | 22/06/1999 | 9 | F | Missing | Missing | No | 37 | 1,43 | 23,5 |
| 100075 | 22/06/1999 | 9 | M | Missing | Missing | No | 40 | 1,46 | 23,5 |
| 100077 | 11/06/1999 | 10 | F | Missing | Missing | No | 50 | 1,49 | 11,2 |
| 100079 | 11/06/1999 | 10 | M | Missing | Missing | No | 37 | 1,4 | 11,2 |
| 100080 | 11/06/1999 | 10 | F | Missing | Missing | No | 46 | 1,49 | 11,2 |
| 100105 | 25/06/1999 | 10 | M | Missing | Missing | Yes | 26 | 1,29 | 13,9 |
| 100107 | 25/06/1999 | 9 | M | Missing | Missing | Yes | 44 | 1,34 | 13,9 |
| 100108 | 25/06/1999 | 10 | F | Missing | Missing | No | 32 | 1,49 | 13,9 |
| 100109 | 25/06/1999 | 10 | F | Missing | Missing | No | 47 | 1,53 | 13,9 |
| 100111 | 25/06/1999 | 9 | F | Missing | Missing | No | 33 | 1,42 | 13,9 |
| 100167 | 04/06/1999 | 11 | F | Missing | Missing | Missing | Missing | Missing | 14,6 |
| 101333 | 21/06/1999 | 11 | F | Missing | Missing | No | 45 | 1,4 | 20 |
| 101513 | 01/06/1999 | 10 | F | No | Missing | No | 33,5 | 1,37 | 12 |
| 101518 | 04/06/1999 | 11 | M | No | Missing | No | 37 | 1,48 | 14,6 |
| 101519 | 04/06/1999 | 10 | F | No | Missing | No | 39 | 1,45 | 14,6 |
| 101520 | 04/06/1999 | 11 | F | No | Missing | Yes | 48 | 1,54 | 14,6 |
| 101521 | 04/06/1999 | 11 | M | Missing | Missing | No | 33 | 1,43 | 14,6 |
| 101537 | 04/06/1999 | 10 | F | Missing | Missing | Missing | 34,6 | 1,51 | 10,9 |
| 101541 | 04/06/1999 | 11 | M | No | Missing | No | 42,6 | 1,54 | 10,9 |
| 101542 | 04/06/1999 | 9 | M | No | Missing | No | 34,6 | 1,42 | 10,9 |
| 101543 | 04/06/1999 | 10 | M | No | Missing | No | 35,4 | 1,4 | 10,9 |
| 101544 | 04/06/1999 | 10 | F | No | Missing | No | 40 | 1,38 | 10,9 |
| 101546 | 12/10/1999 | 8 | M | Missing | Missing | No | 27 | 1,29 | 15,6 |
| 101554 | 01/06/1999 | 9 | F | No | Missing | No | 33 | 1,39 | 12 |
| 101555 | 01/06/1999 | 9 | M | No | Missing | No | 39 | 1,46 | 12 |
| 101556 | 01/06/1999 | 9 | F | No | Missing | No | 27 | 1,4 | 12 |

MOYENNES avec une variable

La commande **Moyennes** donne une description des variables numériques continues, avec les valeurs minimale et maximale, la médiane, les quartiles et le mode, la moyenne et la variance (voir les définitions en annexe page 67).

- ☞ Cliquez sur la commande **Moyennes** depuis la fenêtre "Analyse", dans le groupe des commandes ■ Statistiques;
- ☞ dans la boîte de dialogue MOYENNES, sélectionnez la variable PM2.5 (la concentration en micro particules) de notre exemple de données, depuis le menu déroulant appelé "Means of" (Moyenne de);
- ☞ cliquez sur **OK**.



⇒ Les statistiques apparaissent dans la fenêtre de sortie des résultats:

C:\Cours_Epi\OUT37.htm

| | Précédent | Suivant | Demier | Histoire | Ouvrir | Signet | Imprimer | Agrandir |
|--------------|-----------|---------|--------|----------|--------|--------|----------|----------|
| 29,3 | | 5 | 0,4% | | 96,0% | | | |
| 29,4 | | 5 | 0,4% | | 96,3% | | | |
| 29,6 | | 5 | 0,4% | | 96,7% | | | |
| 29,9 | | 10 | 0,7% | | 97,4% | | | |
| 31 | | 5 | 0,4% | | 97,8% | | | |
| 31,2 | | 5 | 0,4% | | 98,2% | | | |
| 31,4 | | 5 | 0,4% | | 98,5% | | | |
| 33 | | 10 | 0,7% | | 99,3% | | | |
| 35 | | 5 | 0,4% | | 99,6% | | | |
| 70,3 | | 5 | 0,4% | | 100,0% | | | |
| Total | | 1366 | 100,0% | | 100,0% | | | |

Obs Total Moyenne Variance Ecart type
1366 22699,2000 16,6173 49,2394 7,0171

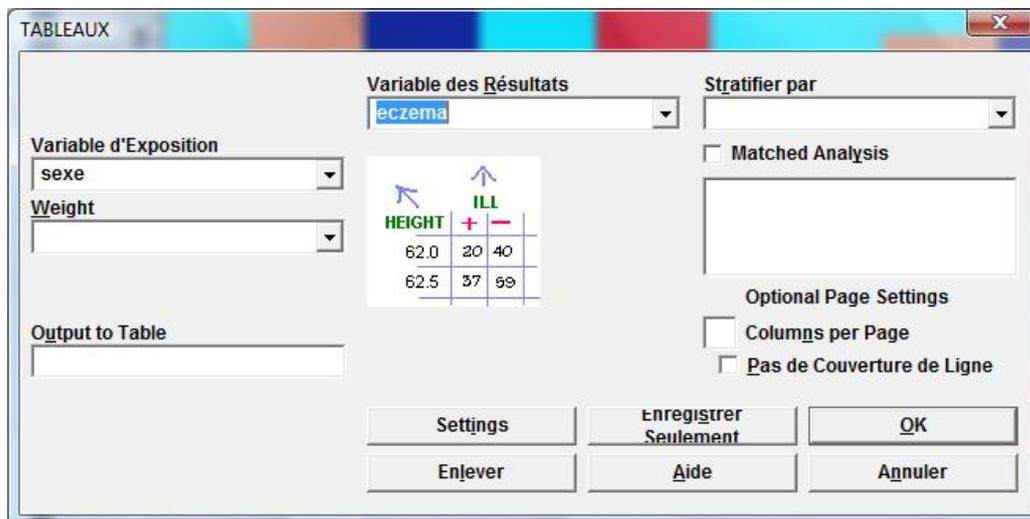
Minimum 25% Médiane 75% Maximum Mode
5,4000 11,5000 14,9000 19,8000 70,3000 10,2000

TABLEAUX

La commande **Tableaux** donne des tableaux croisés avec les statistiques pour tester l'association entre deux variables qualitatives (voir en annexe pages 68 à 70). Elle compte le nombre d'observations qui ont la même valeur pour ces deux variables.

Dans notre exemple de données, pour étudier l'association entre l'eczéma et le sexe:

- ☞ Cliquez sur la commande **Tableaux**, dans le groupe des commandes **■ Statistiques**;
- ☞ dans la boîte de dialogue TABLEAUX, sélectionnez la variable "sexe" depuis le menu déroulant appelé "**Variable d'Exposition**", et la variable "eczema" depuis le menu déroulant appelé "**Variable des Résultats**";
- ☞ Cliquez sur **OK**.



⇒ Les résultats apparaissent dans la fenêtre de sortie:

[Procédure Suivante](#)

[Suite](#)

| ECZEMA | | | |
|--------------|-------|-------|-------|
| sexe | Yes | No | TOTAL |
| F | 86 | 601 | 687 |
| % > | 12,5 | 87,5 | 100,0 |
| % v | 58,5 | 49,6 | 50,6 |
| M | 61 | 610 | 671 |
| % > | 9,1 | 90,9 | 100,0 |
| % v | 41,5 | 50,4 | 49,4 |
| TOTAL | 147 | 1211 | 1358 |
| % > | 10,8 | 89,2 | 100,0 |
| % v | 100,0 | 100,0 | 100,0 |

| Table d'Analyse Unique | | | |
|---|-----------------------|-----------------------------|------------|
| | Point | 95% Intervalle de Confiance | |
| | Estimer | Inferieur | Supérieur |
| PARAMETRES: Basé sur Odds | | | |
| Odds Ratio (cross product) | 1,4309 | 1,0115 | 2,0243 (T) |
| Odds Ratio (MLE) | 1,4306 | 1,0121 | 2,0305 (M) |
| | | 0,9980 | 2,0606 (F) |
| PARAMETRES: Basé sur Risk | | | |
| Risk Ratio (RR) | 1,3770 | 1,0096 | 1,8781 (T) |
| Risk Difference (RD%) | 3,4273 | 0,1326 | 6,7220 (T) |
| (T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact) | | | |
| TESTS STATISTIQUES | | | |
| | Chi-quarré 1-tailed p | 2-tailed p | |
| Chi-quarré - non corrigé | 4,1306 | 0,0421143314 | |
| Chi-quarré - Mantel-Haenszel | 4,1276 | 0,0421900966 | |
| Chi-quarré - corrigé (Yates) | 3,7832 | 0,0517697792 | |
| Mid-p exact | 0,0212617287 | | |
| Fisher exact | 0,0256913870 | | |

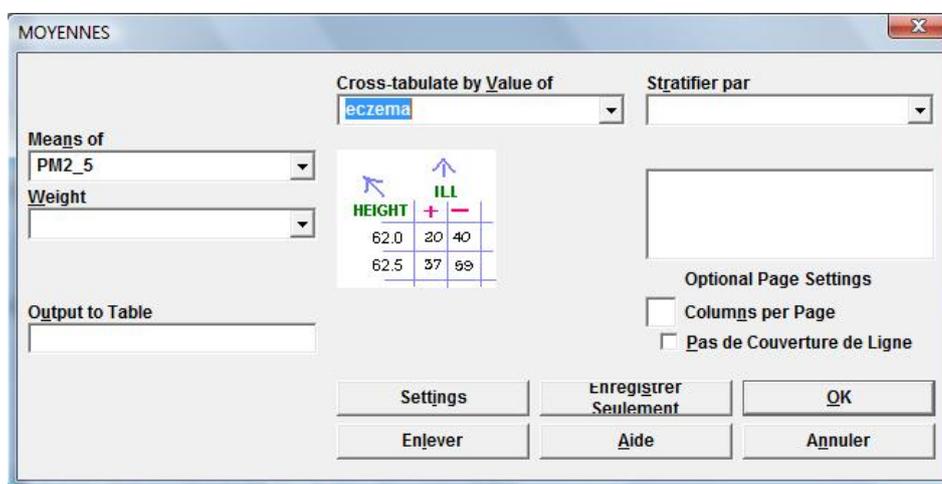
Remarquez que dans la fenêtre d'édition des commandes, TABLES est suivi d'abord par la variable sexe (le facteur "d'exposition"), puis par la variable eczéma (la "maladie"). Ainsi le risque relatif calculé représente le risque d'avoir de l'eczéma chez les filles par rapport aux garçons. Comme la valeur n'est pas incluse dans l'intervalle de confiance à 95 % du risque relatif (RR=1,38; IC 1,01 – 1,88) et de l'odds ratio (OR=1,43; IC 1,01 – 2,02), l'association entre l'eczéma et le sexe est statistiquement significative. De même, les valeurs p du Chi 2 sont inférieures à 0,05. Donc on peut dire que dans cette étude, les filles ont une probabilité plus élevée que les garçons d'avoir de l'eczéma.

MOYENNES avec deux variables (ANOVA)

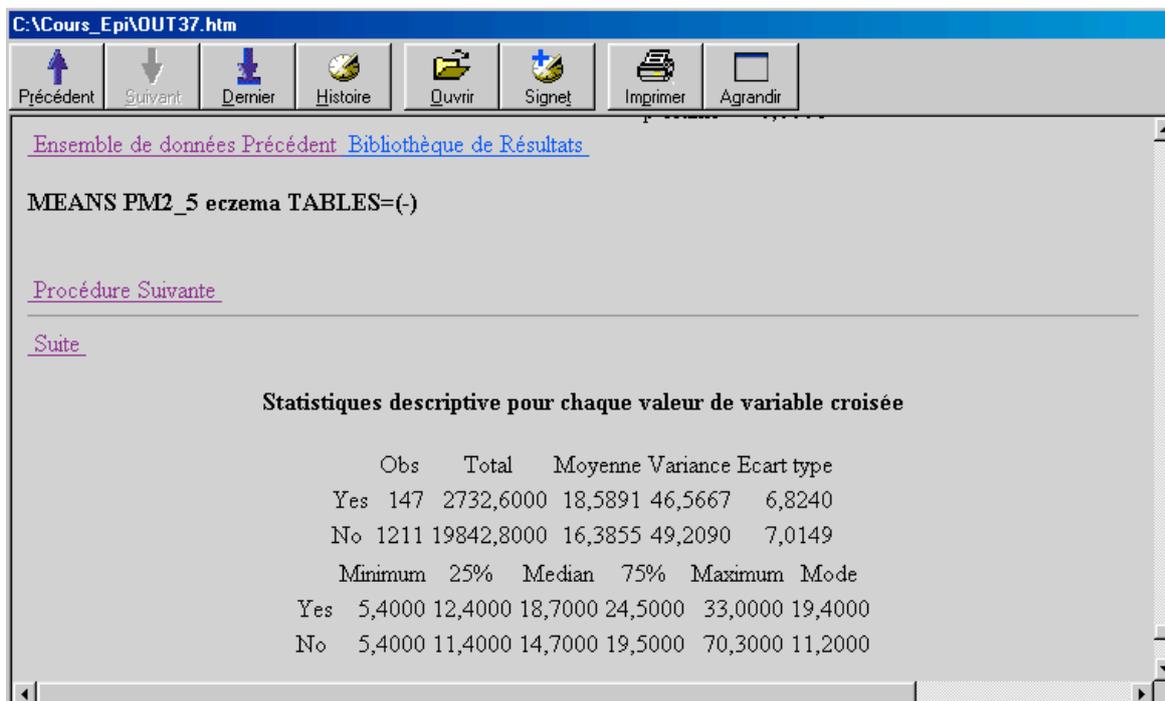
La commande TABLEAUX s'utilise avec les variables qualitatives, avec par exemple des valeurs binaires comme oui/non (comme la variable ECZEMA) ou Masculin et Féminin (comme la variable SEXE). Les variables quantitatives, comme AGE et PM2_5, nécessitent un autre type d'analyse. La commande **Moyennes** permet de comparer les valeurs d'une variable quantitative entre les groupes définis par les catégories d'une variable qualitative en effectuant une analyse de la variance (ANOVA).

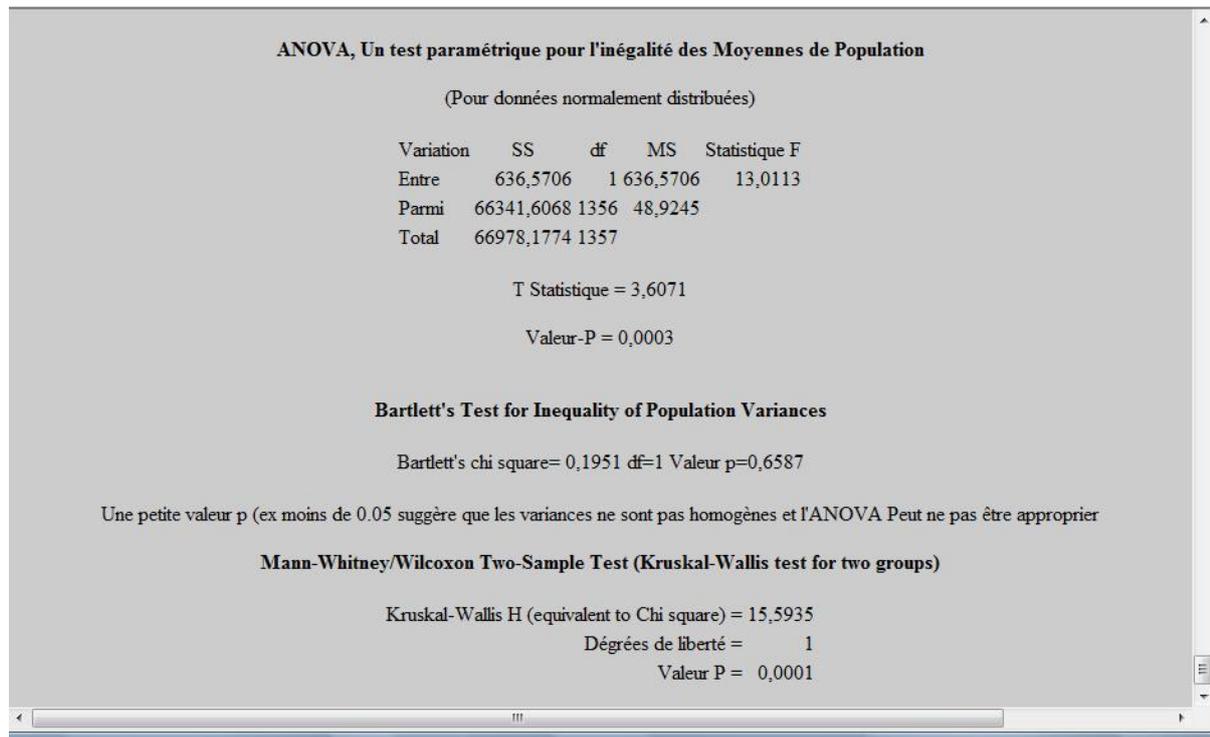
Pour comparer les valeurs des PM2.5 (les concentrations en micro particules dans l'air) entre les enfants qui ont de l'eczéma et ceux qui n'en ont pas:

- ☞ Cliquez sur la commande **Moyennes**, dans le groupe des commandes **Statistiques**;
- ☞ dans la boîte de dialogue MOYENNE, sélectionnez la variable "PM2_5" depuis le menu déroulant appelé "**Moyenne de**", et sélectionnez la variable "eczema" depuis le menu déroulant "**Tableaux croisés par Valeur de**";
- ☞ cliquez sur **OK**.



⇒ Les moyennes apparaissent dans la fenêtre de sortie, avec d'autres statistiques:





La moyenne des concentrations dans l'air des salles de classe des microparticules PM2.5 est de 18,6 $\mu\text{g}/\text{m}^3$ pour les enfants avec de l'eczéma, et de 16,4 $\mu\text{g}/\text{m}^3$ pour les enfants sans eczéma. Les valeurs p pour les tests ANOVA (test paramétrique) et de Kruskal-Wallis (test non paramétrique) sont très inférieures à 0,05, donc on peut conclure que les différences d'exposition aux PM2.5 sont significatives entre les enfants avec de l'eczéma et sans. Les enfants qui sont plus exposés aux PM2.5 ont une probabilité plus élevée d'avoir de l'eczéma.

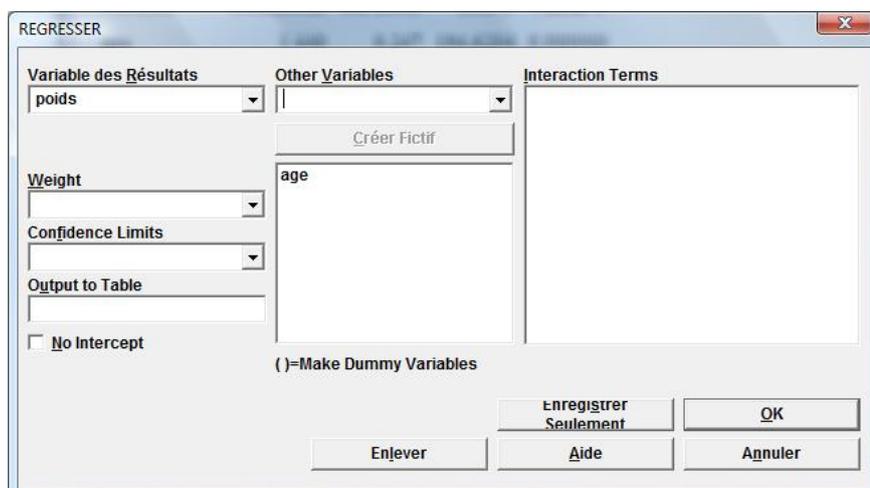
Le test de Bartlett nous renseigne sur lequel des deux tests de comparaison utiliser. Si les variances entre les groupes ne sont pas les mêmes (ne sont pas homogènes), il faut utiliser le test non paramétrique (le test de Kruskal-Wallis) plutôt que le test paramétrique (ANOVA). Ici les variances sont statistiquement égales.

REGRESSION LINEAIRE

La commande REGRESSION LINEAIRE permet de tester l'association entre deux ou plus de deux variables quantitatives. Elle calcule les coefficients de la régression linéaire, et le coefficient de corrélation linéaire entre les variables.

Dans notre exemple de données, pour tester l'association entre le poids et l'âge, et calculer la droite de régression du poids par l'âge (relation linéaire de la forme : $\text{poids} = a * \text{age} + b$):

- ☞ Cliquez sur la commande **Régression Linéaire**, dans le groupe des commandes **Statistiques Avancées**;
- ☞ dans la boîte de dialogue REGRESSER, sélectionnez la variable "Poids" depuis le menu déroulant appelé "**Variable des Résultats**", et sélectionnez la variable "Age" depuis le menu déroulant appelé "**Other Variables**";
- ☞ cliquez sur **OK**.



⇒ Les coefficients de la régression apparaissent dans la fenêtre de sortie, avec d'autres statistiques:

REGRESS poids = age

[Procédure Suivante](#)

Régression Linéaire

| Variable | Coefficient | Std Error | Test-F | Valeur-P |
|----------|-------------|-----------|----------|----------|
| age | 3,440 | 0,247 | 194,6204 | 0,000000 |
| CONSTANT | 2,105 | 2,472 | 0,7251 | 0,394464 |

Correlation Coefficient: $r^2 = 0,13$

| Source | df | Somme des Carrés | Moyenne Carré | Statistique-F |
|------------|------|------------------|---------------|---------------|
| Régression | 1 | 11701,244 | 11701,244 | 194,620 |
| Résiduels | 1361 | 81827,996 | 60,123 | |
| Total | 1362 | 93529,240 | | |

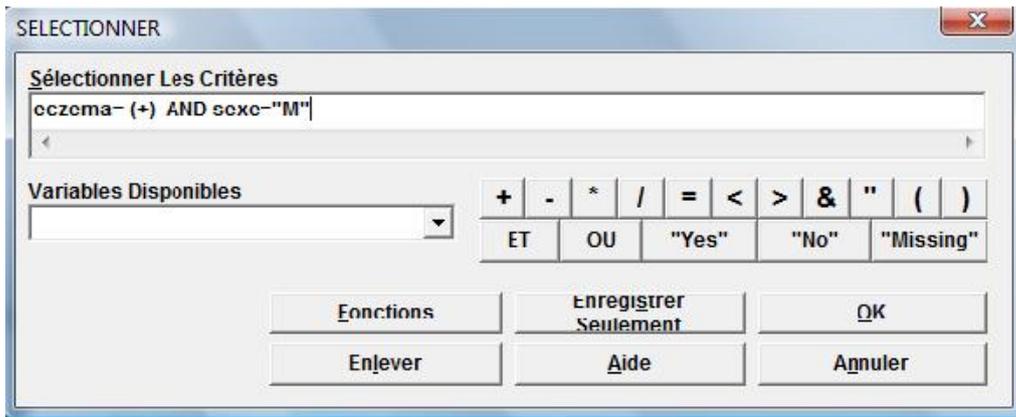
Le coefficient de corrélation linéaire entre le poids et l'âge est de 0,13, les coefficients a et b de la droite de régression valent respectivement 3,44 et 2,11 ($\text{poids} = 3,44 * \text{age} + 2,11$).

SELECTIONNER, Annuler SELECTION

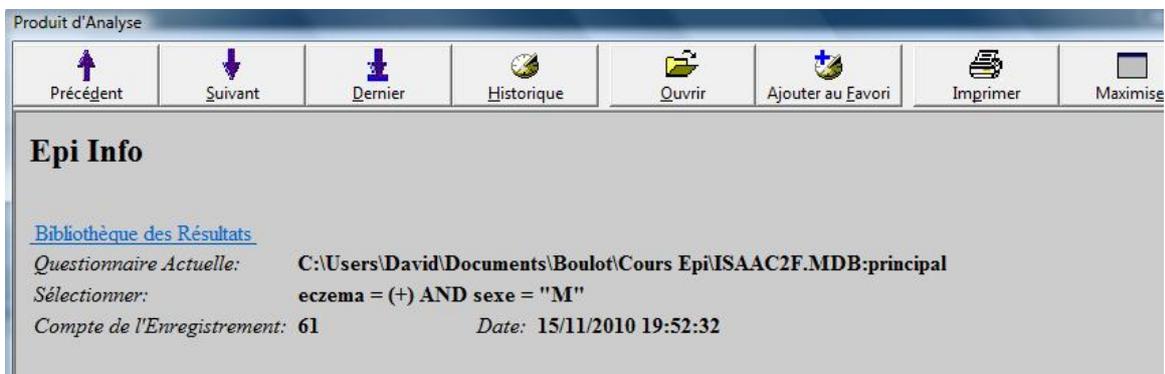
La commande **Sélectionner** vous permet de sélectionner des observations particulières selon vos critères, pour les analyser. La commande **Annuler Sélection** supprime l'effet de la sélection.

Avec nos données, pour sélectionner seulement les garçons avec de l'eczéma:

- ☞ cliquez sur la commande **Select**, dans le groupe des commandes ■ Sélectionner/Si;
- ☞ dans la boîte de dialogue SELECT, dans le champ de saisie appelé "**Critères de sélection**", écrivez à l'aide des boutons ou au clavier: **eczema=(+) AND sexe="M"** ;
- ☞ Cliquez sur **OK**.

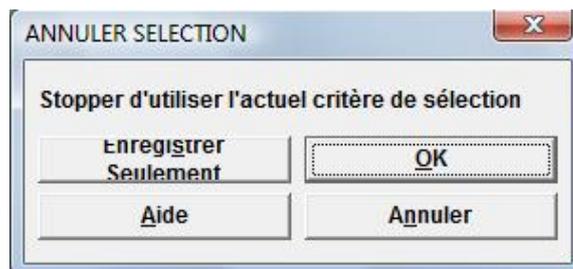


⇒ Dans la table de données, 61 enfants sont sélectionnés.



Pour annuler la sélection:

- ☞ Cliquez sur la commande **Annuler Sélection**;
- ☞ Cliquez sur **OK** dans la boîte de dialogue qui apparaît.

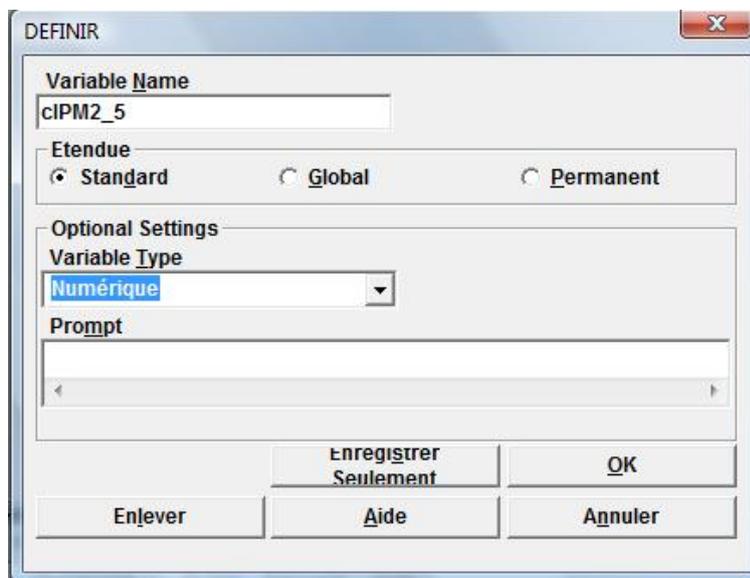


DEFINIR, Annuler DEFINIR

Le commande **Définir** permet de créer de nouvelles variables pour recoder ou combiner des variables existantes. La commande **Annuler Définir** supprime les nouvelles variables.

Pour créer, dans notre table de données, une nouvelle variable appelée “cIPM2_5” :

☞ cliquez sur la commande **Définir** dans le groupe des commandes ■ Variables;



☞ dans la boîte de dialogue DEFINIR, qui apparaît, dans le champ de saisie appelé “Variable Name”, écrivez le nom de la nouvelle variable: **cIPM2_5**;

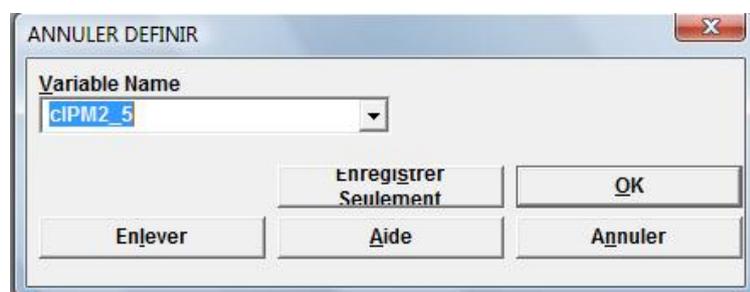
☞ dans le menu déroulant appelé “Variable Type” sélectionnez le type de la variable créée, ici **Numérique**.

☞ Cliquez sur **OK**.

Pour supprimer la nouvelle variable “cIPM2_5” :

☞ cliquez sur la commande **Annuler définir** dans le groupe des commandes ■ Variables;

☞ dans la boîte de dialogue ANNULER DEFINIR, qui apparaît, dans le champ de saisie appelé “Nom de la variable”, écrivez le nom de la variable à supprimer: **cIPM2_5**;



☞ Cliquez sur **OK**.

RECODER

Le commande **Recoder** permet de recoder une variable existante et de mettre le résultat dans une nouvelle variable.

Pour recoder la variable "PM2_5" dans la variable "cIPM2_5" :

☞ cliquez sur la commande **Recoder** dans le groupe des commandes **Variables**;

RECODER

From: [] To: []

Dates must be in US format

| Valeur (vide = autre) | a Valeur (si c'est le c: | Valeur Recodée |
|-----------------------|--------------------------|----------------|
| | | |

Remplir Les Intervalles Enregistrer Seulement OK

Enlever Aide Annuler

☞ dans la boîte de dialogue RECODER, en haut dans le champ de saisie "From" (De), écrivez le nom de la variable à recoder **PM2_5** et dans le champ "To" (à), mettez le nom de la nouvelle variable **cIPM2_5** ;

RECODER

From: PM2_5 To: cIPM2_5

Dates must be in US format

| Valeur (vide = autre) | a Valeur (si c'est le c: | Valeur Recodée |
|-----------------------|--------------------------|----------------|
| LOVALUE | 15 | 1 |
| 15 | HIVALUE | 2 |

Remplir Les Intervalles Enregistrer Seulement OK

Enlever Aide Annuler

- ☞ dans les champ de saisie en dessous mettez l'intervalle des valeurs qui vont prendre un nouveau code, la valeur inférieure dans la colonne "**Valeur**", la valeur supérieure dans la colonne "**A la valeur**", puis le nouveau code dans la colonne "**Valeur Recodée**", en appuyant sur la touche Entrée à chaque fois.
- ☞ Pour finir, cliquez sur le bouton **OK**.

Vous avez défini ainsi une nouvelle variable en deux classes, qui vaudra 1 pour les valeurs basses de PM2_5 (entre la plus petite valeur, mot clé LOVALUE et 15), et 2 pour les valeurs élevées (entre 15 et la plus grande valeur, mot clé HIVALUE).

The screenshot shows a software window titled "FREQ ciPM2_5". At the top, there is a toolbar with icons for navigation and actions: Précédent, Suivant, Dernier, Historique, Ouvrir, Ajouter au Favori, Imprimer, and Maximiser. Below the toolbar, there is a section titled "Procédure Suivante" with a link "En avant". The main content area displays a table with the following data:

| ciPM2_5 | Fréquence | Pourcentage | Pourcentage Cum |
|--------------|-------------|---------------|-----------------|
| 1 | 698 | 51,1% | 51,1% |
| 2 | 668 | 48,9% | 100,0% |
| Total | 1366 | 100,0% | 100,0% |

Below the table, there is a section titled "95% Conf Limits" with the following data:

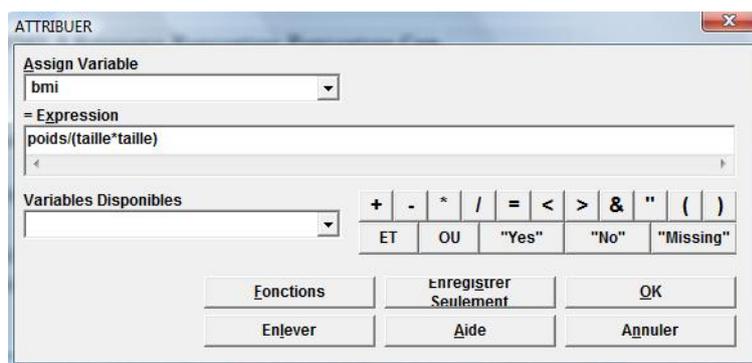
| | | |
|---|-------|-------|
| 1 | 48,4% | 53,8% |
| 2 | 46,2% | 51,6% |

ATTRIBUER

Le commande **Attribuer** permet d'assigner à une nouvelle variable le résultat d'un calcul arithmétique à partir d'une ou de plusieurs variables existantes.

Pour créer une variable correspondant à l'indice de masse corporelle (bmi en anglais) des enfants, qui correspond au poids en kg divisé par le carré de la taille en cm, après avoir défini, avec la commande DEFINIR (voir page 43), une variable appelée "bmi" :

☞ cliquez sur la commande **Attribuer** dans le groupe des commandes ■ Variables;



☞ dans le champ "**Assign Variable**", sélectionnez depuis la liste déroulante la variable **bmi**, et dans le champ "**=Expression**" tapez **poids/(taille*taille)**, puis cliquez sur le bouton **OK**;

Vous avez défini ainsi une nouvelle variable qui contiendra le résultat du calcul de l'indice de masse corporelle à partir du poids et de la taille des enfants, dont on peut déterminer la moyenne.

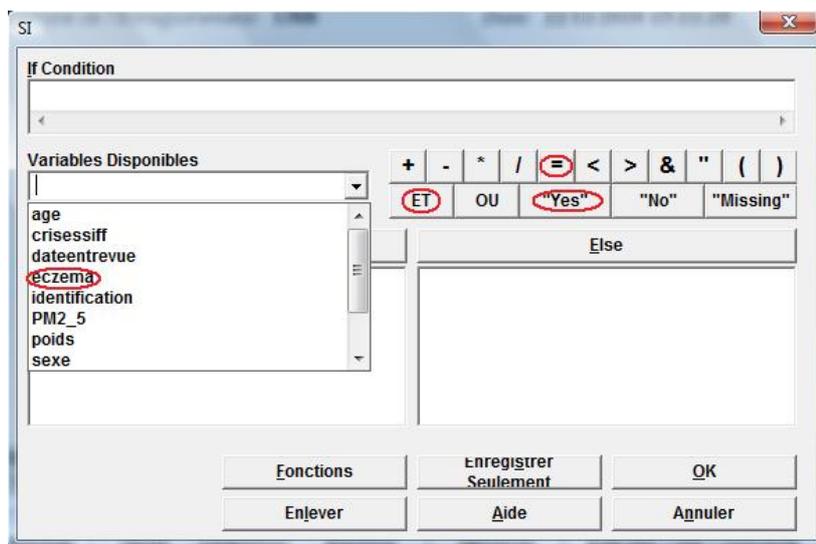
| Obs | Total | Moyenne | Variance | Std Dev | |
|---------|------------|---------|----------|---------|---------|
| 1363 | 24490,4772 | 17,9681 | 8,4465 | 2,9063 | |
| Minimum | 25% | Médiane | 75% | Maximum | Mode |
| 12,2449 | 15,8932 | 17,3469 | 19,4637 | 30,2618 | 15,3061 |

SI, ATTRIBUER

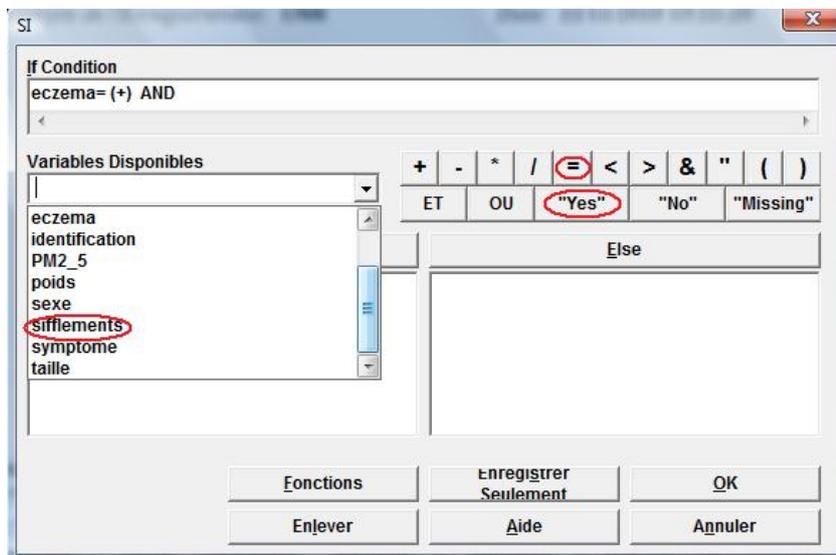
Le commande **Si**, combinée à la commande **Attribuer**, permet de coder une nouvelle variable suivant le contenu d'une ou de plusieurs variables existantes.

Avec notre table de données, pour créer une variable indiquant si un sujet cumule les symptômes de sifflements et d'eczéma, après avoir défini avec la commande DEFINIR (voir page 43) une variable appelée "symptome" :

☞ cliquez sur la commande **Si** dans le groupe des commandes ■ Sélectionner/Si;

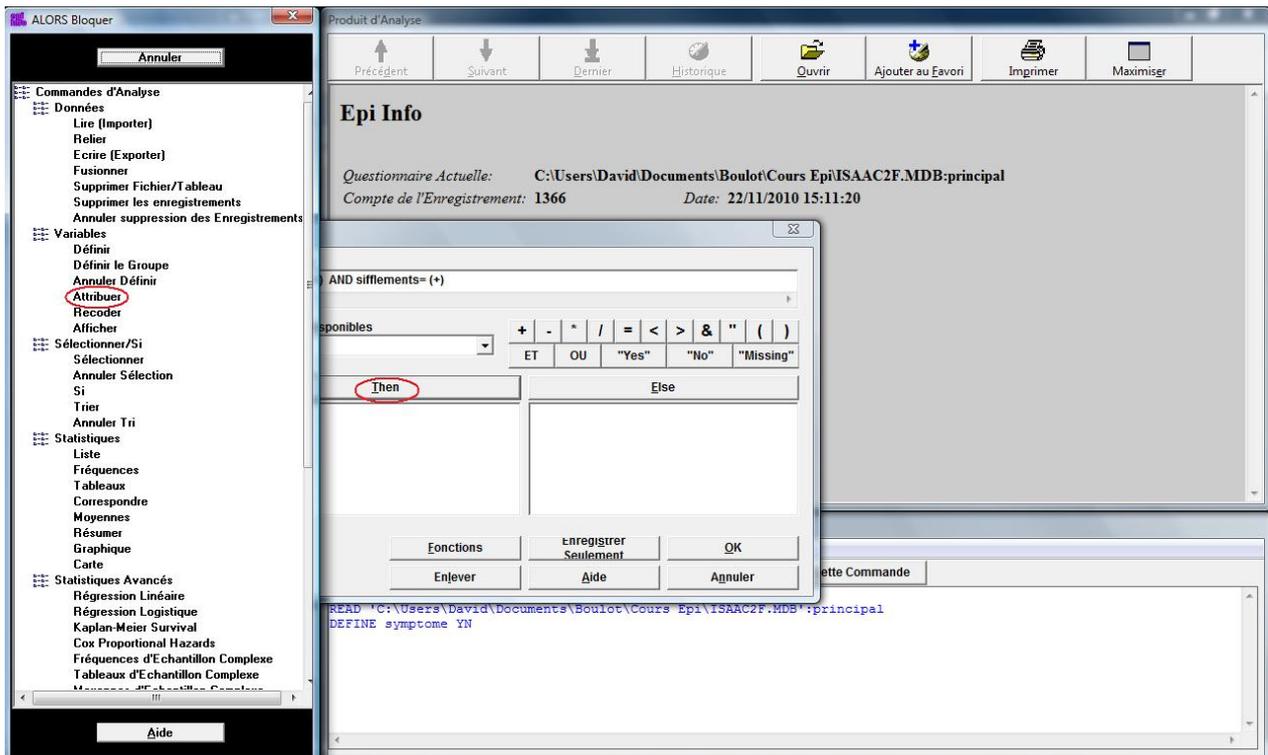


☞ Dans la boîte de dialogue SI, dans le champ "If Condition", sélectionnez depuis la liste déroulante en dessous "Variables disponibles", la variable **EczeMa** , et à partir du clavier virtuel de droite, cliquez sur les boutons =, "Yes", ET.

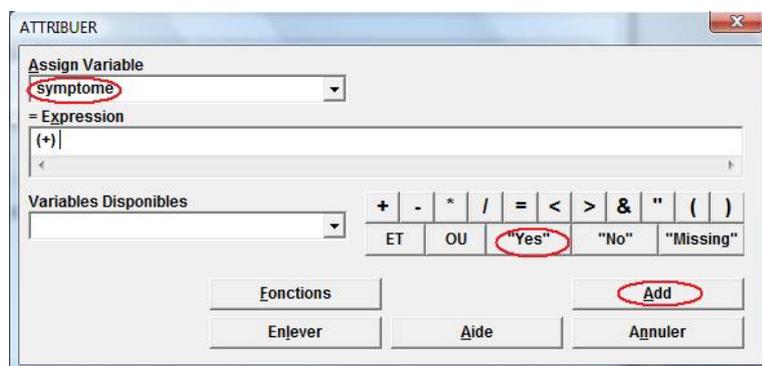


☞ Dans le champ "If Condition", sélectionnez depuis la liste déroulante "Variables disponibles", la variable **Sifflements**, et à partir du clavier virtuel de droite, cliquez sur les boutons =, "Yes".

☞ Cliquez sur le bouton **Then**, apparaît alors une liste de commandes sur la gauche;

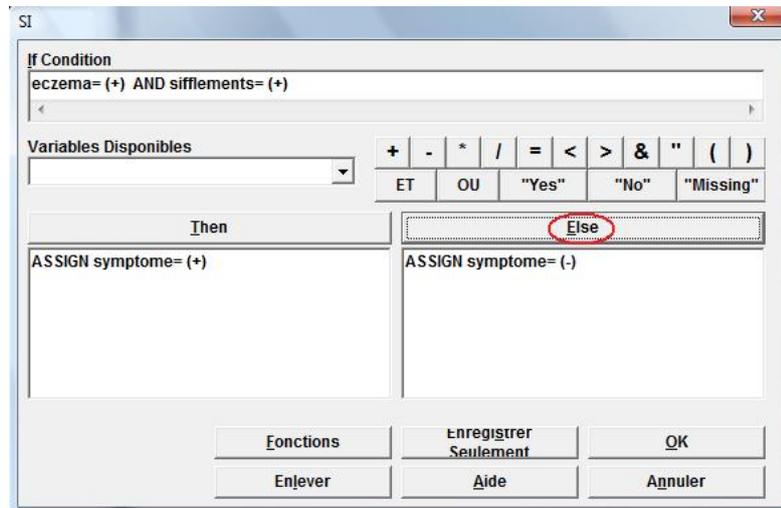


☞ choisissez la commande **Attribuer** ;



☞ dans le champ "Assign Variable", sélectionnez depuis la liste déroulante la variable **Symptome**, et à partir du clavier virtuel de droite, cliquez sur le bouton "Yes", puis cliquez sur le bouton **Add** (ajouter en anglais);

☞ dans la boîte de dialogue SI qui réapparaît, cliquez sur le bouton **Else** (autrement en anglais), puis dans la liste des commandes qui ressurgit, choisissez la commande **Attribuer**, et comme précédemment attribuer la valeur "No" à la variable **symptome** ;



☞ Pour finir, cliquez sur le bouton **OK** dans la boîte de dialogue SI.

Vous avez défini ainsi une nouvelle variable "symptome" qui vaudra "Yes" (oui) pour les sujets qui auront déclaré de l'eczéma et des sifflements, ou "No" (non) autrement.

| FREQ symptome | | | | |
|------------------------------------|-------------|---------------|-----------------|--|
| Procédure Suivante | | | | |
| En avant | | | | |
| symptome | Fréquence | Pourcentage | Pourcentage Cum | |
| Yes | 27 | 2,0% | 2,0% | |
| No | 1339 | 98,0% | 100,0% | |
| Total | 1366 | 100,0% | 100,0% | |
| 95% Conf Limits | | | | |
| Yes | 1,3% | 2,9% | | |
| No | 97,1% | 98,7% | | |

GRAPHIQUE

Le module ANALYSE permet de réaliser différents types de graphiques comme des histogrammes, des représentations en X/Y, des camemberts, des « boîtes de distribution » de valeurs, directement à partir des données.

SECTEURS, BARRES

Ces types de graphiques permettent de résumer les variables qualitatives.

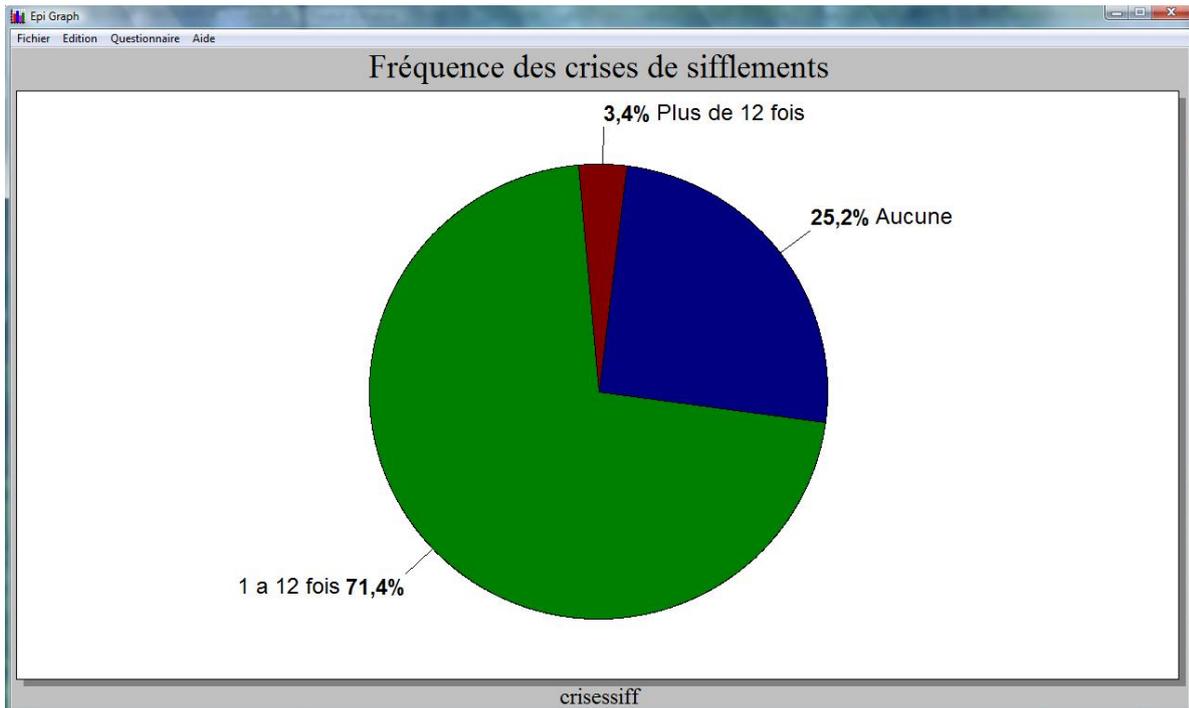
- ☞ Cliquez sur la commande **Graphique** depuis la fenêtre "Analyse", dans le groupe des commandes ■ Statistiques;
- ☞ dans la boîte de dialogue GRAPHIQUE, sélectionnez **Secteurs** depuis le menu déroulant intitulé "Type de Graphique";
- ☞ sélectionnez la variable "crisessiff" depuis le menu déroulant "Variable(s) Principale(s)";
- ☞ ajoutez un titre si vous voulez, dans le champ de saisie appelé "1er Titre";
- ☞ cliquez sur **OK**.

The screenshot shows the 'GRAPHIQUE' dialog box with the following settings:

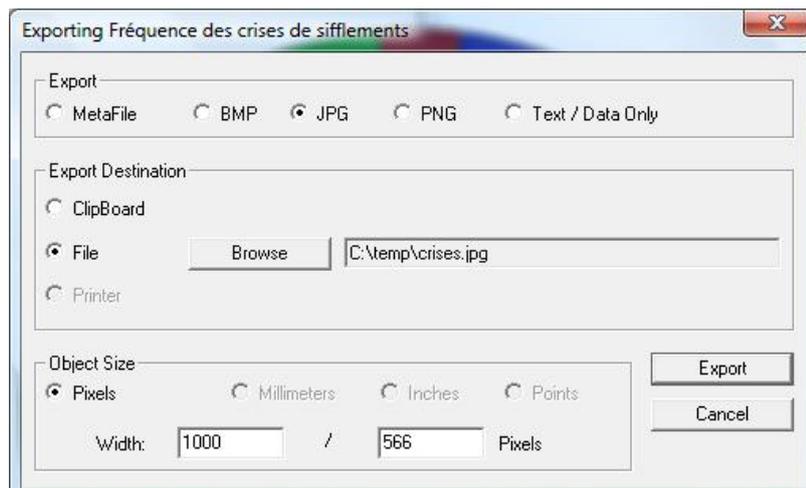
- Type de Graphique:** Secteurs
- 1er Titre / 2nd Titre:** Fréquence des crises de sifflements
- 3D:**
- Axe-X:** Variable(s) Principales: crisessiff
- Axe-Y:** Afficher la Valeur De: Compte
- Séries:** Barre pour chaque: Pas applicable; Un Graphique pour Chaque: 1
- Echelle:** [] []
- Afficher la fenêtre d'aperçu:** Durant l'interaction; Lorsque PGM est en cours
- Graphiques Multiples par Page:** Horizontal: 1; Vertical: 1

Buttons: Parcourir, Enregistrer Seulement, OK, Enlever, Aide, Annuler

- ☞ Le graphique apparaît alors dans une nouvelle fenêtre intitulée EPI GRAPH (25,2 % ont répondu aucun à la question sur la fréquence des sifflements, 71,4 % 1 à 12 fois, et 3,4 % plus de 12 fois).



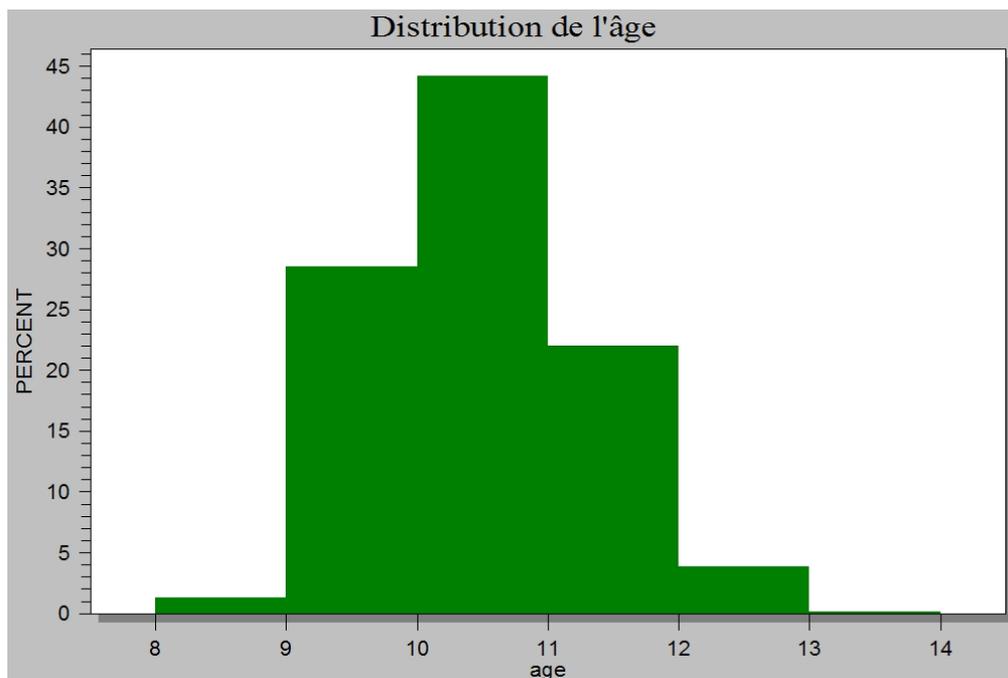
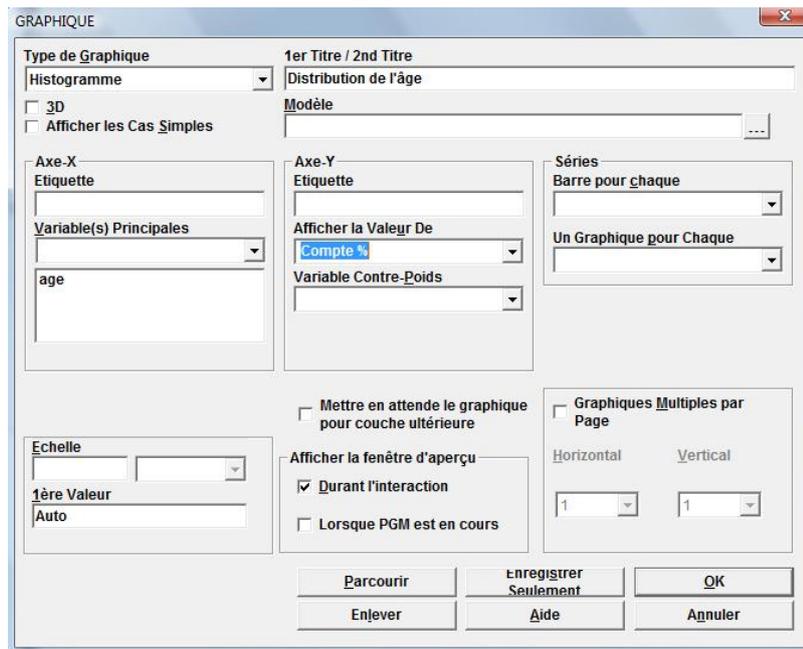
- ☞ Dans la fenêtre EPI GRAPH, depuis le menu déroulant **Fichier** en choisissant "**Quitter**", le graphique apparaîtra dans la fenêtre de sortie des résultats.
- ☞ Le graphique peut être enregistré à part en tant que fichier image, en choisissant, dans la fenêtre EPI GRAPH, depuis le menu **Fichier**, "**Exporter**", une nouvelle fenêtre apparaît;
- ☞ on choisit ensuite le format d'exportation comme JPG, dans le champ appelé "**Export**", puis un nom de fichier dans le champ appelé "**Export Destination**", option "**File**";
- ☞ Pour finir, on clique sur le bouton **Export**, en bas à droite de la fenêtre.



HISTOGRAMME

Pour résumer les variables quantitatives on utilise HISTOGRAMME.

- ☞ Cliquez sur la commande **Graphique**, dans le groupe des commandes **Statistiques**;
- ☞ dans la boîte de dialogue GRAPHIQUE, sélectionnez **Histogramme** depuis le menu déroulant “**Type de graphique**”;
- ☞ sélectionnez la variable "Age" depuis le menu déroulant “**Variable(s) Principale(s)**” de la partie “**Axe-X**”;
- ☞ sélectionnez l’option **Compte %** depuis le menu déroulant “**Afficher la valeur De**” de la partie “**Axe-Y**”;
- ☞ ajoutez éventuellement un titre dans le champ de saisie “**1er Titre**” (Distribution de l’âge);
- ☞ Cliquez sur **OK**.



HISTOGRAMME en définissant des classes d'intervalle égal

HISTOGRAMME permet aussi de représenter la distribution des valeurs en choisissant des intervalles (comme le poids de 15 à 75 kg, en classes de 10 kg d'intervalle).

- ☞ Cliquez sur la commande **Graphique**, dans le groupe des commandes **Statistiques**;
- ☞ dans la boîte de dialogue GRAPHIQUE, sélectionnez **Histogramme** depuis le menu déroulant "**Type de Graphique**";
- ☞ sélectionnez la variable "Poids" depuis le menu déroulant "**Variable(s) Principale(s)**";
- ☞ sélectionnez l'option **Compte %** depuis le menu déroulant "**Axe-Y Afficher la Valeur De**";
- ☞ ajoutez éventuellement un texte dans le champ de saisie "**Axe-Y Etiquette**" (Fréquence %);
- ☞ donnez la largeur de l'intervalle des classes dans le champ de saisie "**Echelle**", en bas à gauche (10);
- ☞ entrez en dessous la borne inférieure de la première classe dans le champ de saisie "**1ère Valeur**" (15);
- ☞ ajoutez éventuellement un titre dans le champ de saisie "**1er Titre**" (Distribution du poids);
- ☞ Cliquez sur **OK**.

GRAPHIQUE

Type de Graphique: Histogramme

1er Titre / 2nd Titre: Distribution du poids

Modèle: []

Axe-X: Etiquette: kg; Variable(s) Principales: poids

Axe-Y: Etiquette: Fréquence %; Afficher la Valeur De: Compte %; Variable Contre-Poids: []

Séries: Barre pour chaque; Un Graphique pour Chaque

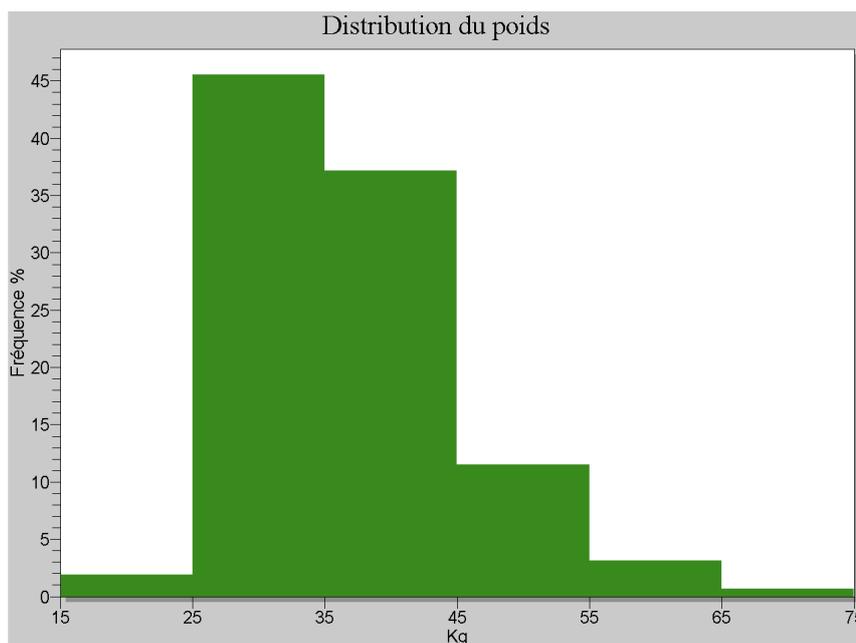
Echelle: 10; 1ère Valeur: 15

Mettre en attente le graphique pour couche ultérieure: []

Graphiques Multiples par Page: Horizontal: 1; Vertical: 1

Afficher la fenêtre d'aperçu: [x] Durant l'interaction; [] Lorsque PGM est en cours

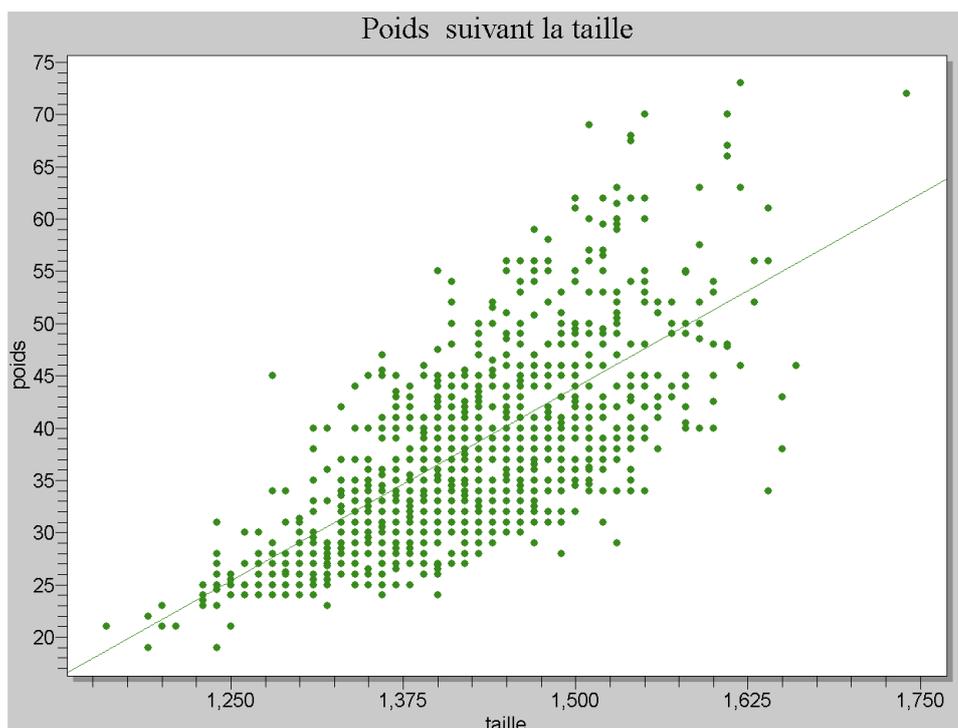
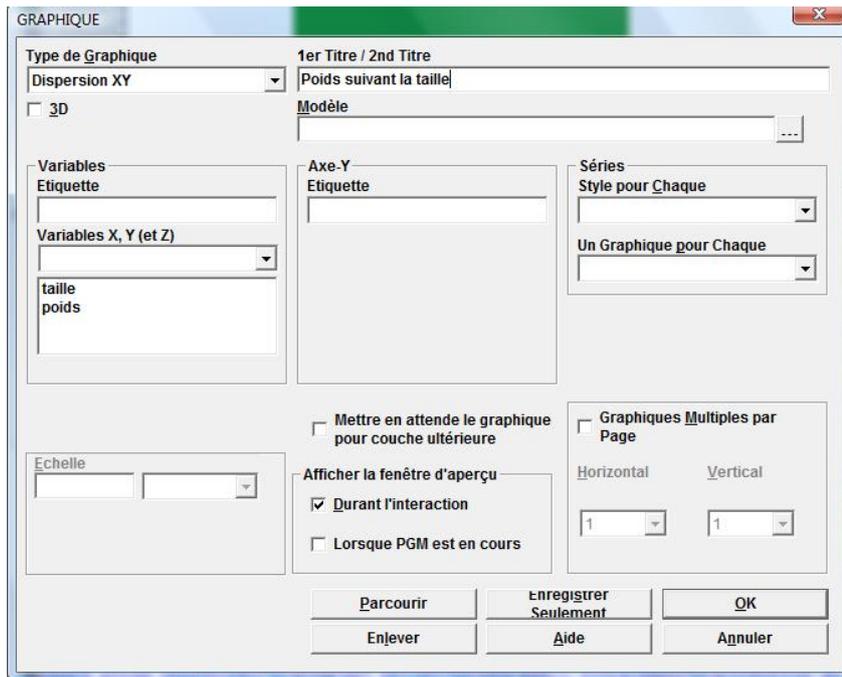
Parcourir, Enregistrer Seulement, OK, Enlever, Aide, Annuler



DISPERSION XY

On peut représenter les valeurs d'une variable continue en fonction d'une autre variable continue pour avoir une idée de leur corrélation éventuelle, avec le type de graphique **DISPERSION XY**.

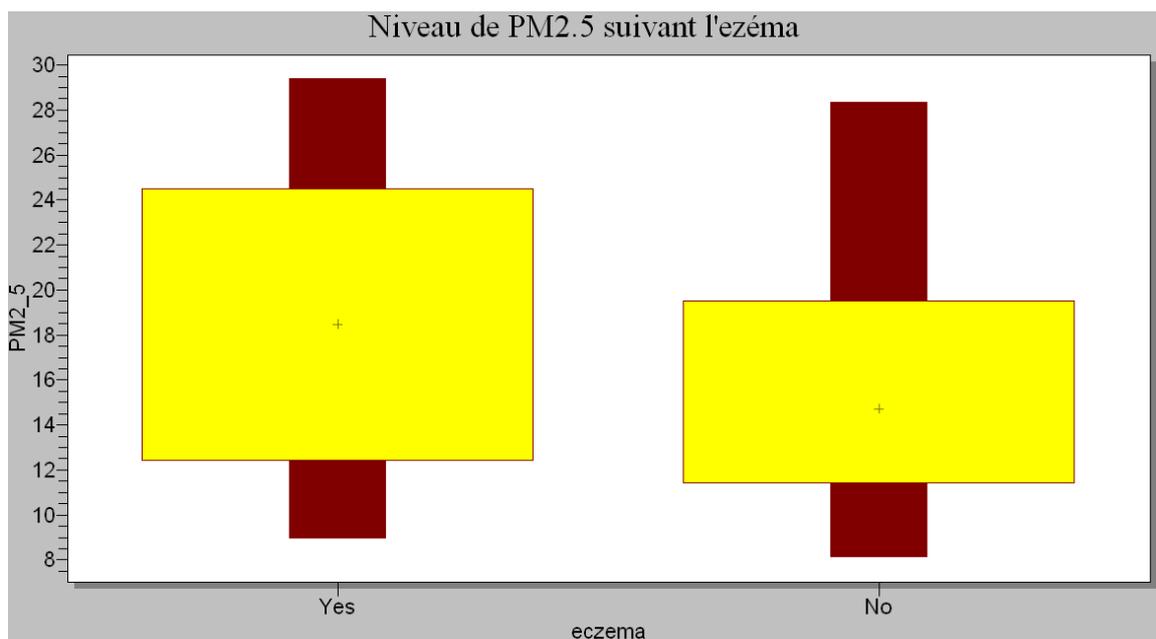
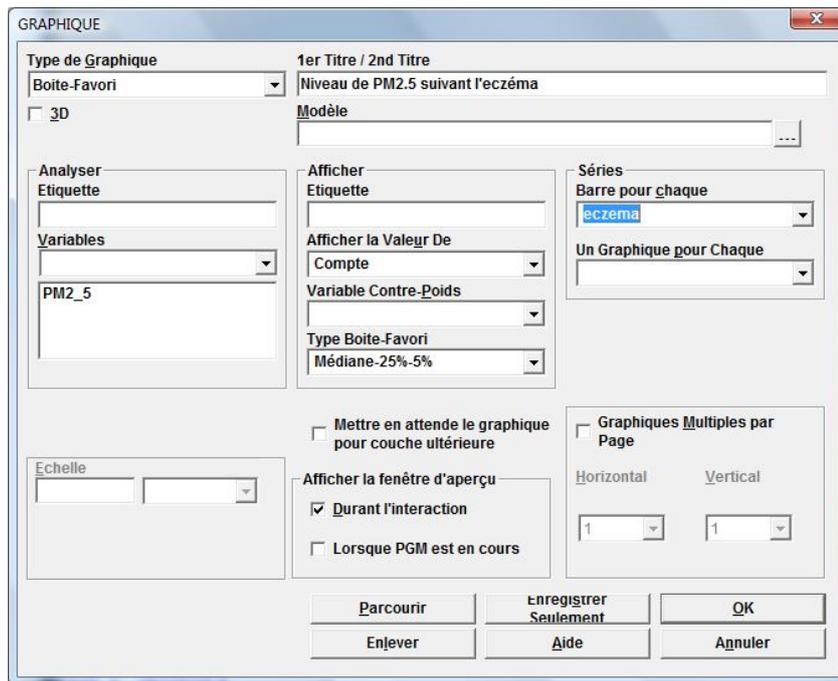
- ☞ Cliquez sur la commande **Graphique**, dans le groupe des commandes **Statistiques**;
- ☞ dans la boîte de dialogue GRAPHIQUE, sélectionnez **Dispersion XY** depuis le menu déroulant "**Type de graphique**";
- ☞ sélectionnez d'abord **Taille** comme variable en X depuis le menu déroulant "**VARIABLES X, Y (et Z)**", et ensuite **Poids** comme variable en Y;
- ☞ ajoutez un titre dans le champ de saisie "**1er Titre**" (Poids suivant la taille) et cliquez sur OK.



BOITE-FAVORI avec deux variables

Pour comparer la répartition des valeurs d'une variable quantitative entre les groupes définis par une variable qualitative, il est intéressant de représenter dans un même graphique la distribution des valeurs de cette variable quantitative dans chaque groupe. Epi Info le permet avec le type de graphique **BOITE-FAVORI**, qui représente sous forme de boîte les valeurs des quartiles d'une distribution.

- ☞ Dans la boîte de dialogue GRAPHIQUE, sélectionnez **Boîte-Favori** depuis le menu déroulant "Type de graphique";
- ☞ sélectionnez d'abord "PM2_5" depuis le menu déroulant "Analyser Variables", et ensuite "eczema" depuis le menu déroulant "Séries Barre pour chaque";
- ☞ choisissez le type de graphique, ici **Médiane-25%-5%**, dans le menu "Type Boîte-Favori";
- ☞ ajoutez un titre dans le champ de saisie "1er Titre" et pour finir cliquez sur OK.



BARRE, avec deux variables

Pour étudier l'association entre deux variables qualitatives, on peut représenter dans un même graphique la fréquence de l'une des variables dans chaque groupe défini par l'autre variable. Epi Info le permet en utilisant le type de graphique **BARRE**.

- ☞ Dans la boîte de dialogue GRAPHIQUE, sélectionnez **Barre** depuis le menu déroulant "Type de Graphique";
- ☞ sélectionnez d'abord "crisesesiff" comme variable dont on veut voir la fréquence des réponses, depuis le menu déroulant "**Variable(s) Principale(s)**", et ensuite "sexe" comme variable définissant les groupes depuis le menu déroulant "**Séries Barre pour chaque**";
- ☞ choisissez le type de fréquence, ici les pourcentages, **Compte%**, dans le menu "Axe-Y Afficher la Valeur De";
- ☞ ajoutez un titre dans le champ de saisie "1er Titre" et cliquez sur OK.

GRAPHIQUE

Type de Graphique: Barre

1er Titre / 2nd Titre: Crises de sifflement par sexe

Axe-X: crisesesiff

Axe-Y: Compte %

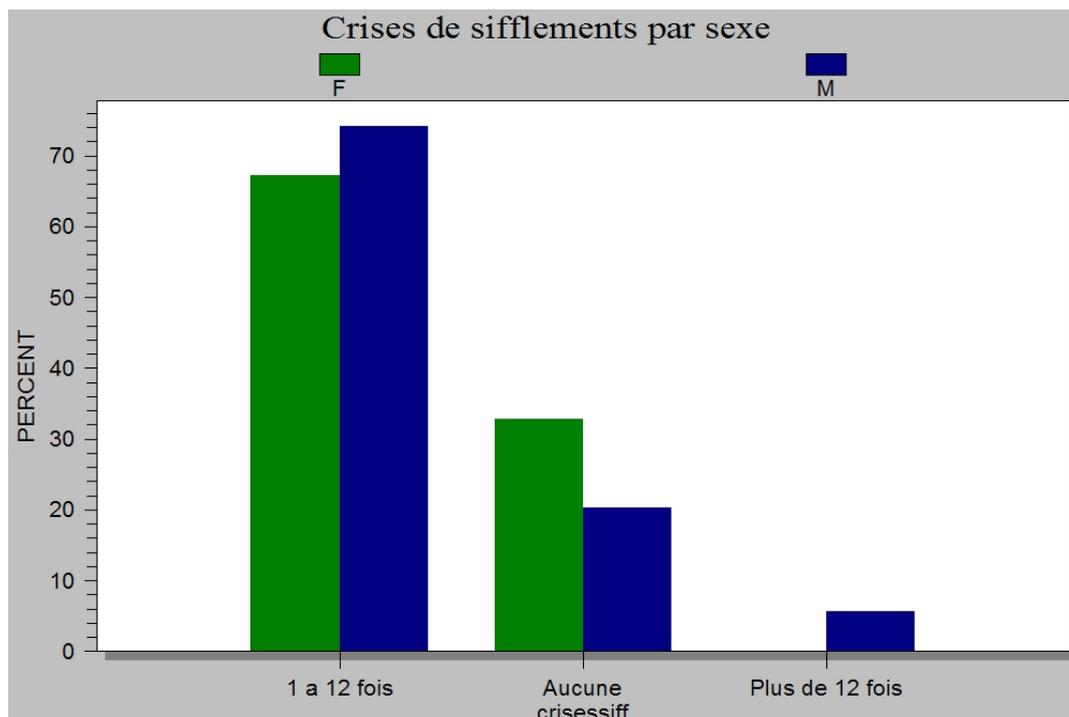
Séries: Barre pour chaque, sexe

Afficher la fenêtre d'aperçu: Durant l'interaction

Echelle: 1

Graphiques Multiples par Page: Horizontal 1, Vertical 1

Buttons: Parcourir, Enregistrer Seulement, OK, Enlever, Aide, Annuler



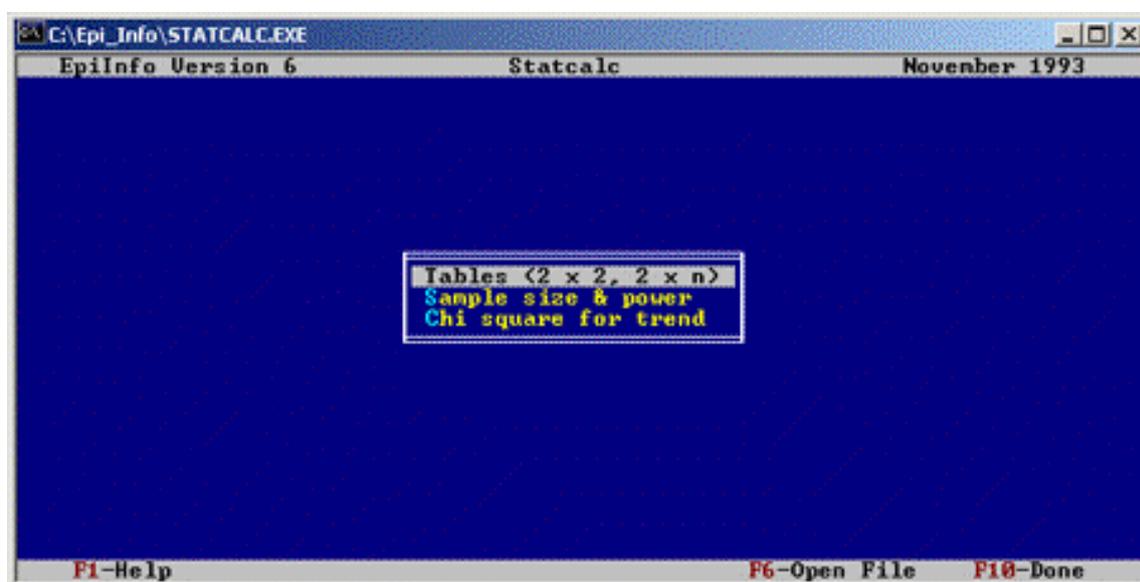
V. Le Module STATCALC

Le module STATCALC permet de calculer des statistiques à partir de données épidémiologiques résumées.

Trois types de calculs sont possibles :

- les statistiques pour les données entrées sous forme de tableau (OR, RR, Chi-2...);
- le calcul du nombre de sujets nécessaire et de la puissance ;
- le calcul d'un Chi-2 de tendance.

☞ Pour lancer **StatCalc**, à partir de la fenêtre d'accueil d'Epi Info, faites apparaître avec la souris le menu déroulant "**Utilitaires**", et sélectionnez "**Statcalc**";



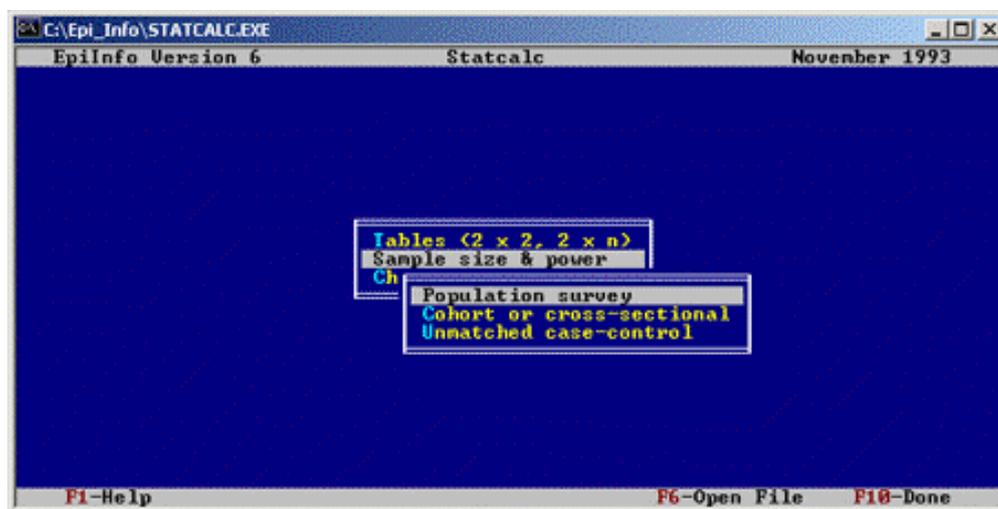
☞ Dans StatCalc pour choisir les différents programmes, il faut utiliser les touches de flèche haut, bas du clavier, et appuyer sur la touche entrée (la souris n'est pas active).

Calcul du nombre de sujets nécessaire et de la puissance

Epi Info permet le calcul du nombre de sujets que l'on doit inclure dans une enquête pour arriver à mettre en évidence une différence entre deux prévalences. Il offre aussi la possibilité d'estimer la puissance du test bilatéral de comparaison de proportions à partir du nombre de sujets d'une enquête.

Les exercices III.2 et III.3 qui sont en annexe (p. 73) vont nous permettre d'illustrer le calcul du nombre de sujets nécessaire et de la puissance.

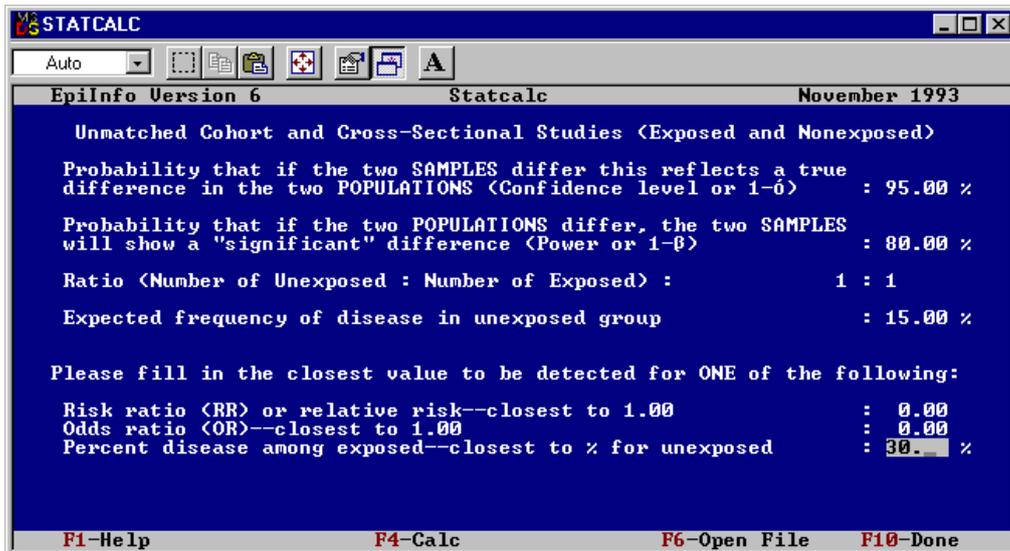
☞ Pour lancer dans StatCalc le programme de calcul du nombre de sujets et de la puissance, on choisit l'option "**Sample size & power**"; apparaît alors une fenêtre avec trois nouvelles options selon le type d'enquête considérée (enquête descriptive, enquête de cohorte ou transversale, enquête cas-témoin);



Calcul du nombre de sujets nécessaires dans une enquête transversale

☞ L'exercice III.2 est au sujet d'une enquête transversale sur la prévalence des avortements spontanés selon une exposition particulière. La deuxième question porte sur le calcul du nombre de sujets nécessaires pour obtenir une puissance de 80 % dans le test de comparaison des prévalences chez les exposés ($p_1=30\%$) et chez les non exposés ($p_0=15\%$).

☞ On choisit donc l'option "**Cohort or cross-sectional**", pour effectuer les calculs pour une enquête transversale, il apparaît alors une nouvelle fenêtre avec les paramètres à rentrer pour le calcul.



Les paramètres à rentrer pour le calcul du nombre de sujets nécessaires sont :

- * **l'amplitude de l'intervalle de confiance** qui correspond à la probabilité de ne pas rejeter l'hypothèse nulle si elle est vraie (H_0 : les prévalences ne sont pas différentes), soit $1 - \alpha$; cette valeur est quasiment toujours fixée à 95 %;
- * **la valeur de la puissance** recherchée pour le test statistique bilatéral, c'est-à-dire la probabilité de rejeter l'hypothèse nulle si l'hypothèse alternative est vraie (H_1 : les prévalences sont différentes), soit $1 - \beta$; cette valeur est souvent de 80 %.
- * **le ratio** entre le nombre de sujets non exposés et le nombre de sujets exposés (pour un nombre de sujets fixés, la puissance du test est la meilleure si le ratio est de 1:1);
- * **la prévalence de la maladie** ou du symptôme considéré **chez les sujets non exposés**;
- * une des valeurs équivalentes à déterminer par l'enquête, le risque relatif (RR), l'odds ratio (OR) ou **la prévalence de la maladie parmi les sujets exposés**.

☞ Pour l'exercice, on laisse les trois premières valeurs telles quelles, en passant d'un champ à l'autre en appuyant sur la touche **Entrée**, puis on tape 15 pour la fréquence de la maladie chez les non exposés, et 30 pour la fréquence chez les exposés.

☞ Pour lancer le calcul, on appuie sur la touche de fonction **F4**, les résultats apparaissent alors sous forme d'un tableau.

| Conf. | Power | Unex:Exp | Disease in Exposed | Risk Ratio | Odds Ratio | Sample Size Unexp. | Sample Size Exposed | Total |
|---------|---------|----------|--------------------|------------|------------|--------------------|---------------------|-------|
| 95.00 % | 80.00 % | 1:1 | 30.00 % | 2.00 | 2.43 | 133 | 133 | 266 |
| 90.00 % | " | " | " | " | " | 108 | 108 | 216 |
| 95.00 % | " | " | " | " | " | 133 | 133 | 266 |
| 99.00 % | " | " | " | " | " | 193 | 193 | 386 |
| 99.90 % | " | " | " | " | " | 276 | 276 | 552 |
| 95.00 % | 80.00 % | " | " | " | " | 133 | 133 | 266 |
| " | 90.00 % | " | " | " | " | 174 | 174 | 348 |
| " | 95.00 % | " | " | " | " | 212 | 212 | 424 |
| " | 99.00 % | " | " | " | " | 293 | 293 | 586 |
| " | 80.00 % | 4:1 | " | " | " | 312 | 78 | 390 |
| " | " | 3:1 | " | " | " | 252 | 84 | 336 |
| " | " | 2:1 | " | " | " | 194 | 97 | 291 |
| " | " | 1:2 | " | " | " | 102 | 205 | 307 |
| " | " | 1:3 | " | " | " | 92 | 276 | 368 |
| " | " | 1:4 | " | " | " | 87 | 347 | 434 |

🔗 Description du tableau de résultats:

| Paramètres entrés par l'utilisateur | | | | | | Paramètres calculés par le programme | | |
|-------------------------------------|-------------------|---|-----------------------------|----------------|------------|--------------------------------------|------------------|----------------|
| Conf. | Power | Unex:Exp | Disease in Exposed | Risk Ratio | Odds Ratio | Unexp. Effectif | Exposed Effectif | Total Effectif |
| Valeur de l'intervalle de confiance | Puissance du test | Ratio entre sujets exposés et non exposés | Prévalence chez les exposés | Risque Relatif | | chez les non exposés | chez les exposés | total |
| 1 - α | 1 - β | | | | | | | |

Sur la première ligne on retrouve les valeurs des paramètres entrés par l'utilisateur, puis sur les lignes suivantes les trois premières valeurs varient, celles de l'intervalle de confiance, celles de la puissance du test et celles du ratio entre nombre de sujets exposés et non exposés de 1:1 à 1:4, et les nombres de sujets sont calculés en fonction des différentes combinaisons de ces valeurs.

On peut, à partir de ce tableau, modifier les paramètres initiaux sur la première ligne, en passant de l'un à l'autre avec les flèches haut \uparrow et bas \downarrow sur le clavier, puis relancer le calcul avec la touche F4.

📖 Avec les paramètres rentrés pour l'exercice, il faut environ 120 sujets par groupe pour trouver avec une puissance de 80 %, une différence significative entre les prévalences de 15 % et de 30 %.

Calcul de la puissance d'un test dans une enquête transversale

📖 La première question de l'exercice III.2 porte sur le calcul de la puissance du test à partir du nombre de sujets (50 dans chaque groupe) et des prévalences à comparer (15% et 30%). StatCalc ne permet pas de calculer directement la puissance, puisqu'il ne donne que le nombre de sujets, mais on peut faire varier dans le tableau de résultats la valeur de la puissance, jusqu'à obtenir un nombre de sujets proche de 50 par groupe.

🔗 Essayez des valeurs de puissance décroissantes dans le tableau de résultats précédent, de 70 à 30 %.

Avec une puissance de 40 %, on trouve 58 sujets par groupe, avec 30 % 45 sujets, donc la puissance du test avec 50 sujets par groupe doit être entre les deux. C'est une puissance de 35 % qui donne 51 sujets par groupe. Donc la puissance du test comparatif avec 50 sujets par groupe est de l'ordre de 35 %.

The screenshot shows the STATCALC software window with the following data table:

| Conf. | Power | Unex:Exp | Disease in Exposed | Risk Ratio | Odds Ratio | Sample Size Unexp. | Sample Size Exposed | Total |
|---------|---------|----------|--------------------|------------|------------|--------------------|---------------------|-------|
| 95.00 % | 35.00 % | 1:1 | 30.00 % | 2.00 | 2.43 | 51 | 51 | 102 |
| 90.00 % | " | " | " | " | " | 37 | 37 | 74 |
| 95.00 % | " | " | " | " | " | 51 | 51 | 102 |
| 99.00 % | " | " | " | " | " | 88 | 88 | 176 |
| 99.90 % | " | " | " | " | " | 144 | 144 | 288 |
| 95.00 % | 80.00 % | " | " | " | " | 133 | 133 | 266 |
| " | 90.00 % | " | " | " | " | 174 | 174 | 348 |
| " | 95.00 % | " | " | " | " | 212 | 212 | 424 |
| " | 99.00 % | " | " | " | " | 293 | 293 | 586 |
| " | 35.00 % | 4:1 | " | " | " | 108 | 27 | 135 |
| " | " | 3:1 | " | " | " | 87 | 29 | 116 |
| " | " | 2:1 | " | " | " | 70 | 35 | 105 |
| " | " | 1:2 | " | " | " | 42 | 84 | 126 |
| " | " | 1:3 | " | " | " | 39 | 116 | 155 |
| " | " | 1:4 | " | " | " | 37 | 148 | 185 |

At the bottom of the window, there are function key shortcuts: F1-Help, F5-Print, F6-Open File, and F10-Done.

Calcul de l'augmentation du risque minimal qu'un test peut détecter

📖 La dernière question de l'exercice III.2 porte sur le calcul de l'augmentation du risque minimale mis en évidence par le test avec un nombre de sujets de 50 dans chaque groupe et une puissance de 80 %. A nouveau StatCalc ne permet pas de calculer directement ce risque, mais on peut l'évaluer en faisant varier dans le tableau de résultats la valeur de la prévalence chez les exposés, jusqu'à obtenir un nombre de sujets proche de 50 par groupe.

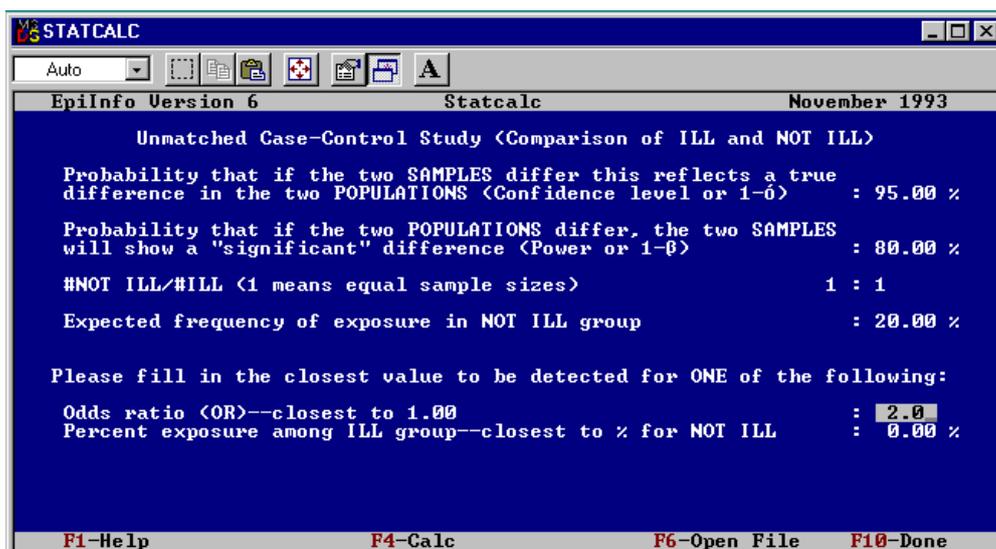
🔍 Essayez des valeurs de fréquences de maladie croissantes dans le tableau de résultats précédent, 40 %, 41 %....

Avec une prévalence de 42 %, on trouve 50 sujets par groupe. Donc le risque relatif minimal que l'on pourra mettre en évidence est de 2,8.

Calcul du nombre de sujets nécessaires dans une enquête cas-témoin

📖 L'exercice III.3 traite d'une enquête cas-témoin sur l'association entre une maladie M et une exposition E. On connaît la prévalence de l'exposition chez les sujets témoins (20 %) et on cherche à mettre en évidence un Odds Ratio de 2 par rapport à la maladie chez les exposés. On pense inclure 200 cas et 200 témoins et on veut savoir si c'est suffisant pour avoir une puissance d'au moins 80 %.

🔍 Pour effectuer les calculs pour une enquête cas-témoin, dans StatCalc, on choisit l'option "Unmatched case-control" du menu "Sample size & power" et il apparaît une nouvelle fenêtre avec les paramètres à rentrer pour le calcul.



A nouveau, les paramètres à rentrer pour le calcul du nombre de sujets nécessaires sont :

- * **l'amplitude de l'intervalle de confiance** qui correspond à la probabilité de ne pas rejeter l'hypothèse nulle si elle est vraie (H_0 : les prévalences ne sont pas différentes), soit $1 - \alpha$; cette valeur est quasiment toujours fixée à 95 %;
- * **la valeur de la puissance** recherchée pour le test statistique bilatéral c'est-à-dire la probabilité de rejeter l'hypothèse nulle si l'hypothèse alternative est vraie (H_1 : les prévalences sont différentes), soit $1 - \beta$; cette valeur est souvent de 80 %;
- * **le ratio** entre le nombre de sujets témoins et le nombre de sujets cas (pour un nombre de sujets fixés, la puissance du test est la meilleure si le ratio est de 1:1);
- * **la prévalence de l'exposition chez les sujets témoins;**
- * une des valeurs équivalentes à déterminer par l'enquête, l'odds ratio (OR) ou **la prévalence de l'exposition parmi les sujets malades** (les cas).

- ☞ Pour l'exercice, on laisse les trois premières valeurs telles quelles, en passant d'un champ à l'autre en appuyant sur la touche **Entrée**, puis on tape 20 pour la fréquence de l'exposition chez les témoins, et 2 pour l'OR que l'on veut mettre en évidence.
- ☞ Pour lancer le calcul, on appuie sur la touche de fonction **F4**, les résultats apparaissent alors sous forme d'un tableau.

STATCALC
EpiInfo Version 6 Statcalc November 1993

Unmatched Case-Control Study (Comparison of ILL and NOT ILL)
Sample Sizes for 20.00 % Exposure in NOT ILL Group

| Conf. | Power | NOT ILL :ILL | Exposure in ILL | Odds Ratio | Sample Size NOT ILL | ILL | Total |
|---------|---------|-----------------|--------------------|---------------|------------------------|-----|-------|
| 95.00 % | 80.00 % | 1:1 | 33.33 % | 2.00 | 186 | 186 | 372 |
| 90.00 % | " | " | 1 | Ⓢ | 150 | 150 | 300 |
| 95.00 % | " | " | | | 186 | 186 | 372 |
| 99.00 % | " | " | | | 270 | 270 | 540 |
| 99.90 % | " | " | | | 389 | 389 | 778 |
| 95.00 % | 80.00 % | " | | | 186 | 186 | 372 |
| " | 90.00 % | " | | | 244 | 244 | 488 |
| " | 95.00 % | " | | | 298 | 298 | 596 |
| " | 99.00 % | " | | | 414 | 414 | 828 |
| " | 80.00 % | 4:1 | | | 444 | 111 | 555 |
| " | " | 3:1 | | | 360 | 120 | 480 |
| " | " | 2:1 | | | 274 | 137 | 411 |
| " | " | 1:2 | | | 142 | 284 | 426 |
| " | " | 1:3 | | | 127 | 381 | 508 |
| " | " | 1:4 | | | 120 | 479 | 599 |

Change values for inputs as desired, then press F4 to recalculate.

F1-Help F5-Print F6-Open File F10-Done

Ce tableau est similaire à celui pour une enquête transversale (avec une colonne en moins, celle avec le risque relatif car il ne peut être calculé pour une enquête cas témoin).

☞ Avec les paramètres rentrés, pour l'exercice, il faut environ 190 sujets par groupe pour trouver avec une puissance de 80 %, un odds ratio significatif de 2. Donc 200 sujets cas et 200 témoins sont suffisants.

INSTALLATION D'EPI INFO V3.5

➤ Configuration requise:

- Windows 98, NT 4.0, 2000, XP ou Vista.
- 32 Mo de mémoire vive. Plus de mémoire (64 Mo minimum) est recommandée pour Windows NT 4.0, ou 2000, et 128 Mo pour Windows XP.
- Un processeur à 300 mégahertz est recommandé.
- Au moins 260 Mo d'espace libre sur le disque dur.

➤ Installation du logiciel

Le programme d'installation du logiciel Epi Info s'appelle "Setup.exe". Il peut être téléchargé depuis mon site <http://www.epivf.fr>

☞ **Pour installer Epi Info Version 3.5, lancez juste le programme "Setup.exe", en cliquant deux fois dessus, et suivez les instructions qui apparaissent à l'écran.**

➤ Installation de la version française

Le programme qui rajoute la langue française à Epi Info s'appelle "francais.exe". Il a été réalisé sous la direction du Dr Jeff K Mathe, MD, MMed OBS/GYN de la CLINIQUE DIEU VIVANT POUR LA SANTE DE LA FEMME ET DE L'ENFANT (CDV-SFE) à Butembo, en République Démocratique du Congo. Vous pouvez l'obtenir sur simple demande à mon adresse électronique david.moreau@epivf.fr, ou en allant sur mon site dédié à Epi Info à l'adresse : <http://www.epivf.fr>

☞ **Pour installer la version française d'Epi Info, placez le fichier "francais.exe" dans le répertoire \TransEXE qui se trouve sous le répertoire \Epi_info;**

☞ **lancez Epi Info;**

☞ **ouvrez le menu "Settings", et sélectionnez "Manage Translations";**

☞ **choisissez "Français" et cliquez sur 'Install';**

☞ **Enfin retournez dans le menu "Settings", et sélectionner "Choose Language". Français doit apparaître dans la liste, choisissez le.**

Centers for Disease Control and Prevention (CDC)

1600 Clifton Road NE

Atlanta GA 30333

USA

Internet: <http://www.cdc.gov/>

Page d'accueil Epi Info du CDC: <http://www.cdc.gov/epiinfo>

Commentaires et suggestions

Envoyez vos commentaires et suggestions pour de futures versions à:

Epi Info™ Helpdesk

Centers for Disease Control and Prevention Atlanta

Division of Integrated Surveillance Systems and Services

1600 Clifton Road, (Mail Stop E-91)

Atlanta, Georgia 30333

USA

E-mail: epiinfo@cdc.gov

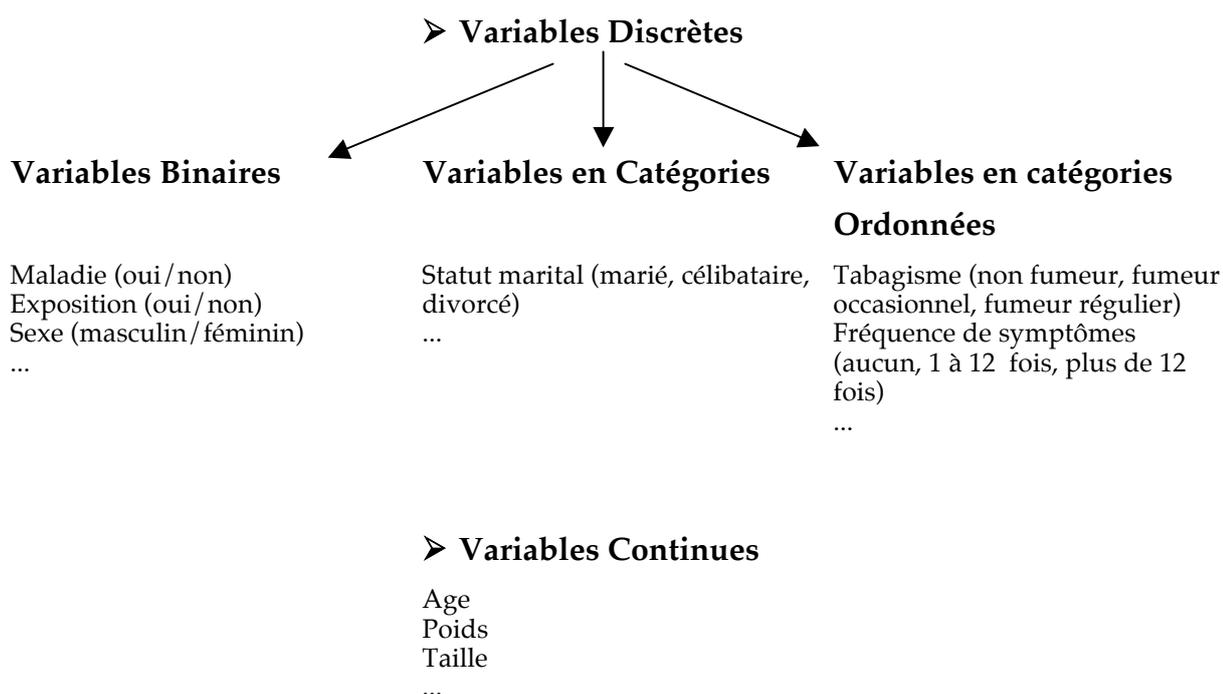
Les types de variables

Les variables peuvent être scindées en deux groupes principaux : les variables discrètes et les variables continues.

Les **variables discrètes ou qualitatives**, ont des codes qui indiquent l'appartenance à un groupe ou une catégorie définie.

Les **variables continues ou quantitatives** ont des valeurs qui correspondent à des mesures provenant d'une échelle continue ordonnée.

Exemples des différents types de variable



Analyse des données

➤ Statistiques descriptives

| Type de variable | Statistiques descriptives | Commandes Epi Info |
|------------------|--|----------------------------|
| Binaire | Fréquences | FREQUENCES |
| En Catégories | Fréquences | FREQUENCES |
| | Graphiques camembert et barres | GRAPHIQUE SECTEURS, BARRES |
| Continue | Moyenne , écart type, médiane, quartiles, mode | MOYENNE |
| | Histogramme | GRAPHIQUE HISTOGRAMME |

➤ Tests statistiques de base pour comparer deux variables

| Type de variable | Mesures d'association | Commandes Epi Info |
|-------------------------|---|---------------------|
| Binaire X Binaire | Chi 2 | TABLEAUX |
| | Odds Ratio | |
| | Risque Relatif | |
| | Excès de Risque | |
| Binaire X Catégories | Chi 2 | TABLEAUX |
| Catégories X Catégories | Chi 2 | TABLEAUX |
| Binaire X Continue | Statistique T (ANOVA) | MOYENNES |
| | Statistique H de Kruskal-Wallis | |
| Catégories X Continue | F statistique (ANOVA) | MOYENNES |
| | Kruskal-Wallis H | |
| Continue X Continue | Coefficient de corrélation R ² | REGRESSION LINEAIRE |
| | Statistique F | |

Il y a deux façons d'estimer l'association entre deux variables :

en calculant un **intervalle de confiance** (d'un odds ratio, d'un risque relatif) (voir p 71) ;

en testant une **hypothèse statistique** (avec un chi-2, les statistiques T, F, H de Kruskal-Wallis...).

Hypothèse statistique

L'hypothèse statistique repose sur le postulat, appelé l'hypothèse nulle, qu'il n'y a pas d'association entre deux variables. Le test statistique consiste à comparer les données observées avec ce qu'elles devraient être (les données attendues) si l'hypothèse nulle était vraie. Le test est résumé par une valeur p qui correspond à la probabilité que les données ressemblent à ce que l'on observe sous l'hypothèse nulle. L'hypothèse nulle de non association est généralement rejetée quand la valeur p est inférieure à 0,05. Dans ce cas l'association entre les deux variables est dite statistiquement significative.

Le calcul des statistiques descriptives

➤ Variables binaires et en catégories

Fréquences :

(nombre d'observations dans une catégorie ou un groupe défini) / (nombre total d'observations)

➤ Variables continues

| Mesure | Définition | Formule |
|--------------------------|--|--|
| Moyenne | Somme des valeurs des observations x divisée par le nombre des observations N | $\bar{x} = \frac{1}{N} \sum x$ |
| Variance | Somme des carrés des différences entre chaque observation et la moyenne, divisée par N-1 | $\sigma^2 = \frac{1}{(N-1)} \sum (x - \bar{x})^2$ |
| Ecart type | Racine carrée de la variance | $\sigma = \sqrt{\frac{1}{(N-1)} \sum (x - \bar{x})^2}$ |
| Minimum | Valeur la plus petite | |
| Quartile 1 (25 %) | Valeur telle que 25% des observations sont inférieures à elle et 75 % sont au-dessus | |
| Médiane | Valeur telle que la moitié des observations est au-dessus et la moitié est au-dessous | |
| Quartile 3 (75%) | Valeur telle que 75% des observations sont inférieures à elle et 25 % sont au-dessus | |
| Maximum | Valeur la plus grande | |
| Mode | Valeur qui est la plus fréquente | |

N : nombre total d'observations

x : valeur de la variable continue pour une observation

Le calcul des Risques pour deux variables binaires

Les mesures de maladie et d'exposition sont souvent analysées avec un tableau de contingence 2X2 :

| | | | | |
|------------|----------|---------|-------|----------------|
| | | Maladie | | |
| | | Oui | Non | |
| Exposition | Présente | a | b | Total a + b |
| | Absente | c | d | c + d |
| Total | | a + c | b + d | N |

a = nombre de sujets exposés et malades
 b = nombre de sujets exposés et non malades
 c = nombre de sujets non exposés et malades
 d = nombre de sujets non exposés et non malades
 N = nombre de sujets dans l'échantillon de l'étude (a+b+c+d)

Définitions :

Prévalence :

probabilité d'un événement dans une population étudiée. Cela correspond au nombre d'apparition d'une maladie spécifique ou d'une condition particulière dans une population à un moment donné, divisé par la population à risque.

Excès de Risque ou différence de risque:

différence entre la prévalence d'une maladie parmi les sujets exposés ($a / (a + b)$) et celle parmi les sujets non exposés ($c / (c + d)$).

Risque Relatif :

rapport entre la prévalence d'une maladie chez les sujets exposés ($a / (a + b)$) et la prévalence chez les sujets non exposés ($c / (c + d)$).

Odds ratio :

rapport entre le rapport des malades sur les non malades parmi les sujets exposés (a/b) et le rapport des malades sur les non malades chez les sujets non exposés (c/d).

Formules de calcul :

| | |
|-----------------------------------|-------------------------------------|
| Prévalence (de la maladie) | $\frac{a + c}{N}$ |
| Excès de Risque | $\frac{a}{a + b} - \frac{c}{c + d}$ |
| Risque Relatif (RR) | $\frac{a / (a + b)}{c / (c + d)}$ |
| Odds ratio (OR) | $\frac{a/b}{c/d} = \frac{ad}{bc}$ |

Le risque relatif, l'excès de risque et l'odds ratio sont des indices mesurant l'association entre une maladie et un facteur d'exposition.

Interprétation de ces indices

Le Risque Relatif (RR) et l'Odds Ratio (OR) sont toujours supérieurs à zéro. Un RR ou un OR supérieur à un ou un excès de risque supérieur à zéro indiquent une augmentation du risque de la maladie associée au facteur d'exposition. Un RR ou un OR égal à un, ou un excès de risque égal à zéro, indiquent qu'il n'y a pas d'association entre la maladie et le facteur d'exposition. Un RR ou un OR inférieur à un, ou un excès de risque négatif, indiquent une diminution du risque de maladie associée au facteur d'exposition.

| Excès de risque | Risque Relatif ou Odds Ratio | Interprétation |
|-----------------|------------------------------|--|
| < 0 | < 1 | L'exposition peut protéger de la maladie |
| 0 | = 1 | La maladie n'est pas liée à l'exposition |
| > 0 | > 1 | L'exposition peut augmenter le risque de maladie |

Pour tester l'association entre une maladie et une exposition, on calcule un **intervalle de confiance** du risque relatif ou de l'odds ratio (voir p 71). Si l'intervalle de confiance n'inclut pas la valeur un, alors l'association est probablement vraie.

Indices et type d'enquête épidémiologique

| Type d'enquête : | Enquête de cohorte | Enquête transversale | Enquête cas-témoin |
|------------------------|--------------------|----------------------|--------------------|
| Mesure d'association : | Risque Relatif | Risque Relatif | Odds Ratio |
| | Odds Ratio | Odds Ratio | |
| | Excès de risque | Excès de risque | |

Le calcul du Chi-2 pour les variables en catégories

Statistique du Chi-2 :

$$\chi^2 = \sum \left[\frac{(O - E)^2}{E} \right]$$

Les O représentent le nombre d'observations dans chaque cellule du tableau de contingence.

Les E représentent le nombre d'observations attendues sous l'hypothèse nulle (pas d'association entre les deux variables).

Pour un tableau de contingence 2X2 :

- O est égal à a, b, c et d

- Les valeurs attendues pour chaque cellules sont :

$$\text{Attendu}(a) = (a + c)(a + b) / N$$

$$\text{Attendu}(b) = (b + d)(a + b) / N$$

$$\text{Attendu}(c) = (a + c)(c + d) / N$$

$$\text{Attendu}(d) = (b + d)(c + d) / N$$

La statistique du Chi-2 est alors :

$$\chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

Valeur p:

Le chi-2 mène à une valeur p qui correspond à la probabilité que l'association observée apparaisse par hasard.

Quand la valeur p est inférieure à 0,05, l'association entre les deux variables en catégories est dite statistiquement significative.

Fluctuation d'échantillonnage et Intervalle de Confiance

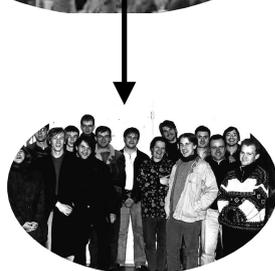
Les mesures de maladies et d'expositions à partir d'un échantillon d'étude sont sujettes à des fluctuations aléatoires. Si l'étude était répétée avec différents échantillons issus d'une même population, les mesures calculées (prévalence, risque relatif...) seraient légèrement différentes d'un échantillon à l'autre. La mesure dans un échantillon est considérée comme la meilleure estimation de la vraie valeur dans la population. On l'appelle la valeur estimée. L'effet des variations aléatoires peut être pris en compte statistiquement, en calculant un intervalle de valeurs autour de la valeur estimée (l'**intervalle de confiance**) qui a une probabilité donnée d'inclure la vraie valeur. La probabilité donnée est appelée le niveau de confiance et est en général de 95%.

De la population cible à l'échantillon d'étude



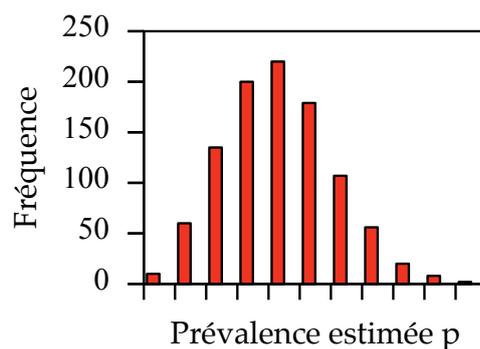
Population cible

Vraie prévalence de la maladie P



Echantillon tiré de la population cible

Prévalence estimée p à partir de 1000 échantillons :



L'éventail des valeurs que la vraie prévalence P peut prendre est appelé l'**intervalle de confiance**. L'intervalle de confiance à 95 % est l'intervalle dans lequel seraient comprises 95 prévalences estimées p sur 100, si on utilisait différents échantillons d'étude tirés de la même population cible.

Types d'erreur des tests d'hypothèses, Puissance d'un test statistique

Definitions

L'hypothèse statistique repose sur la supposition, appelée l'**hypothèse nulle**, qu'il n'y a pas d'association entre deux variables (voir p 66). Le test statistique est résumé par la **valeur p**, et l'hypothèse nulle est rejetée quand la valeur p est inférieure à une valeur fixée à l'avance appelée α (généralement elle est égale à 0,05).

La probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie, est appelée l'**erreur α** ou l'**erreur de type I**.

Il y a aussi la probabilité d'accepter l'hypothèse nulle alors qu'elle est fautive, appelée l'**erreur β** ou l'**erreur de type II**.

La **puissance** d'un test est la probabilité de rejeter l'hypothèse nulle qui est vraiment fautive, donc de tirer une bonne conclusion. Elle est appelée **$1 - \beta$** , et est généralement fixée à 80 %.

L'**intervalle de confiance** est la probabilité de ne pas rejeter l'hypothèse nulle si elle est vraie, donc de tirer une bonne conclusion. Il est égal à $1 - \alpha$.

Résumé sur les deux types d'erreur, l'intervalle de confiance et la puissance:

| | | Réalité | |
|---------------------------------|---|---|--|
| | | Il existe une association (l'hypothèse nulle est fautive) | Il n'existe PAS d'association (l'hypothèse nulle est vraie) |
| Conclusion tirée de l'étude. | Rejet de l'hypothèse nulle (il existe une association) | Conclusion exacte Puissance du test Probabilité= $1-\beta$ | Erreur de type I Probabilité= α |
| | Acceptation de l'hypothèse nulle (il n'existe PAS d'association) | Erreur de type II Probabilité= β | Conclusion exacte Intervalle de Confiance Probabilité= $1-\alpha$ |

Exercices sur la puissance

III. Exercices sur les calculs de puissance et de nombre de sujets nécessaires

2. On veut comparer la fréquence des avortements spontanés selon l'exposition pendant la grossesse à une nuisance particulière. On dispose de deux échantillons, l'un de 50 grossesses exposées, l'autre de 50 grossesses non exposées prises comme témoins. Compte tenu des données de la littérature, on peut considérer que la fréquence de référence des avortements spontanés est égale à 15%. On veut calculer la puissance de l'investigation si l'exposition étudiée était associée à une fréquence de 30%. Ce choix d'un doublement est arbitraire, mais on peut considérer que l'enquête mérite d'être entreprise si elle a au moins une bonne puissance pour cette situation.

- Estimer la puissance dans ce cas précis.

On suppose maintenant qu'on n'a pas fixé le nombre de grossesses exposées et non exposées, mais que l'on souhaite au minimum une puissance de 80% avec les valeurs postulées de P_0 (15%) et de P_1 (30%).

- Calculer le nombre de grossesse à inclure pour obtenir une telle puissance ? On suppose qu'on inclut autant de grossesses exposées que de grossesses non exposées.

On suppose à nouveau qu'on a 50 observations par groupe.

- Quelle est l'augmentation minimale du risque d'avortement chez les sujets exposés qu'on pourra mettre en évidence avec une « bonne » puissance (par exemple 80%) ? On suppose que P_0 est toujours égale à 15%.

3. Une enquête cas-témoin a été mise en place pour examiner l'association entre une exposition notée E et la survenue d'une maladie notée M. Les informations disponibles au moment du protocole indiquent qu'une proportion $P_{E0}=20\%$ des sujets de la population source peuvent être considérés comme ayant été exposés un jour ou l'autre à E. Le nombre de sujets nécessaires a été calculé pour mettre en évidence un doublement du risque de survenue de M chez les exposés, c'est à dire $OR=2$. On a décidé d'inclure dans l'étude 200 cas et 200 témoins recrutés sur 18 mois.

- Le nombre de sujets que l'on souhaite inclure est-il suffisant pour garantir une puissance au moins égale à 80% ?

Après deux ans de recrutement, l'enquête est considérée comme terminée et on peut établir le bilan suivant :

| | Cas | Témoins |
|-------------------------|-----|---------|
| Exposés | 18 | 18 |
| Non exposés | 108 | 162 |
| Informations manquantes | 21 | 32 |
| Total | 147 | 212 |

- Calculer l'OR mesurant l'association entre E et M et donner son intervalle de confiance.
- L'OR diffère-t-il significativement de 1 ?
- Effectuer un calcul de puissance a posteriori sous l'hypothèse d'un doublement du risque de M chez les sujets exposés ($OR=2$).
- Interpréter les résultats obtenus.