



**STATISTICA
Version 7.1**

Licence d'utilisation accordée à :
Henri Born
ULG
AXXF509B307919AR

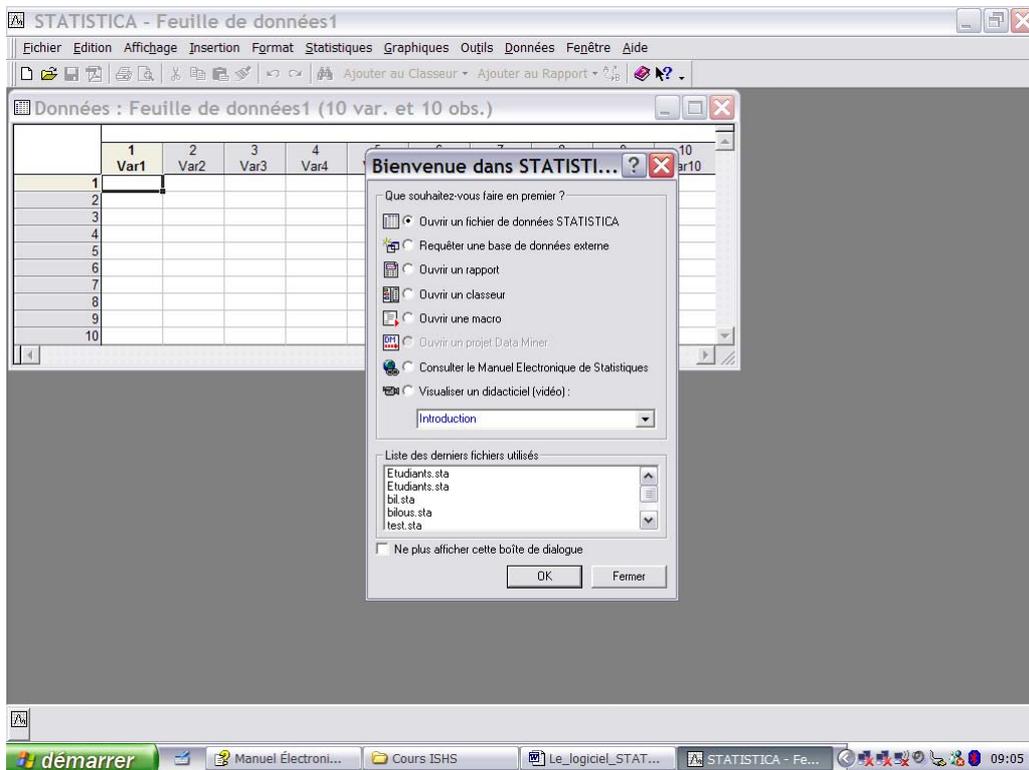
Copyright © Statsoft, Inc. 1984-2005
Tous droits réservés. Ce programme
est protégé par les lois américaines
et internationales sur les droits d'auteur.

TABLE DES MATIERES

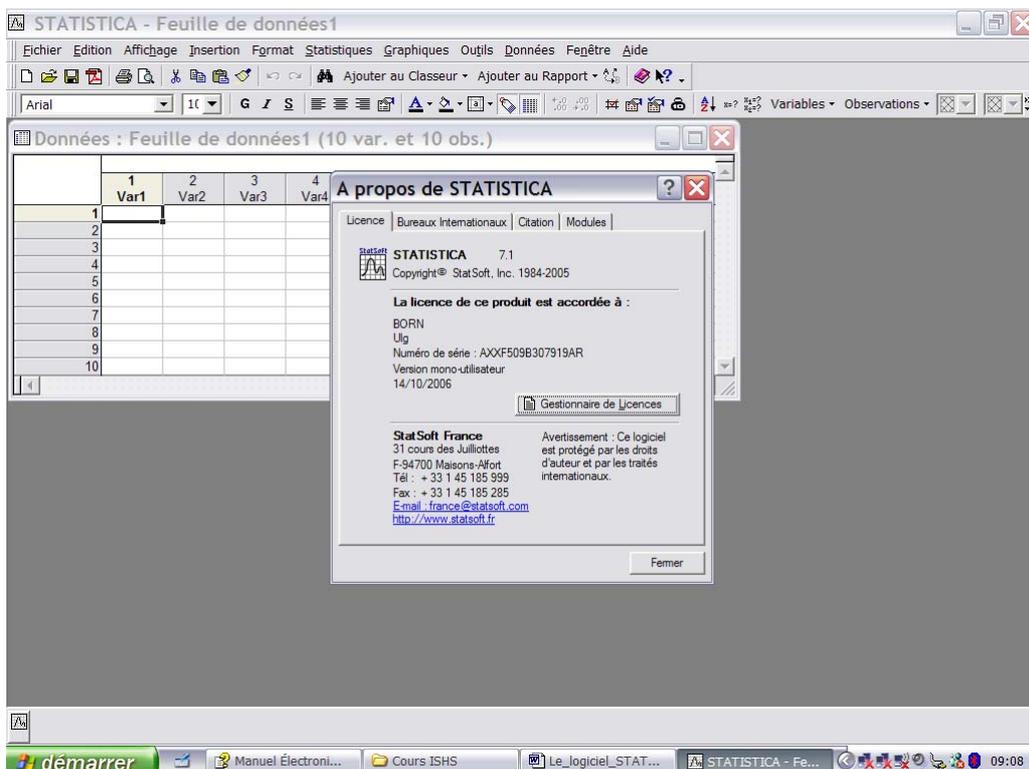
1.	Démarrage du programme	1
2.	Standard d'une fenêtre Windows	1
3.	Utilisation du logiciel STATISTICA	3
A.	<u>Gestion des données</u>	4
1.	Création d'une feuille de données	7
2.	Modifications dans la feuille de données	12
a)	Suppression de variables	12
b)	Déplacement de variables	14
c)	Addition de variables	15
d)	Addition d'observations	16
e)	Copier-coller de variables et d'observations	17
f)	Affichages détaillés	19
g)	Etiquettes associées aux valeurs	20
h)	Noms d'observations	24
i)	Tris	25
j)	Variables de type « Date »	28
B.	<u>Traitements statistiques & graphiques</u>	31
1.	Statistiques rapides	31
2.	Statistiques de blocs	34
3.	Statistiques élémentaires	39
a)	Statistiques descriptives	39
	Affichages des résultats	43
	Approfondissements	49
	Représentation graphique des résultats	53
	Commentaire général sur les graphiques	58
b)	Tables des fréquences	64
c)	Tableaux à plusieurs entrées	73
d)	Corrélation	80
4.	Analyse par groupes d'observations	85
C.	<u>Traitements avancés sur les variables</u>	90
1.	Les expressions	90
2.	Recodification des variables	94
3.	Les filtres de sélection	98

Démarrage du programme

Le lancement du programme se fait par le bouton « démarrer » de la barre des tâches :

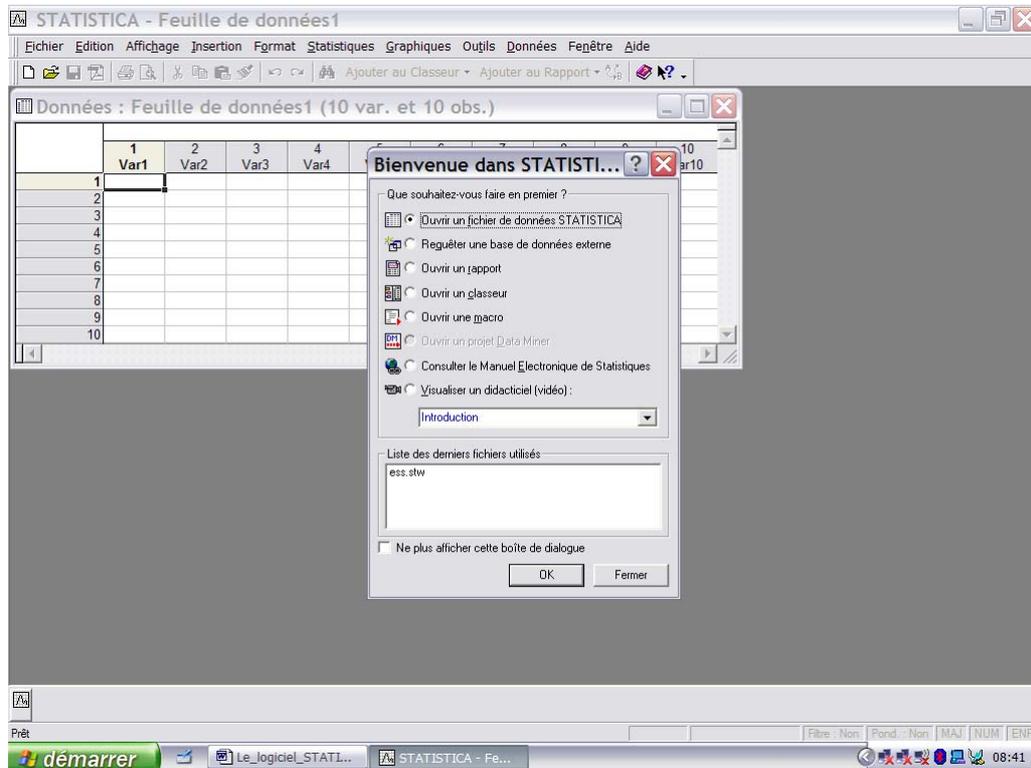


La présentation de l'écran de départ peut-être adaptée en modifiant les paramètres (cf. infra). Le menu « Aide » dans lequel on choisit « A propos de STATISTICA » affiche la version du logiciel et les caractéristiques de la licence :



1. Les standards d'une fenêtre Windows

Lorsqu'une application est active en mémoire centrale (RAM), l'ensemble des informations affichées à l'écran sont intégrées dans une **fenêtre générale standardisée** qu'on appelle la fenêtre Windows ; cette fenêtre est composée de **trois éléments** : la première ligne s'appelle le nom de la fenêtre, la deuxième est la barre des menus et la dernière la fenêtre des résultats.

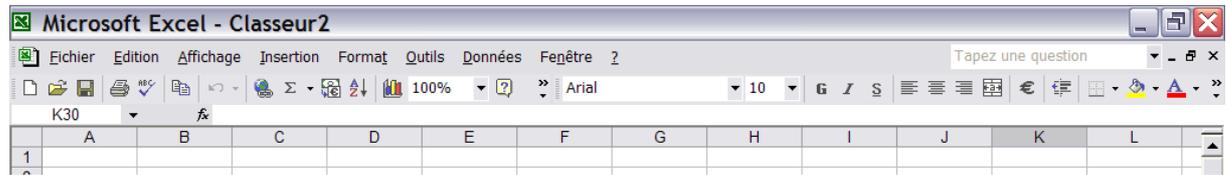


- STATISTICA-Feuille de données1 : nom de la fenêtre
« Feuille de données1 » est le nom par défaut donné par Statistica au fichier dans lequel il enregistrera les données. C'est aussi le nom de la feuille de données actives pour le logiciel ; à chaque ouverture d'une feuille, le nombre qui constitue le dernier caractère s'incrémente automatiquement d'une unité.
- FICHIER – EDITION... : Barre des Menus de l'application
Tous les logiciels de la famille Windows affichent cette deuxième ligne de fenêtre ; par exemple, en WORD :



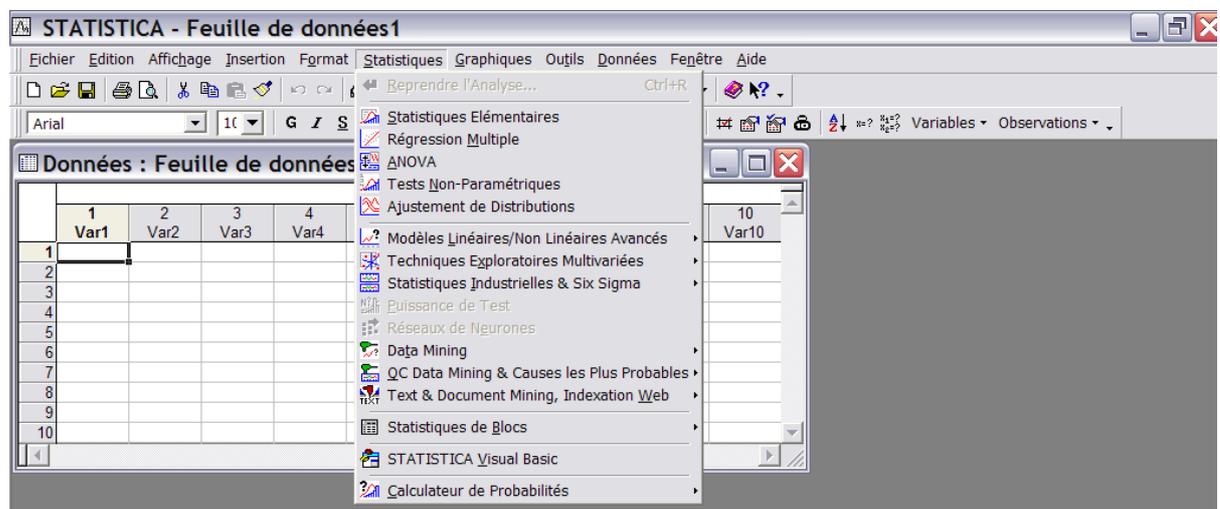
Les cinq premiers mots de la barre des menus sont identiques, de même que les deux derniers (Fenêtre-Aide ou ?)

Tous les mots intermédiaires sont spécifiques au logiciel utilisé. Par comparaison, en EXCEL :



(Question : que signifient « Document2 » dans Word et Classeur2 dans Excel ?)

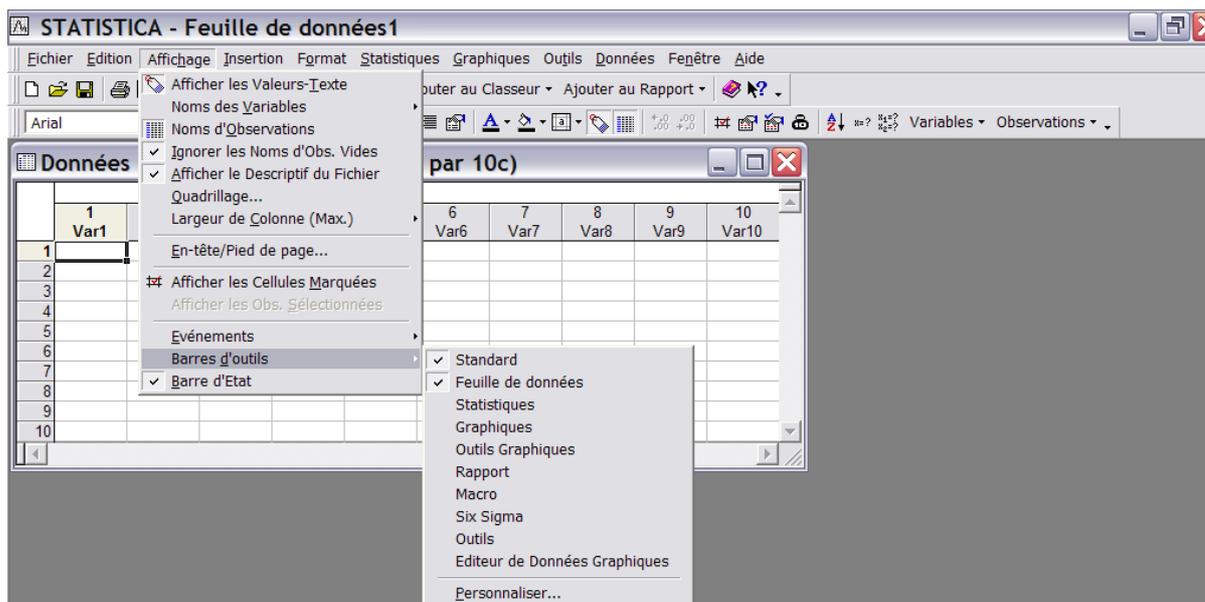
Un clic sur un mot de la barre des menus permet d'afficher les choix de traitement que celui-ci permet, par exemple en STATISTICA, le menu « Statistiques » permet l'accès aux procédures de traitements spécifiques au logiciel :



Il convient de faire remarquer que l'ouverture d'un sous-menu par activation d'un mot de la barre des Menus affiche des fonctionnalités différentes suivant le logiciel utilisé. Ainsi, le Menu « Fichier » en Word et en STATISTICA offre des choix pas forcément identiques même si le nom du Menu ouvert est le même.

- La fenêtre des résultats (la plus importante en taille) contient les informations gérées par l'application. Ci-dessus, en Statistica, il s'agit de l'élément intitulé « Données : Feuille de Données1 (10v par 10c) »

Tous les autres éléments affichables sont optionnels et accessibles grâce au Menu « Affichage », dont la fonctionnalité « Barre d'outils » est omniprésente dans tous les logiciels de la famille Windows.



Dans ce sous-menu d'affichage, on trouve l'option « Barre d'état » qui ajoute une ligne en bas de la fenêtre Windows ; cette ligne informe l'utilisateur de l'état d'une application en cours d'exécution.

Voici une exemple en Word :



Le curseur-souris est à la première ligne , première colonne de la page 1, section 1 ; le document ne comporte qu'une seule page. La marge est de 2,4cm et la langue est le néerlandais.

2. Utilisation du logiciel STATISTICA

STATISTICA rassemble un ensemble de fonctionnalités qui peuvent être regroupées en deux parties :

- Les fonctionnalités relatives à la gestion des données (en input et en output)
- Les fonctionnalités relatives aux traitements statistiques et graphiques de ces données

n.b. : On trouve également, comme dans tous les logiciels des fonctionnalités spécifiques de paramétrage

A. LA GESTION DES DONNEES EN STATISTICA

Tous les outils proposés vont permettre à l'utilisateur du logiciel de créer un fichier de données spécifiques à STATISTICA (le suffixe est « sta ») sur lequel les traitements statistiques et graphiques pourront être appliqués. Ce fichier sera soit encodé dans l'application même, soit importé d'un autre environnement (le cas le plus fréquent). Parmi les fonctionnalités, on trouve des modules d'encodage, de transformations de données, d'enregistrement, d'importation, d'exportation, d'impression de rapports, etc... Ces fonctionnalités sont accessibles dans les Menus suivants : Fichier – Edition – Outils – Données

Prenons l'exemple suivant :

On a rassemblé les notes de 10 étudiants concernant trois examens (Informatique – Droit – Sociologie), on dispose également des noms des étudiants, de leur année de naissance, s'il s'agit d'étudiants masculins ou féminins et s'ils apprécient leurs études.

A ce stade, bien qu'on dispose d'informations c'est-à-dire d'un certain degré de connaissance à propos d'un phénomène, nous n'avons pas encore de données qui permettent un traitement informatique. La structure passe par la définition des entités d'informations qu'on appelle aussi champs ou rubriques (fields) auxquelles on associera des repères (noms) et des caractéristiques (dont une essentielle consiste à savoir si elle peut être l'objet de calculs arithmétiques).

Dans l'exemple :

Entité	Nom
Nom d'étudiant	ETUDIANT
Note à l'examen d'Informatique	INFORMATIQUE
Note à l'examen de Droit	DROIT
Note à l'examen de Sociologie	SOCIOLOGIE
Sexe	SEXE
Degré d'appréciation des études	APPRECIATION
Année de naissance	NAISSANCE

Tableau 1

Un étudiant particulier pourrait être :

NOM : Jane

INFORMATIQUE : 14

DROIT : Bien

SOCIOLOGIE : 15

SEXE : F

APPRECIATION : 2

Il est donc nécessaire de connaître les règles d'attributions qui déterminent les différents niveaux de chacune des rubriques ; celles-ci sont résumées dans le tableau suivant :

Nom	Caractéristique
ETUDIANT	Texte (maximum 30 caractères)
INFORMATIQUE	Valeur numérique entière comprise entre 0 et 20
DROIT	Très Bien – Bien - Insuffisant
SOCIOLOGIE	Valeur numérique avec une décimale
SEXE	H - F
APPRECIATION	Texte correspondant à une valeur comprise entre 1 et 4 1 : Très Bonne 2 : Bonne 3 : Passable 4 : Mauvaise
NAISSANCE	Valeur de l'année

Tableau 2

On s'aperçoit que trois rubriques sont susceptibles de faire l'objet d'opérations arithmétiques : les notes des cours d'INFORMATIQUE et de SOCIOLOGIE, ainsi que l'année de naissance, on leur attribue la caractéristique « numérique » ; pour le reste, il s'agit de rubriques « caractères ».

Dans le cas de rubriques à niveaux ordonnés (DROIT, APPRECIATION), et pour des facilités évidentes d'encodage, on remplacera plutôt le texte par une valeur correspondant au niveau de l'échelle ; pour la rubrique DROIT, « Très Bien » sera égal à 3, « Bien » à 2 et « Insuffisant » à 1. Pour la rubrique APPRECIATION, les codes numériques sont indiqués dans le tableau.

Ce remplacement n'a pas pour conséquence de pouvoir réaliser des opérations arithmétiques sur ces rubriques.

Voici les données complètes de l'exemple :

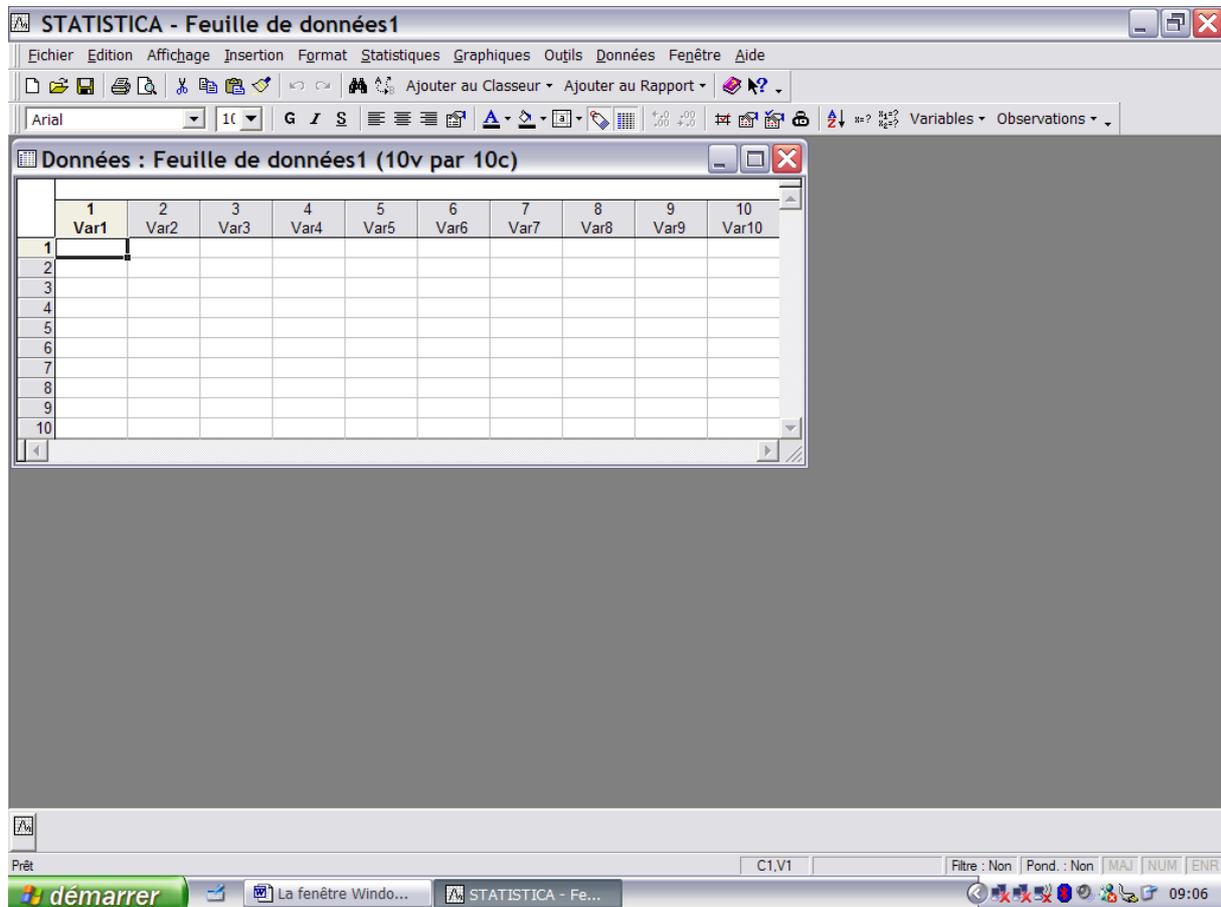
NOM	INFORMATIQUE	DROIT	SOCIOLOGIE	SEXE	APPRECIATION	NAISSANCE
Jane	12	2	14.5	F	2	1985
Helen	14	1	19	F	2	1986
Mike	10	3	12	H	3	1986
Tom	18	1	14	H	1	1983
Jim		2		H	3	1985
Paul	13	2	11.5	H	2	1984
Margret	15	1	16	F	2	1982
Kay	14	3	15	F	1	1984
Lucile	12		14	F		1985
Emma	18	2	17.5	F	1	1982

Tableau 3

n.b. : les cellules vides correspondent à des informations manquantes

1. Création d'une feuille de données

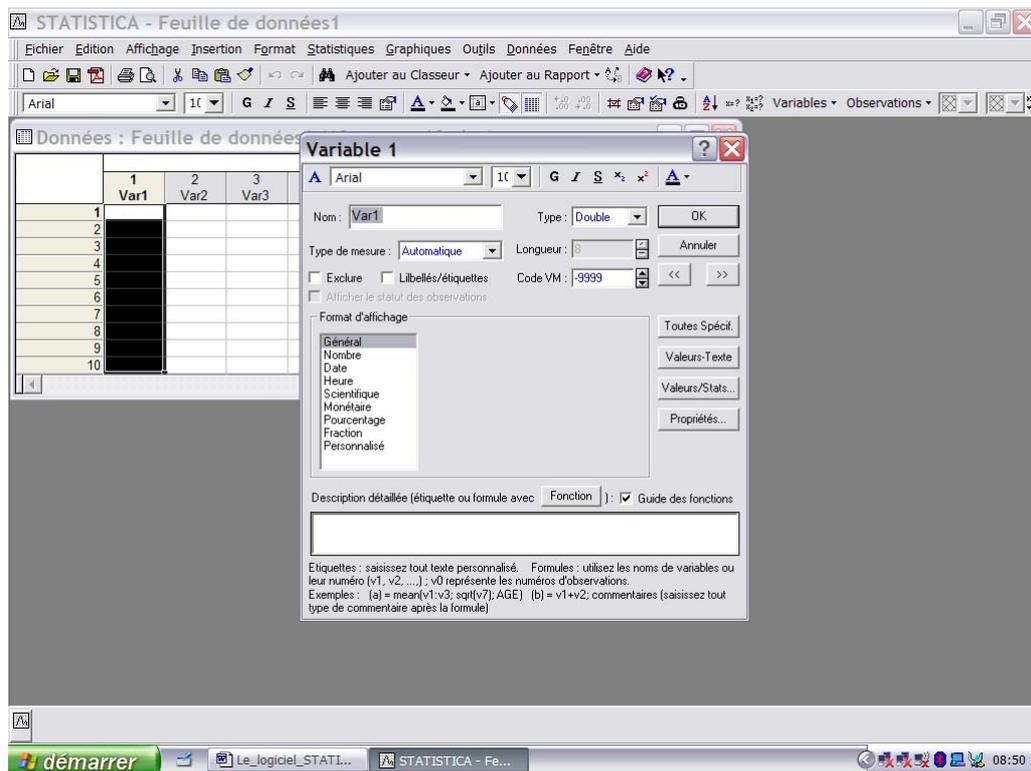
L'exécution du programme STATISTICA affiche la fenêtre suivante :



La fenêtre « Données : Feuille de données1 (10v par 10c) » représente un tableau dans lequel les observations se trouvent en lignes (numérotées 1 à 10) et les rubriques en colonnes (numérotées 1 à 10 et nommées Var1 à Var10). Cette fenêtre est la fenêtre utilisée pour l'encodage des données.

La cellule active (point de rencontre d'une ligne et d'une colonne) est marquée par un rectangle en gras ; on déplace celle-ci au moyen de la souris ou des touches « flèches cardinales » ou encore au moyen de « Enter », voire des touches « Tabulations ». Une entrée de données au clavier remplit la cellule active qu'on efface par l'entrée d'un espace.

Un double clic sur le nom d'une colonne affiche la carte d'identité de la variable (en Statistique, une rubrique ou un champs dans un fichier s'appelle une variable) ; ainsi, pour Var1, cela fournit les informations suivantes ; on obtient le même résultat en choisissant « Spécifications de la variable » du menu « Données » :



La première ligne détaille les propriétés typographiques utilisées pour l’affichage du nom de la variable (ci-dessus, Var1) ; la zone qui commence ensuite est plus intéressante :

- Nom : nom de la variable, par défaut, STATISTICA attribue comme nom de variable le texte « Var » suivi d’une valeur numérique commençant à la valeur 1 et s’incrémentant de 1 dans l’ordre des colonnes.
- Type : la variable contient des valeurs numériques ou du texte (par défaut,) ; quatre types sont possibles : « Double » signifie valeur numérique décimale (valeur prise par défaut) « texte » signifie variable contenant des caractères, « entier » impose des valeurs numériques sans décimales et « octet » idem que « entier » mais dont la valeur ne peut pas dépasser 255
- Type de Mesure : correspond à la typologie des variables statistiques (Continue - Catégorielle – Ordinale) autres choix : Non-Spécifié ou Automatique ; cette fonctionnalité permet de ne retenir dans la liste de sélection des variables celles qui ont le type approprié pour l’analyse
- Longueur : le nombre de caractères retenus pour une variable de type texte (par défaut : 8)
- Exclure : permet de supprimer l’apparition du nom de la variable dans les menus de choix de variables
- Libellés/Etiquettes : permet de définir une variable qui contiendra des étiquettes qui seront affichées dans les graphiques par exemple.

- Code des VM : valeur numérique indiquant la valeur numérique des observations manquantes (missing values); par défaut, STATISTICA considère que toutes les valeurs égales à -9999 ne participent pas aux calculs statistiques. Le code VM des variables de type texte est l'espace qui peut-être utilisé également pour les variables numériques.
- Afficher le statut des observations : permet d'associer des symboles aux observations pour une meilleure compréhension des résultats graphiques

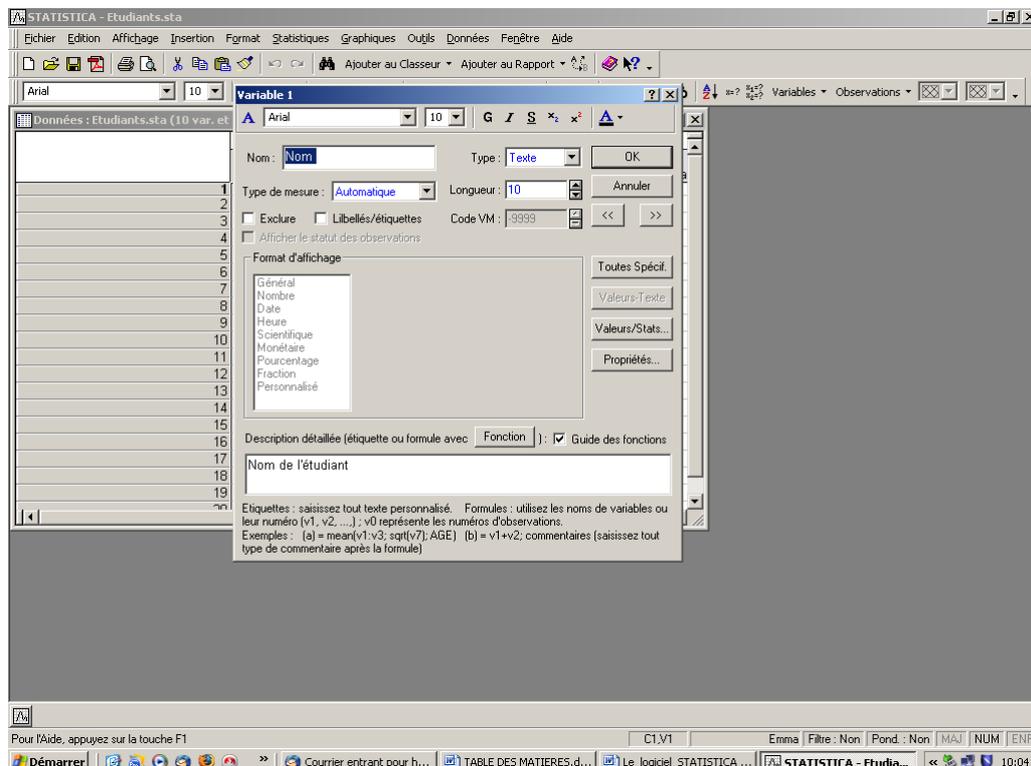
Sous « Annuler », on trouve les boutons << et >> qui permettent d'afficher les cartes d'identité des différentes colonnes du tableau.

Le bloc « Format d'Affichage » fournit une liste de toutes les possibilités qui contrôlent l'édition des valeurs des variables.

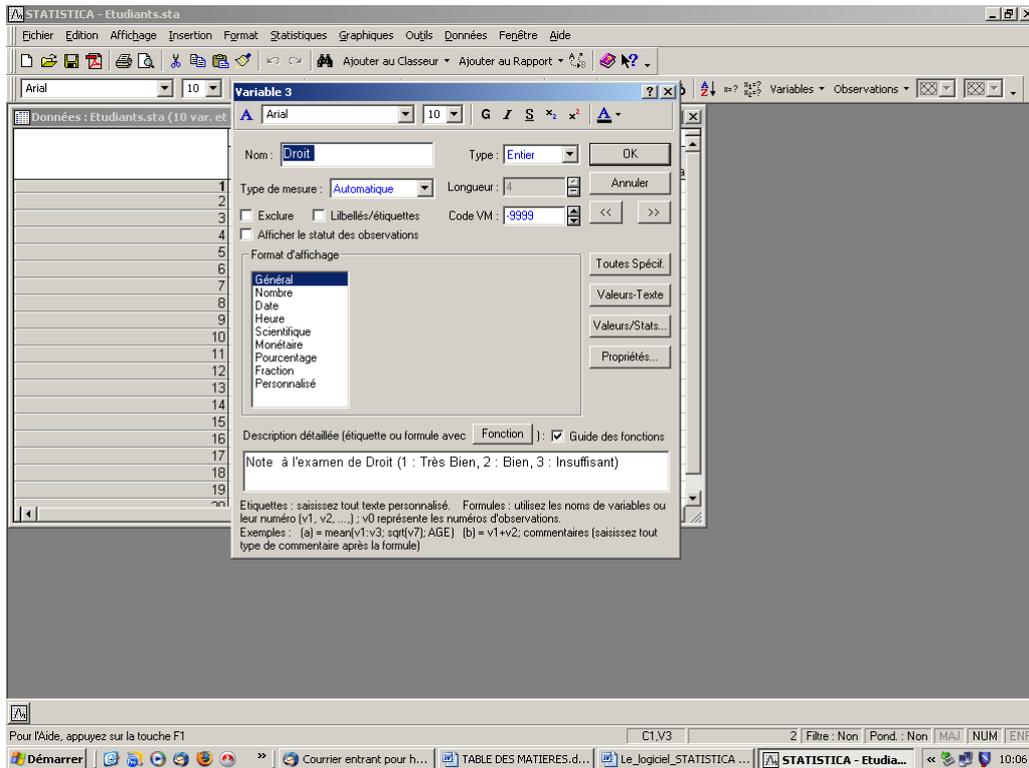
Le cadre « Description détaillée » permet d'encoder des informations et des commentaires détaillés à propos de la variable. Si le premier caractère est le signe « = », on définit alors la variable comme étant le résultat de l'évaluation d'une expression (cf. Infra).

Dans notre exemple, la première tâche consiste à spécifier les différentes variables qui seront encodées et leur type d'après le tableau 2, par exemple, pour les variables NOM, DROIT et SOCIOLOGIE :

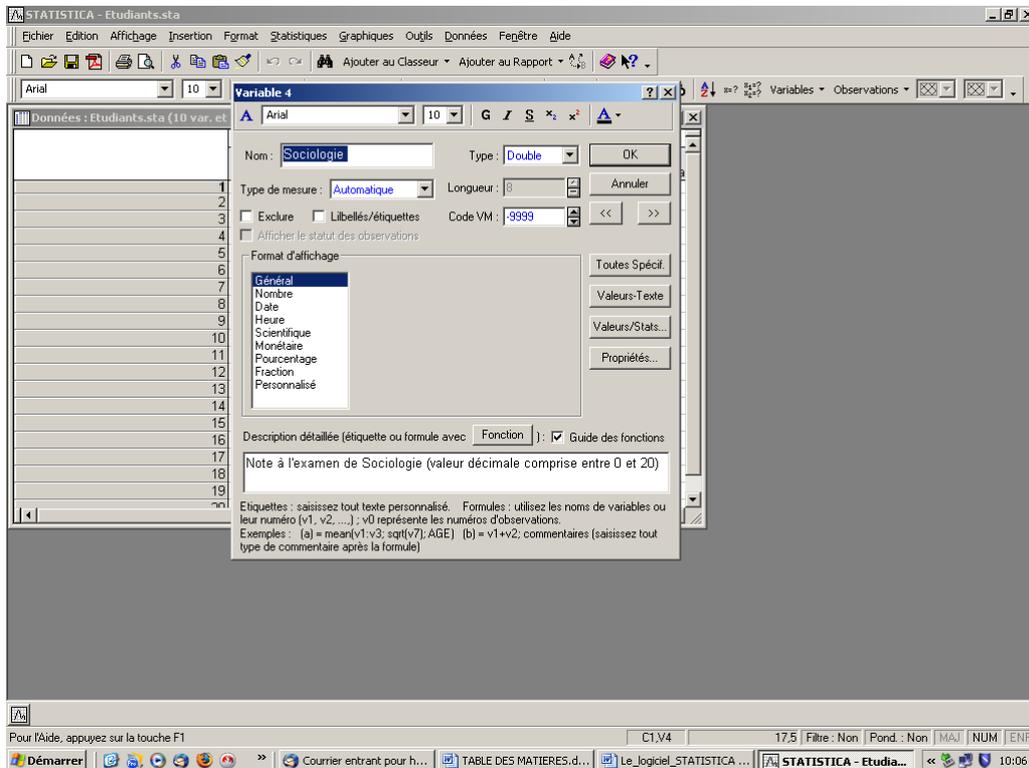
Pour la variable NOM :



Pour la variable DROIT :



Et pour la variable SOCIOLOGIE :

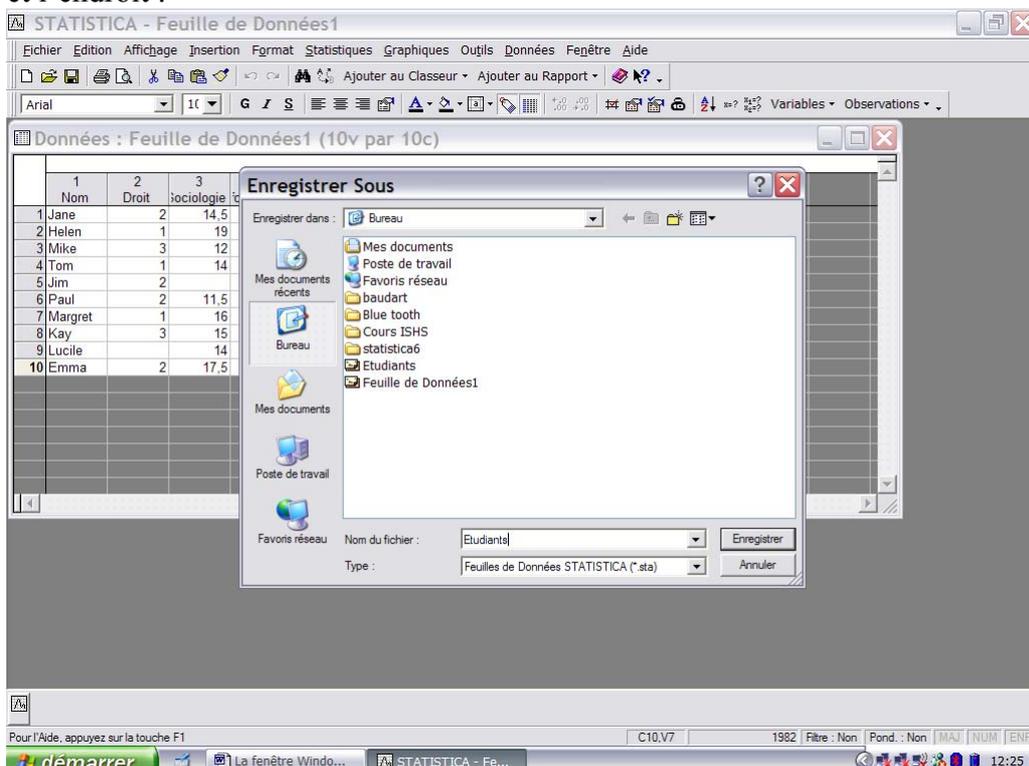


Lorsque toutes les spécifications de variables ont été établies, l'encodage peut commencer et le tableau en format STATISTICA est alors le suivant :

Données : Feuille de Données1 (10v par 10c)

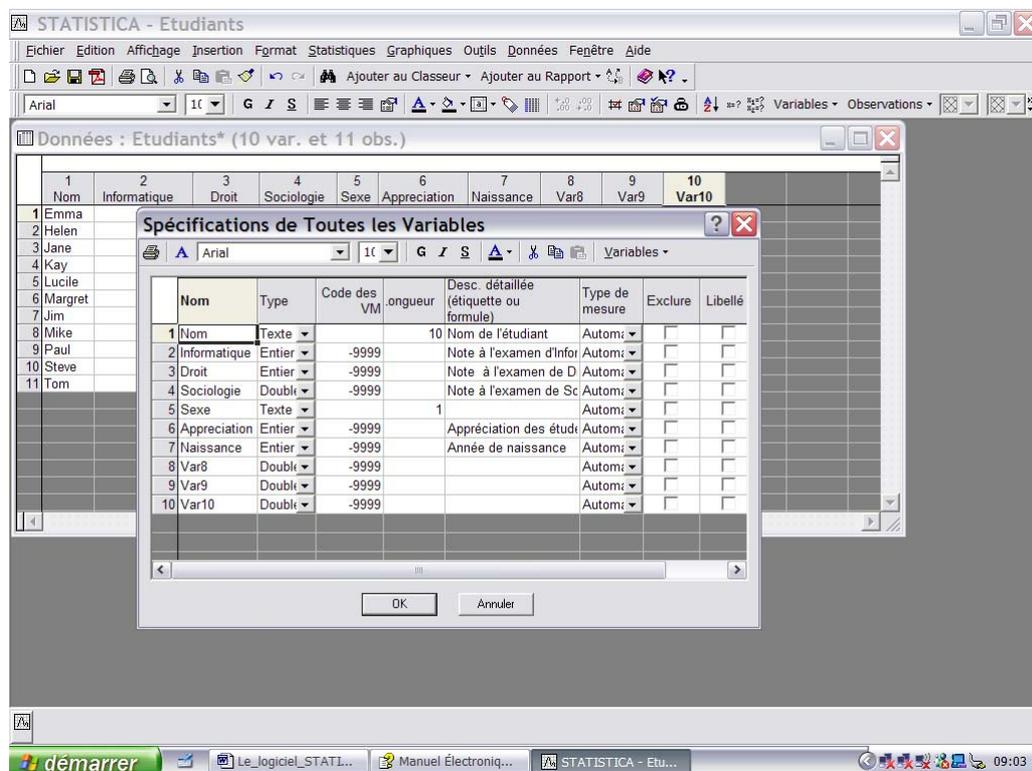
	1	2	3	4	5	6	7	8	9	10				
	Nom	Droit	sociologie	formatique	Sexe	appreciation	Naissance	Var8	Var9	Var10				
1	Jane	2	14,5		12 F		2	1985						
2	Helen	1	19		14 F		2	1986						
3	Mike	3	12		10 H		3	1986						
4	Tom	1	14		18 H		1	1983						
5	Jim	2			H		3	1985						
6	Paul	2	11,5		13 H		2	1984						
7	Margret	1	16		15 F		2	1982						
8	Kay	3	15		14 F		1	1984						
9	Lucile		14		12 F			1985						
10	Emma	2	17,5		18 F		1	1982						

Une règle de bonne pratique consiste à enregistrer un fichier qui a été l'objet de modifications ; nous allons enregistrer nos données dans un fichier dont le nom est « Etudiant » sur le Bureau de Windows dans un format spécifique à STATISTICA. Dans le Menu « Fichier », on sélectionne « Enregistrer sous » et on indique le nom du fichier et l'endroit :



Il convient de remarquer que la feuille de données STATISTICA porte désormais le nom donné au fichier.

Le Menu « Données » dans lequel on choisit « Spéc. de toutes les Variables » fournit les spécifications complètes du tableau des données :



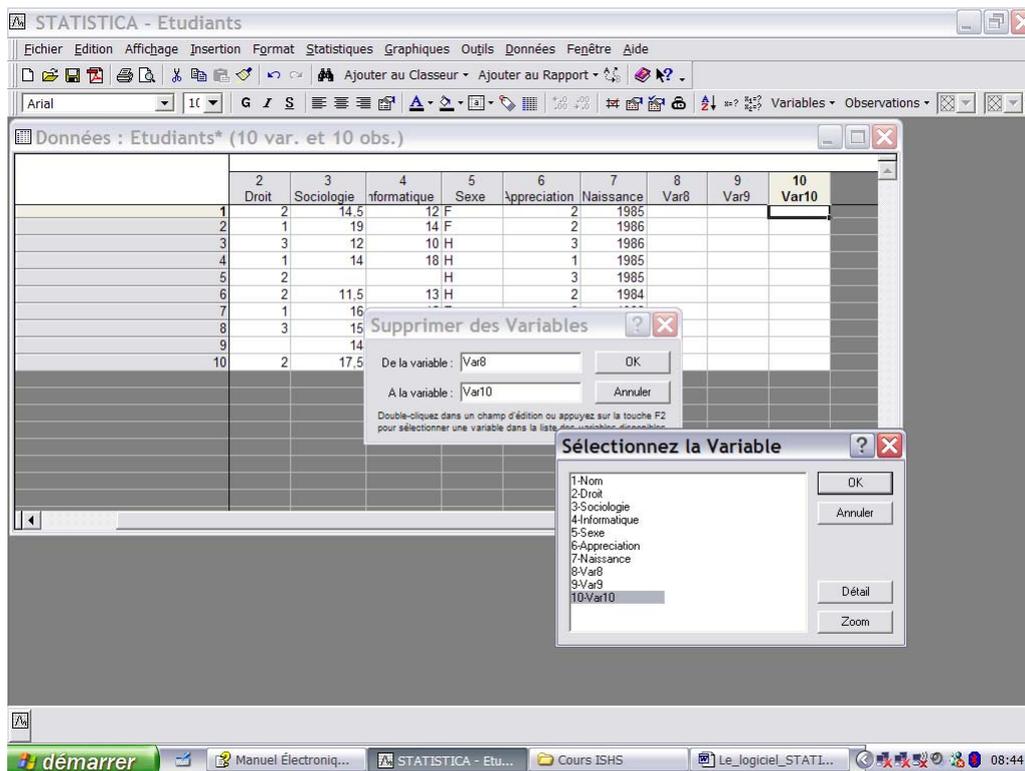
Dans la version française du logiciel, les valeurs décimales sont obligatoirement encodées avec la virgule décimale et non le point ; on peut utiliser des noms de variables identiques car STATISTICA repère et gère les variables en fonction d'un numéro d'ordre qu'il leur attribue. Une fois remplie, une cellule peut être l'objet d'une modification quia affecte une partie de son contenu ; pour éviter de réencoder complètement sa valeur, on « édite » son contenu par un double clic souris qui permet ensuite de déplacer le curseur à l'intérieur de cette cellule et de la modifier.

Une sélection de colonne consiste en un clic sur son nom ; un double clic sur le nom de colonne affiche les spécifications de la variable. Une sélection d'une ligne consiste en un clic sur son numéro.

2. Modifications dans la feuille de données

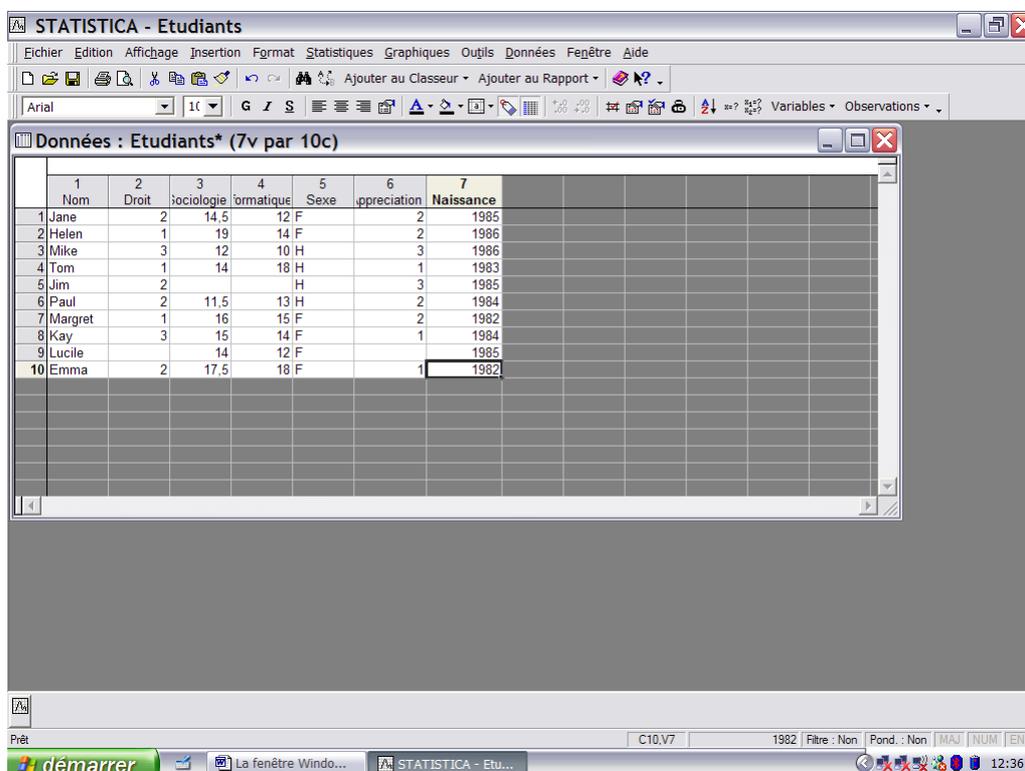
a. Suppression de variables

A la création d'une nouvelle feuille de données, STATISTICA propose un tableau de 10 observations et 10 variables ; dans notre exemple, on peut se passer des colonnes Var8, Var 9 et Var 10. Pour effectuer cette opération, on choisit dans le menu « Edition » la fonctionnalité « Supprimer » puis « Variables » qui affiche l'écran suivant (on a aussi utilisé la touche de fonction F2)



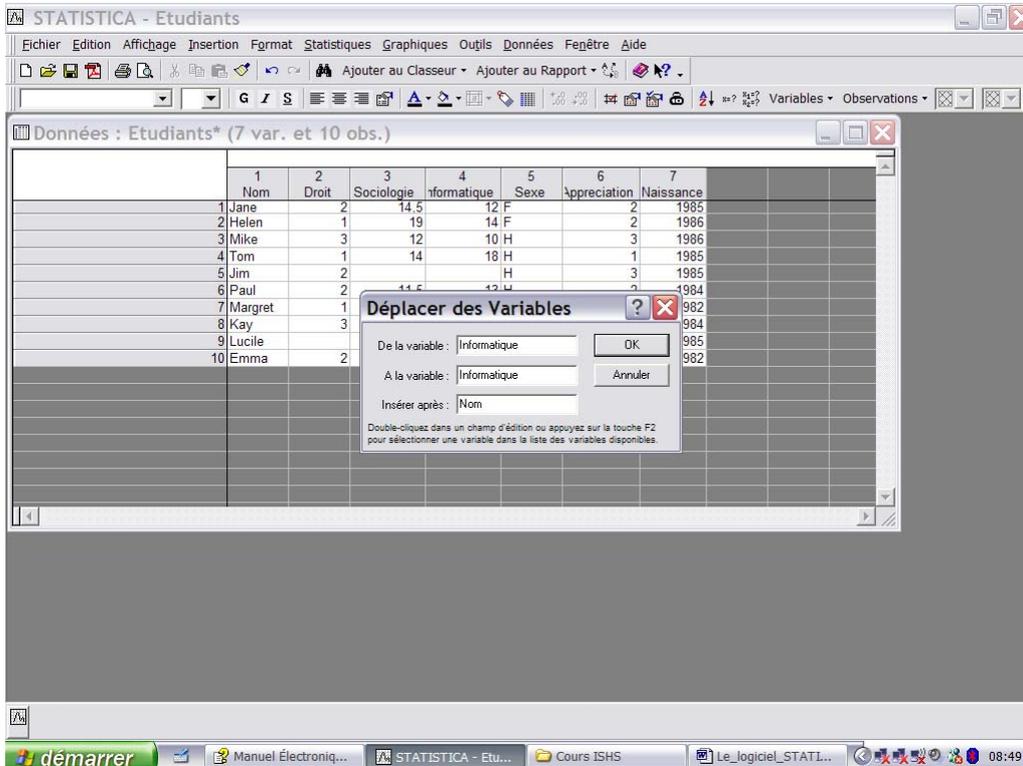
On a alors le choix de supprimer variable par variable ou un groupe de variables à la fois, nous avons choisi cette possibilité en demandant de supprimer de la variable Var8 à Var10.

La feuille de données se présente finalement comme suit :

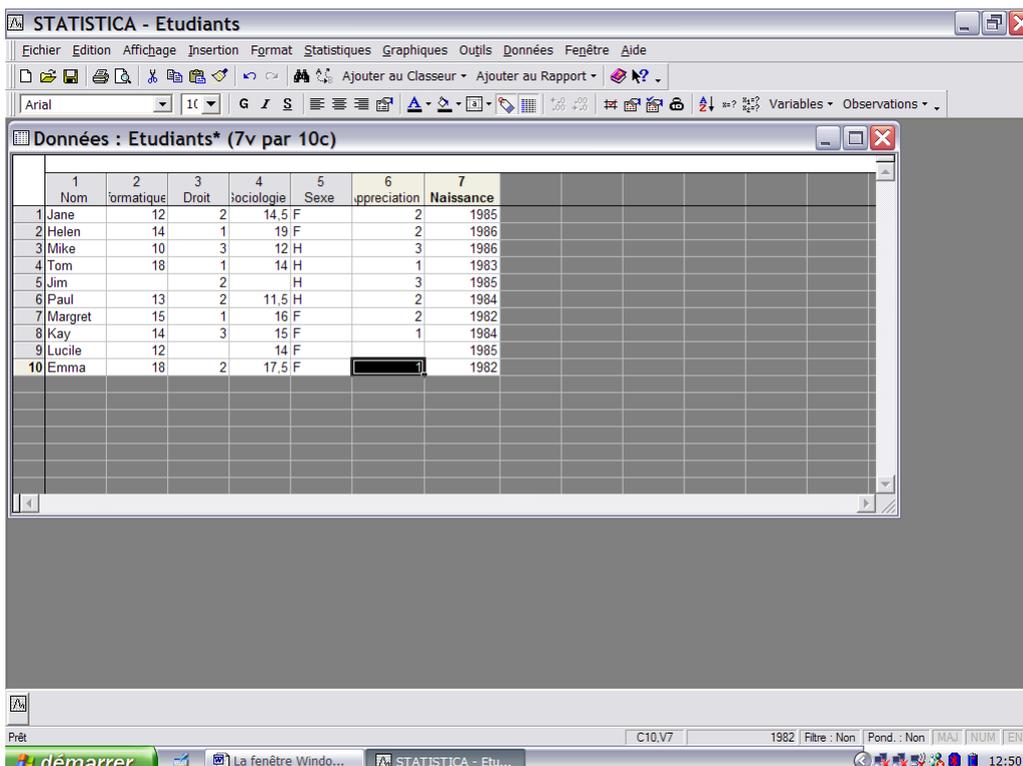


que l'on enregistre.

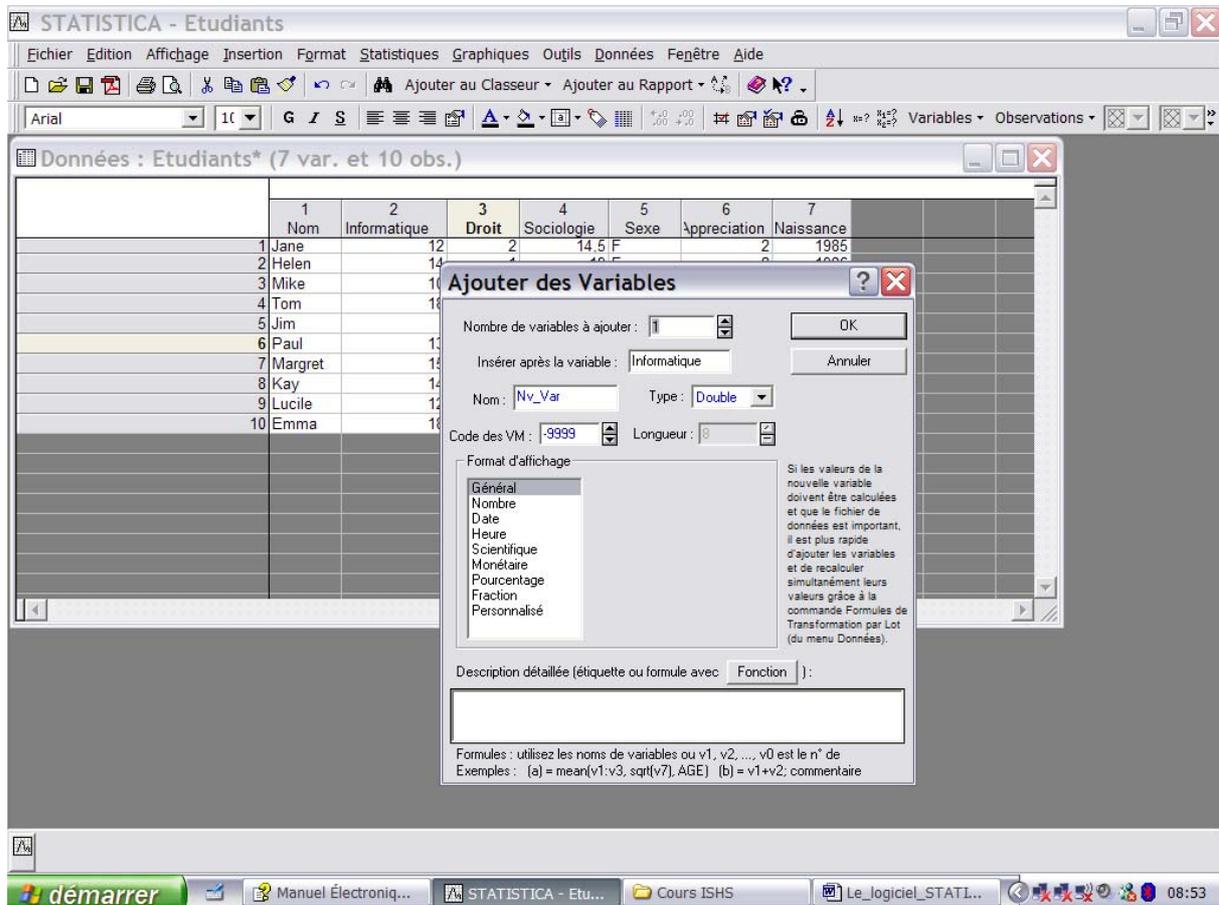
- b. L'ordre des variables de la feuille de données n'est pas tout à fait conforme à celui du Tableau 3, pour cela, il convient de déplacer la variable « Informatique » avant la variable « Droit», pour cela, on choisit dans le menu « Edition » l'option « Déplacer » puis « Variable », on spécifie la variable à déplacer (ou le groupe s'il s'agit d'en déplacer plusieurs) et l'endroit d'insertion :



pour retrouver, en format STATISTICA, les données de notre exemple de départ :

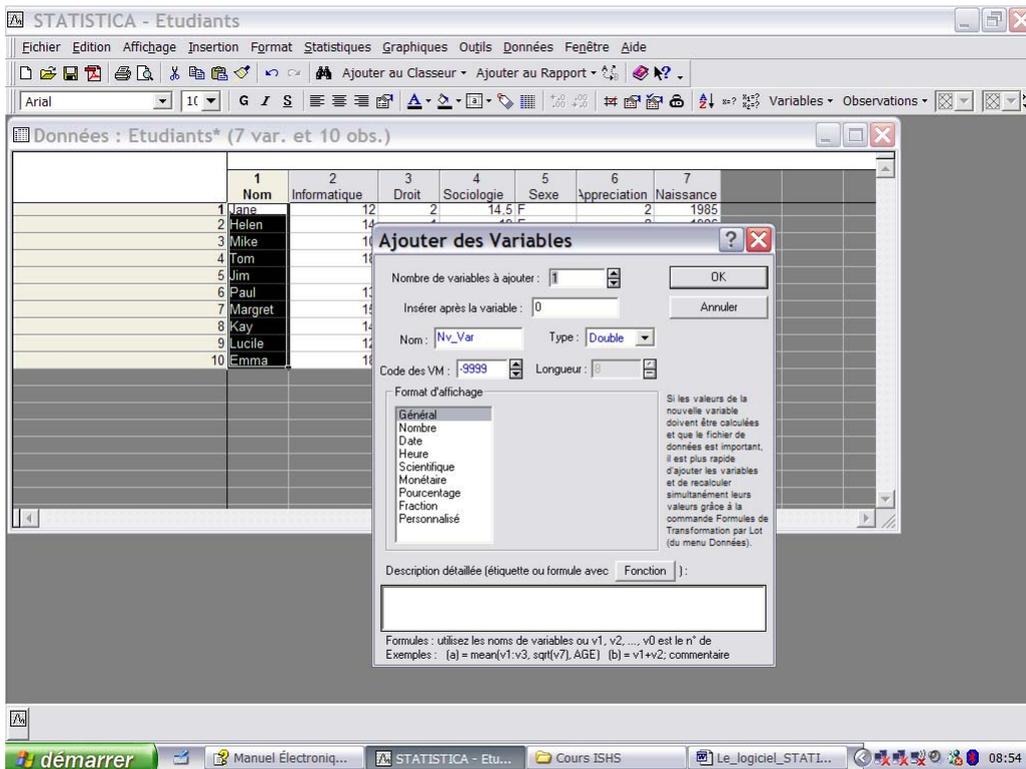


- c. L'addition de nouvelles colonnes est possible grâce au menu « Insertion » et l'option « Ajouter des Variables » qui affiche un écran :

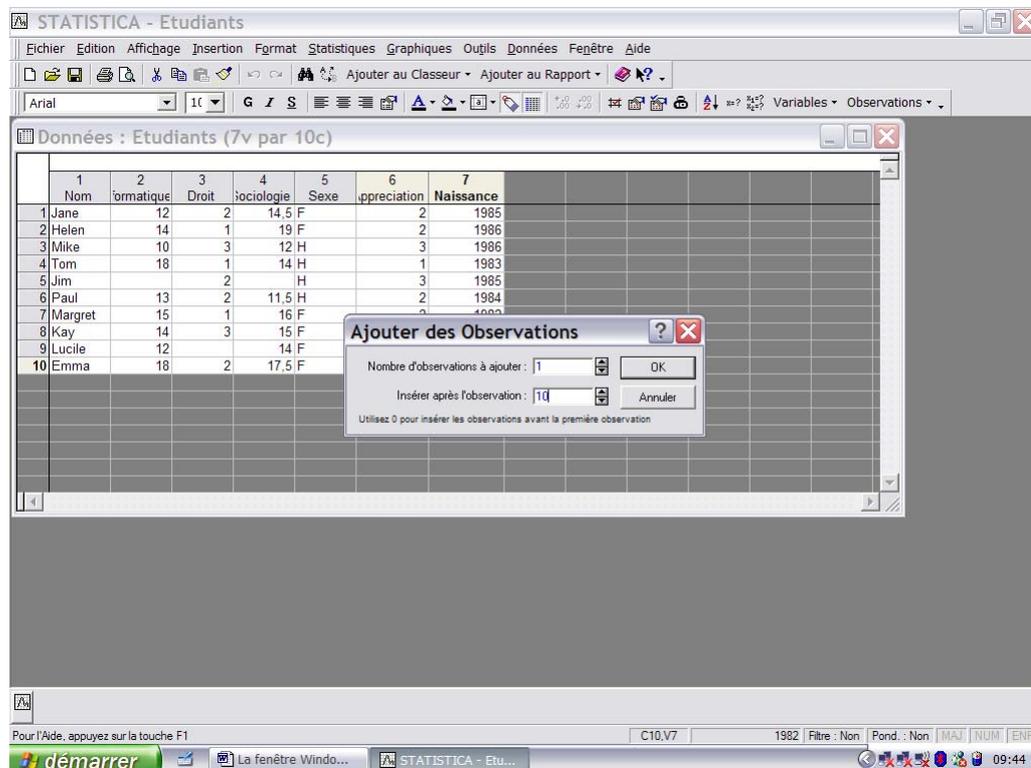


Dans cet exemple, on peut constater que l'ajoute de variable affiche le tableau des spécifications de la (ou des) variable additionnelle(s), l'endroit du tableau où elle(s) sera(ont) insérées. Le nom est choisi par défaut comme étant « Nv_Var » éventuellement complété par un chiffre qui s'incrémente automatiquement de 1 en cas de création de plusieurs variables.

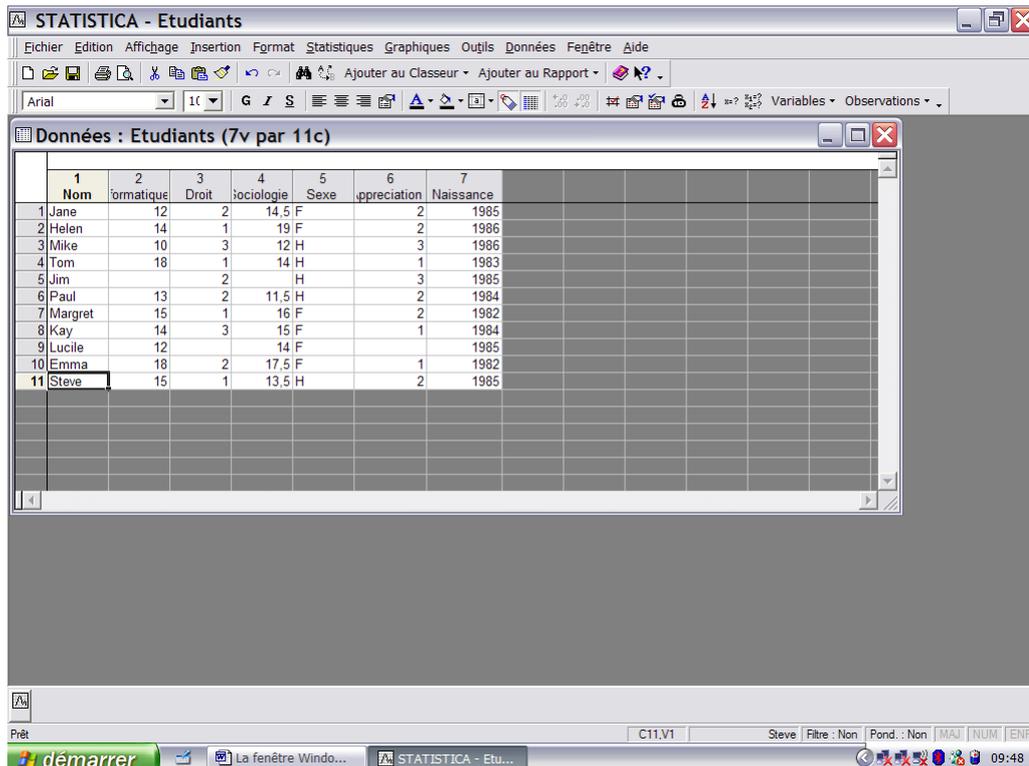
Si on désire insérer une ou plusieurs variables avant la première colonne (Nom), on clique une fois sur la première colonne et on utilise comme indiqué « Insertion » puis « Ajouter des variables » :
(on note que l'insertion est réalisée après la variable fictive 0)



d. On peut compléter les données en ajoutant des observations supplémentaires ; le menu « Insertion » complété par l'option « Ajouter des observations » affiche l'écran suivant :



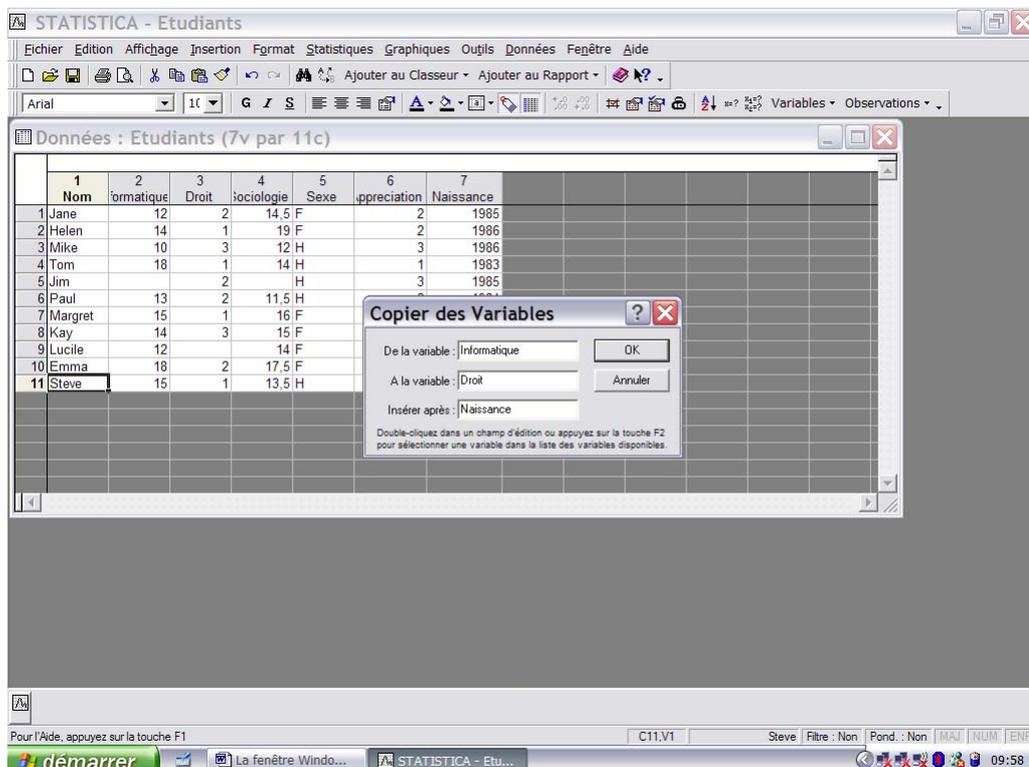
complète la feuille de données en ajoutant une 11^{ème} ligne correspondant à un étudiant supplémentaire :



n.b. : on remarque que la dimension de la feuille est passé à (7v par 11c)

e. Les fonctionnalités d'addition de variables ou d'observations du menu « Insertion » peuvent être complétées par la fonctionnalité « Copier - Coller ».

Par exemple, si on désire ajouter deux variables après la colonne « Naissance » par recopie des variables « Informatique » et « Droit » :



fournira la feuille de données complétée de la manière suivante :

	1	2	3	4	5	6	7	8	9
	Nom	formatique	Droit	sociologie	Sexe	appreciation	Naissance	formatique	Droit
1	Jane	12	2	14.5	F	2	1985	12	2
2	Helen	14	1	19	F	2	1986	14	1
3	Mike	10	3	12	H	3	1986	10	3
4	Tom	18	1	14	H	1	1983	18	1
5	Jim	2	2		H	3	1985		2
6	Paul	13	2	11.5	H	2	1984	13	2
7	Margret	15	1	16	F	2	1982	15	1
8	Kay	14	3	15	F	1	1984	14	3
9	Lucile	12		14	F		1985	12	
10	Emma	18	2	17.5	F	1	1982	18	2
11	Steve	15	1	13.5	H	2	1985	15	1

La recopie de lignes fonctionne de manière tout à fait symétrique : on recopie les trois premières lignes après la onzième grâce au menu « Insertion » « Copier des observations » :

	1	2	3	4	5	6	7	8	9
	Nom	formatique	Droit	sociologie	Sexe	appreciation	Naissance	formatique	Droit
1	Jane	12	2	14.5	F	2	1985	12	2
2	Helen	14	1	19	F	2	1986	14	1
3	Mike	10	3	12	H	3	1986	10	3
4	Tom	18	1	14	H	1	1983	18	1
5	Jim	2	2		H	3	1985		2
6	Paul	13	2	11.5	H	2	1984	13	2
7	Margret	15	1	16	F	2	1982	15	1
8	Kay	14	3	15	F	1	1984	14	3
9	Lucile	12		14	F		1985	12	
10	Emma	18	2	17.5	F	1	1982	18	2
11	Steve	15	1	13.5	H	2	1985	15	1

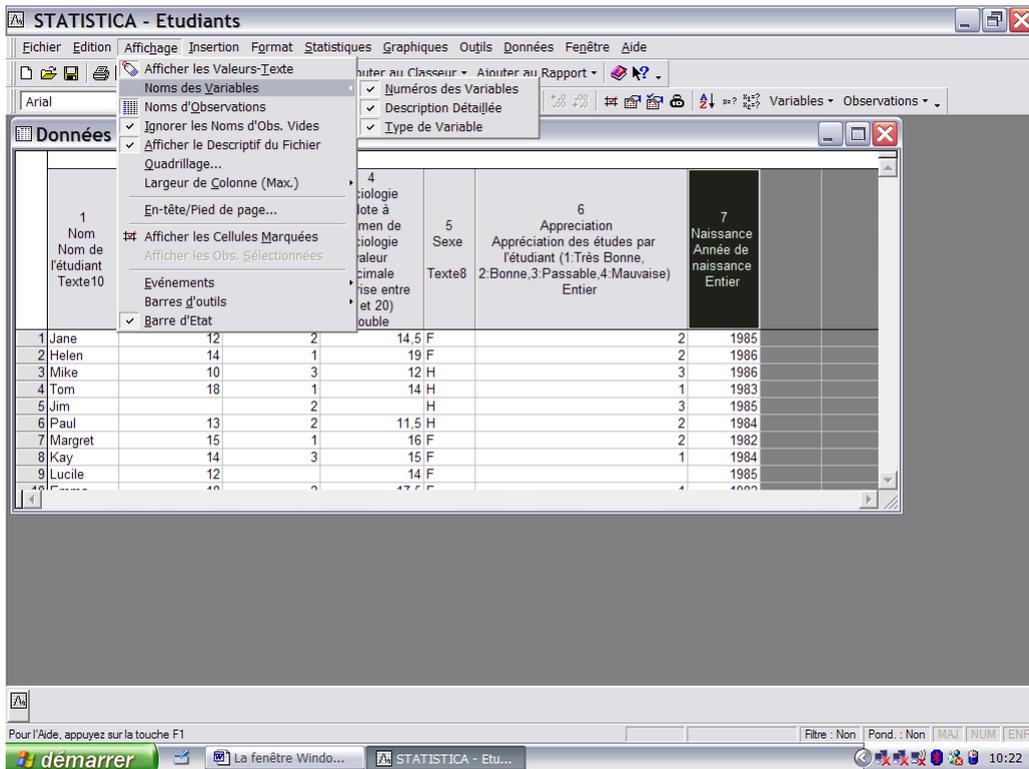
On obtient finalement une feuille de données :

	1	2	3	4	5	6	7	8	9									
	Nom	Informatique	Droit	Sociologie	Sexe	Appréciation	Naissance	Informatique	Droit									
1	Jane	12	2	14,5	F	2	1985	12	2									
2	Helen	14	1	19	F	2	1986	14	1									
3	Mike	10	3	12	H	3	1986	10	3									
4	Tom	18	1	14	H	1	1983	18	1									
5	Jim		2		H	3	1985		2									
6	Paul	13	2	11,5	H	2	1984	13	2									
7	Margret	15	1	16	F	2	1982	15	1									
8	Kay	14	3	15	F	1	1984	14	3									
9	Lucile	12		14	F		1985	12										
10	Emma	18	2	17,5	F	1	1982	18	2									
11	Steve	15	1	13,5	H	2	1985	15	1									
12	Jane	12	2	14,5	F	2	1985	12	2									
13	Helen	14	1	19	F	2	1986	14	1									
14	Mike	10	3	12	H	3	1986	10	3									

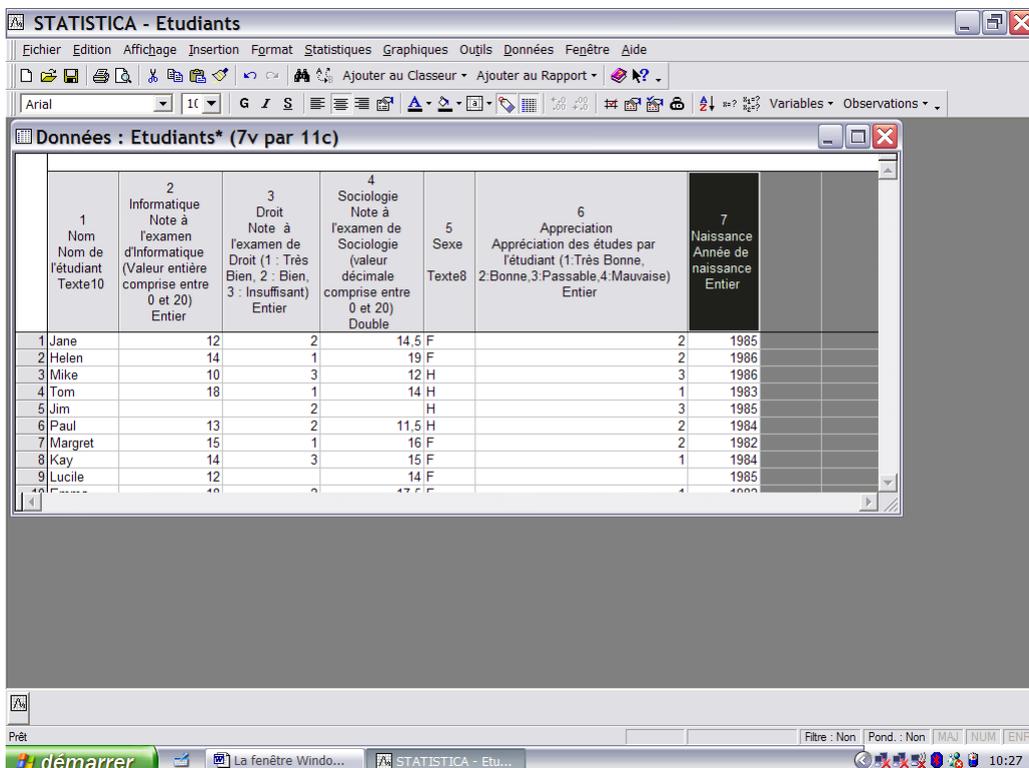
où l'on constate à nouveau que les dimensions ont été adaptées.

- f. L'affichage (ne pas confondre avec la longueur déclarée des variables !) des noms de colonnes est réalisé par défaut dans des cellules de 8 caractères, ce qui amène à tronquer les noms de variables qui dépassent cette longueur (dans notre exemple, la colonne « Informatique ») ; l'adaptation de la largeur des colonnes se fait aisément grâce au pointeur souris que l'on positionne sur la frontière de cellule et qu'on déplace en enfonçant le bouton droit.

Le menu « Affichage » et l'option « Nom des variables » dans lequel on sélectionne « Description détaillée » et « Type de variable » permet d'obtenir dans la feuille de données même toutes les informations spécifiques aux colonnes du tableau, par exemple :



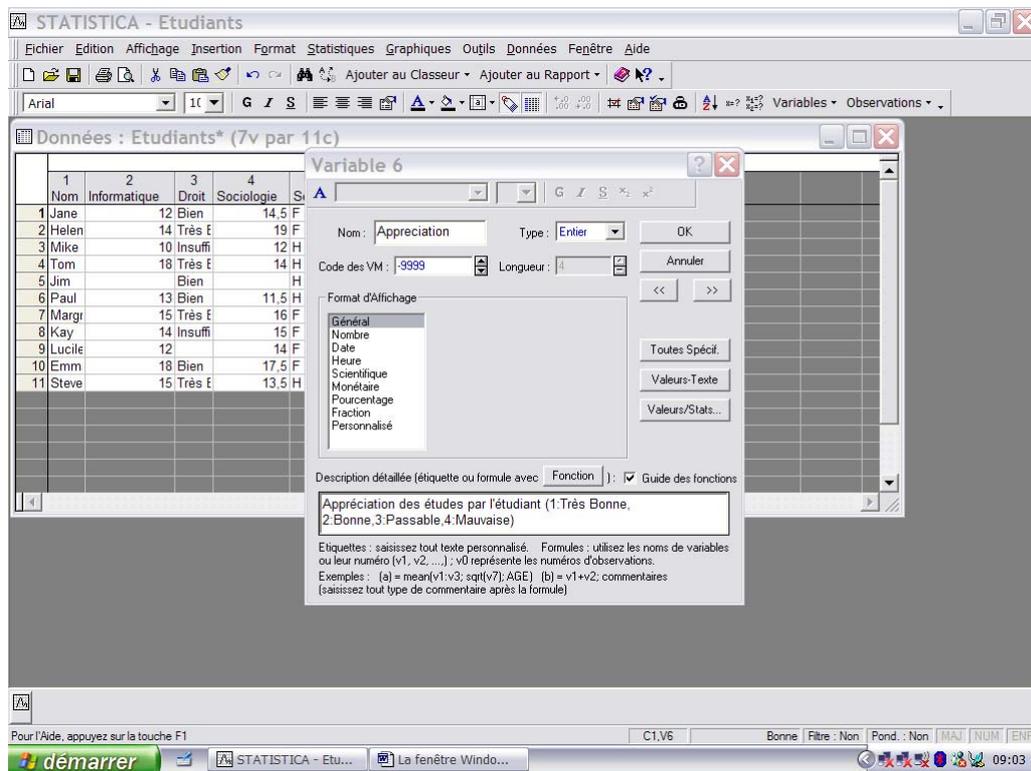
permet d'obtenir la feuille de données complétée par toutes les spécifications insérées lors de la création des variables du tableau :



- g. Les variables « Droit » et « Appréciation » ont été encodées en utilisant des valeurs numériques entières qui remplacent en réalité les véritables scores attribués aux différents états de ces variables ; pour rappel :

- « Droit » prend 3 valeurs : 1 = Très Bien, 2 = Bien, 3 = Insuffisant
- « Appreciation » prend 4 valeurs : 1 = Très Bonne, 2 = Bonne, 3 = Passable, 4 = Mauvaise

Lorsqu'on réalise des traitements statistiques sur ces variables, il est très intéressant de conserver ces valeurs de départ pour la bonne lisibilité des résultats ; par exemple, à la question portant sur l'évaluation des leurs études, observer que « 5 étudiants ont répondu Très Bonne » est évidemment plus parlant que « 5 étudiants ont répondu 1 ». STATISTICA permet d'associer des valeurs numériques et des valeurs de type texte dans les spécifications des variables. Pour la variable « Appreciation », on édite ses spécifications par un double clic sur le nom de colonne (ou par la menu « Données » et l'option « Spécification de la variable » après sa sélection par un clic souris), pour la variable « Appreciation » :



On choisit ensuite « Valeurs-Texte » et on remplit les colonnes « Valeurs-texte », « V. Numériques » comme indiqué sur l'écran. La colonne « Etiquette » permet d'ajouter des détails supplémentaires concernant les modalités de la variable. Le bouton « Trier » permet d'ordonner les différentes valeurs entrées en ordre croissant ou décroissant. Si plusieurs variables du fichier ont des modalités identiques (p. ex. « Très Bien » - « Bien » - « Mauvais »), les boutons « Copier depuis... » et « Appliquer à... » permettent d'éviter de devoir réencoder les valeurs pour chacune en récupérant ou en appliquant les étiquettes déjà utilisées.

Dans la feuille de données, on peut afficher les valeurs-texte des variables par le menu « Affichage » et « Afficher les valeurs-textes » :

Données : Etudiants* (7v par 11c)

1	2	3	4	5	6	7
Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance
1	Jane	12 Bien	14,5	F	Bonne	1985
2	Helen	14 Très É	19	F	Bonne	1986
3	Mike	10 Insuffi	12	H	Passable	1986
4	Tom	18 Très É	14	H	Très Bonne	1983
5	Jim	Bien		H	Passable	1985
6	Paul	13 Bien	11,5	H	Bonne	1984
7	Margr	15 Très É	16	F	Bonne	1982
8	Kay	14 Insuffi	15	F	Très Bonne	1984
9	Lucile	12	14	F		1985
10	Emm	18 Bien	17,5	F	Très Bonne	1982
11	Steve	15 Très É	13,5	H	Bonne	1985

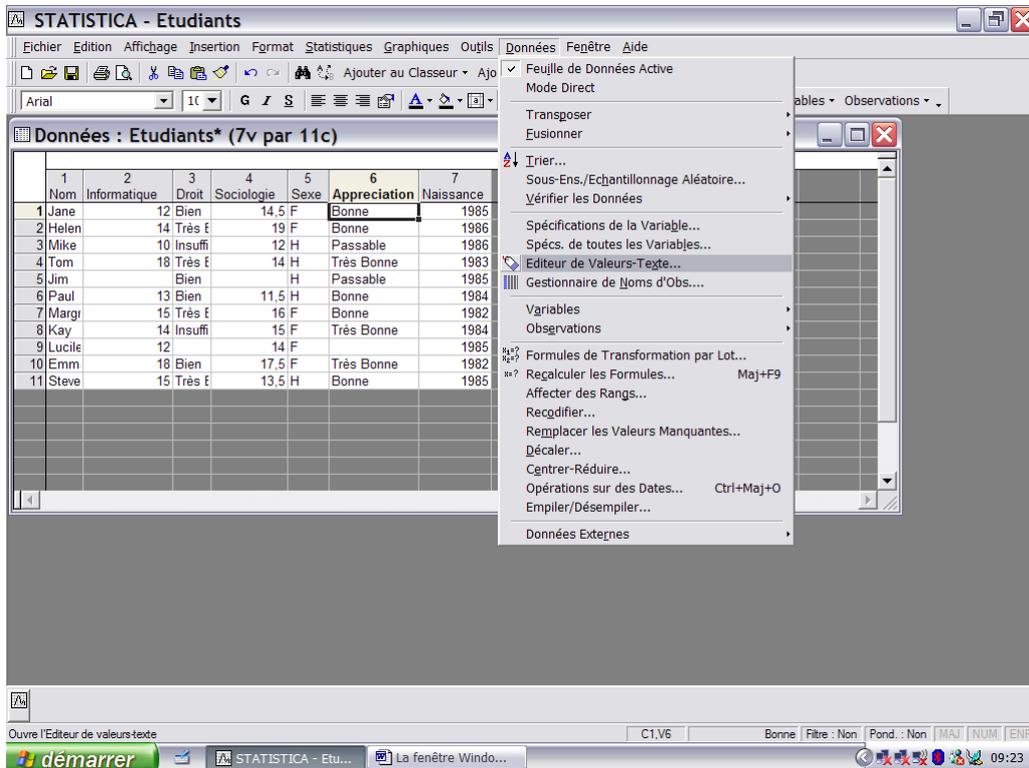
Les colonnes « Droit » et « Appreciation » pour lesquelles on a utilisé cette fonctionnalité affichent ainsi des valeurs « plus parlantes ».

On revient à l’affichage classique des valeurs numériques en désélectionnant l’option « Afficher les valeurs-texte ». On remarquera aussi que les valeurs-textes de la colonne « Droit » ne sont pas entièrement affichées par manque de place dans les cellules de cette colonne ; on résout cette difficulté par le menu « Format » et l’option « Variable » « Ajustement automatique » :

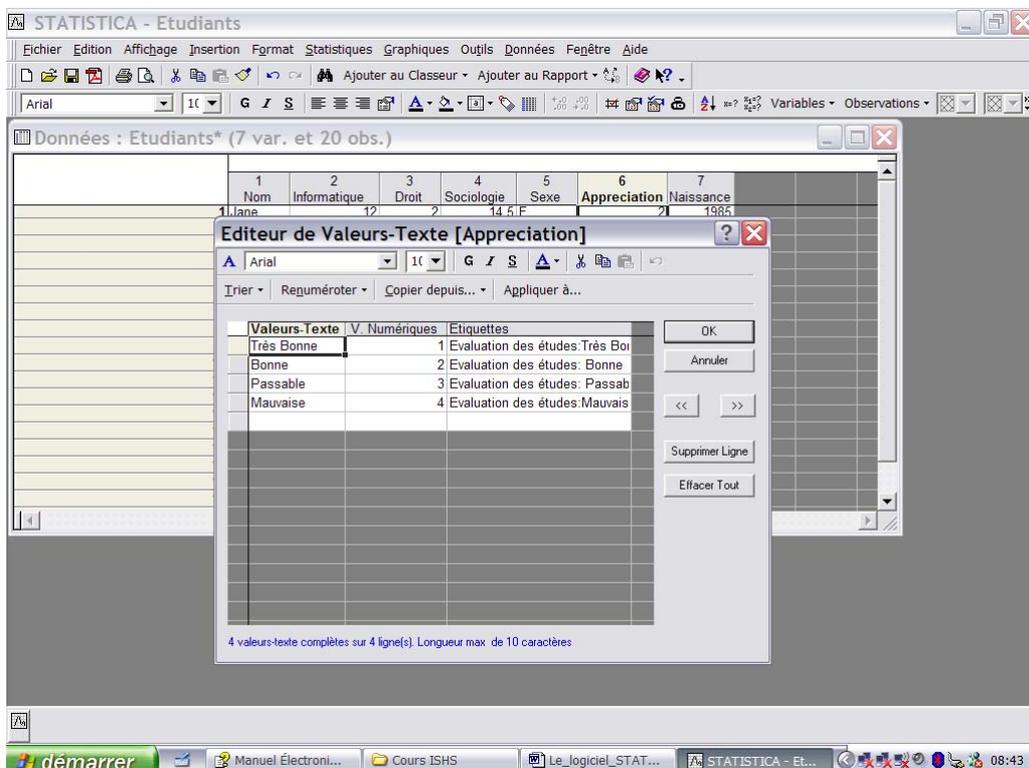
Données : Etudiants* (7v par 11c)

1	2	3	4	5	6	7
Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance
1	Jane	12 Bien	14,5	F	Bonne	1985
2	Helen	14 Très Bien	19	F	Bonne	1986
3	Mike	10 Insuffisant	12	H	Passable	1986
4	Tom	18 Très Bien	14	H	Très Bonne	1983
5	Jim	Bien		H	Passable	1985
6	Paul	13 Bien	11,5	H	Bonne	1984
7	Margr	15 Très Bien	16	F	Bonne	1982
8	Kay	14 Insuffisant	15	F	Très Bonne	1984
9	Lucile	12	14	F		1985
10	Emm	18 Bien	17,5	F	Très Bonne	1982
11	Steve	15 Très Bien	13,5	H	Bonne	1985

n.b. : L'option « Editeur de valeurs-texte » du menu « Données » permet d'afficher directement les codes textes sans passer par l'édition des spécifications de la variable :

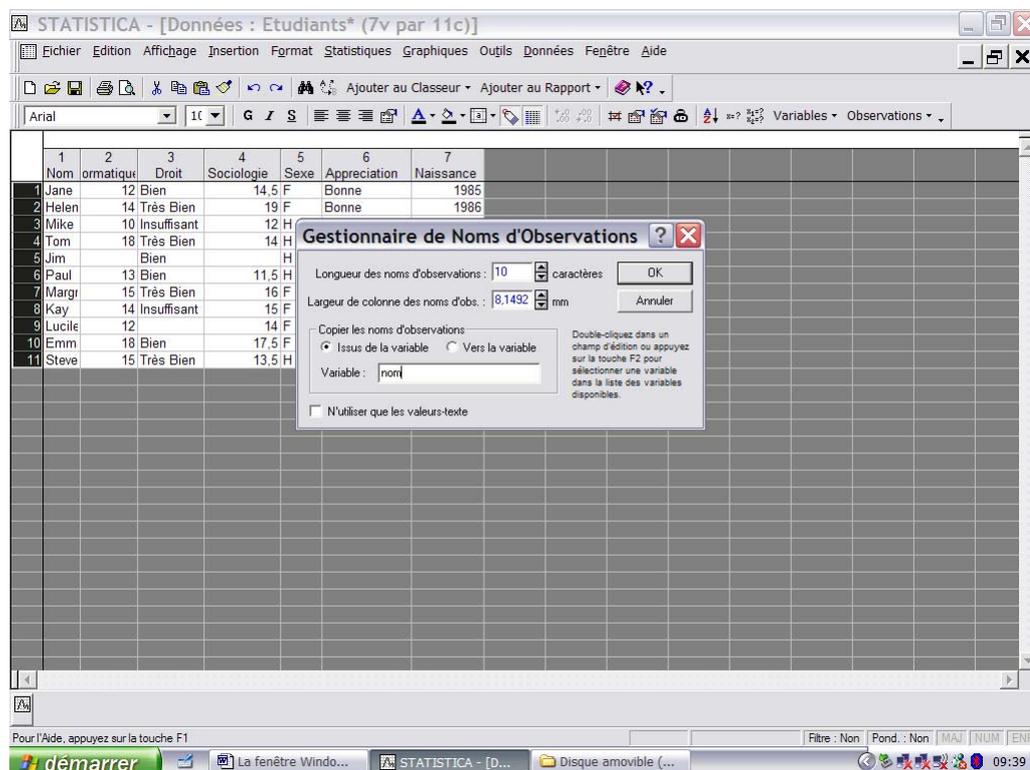


qui affiche l'écran :

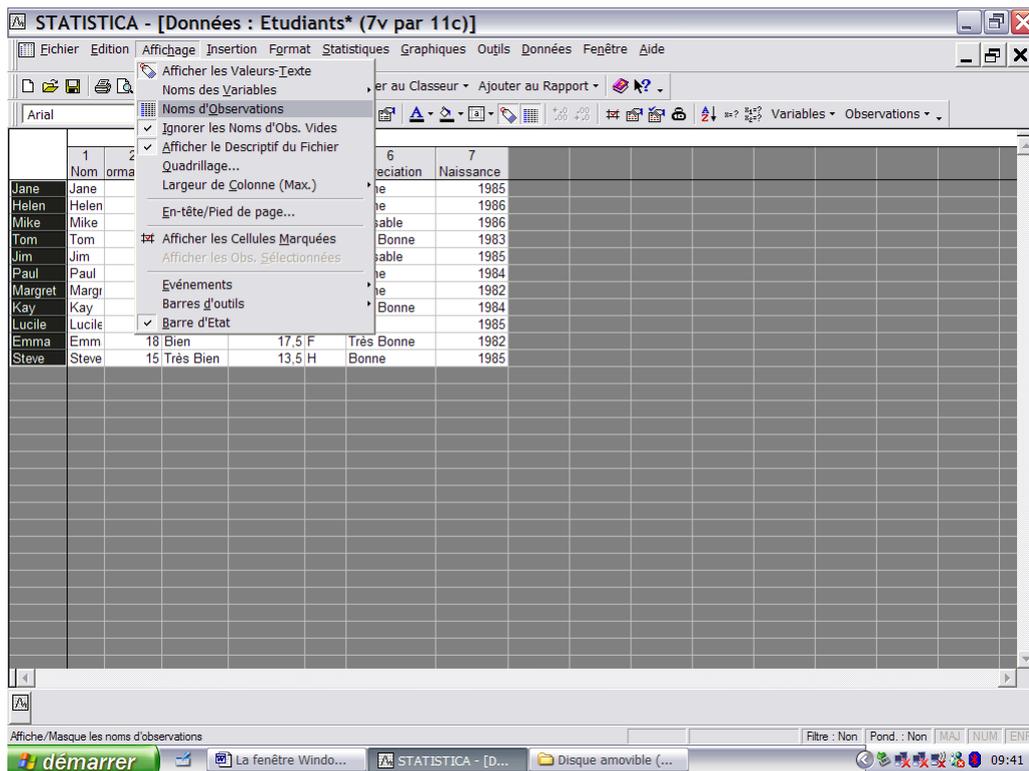


Le bouton « Copier depuis... » permet d'utiliser des valeurs-texte déjà encodées pour une autre variable et « Appliquer à... » affecte les valeurs-texte qui viennent d'être encodées dans l'Editeur à d'autres variables du fichier. Ces deux fonctionnalités présentent un grand intérêt dans le cadre de la réalisation d'enquêtes qui prévoient des questions à choix multiples codés de manière identique (ex. 'Oui'/'Non', 'Toujours'/'Souvent'/'Parfois'/'Jamais', ...)

- h. Par défaut, les observations de la feuille de données (lignes), sont numérotées à partir de 1 en séquence jusqu'à la dernière observation. STATISTICA permet définir un nom de ligne sous la forme de caractères (par exemple, le code d'identification d'un formulaire d'enquête) ; le menu « Données » puis « Gestionnaire de Noms d'Obs... » permet cette possibilité. Par exemple, si on veut définir les noms de lignes comme étant les noms des étudiants :



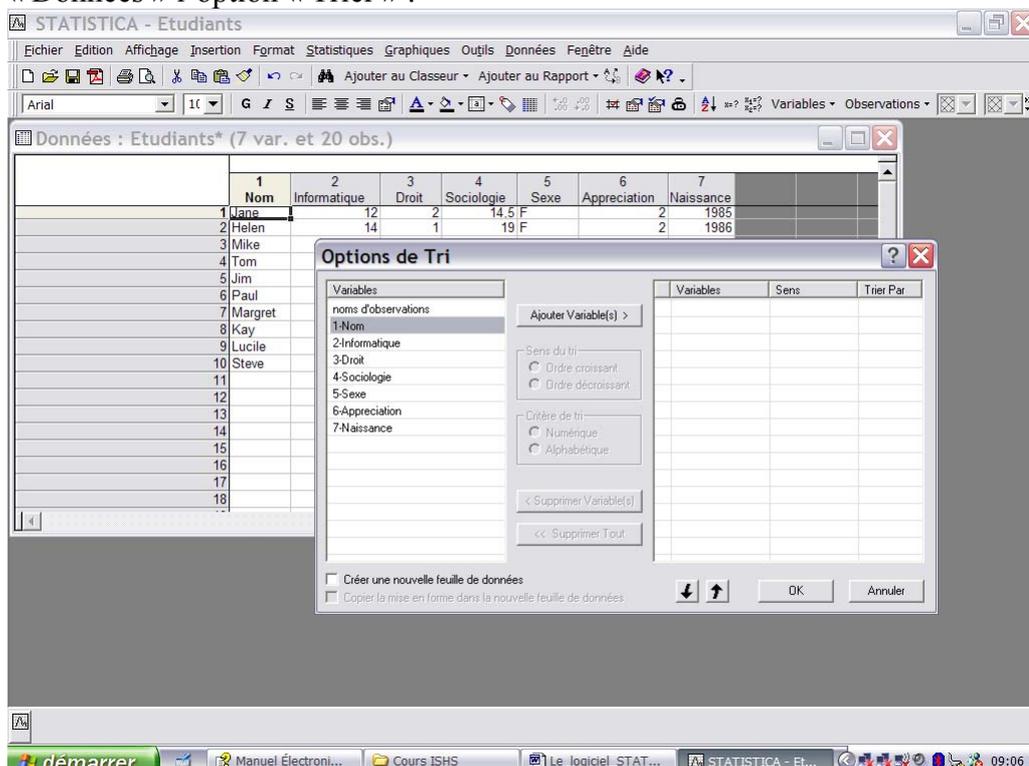
On affiche ensuite les noms de lignes :



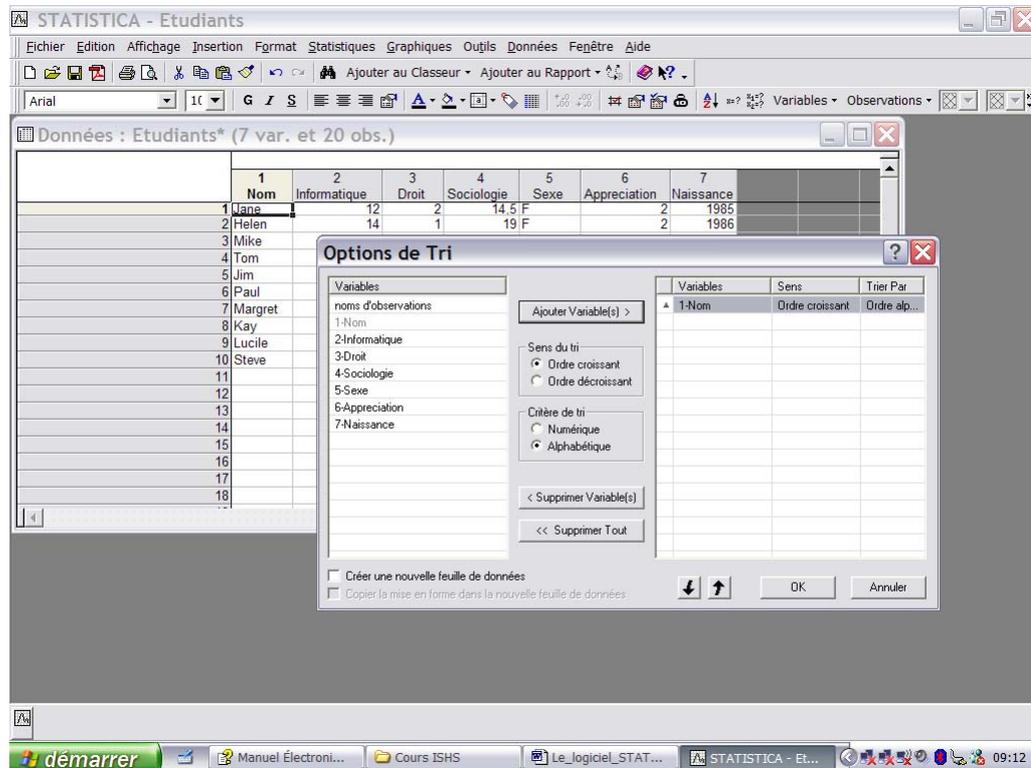
La suppression des noms peut se faire par le menu « Edition » et l’option « Supprimer » dans laquelle on sélectionne « Supprimer tous les noms d’Obs. ».

i. Tris

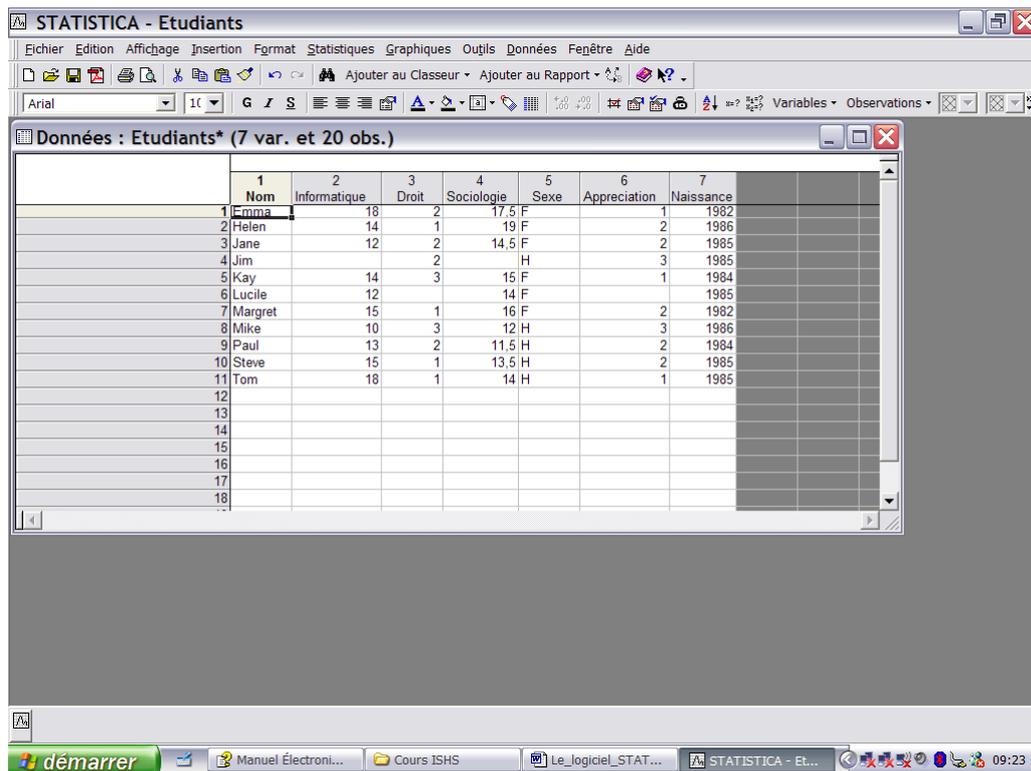
Les données sont généralement encodées « dans l’ordre où elles se présentent » c’est-à-dire sans ordre particulier ; dans notre exemple, on peut vouloir ordonner les observations selon les noms des étudiants par ordre alphabétique croissant. On utilise dans le menu « Données » l’option « Trier » :



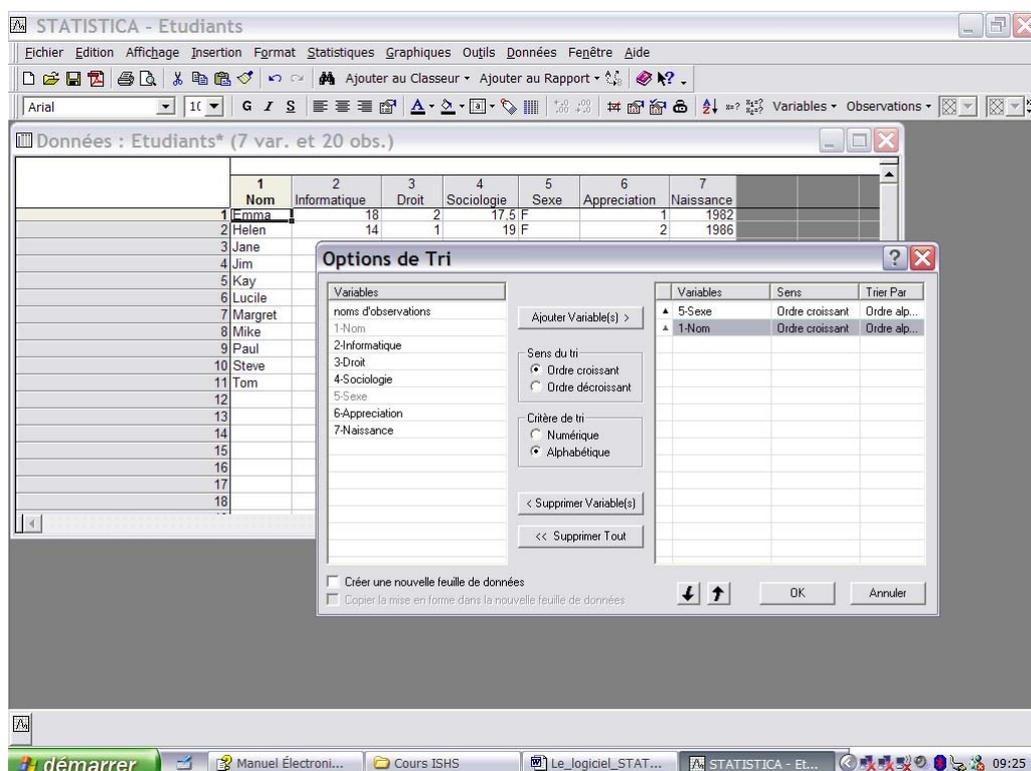
La variable « Nom » apparaît dans ce cas la première variable sélectionnée (cellule active en première colonne). Elle constitue la première clé de tri c'est-à-dire la première zone à partir de laquelle les informations seront ordonnées.
En utilisant le bouton « Ajouter Variable(s) » :



Par défaut, les modalités seront ordonnées par valeur croissante et le type de la variable choisie comme clé de tri sera pris en compte.
L'ordre de tri alphabétique respecte la codification ASCII : les champs alphabétique sont comparés caractère par caractère à partir de la gauche et l'évaluation est réalisée à partir de la valeur entière des codes ASCII. Le tri du fichier selon le nom a pour résultat :



La fonctionnalité de tri peut utiliser plusieurs critères (plusieurs clés) ; ainsi, on peut vouloir créer deux groupes d'observations en distinguant la colonne « Sexe » et au sein des deux groupes, trier par ordre alphabétique des noms. Voici le menu « Tri », dans lequel on a d'abord sélectionné la variable Sexe (clé majeure) ensuite le Nom (clé mineure):



Et le résultat :

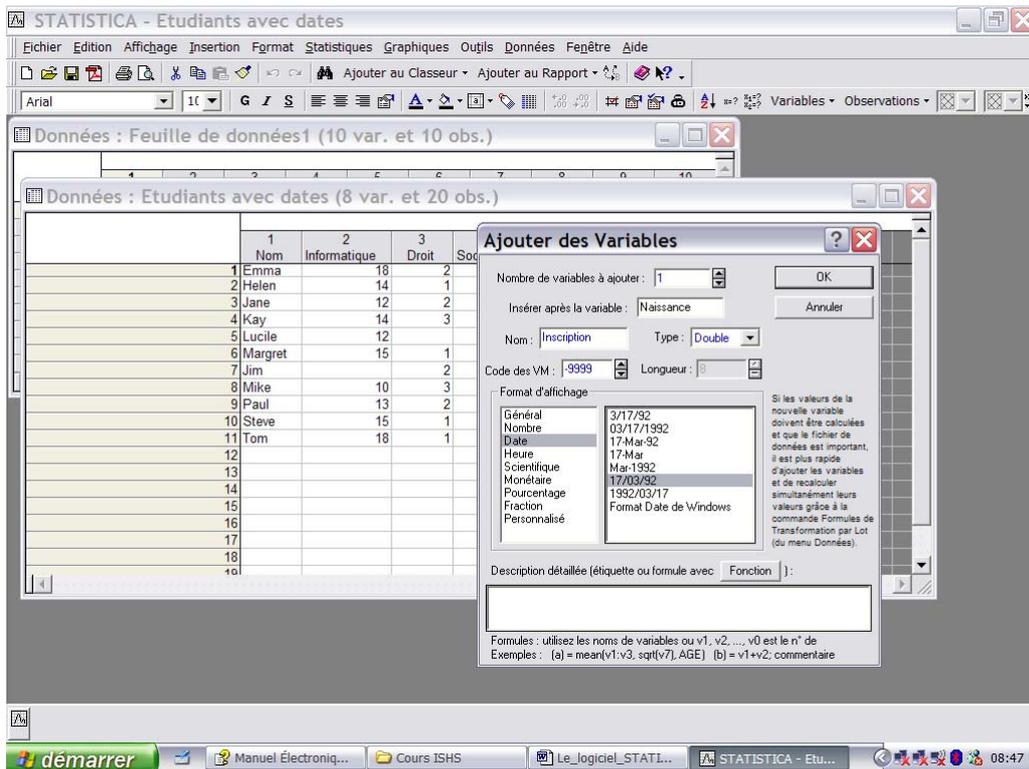
Données : Etudiants* (7 var. et 20 obs.)

	1	2	3	4	5	6	7			
	Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance			
1	Emma	18	2	17,5	F		1	1982		
2	Helen	14	1	19	F		2	1986		
3	Jane	12	2	14,5	F		2	1985		
4	Kay	14	3	15	F		1	1984		
5	Lucile	12		14	F			1985		
6	Margret	15	1	16	F		2	1982		
7	Jim		2		H		3	1985		
8	Mike	10	3	12	H		3	1986		
9	Paul	13	2	11,5	H		2	1984		
10	Steve	15	1	13,5	H		2	1985		
11	Tom	18	1	14	H		1	1985		
12										
13										
14										
15										
16										
17										
18										

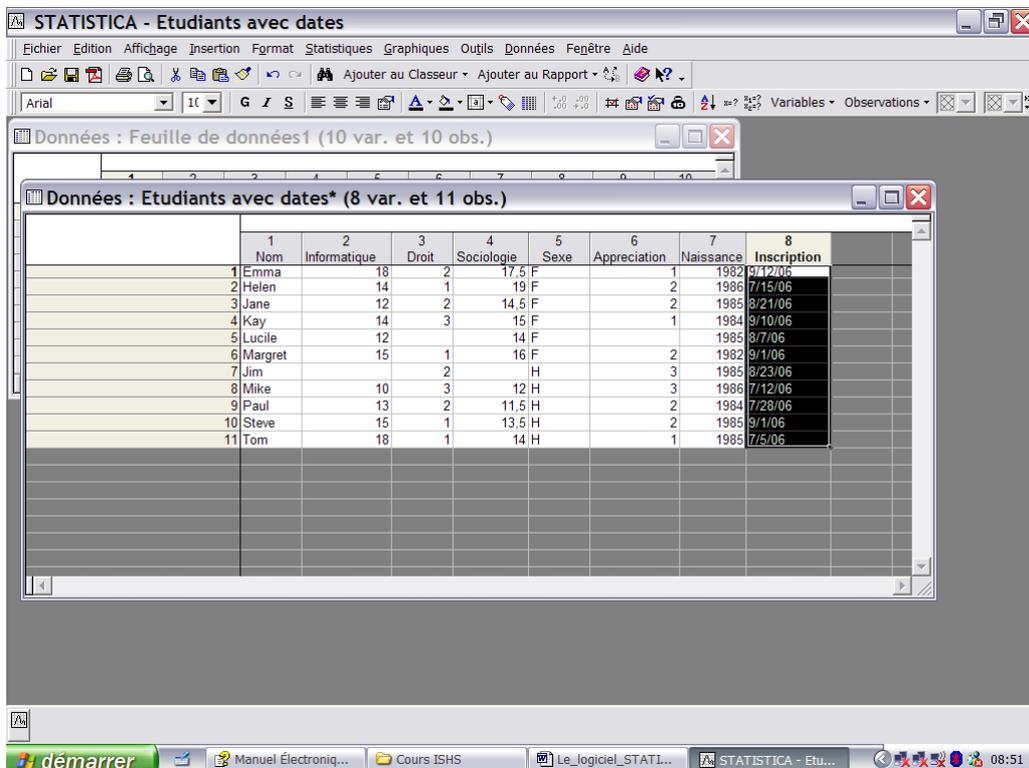
Les boutons « Flèche vers le haut » et « Flèche vers le bas » permettent de modifier l'ordre des clés retenues ; enfin, si on désire conserver l'ordre de départ du fichier de données, on peut enregistrer les données dans un nouveau fichier une fois qu'elles auront été triées (Case « Créer une nouvelle feuille de données ») en y sauvegardant aussi toutes les caractéristiques du fichier de départ (Case « Copier la mise en forme... »).

j. Variables de type « Date »

L'insertion de ce type de variables qui ne sont pas numériques ou de type alphabétique se réalise est permise par STATISTICA via l'insertion de variable :



On sélectionne dans « Format d'affichage » l'option « Date » puis la forme choisie pour la date, l'encodage de la nouvelle variable doit évidemment respecter le format choisi au risque de provoquer un message d'erreur :



Le stockage des dates en STATISTICA utilise le format de la date Julienne en l'occurrence le nombre de jours écoulés depuis une origine qui est pour STATISTICA le 1^{er} Janvier 1900.

L'affichage du compteur peut se faire par double-clic sur le nom et en choisissant le format « Nombre », cela donne dans notre cas :

The screenshot shows the STATISTICA software window titled "STATISTICA - Etudiants avec dates". A data table is displayed with the following columns and rows:

	1	2	3	4	5	6	7	8
	Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance	Inscription
1	Emma	18	2	17,5	F	1	1982	38972
2	Helen	14	1	19	F	2	1986	38913
3	Jane	12	2	14,5	F	2	1985	38950
4	Kay	14	3	15	F	1	1984	38970
5	Lucile	12		14	F		1985	38936
6	Margret	15	1	16	F	2	1982	38961
7	Jim		2		H	3	1985	38952
8	Mike	10	3	12	H	3	1986	38910
9	Paul	13	2	11,5	H	2	1984	38926
10	Steve	15	1	13,5	H	2	1985	38961
11	Tom	18	1	14	H	1	1985	38903

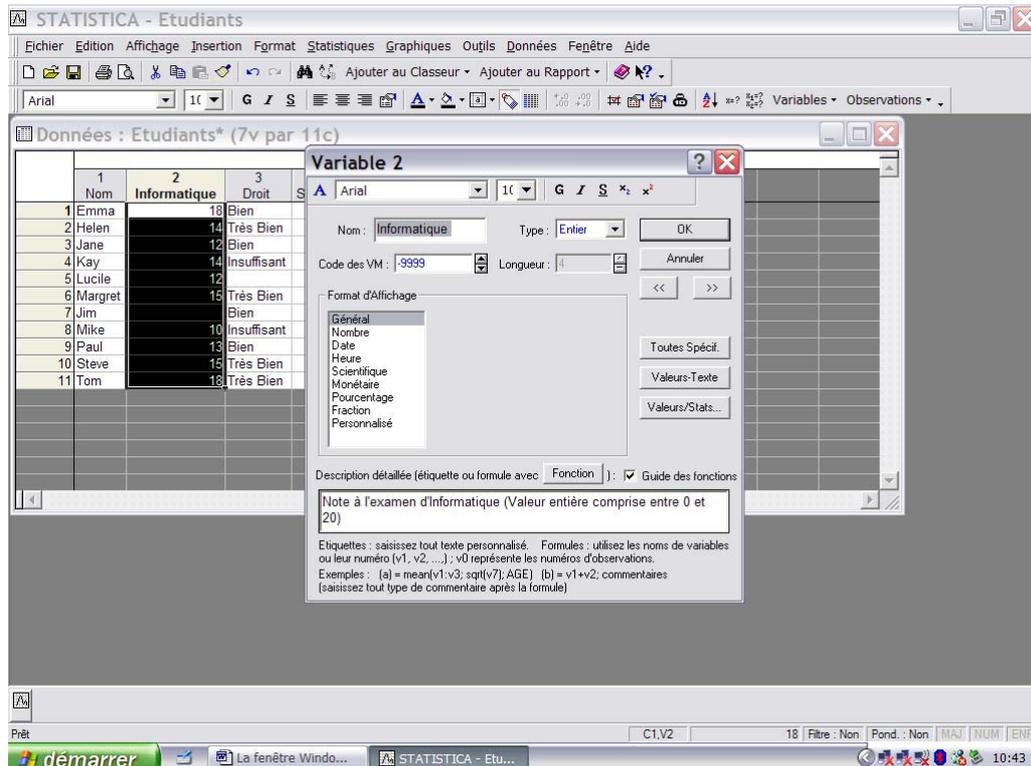
La valeur 38972 correspond au nombre de jours écoulés entre le 1^{er} janvier 1900 et le 12 septembre 2006 (étudiant « Emma »).

Les tris respectent évidemment ce format particulier et des calculs de délais sont possibles (cf. infra).

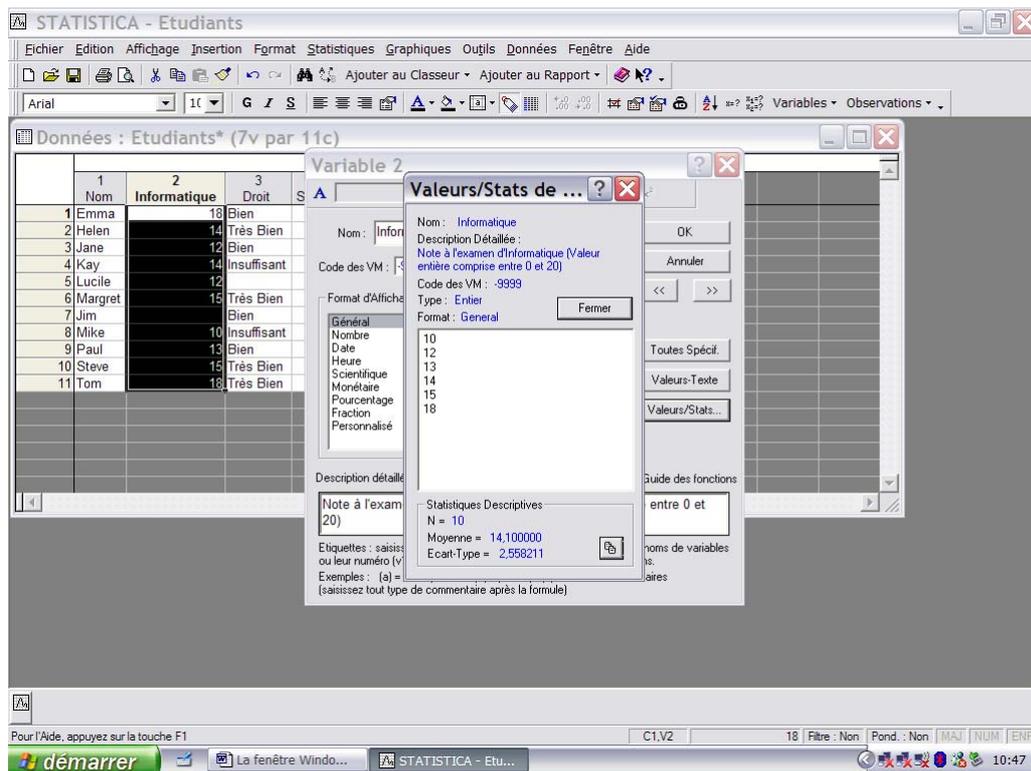
B. TRAITEMENTS STATISTIQUES ET GRAPHIQUES

1. Statistiques rapides

Les données de l'exemple, une fois encodées, peuvent être l'objet des premières analyses au moyen des menus « Statistiques » et « Graphiques » mais on peut obtenir les premiers résumés statistiques quand on affiche les spécifications des variables (double clic sur le nom de la colonne) ; ainsi, pour la variable qui contient les notes des étudiants pour le cours d'Informatique :



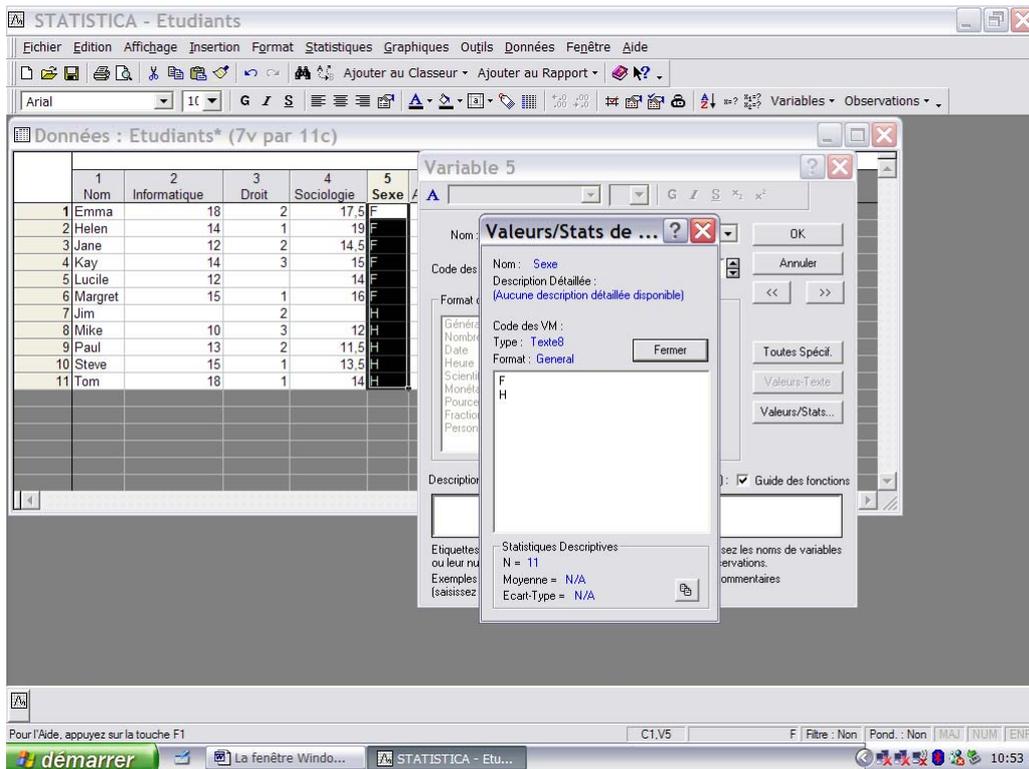
Le bouton intitulé « Valeurs/Stats » permet d'obtenir les résultats suivants pour la variable :



On retrouve en plus des spécifications générales de la variable « Informatique » :

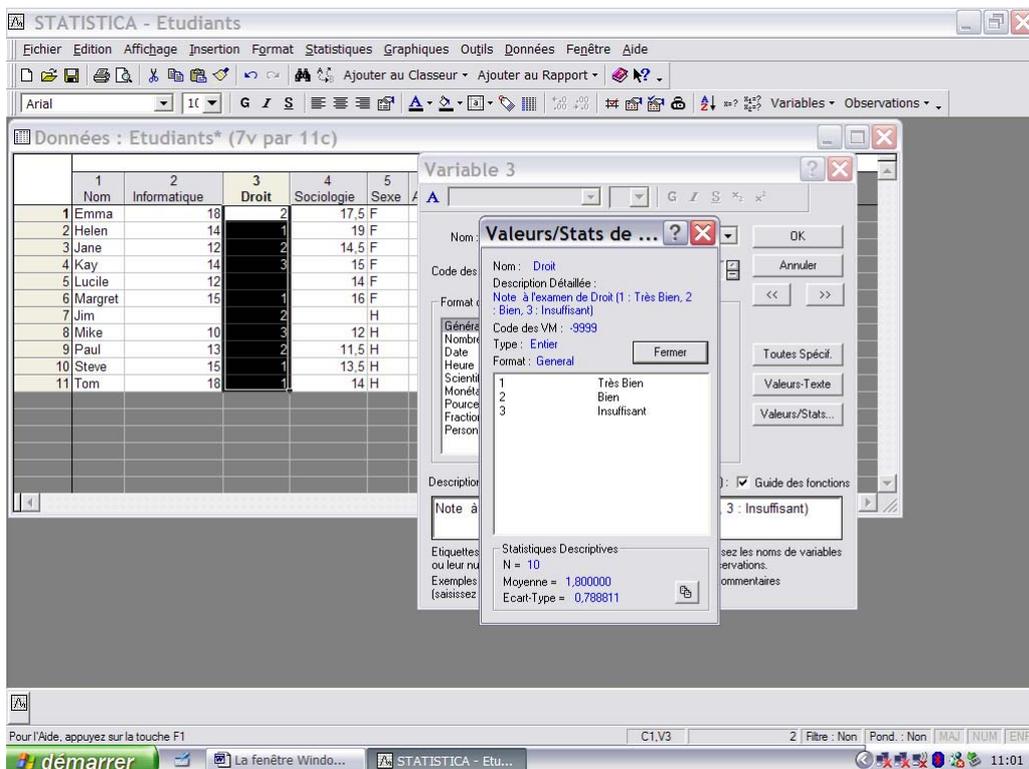
- un tableau qui liste ses différentes valeurs qui peut s'avérer très utile pour vérifier les différentes valeurs encodées et repérer ensuite les valeurs anormales (par exemple, une note qui serait supérieure à 20)
- le nombre d'observations (N) dans lequel, les valeurs manquantes n'interviennent pas (ici, il s'agit des 11 lignes moins une valeur manquante pour l'étudiant « Jim », donc N=10)
- la moyenne arithmétique des 10 valeurs (14,1)
- l'écart-type de ces valeurs (2,558211)

Pour une variable de type texte, par exemple la variable « Sexe », le résultat sera bien évidemment adapté :

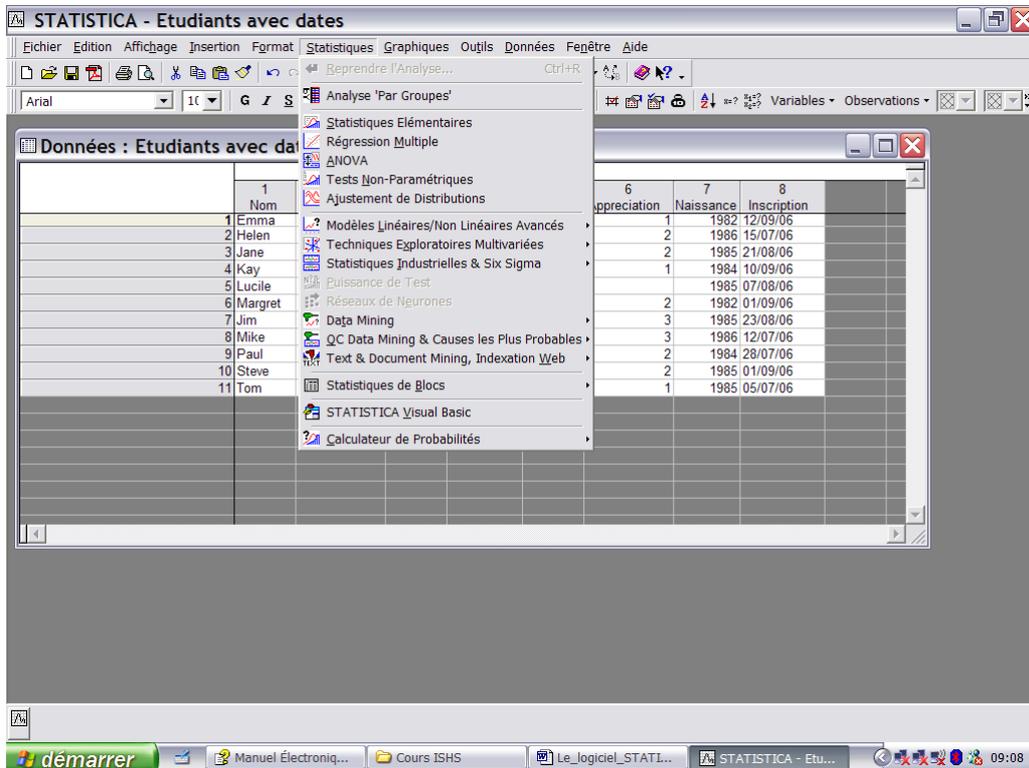


Il y a dans ce cas 11 valeurs encodées qui sont soit « F », soit « H », la moyenne et l'écart-type ne peuvent être calculés (N/A signifie « Non Available » soit « Non Disponible »).

Par ailleurs, cette statistique est disponible pour la variable « Droit » comme l'indique l'écran suivant ; ce qui n'a pas beaucoup de sens sur le plan statistique mais STATISTICA effectue ces calculs puisque la variable a été déclarée de type numérique entier :



Le menu « Statistiques » se décline sous la forme d'un ensemble d'outils très riches dont nous utiliserons deux options : Statistiques Élémentaires et Statistiques de Blocs.

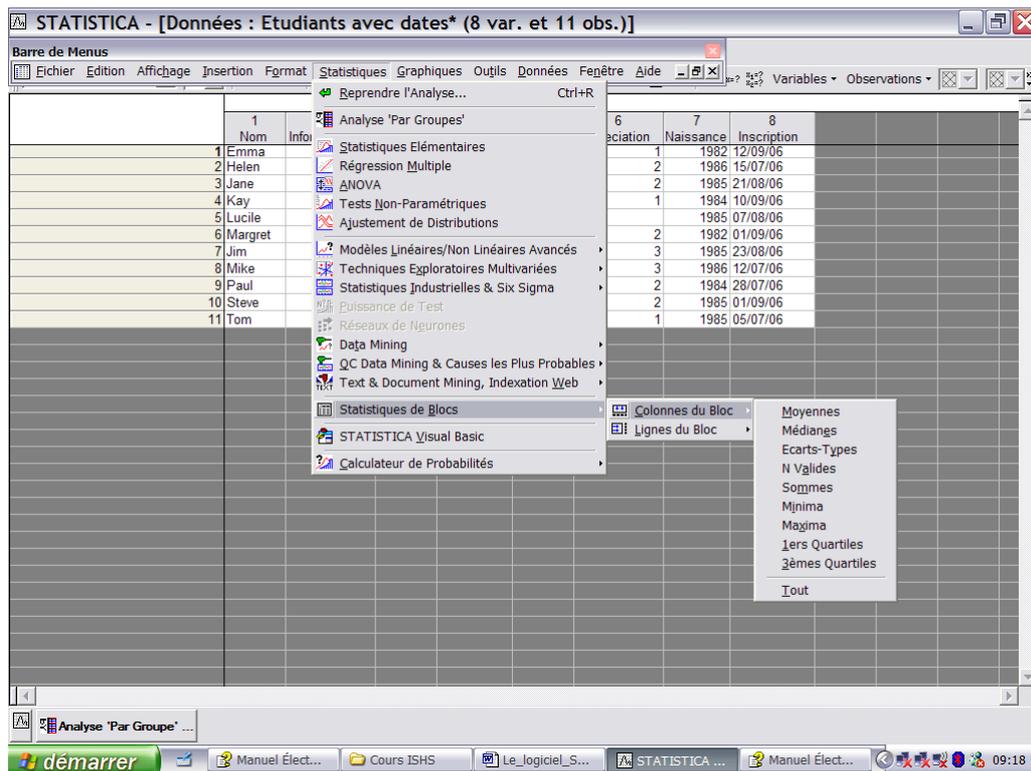


2. L'option « Statistiques de Blocs »

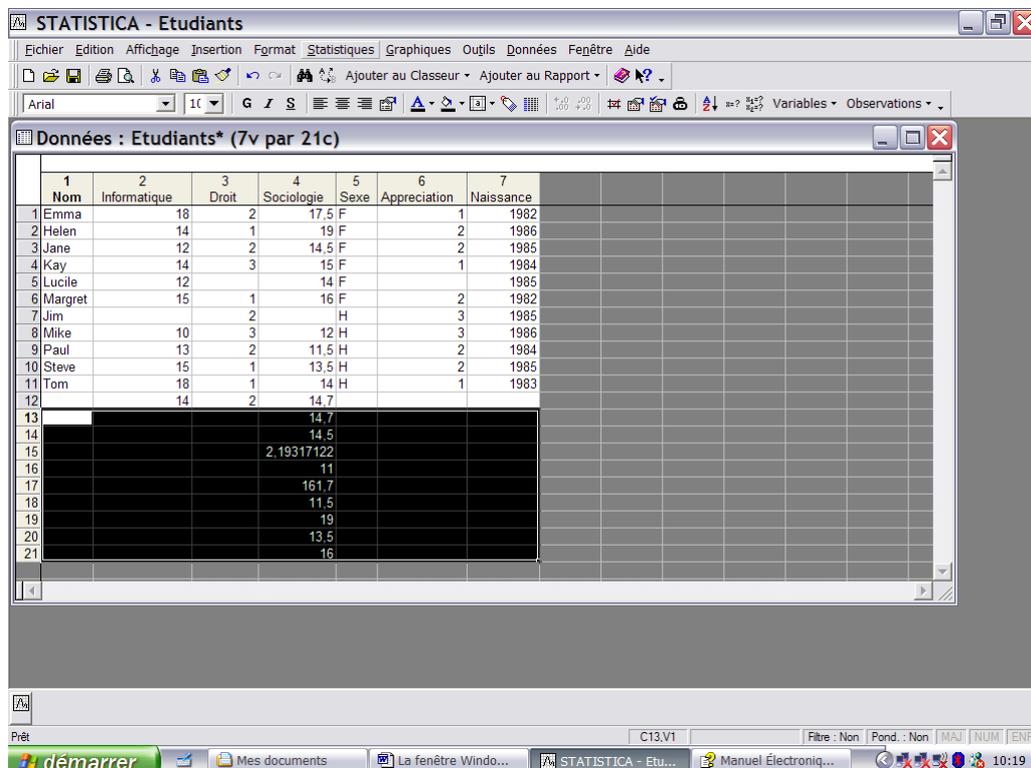
Un Bloc est défini comme un sous-ensemble de cellules de la feuille de données ; pour définir un bloc, on sélectionne simplement l'ensemble des cellules qui le composent par le clic gauche maintenu enfoncé.

On peut créer un bloc dans le sens des observations (on définit un sous-ensemble d'individus), dans le sens des colonnes (on définit un sous-ensemble de variables) ou dans le sens lignes-colonnes (on définit un groupe d'individus observés sur un certain nombre de variables).

Supposons qu'on sélectionne la variable « Sociologie », on calculera ensuite les statistiques pour cette colonne au moyen du menu « Statistiques » et des options « Statistiques de Blocs » et « Colonnes du Bloc » :



L'option « Tout » permet d'obtenir la plupart des résultats statistiques de base qui constituent autant de lignes supplémentaires ajoutées à la fin de la feuille de données :



Pour interpréter ce dernières, il est nécessaire de sélectionner « Noms d'observations » dans le menu « Affichage » :

STATISTICA - Etudiants

Données : Etudiants* (7v par 20c)

	1	2	3	4	5	6	7
	Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance
Emma	Emma	18	2	17,5	F	1	1982
Helen	Helen	14	1	19	F	2	1986
Jane	Jane	12	2	14,5	F	2	1985
Kay	Kay	14	3	15	F	1	1984
Lucile	Lucile	12		14	F		1985
Margret	Margret	15	1	16	F	2	1982
Jim	Jim		2		H	3	1985
Mike	Mike	10	3	12	H	3	1986
Paul	Paul	13	2	11,5	H	2	1984
Steve	Steve	15	1	13,5	H	2	1985
Tom	Tom	18	1	14	H	1	1983
MOYENNE obs. 1-11				14,7			
MEDIANE obs. 1-11				14,25			
EC-TYPE obs. 1-11				2,31180545			
N_VALIDES obs. 1-11				10			
SOMME obs. 1-11				147			
MIN obs. 1-11				11,5			
MAX obs. 1-11				19			
QUARTILE_1 obs. 1-11				13,5			
QUARTILE_3 obs. 1-11				16			

Les noms des lignes ajoutées indiquent les mesures statistiques effectuées ; ils indiquent en outre les observations sur lesquelles les calculs ont été réalisés ainsi que le nombre de données valides (le nombre de lignes définies dans le bloc diminué du nombre de valeurs manquantes).

On peut sélectionner un groupe d'observations pour le calcul des statistiques de bloc ; par exemple, si on désire calculer les mêmes paramètres pour le cours de « Sociologie » mais pour les Filles seulement, on définit le Bloc (après avoir supprimé les lignes qui viennent d'être créées) :

STATISTICA - Etudiants

Données : Etudiants* (7v par 11c)

	1	2	3	4	5	6	7
	Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance
Emma	Emma	18	2	17,5	F	1	1982
Helen	Helen	14	1	19	F	2	1986
Jane	Jane	12	2	14,5	F	2	1985
Kay	Kay	14	3	15	F	1	1984
Lucile	Lucile	12		14	F		1985
Margret	Margret	15	1	16	F	2	1982
Jim	Jim		2		H	3	1985
Mike	Mike	10	3	12	H	3	1986
Paul	Paul	13	2	11,5	H	2	1984
Steve	Steve	15	1	13,5	H	2	1985
Tom	Tom	18	1	14	H	1	1983

Les mêmes choix que dans le calcul précédent donnent les résultats :

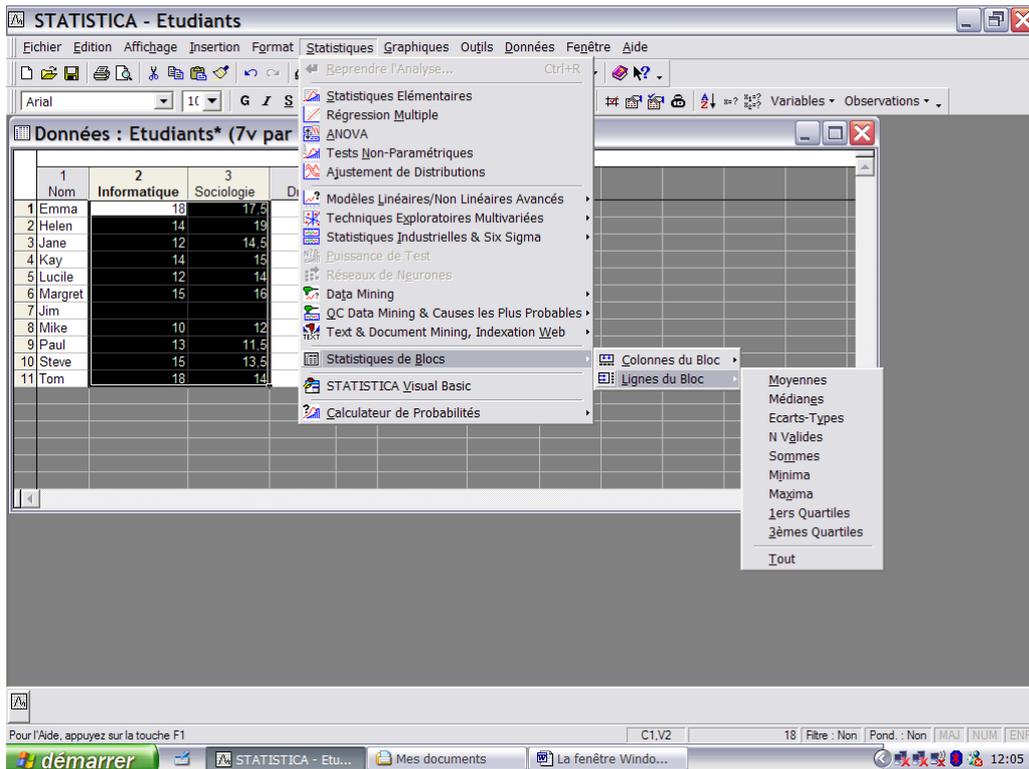
	1	2	3	4	5	6	7
	Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance
Emma	Emma	18	2	17,5	F	1	1982
Helen	Helen	14	1	19	F	2	1986
Jane	Jane	12	2	14,5	F	2	1985
Kay	Kay	14	3	15	F	1	1984
Lucile	Lucile	12		14	F		1985
Margret	Margret	15	1	16	F	2	1982
Jim	Jim		2		H	3	1985
Mike	Mike	10	3	12	H	3	1986
Paul	Paul	13	2	11,5	H	2	1984
Steve	Steve	15	1	13,5	H	2	1985
Tom	Tom	18	1	14	H	1	1983
MOYENNE obs. 1-6				16			
MEDIANE obs. 1-6				15,5			
EC-TYPE obs. 1-6				1,92353841			
N_VALIDES obs. 1-6				6			
SOMME obs. 1-6				96			
MIN obs. 1-6				14			
MAX obs. 1-6				19			
QUARTILE_1 obs. 1-6				14,5			
QUARTILE_3 obs. 1-6				17,5			

Les calculs effectués ne concernent plus que le groupe des Filles. Si on désire effectuer les mêmes opérations pour le groupe des Garçons, on redéfinit un Bloc adéquat et les résultats s'affichent sous la forme de lignes supplémentaires par rapport aux lignes de la dernière feuille de données (qui contient les résultats pour les Filles) :

	1	2	3	4	5	6	7
	Nom	Informatique	Droit	Sociologie	Sexe	Appreciation	Naissance
Emma	Emma	18	2	17,5	F	1	1982
Helen	Helen	14	1	19	F	2	1986
Jane	Jane	12	2	14,5	F	2	1985
Kay	Kay	14	3	15	F	1	1984
Lucile	Lucile	12		14	F		1985
Margret	Margret	15	1	16	F	2	1982
Jim	Jim		2		H	3	1985
Mike	Mike	10	3	12	H	3	1986
Paul	Paul	13	2	11,5	H	2	1984
Steve	Steve	15	1	13,5	H	2	1985
Tom	Tom	18	1	14	H	1	1983
MOYENNE obs. 1-6				16			
MEDIANE obs. 1-6				15,5			
EC-TYPE obs. 1-6				1,92353841			
N_VALIDES obs. 1-6				6			
SOMME obs. 1-6				96			
MIN obs. 1-6				14			
MAX obs. 1-6				19			
QUARTILE_1 obs. 1-6				14,5			
QUARTILE_3 obs. 1-6				17,5			
MOYENNE obs. 7-11				12,75			
MEDIANE obs. 7-11				12,75			
EC-TYPE obs. 7-11				1,19023807			
N_VALIDES obs. 7-11				4			
SOMME obs. 7-11				51			
MIN obs. 7-11				11,5			
MAX obs. 7-11				14			

La statistique de bloc s'applique également « dans le sens des lignes », on peut donc calculer les paramètres statistiques observations par observations. Supposons qu'on désire connaître la moyenne des notes obtenues en « Informatique » et en « Sociologie » pour chaque étudiant : on doit d'abord définir le bloc qui contient ces deux colonnes, étant donné que les cellules d'un bloc doivent être adjacentes, on déplace la colonne « Sociologie » à gauche de « Informatique » et on définit le bloc correspondant à ces deux colonnes pour tous les étudiants.

Ensuite, le menu « Statistiques » et les options « Statistiques de Blocs » et « Lignes du bloc » permettent le calcul attendu :



Voici le résultat :

The screenshot shows the STATISTICA software window titled "STATISTICA - Etudiants". The main data window is titled "Données : Etudiants* (8v par 11c)". It displays a table with 11 rows of student data. The columns are: 1. Nom, 2. Informatique, 3. Sociologie, 4. Droit, 5. Sexe, 6. Appreciation, 7. Naissance, and 8. MOYENNE vars 2-3. The data is as follows:

	1 Nom	2 Informatique	3 Sociologie	4 Droit	5 Sexe	6 Appreciation	7 Naissance	8 MOYENNE vars 2-3
1	Emma	18	17,5	2	F		1982	17,75
2	Helen	14	19	1	F		1986	16,9
3	Jane	12	14,5	2	F		1985	13,25
4	Kay	14	15	3	F		1984	14,5
5	Lucile	12	14		F		1985	13
6	Margret	15	16	1	F		1982	15,5
7	Jim			2	H		1985	
8	Mike	10	12	3	H		1986	11
9	Paul	13	11,5	2	H		1984	12,25
10	Steve	15	13,5	1	H		1985	14,25
11	Tom	18	14	1	H		1983	16

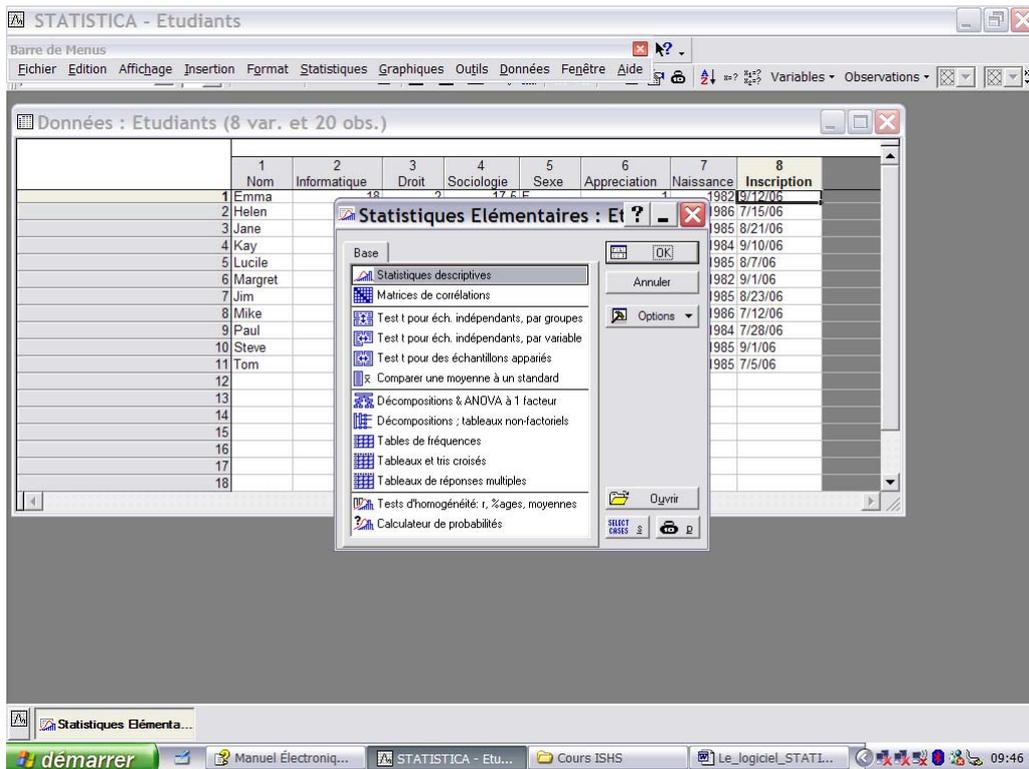
Etant donné que le calcul de la moyenne arithmétique est réalisé étudiant par étudiant (l'étudiant « Emma » a obtenu une moyenne de 17,75) pour les deux examens retenus), il apparaît une variable supplémentaire dans la feuille de données dont le nom attribué par STATISTICA est « MOYENNE vars 2-3 » qui peut à son tour faire l'objet de calculs. Si une donnée est manquante dans l'une des colonnes du bloc, le résultat est aussi manquant.

Il faut être attentif que le sauvetage de la feuille de données enregistrera toutes les lignes et colonnes qui auront été ajoutées par la procédure de Statistiques de Bloc.

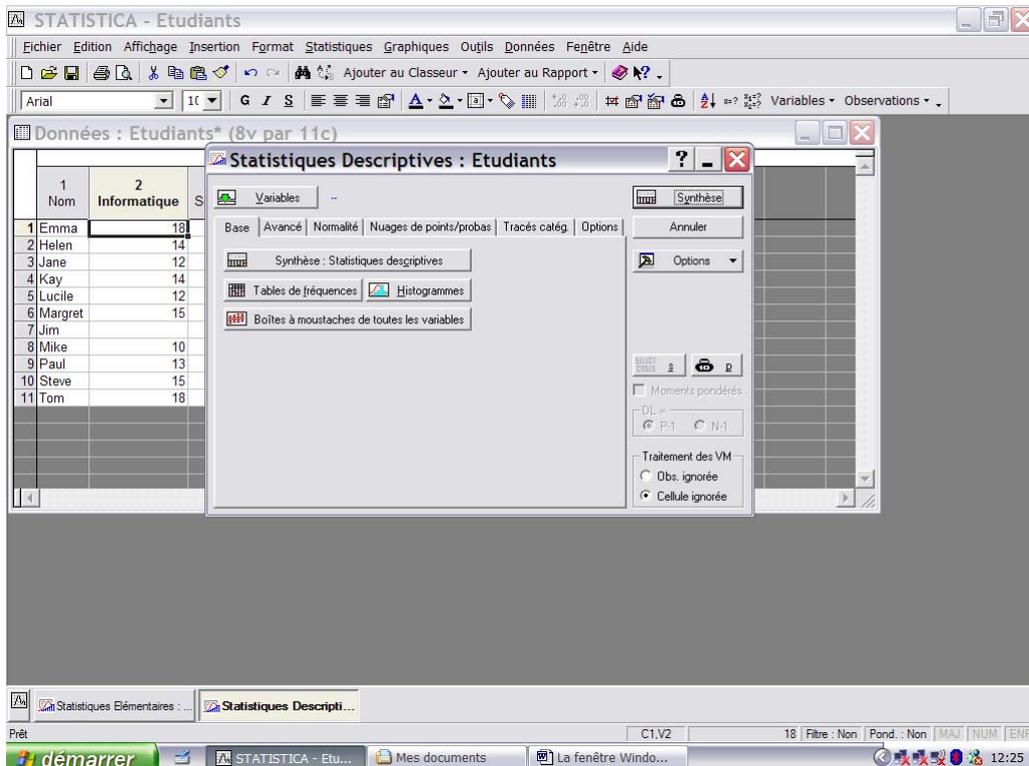
3. L'option « Statistiques élémentaires »

a. Statistiques descriptives

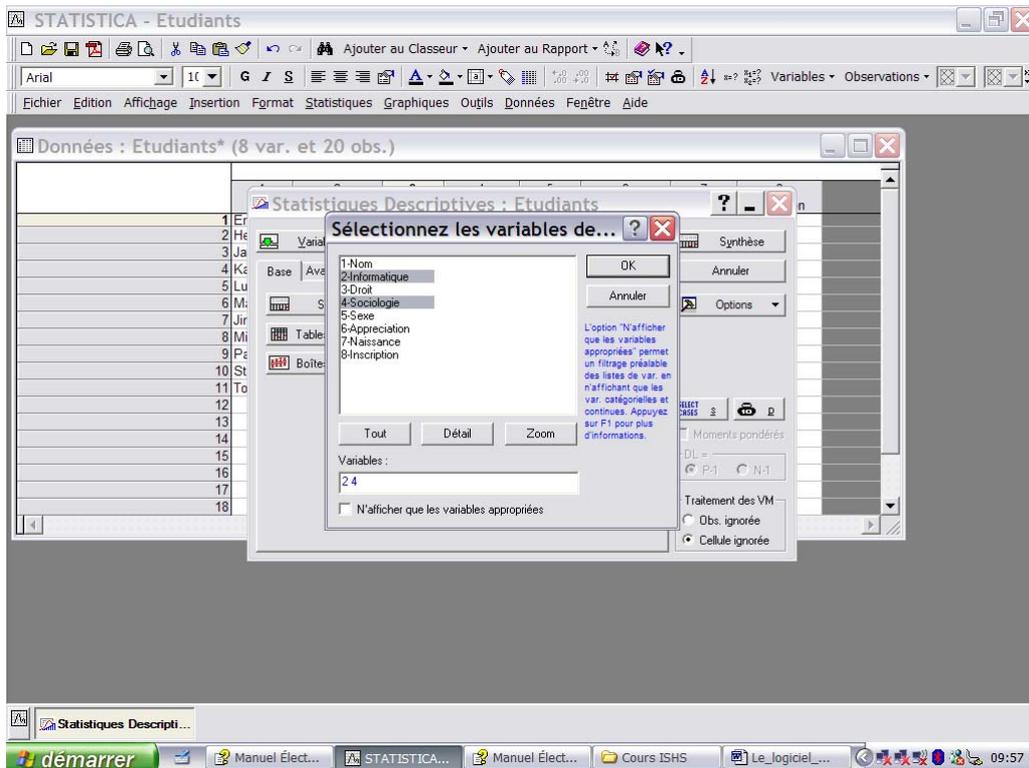
Grâce à cette option, STATISTICA permet d'accéder à tous les outils classiques de la statistique descriptive et inférentielle en proposant un premier menu :



dans lequel on choisit d'abord l'option « Statistiques descriptives » :



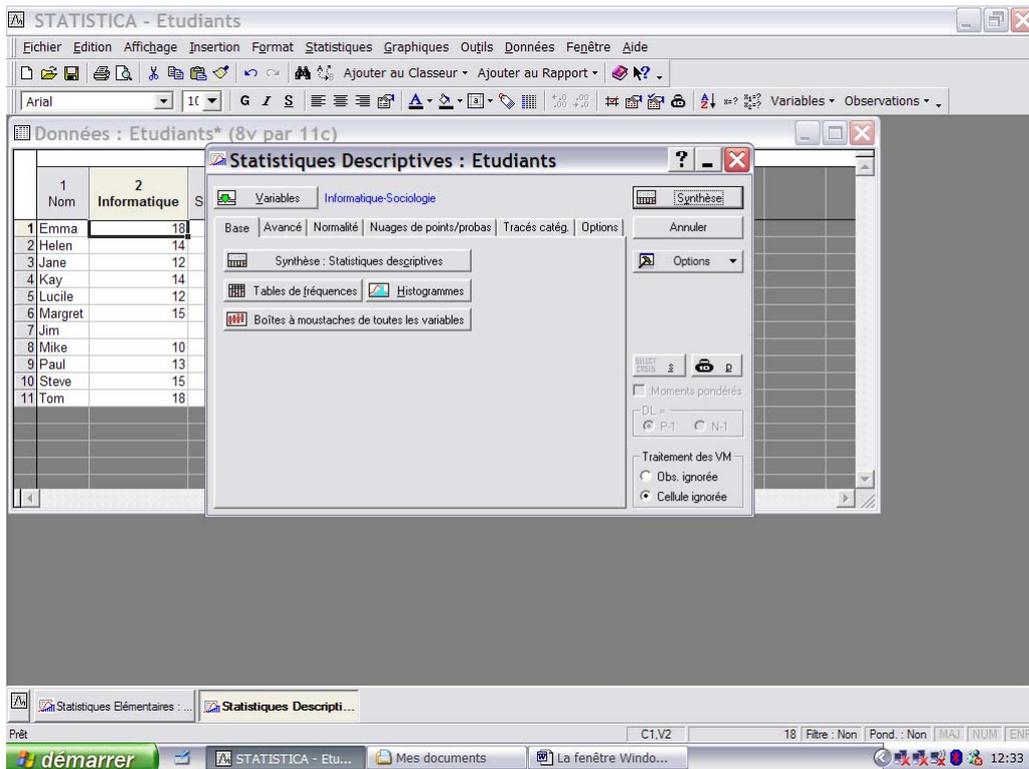
On commence par définir les variables (bouton « Variables ») pour lesquelles on désire procéder aux calculs de synthèse, par exemple « Informatique » et « Sociologie » :



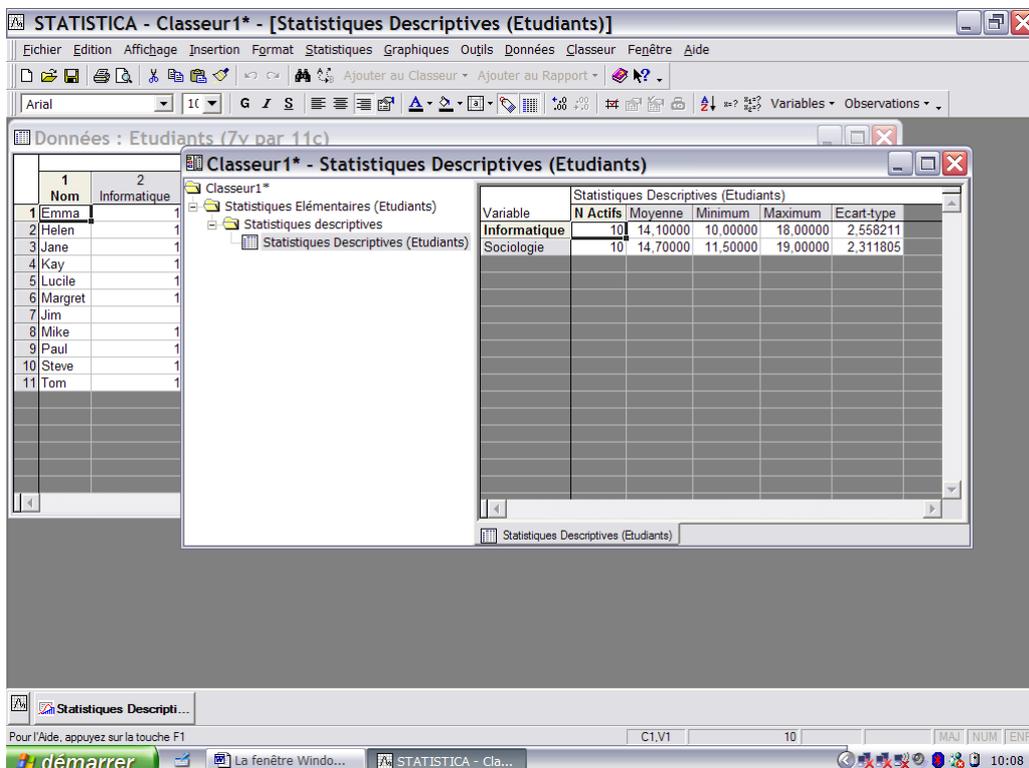
n.b. : La sélection des variables peut se faire en utilisant les standards de sélection Windows (CTRL clic gauche souris et majuscule clic gauche souris)

La case « N'afficher que les variables appropriées » permet de limiter la liste des variables à celles qui peuvent faire l'objet des traitements statistiques de ce module, en l'occurrence les variables de ratio (appelées ici variables continues) et les variables discrètes se prêtant à certains résumés, notamment les tables de fréquences ; pour bénéficier de cette option, il est nécessaire de choisir le type de mesure adéquat lors de la création de la variable. Par ailleurs, les variables dont le type est « Date » sont systématiquement retenues...

La sélection par « OK » permet de retrouver l'écran précédent dans lequel apparaissent les noms des variables sélectionnées (Informatique-Sociologie à droite du bouton « Variables ») :



Avant d'entrer dans le détail de toutes les options des statistiques descriptives, on peut accéder aux résultats par le bouton « Synthèse » :



Affichage des résultats

On peut noter que les résultats statistiques apparaissent dans la partie droite de l'écran, appelée « Fenêtre des résultats ». Par défaut, STATISTICA propose pour les variables sélectionnées leurs valeurs minimales et maximales, la moyenne arithmétique, l'écart-type et le nombre d'observations (càd celles qui ne sont pas manquantes) appelé N Actifs.

A gauche, apparaît l'arborescence d'enregistrement des résultats : lors de l'appel d'une procédure de traitement statistique, il y a création d'un fichier « Classeur » suivi d'un numéro. Tous les résultats sont rangés dans ce classeur qui peut faire l'objet d'un enregistrement ; à une prochaine ouverture du logiciel, un double clic sur l'icône du classeur rétablit la fenêtre des résultats passés.

La structure de l'arborescence du Classeur correspond parfaitement aux noms des options choisies dans le menu « Statistiques » ; pour réaliser les calculs dont les résultats apparaissent à droite, nous avons sélectionné « Statistiques élémentaires » ensuite « Statistiques Descriptives ». Si on désire réutiliser cette procédure, nous avons la possibilité de « Reprendre l'analyse » (menu « Statistiques » option « Reprendre l'analyse »), dans ce cas, les nouveaux résultats apparaissent comme étant une suite du Classeur ; si nous choisissons de faire une nouvelle analyse, il y a création d'une nouvelle subdivision de l'arborescence :

Exemple 1 : - option « Reprendre l'analyse » : nous recalculons les paramètres sur la colonne « Informatique » seulement

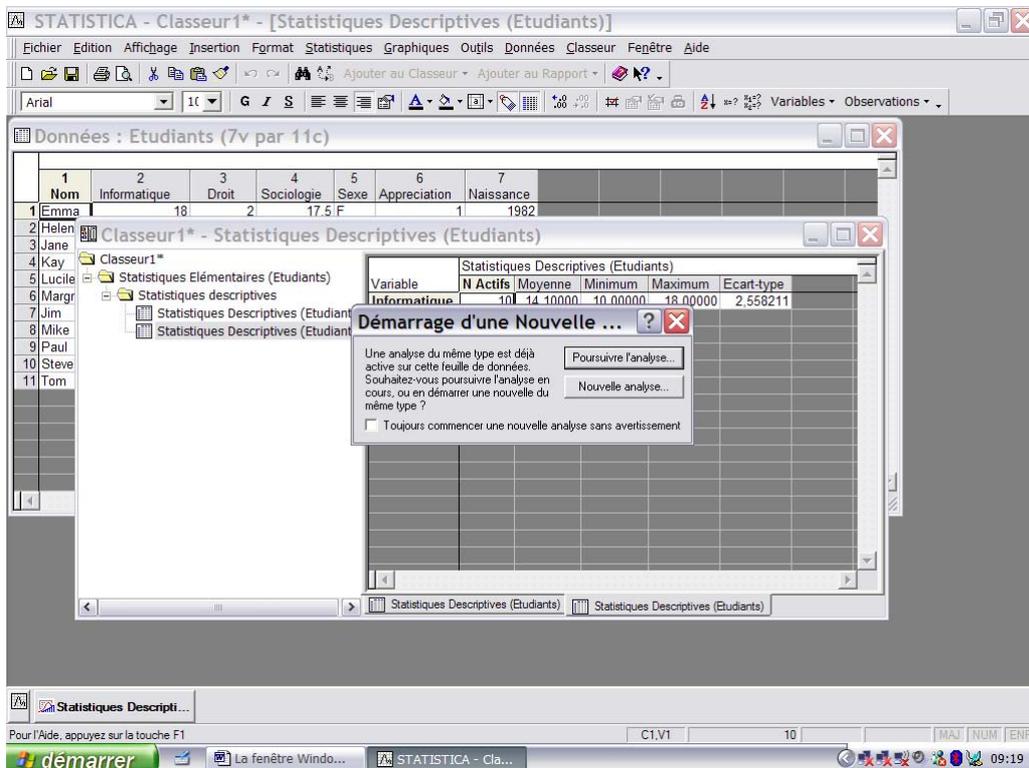
The screenshot shows the STATISTICA software interface. The main window displays a data table with columns: 1 Nom, 2 Informatique, 3 Droit, 4 Sociologie, 5 Sexe, 6 Appreciation, 7 Naissance. The data rows are: 1 Emma (18, 2, 17.5, F, 1, 1982), 2 Helen, 3 Jane, 4 Kay, 5 Lucile, 6 Margr, 7 Jim, 8 Mike, 9 Paul, 10 Steve, 11 Tom.

An overlay window titled "Classeur1* - Statistiques Descriptives (Etudiants)" is open, showing a table of descriptive statistics for the variable "Informatique":

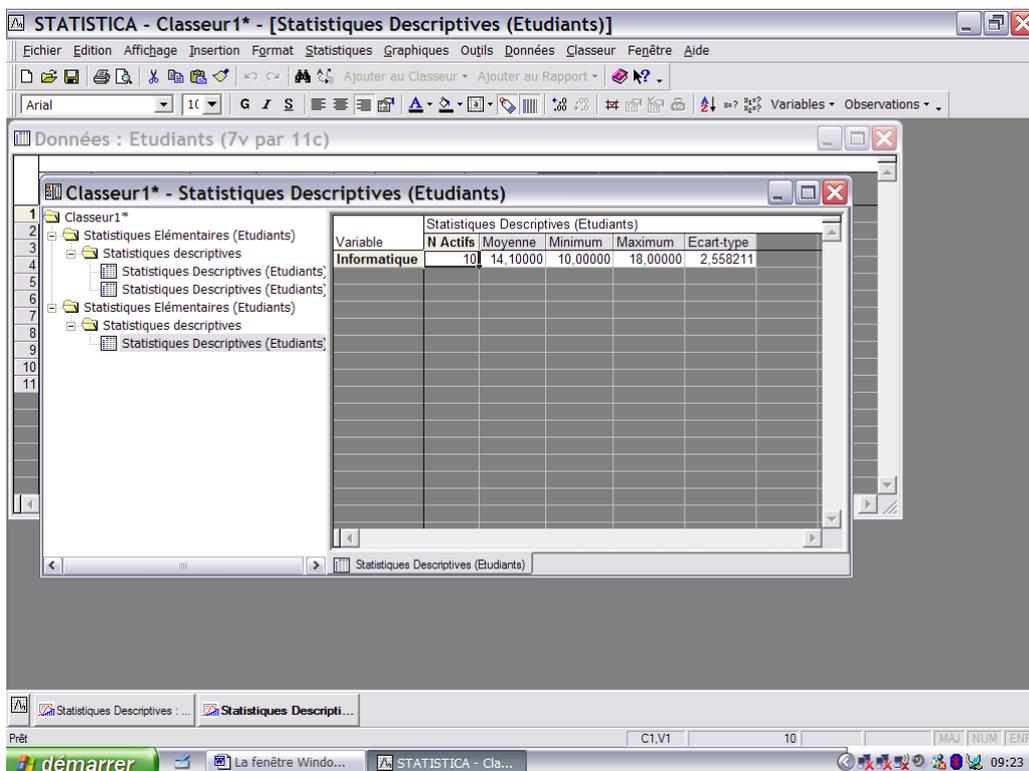
Variable	N Actifs	Moyenne	Minimum	Maximum	Ecart-type
Informatique	10	14.10000	10.00000	18.00000	2.558211

On constate que l'arborescence des résultats (partie à gauche) a été complétée par une ligne supplémentaire correspondant à l'analyse réalisée sur la variable « Informatique » (résultats à droite).

Exemple 2 : - option « Nouvelle analyse » : nous recalculons les paramètres sur la colonne « Informatique » seulement à partir du menu « Statistiques » option « Statistiques élémentaires »

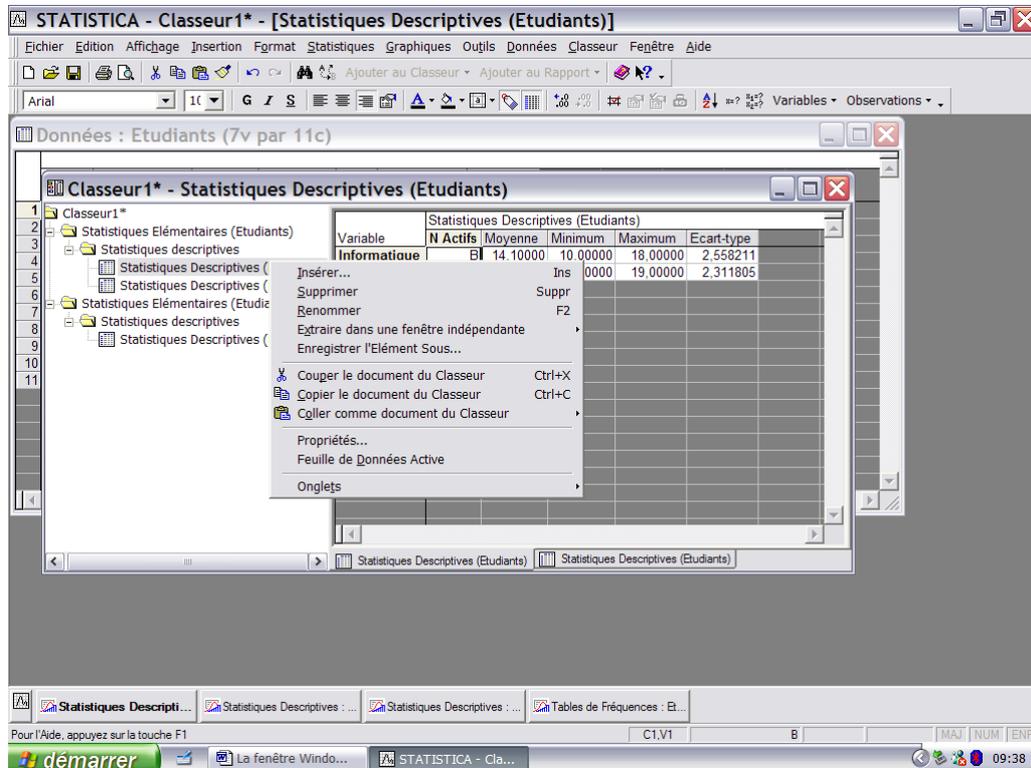


Avec pour résultat, la création d'une subdivision dans l'arborescence :

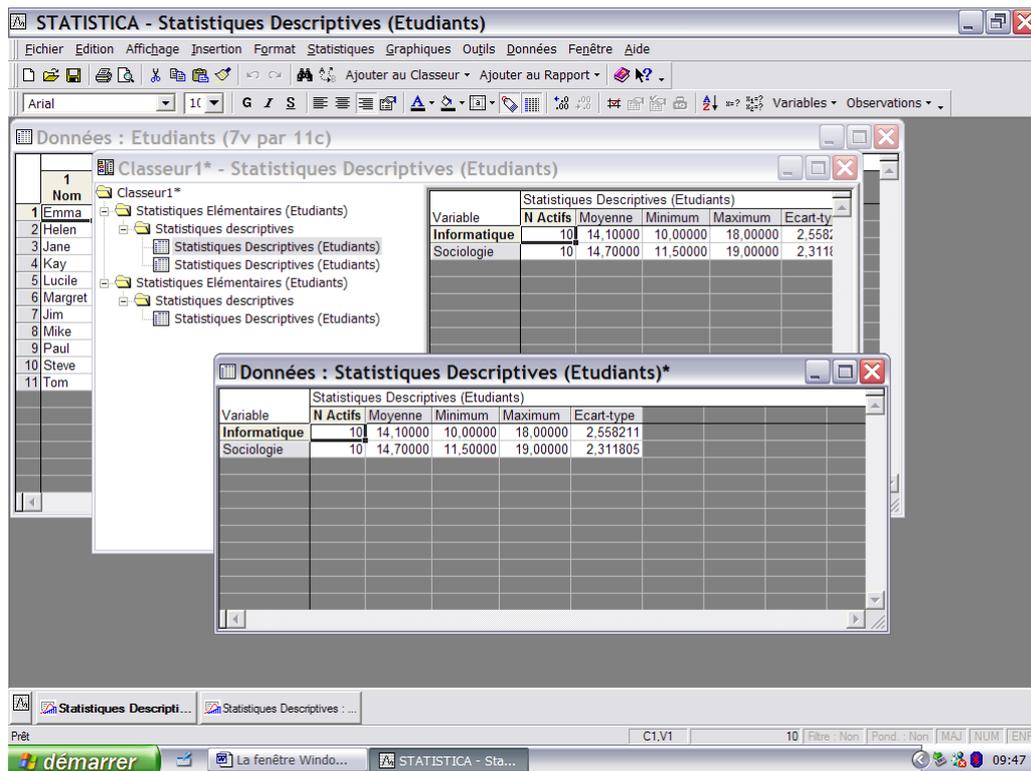


L'existence de l'arborescence dans la partie de droite est fort utile si on désire : pour accéder à des résultats antérieurs, il suffit d'un clic (gauche) pour le faire réapparaître dans la partie de droite (l'utilisation des onglets dans le bas de fenêtre produit le même résultat).

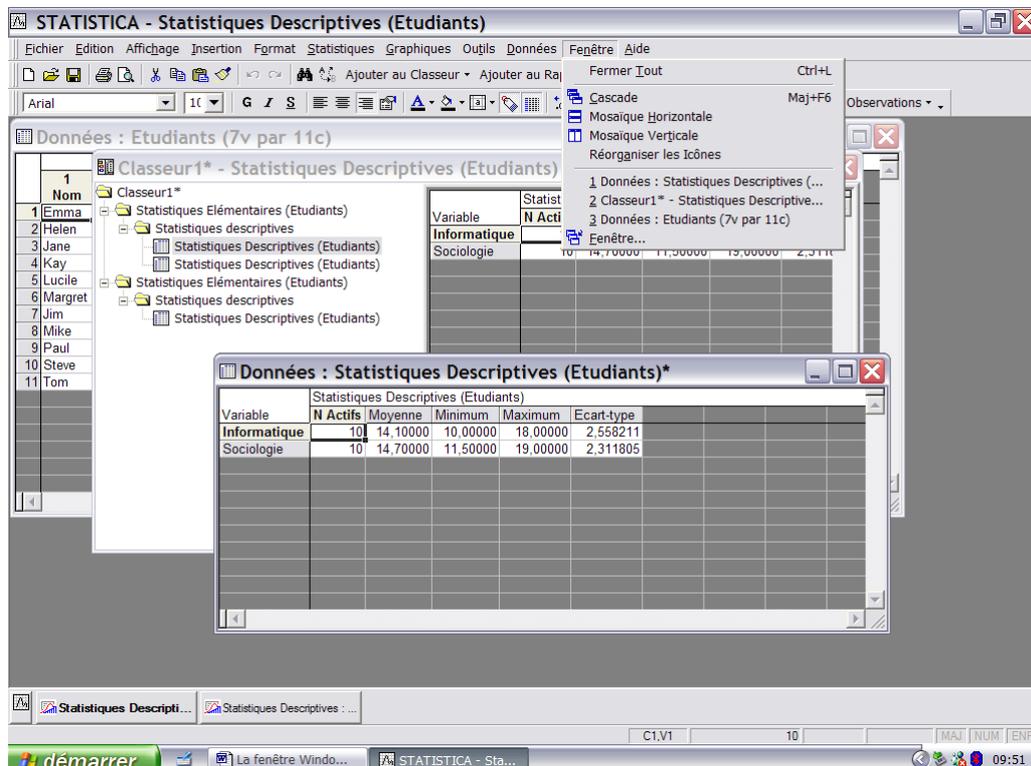
L'arborescence peut faire l'objet de traitements spécifiques en déplaçant le curseur sur l'élément spécifique et en faisant apparaître le menu contextuel (clic droit) :



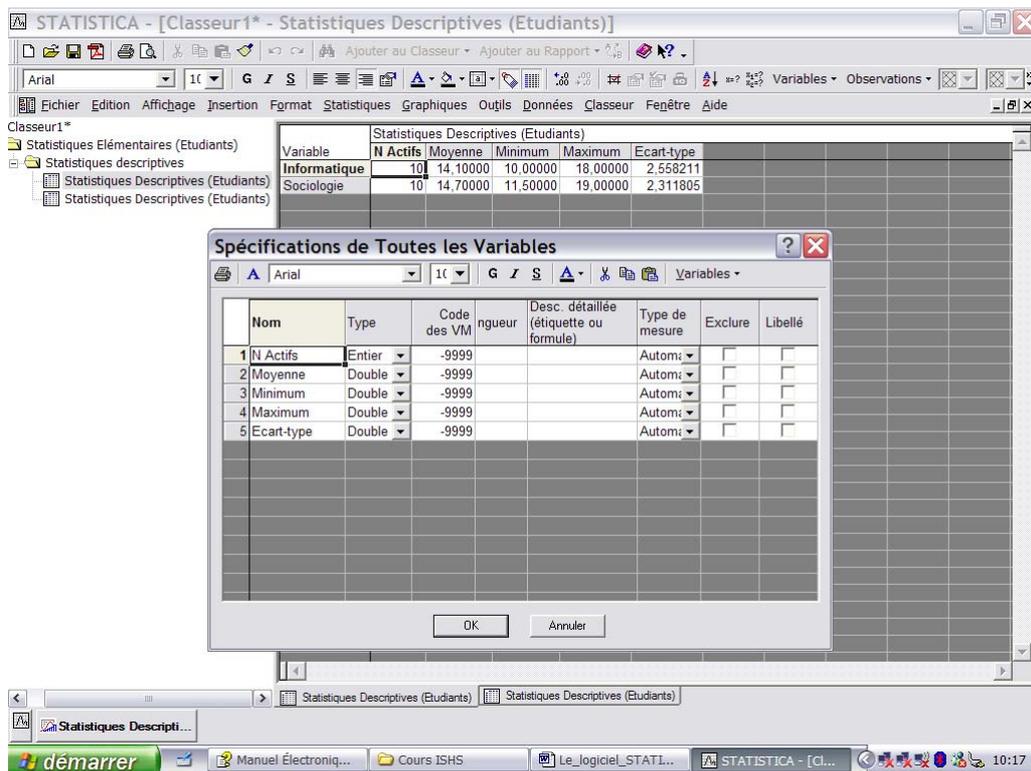
Les trois premières options ne posent pas de problème particulier ; l'option « Extraire dans une fenêtre indépendante » est extrêmement intéressante parce qu'elle permet de créer une feuille de données qui contiendra les résultats statistiques mais qui pourront dès lors être l'objet de calculs. Si on choisit cette possibilité, (avec l'option « copie » qui conserve la copie dans le Classeur), une nouvelle feuille de données apparaît à l'écran et la barre des menus contient une entrée supplémentaire (« Classeur ») :



Si on désire un état des feuilles actives, on choisit le menu « Fenêtre » :

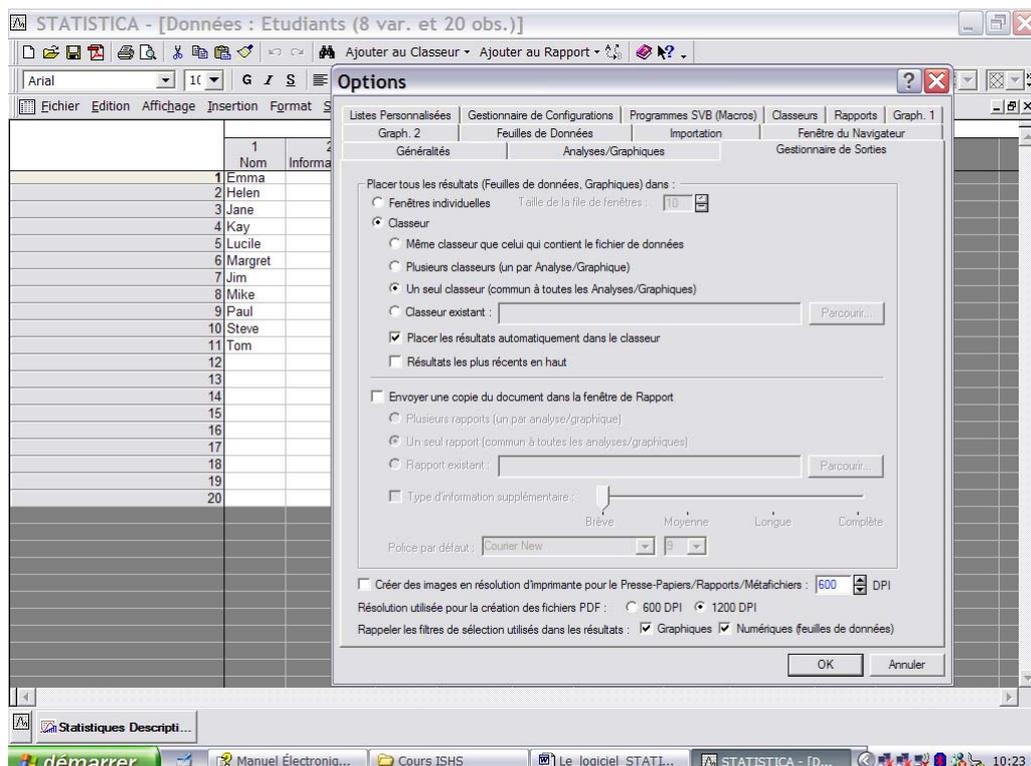


La feuille de données « Statistiques Descriptives (Etudiants) » apparaît en premier dans l'ordre des fenêtres, c'est désormais à ce fichier de données (mais qui sont en réalité des résultats !) que vont s'appliquer les fonctionnalités STATISTICA. Par exemple, l'option « Spécifications de toutes les variables » du menu « Données » :



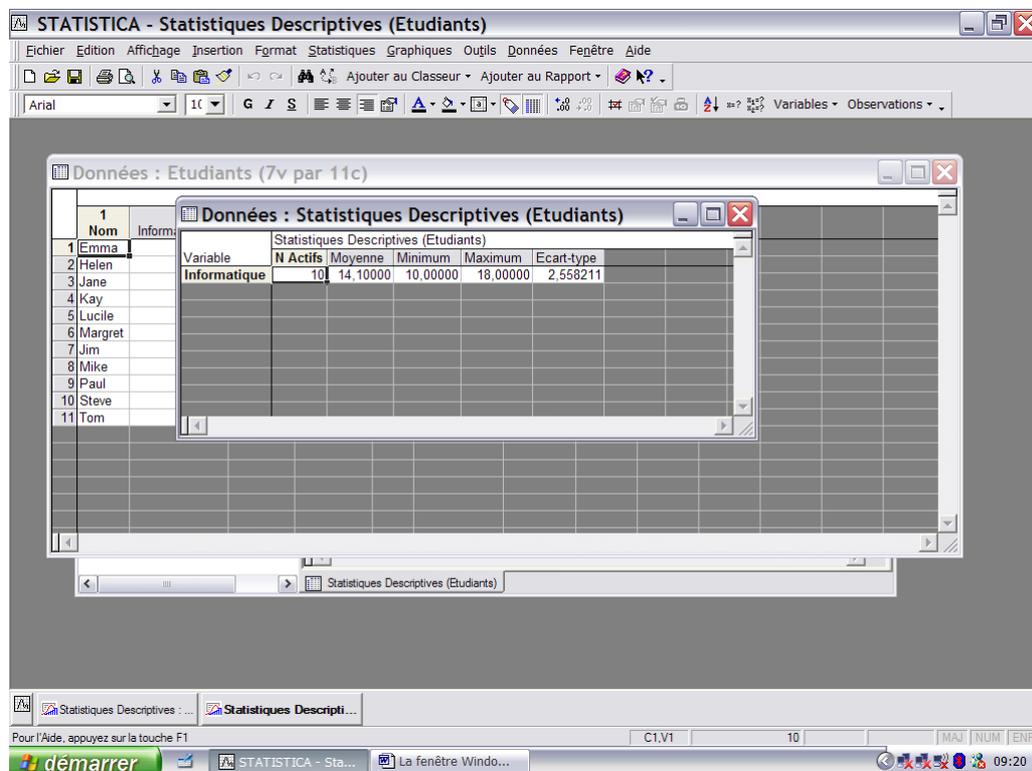
Pour redéfinir le fichier de départ (« Etudiant ») comme feuille de données actives, on clique simplement sur le nom de la fenêtre (ou encore on sélectionne son nom dans le menu « Fenêtre » ci-dessus).

La présentation des résultats dans une fenêtre partagée par une arborescence de classement à gauche dépend en réalité d'un paramétrage spécifique de STATISTICA ; le menu « Outils » dans lequel on sélectionne « Options » permet d'obtenir un état des différents paramètres en vigueur, l'onglet « Gestionnaire de Sorties » permet de gérer la fenêtre de résultats :



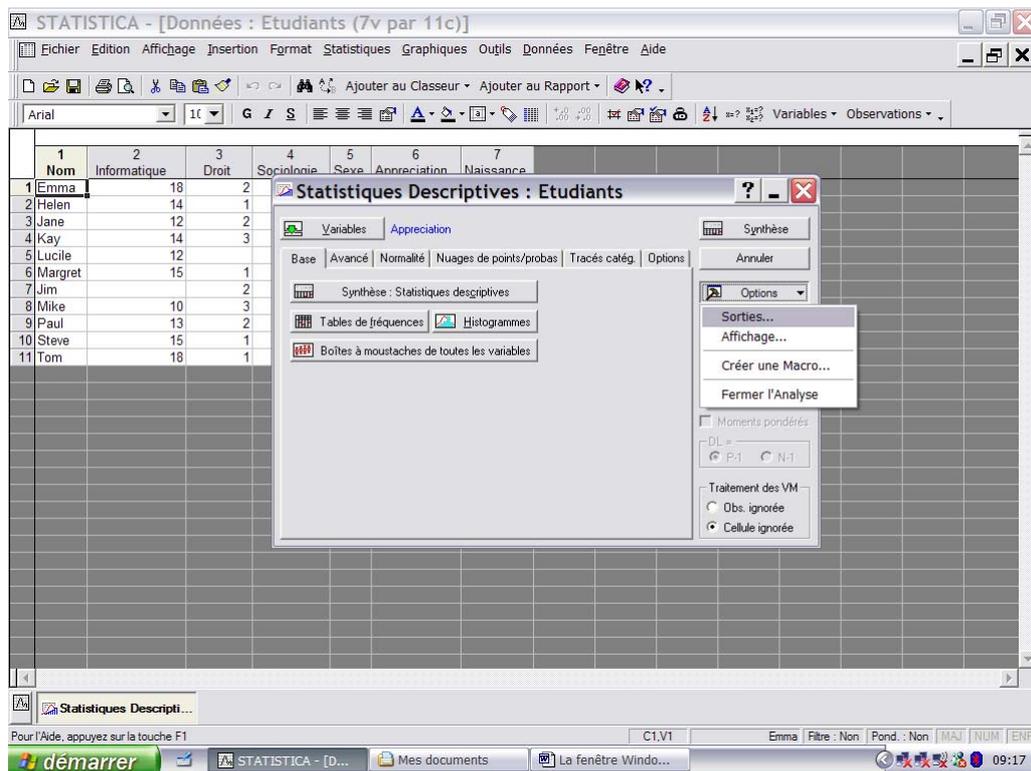
Dans l'écran ci-dessus, la sélection du bouton « Classeur » garantit la présentation des résultats comme ci-dessus.

La sélection du bouton radio « Fenêtres Individuelle » a pour conséquence la suppression de l'arborescence dans la fenêtre des résultats. Par exemple, si on recalcule les statistiques élémentaires sur la variable « Informatique » :



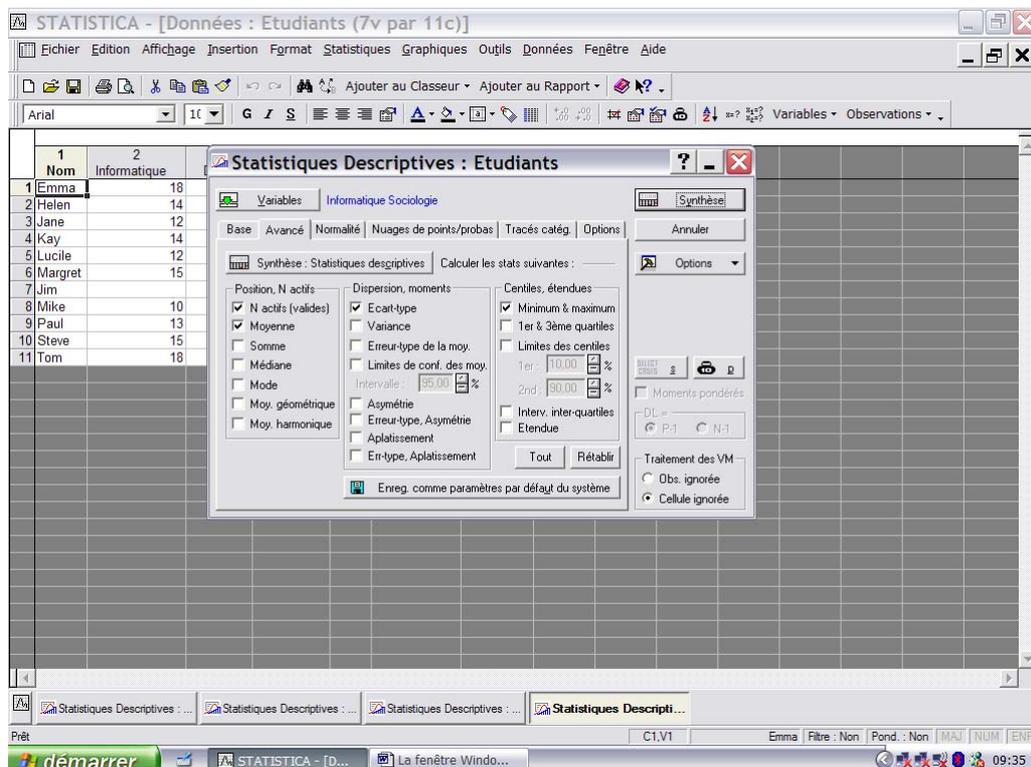
La feuille de résultats devient automatiquement la feuille de données actives.

La gestion des sorties est accessible immédiatement à partir de l'écran « Statistiques Descriptives » via le bouton « Options » dans lequel on sélectionne « Sorties » :

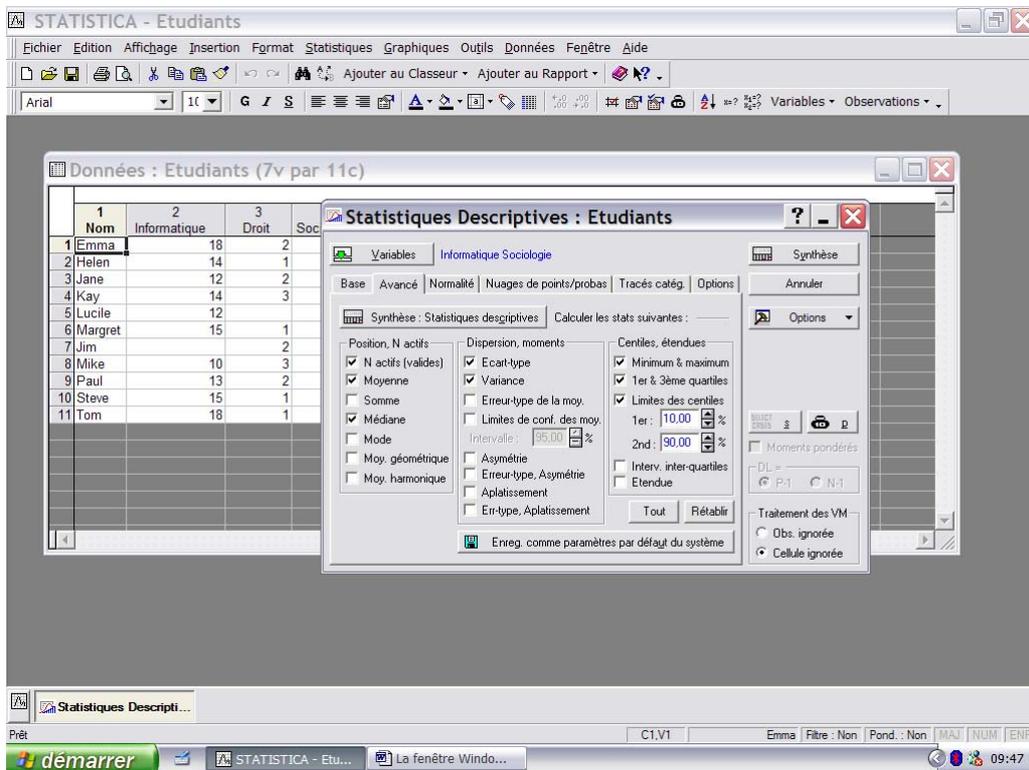


Approfondissements

Comme on l'a vu, la fonctionnalité « Statistiques Descriptives » propose par défaut cinq paramètres statistiques pour synthétiser les informations concernant les variables sous analyse (cf page 37). La sélection de l'onglet « Avancé » permet de préciser de façon très détaillée les analyses que l'utilisateur désire calculer :

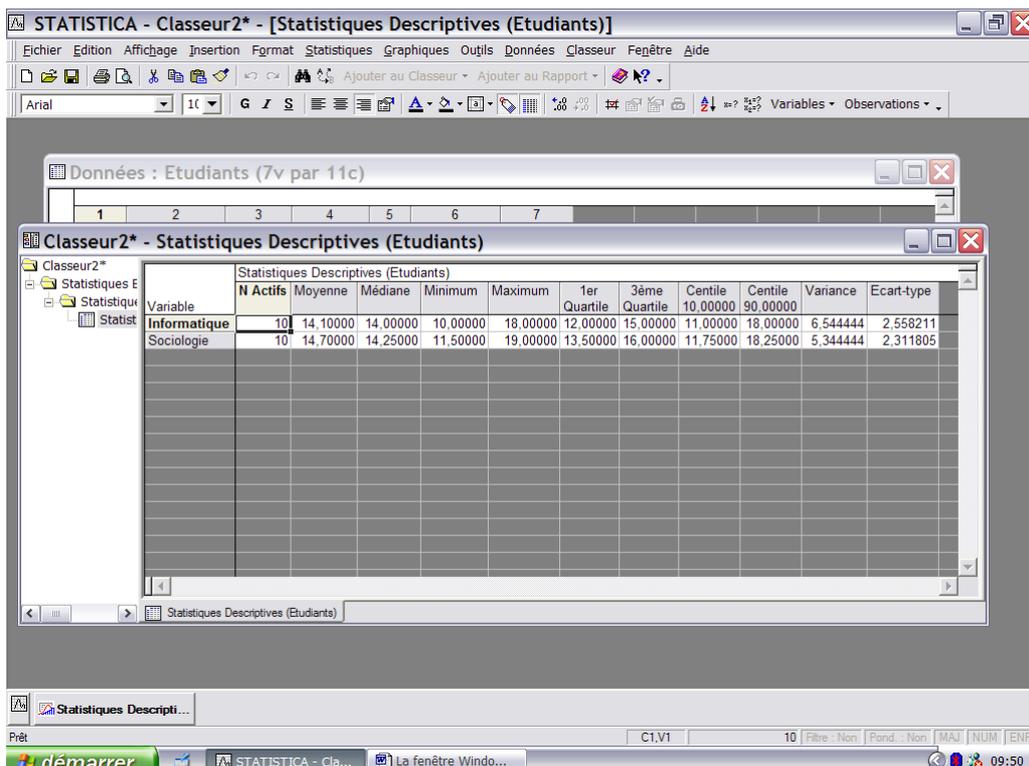


Pour les variables sélectionnées (notes à l'examen d'Informatique et à celui de Sociologie), étant donné qu'il s'agit de variables numériques de ratio, toutes les options de calcul sont permises, en bas de fenêtre, elles sont enregistrables comme valeurs par défaut, par exemple :



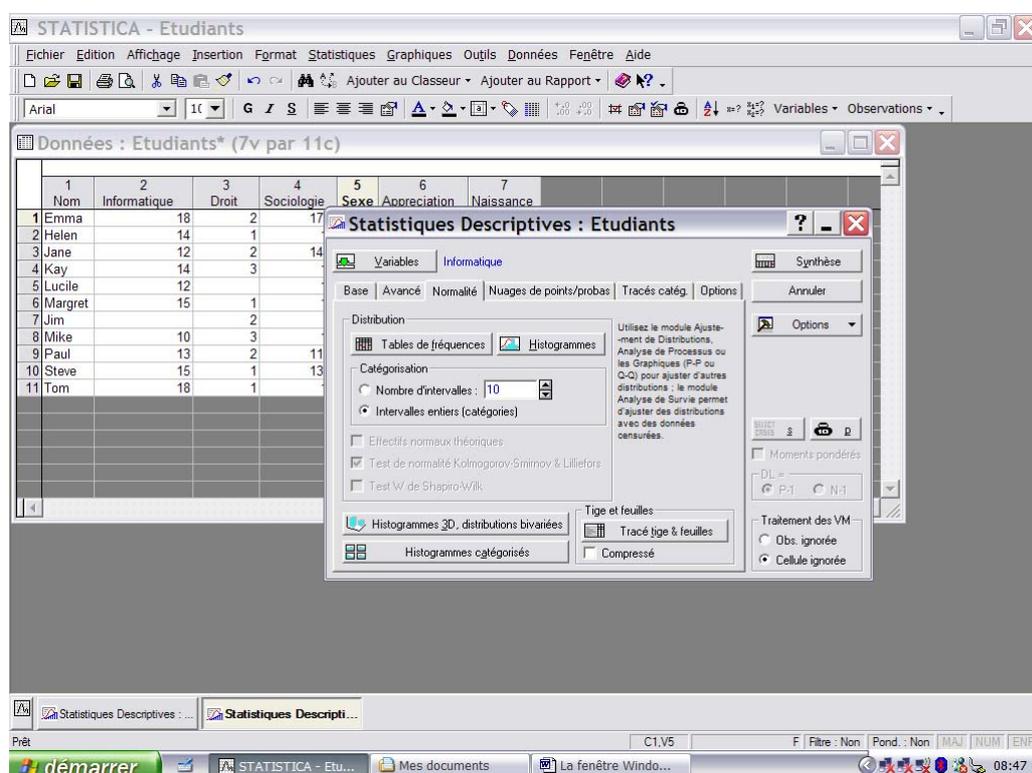
n.b. : le bouton « Tout » permet la sélection globale, le bouton « Rétablir » rétablit les options par défaut

La fenêtre des résultats se présente sous la forme tout à fait classique :



S'il s'agit à présent de procéder à une analyse statistique des variables concernant les notes de l'examen de Droit et l'appréciation que portent les étudiants à leurs études, on se souviendra qu'il s'agit de variables ordinales pour lesquelles des outils de synthétisation spécifique existent. Les options qui permettent de calculer la moyenne arithmétique, la variance (écart-type) sont bien évidemment exclues. S'il s'agit d'analyser la variable nominale « Sexe », le nombre d'options diminue encore ; STATISTICA ne propose évidemment pas le choix adapté.

L'onglet « Normalité » permet de définir un certain nombre d'options relatives à l'étude de la distribution des valeurs d'une variable à laquelle on s'intéresse. Par exemple, si on désire plus d'informations à propos des différentes notes obtenues (valeurs entières) par les étudiants à l'examen d'Informatique :



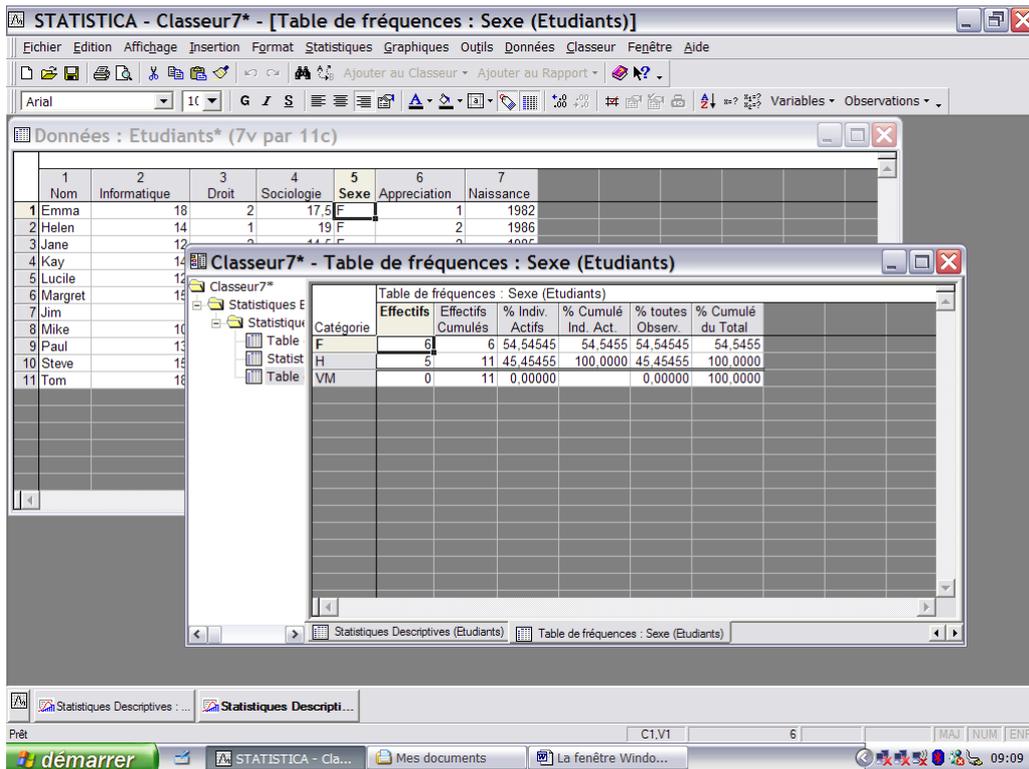
Le bouton « Synthèse » produit les mêmes résultats quant aux paramètres statistiques mais le bouton « Tables de fréquences » procède au comptage des différentes notes attribuées :

The screenshot shows the STATISTICA software interface with a window titled 'Table de fréquences : Informatique: Note à l'examen d'Info...'. The table displays the following data:

Catégorie	Effectifs	Effectifs Cumulés	% Indiv. Actifs	% Cumulé Ind. Act.	% toutes Observ	% Cumulé du Total
10	1	1	10,00000	10,0000	9,09091	9,0909
12	2	3	20,00000	30,0000	18,18182	27,2727
13	1	4	10,00000	40,0000	9,09091	36,3636
14	2	6	20,00000	60,0000	18,18182	54,5455
15	2	8	20,00000	80,0000	18,18182	72,7273
18	2	10	20,00000	100,0000	18,18182	90,9091
VM	1	11	10,00000		9,09091	100,0000

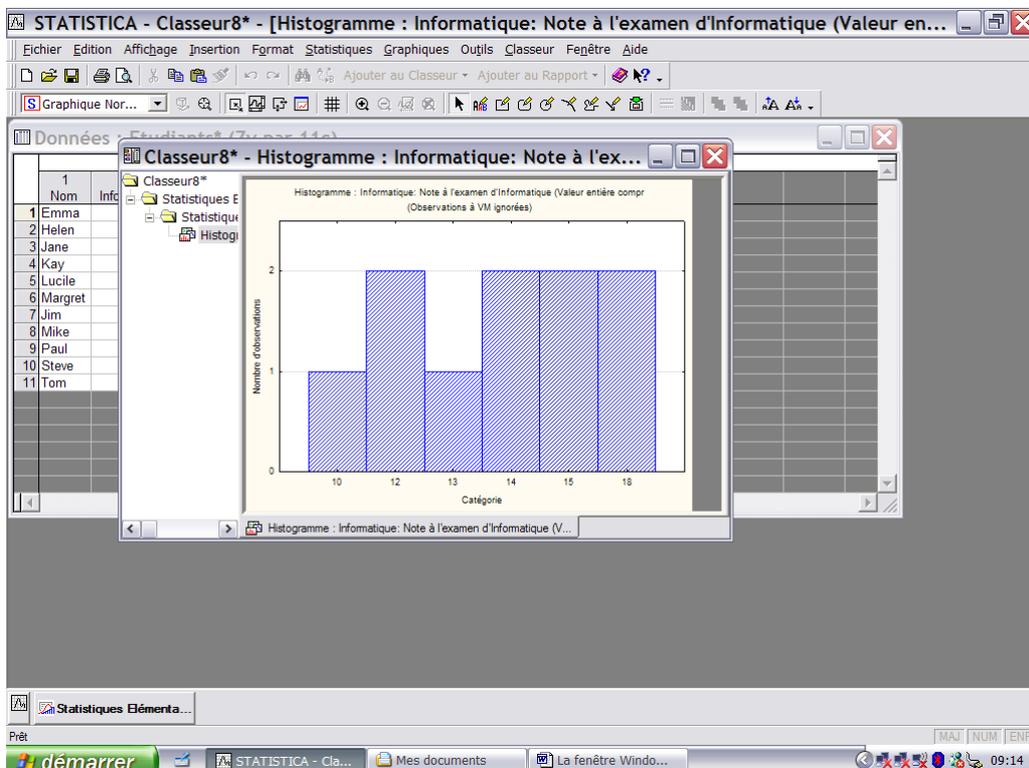
Les différentes notes de l'examen attribuées à l'examen d'Informatique apparaissent dans la colonne « Catégorie », il y a une valeur manquante (VM), en l'occurrence celle de l'étudiant Jim. Le comptage des différentes notes se trouve dans « Effectifs » dont le cumul apparaît dans « Effectifs Cumulés » (6 étudiants ont obtenu une note inférieure ou égale à 14). La colonne « % Indiv. Actifs » est le rapport entre les « Effectifs » et le nombre de valeurs non manquantes (10) ; les « % Cumulé Ind. Act. » est leur valeur cumulée (60% des étudiants ont une note inférieure ou égale à 14). Les deux dernières colonnes donnent les mêmes pourcentages mais calculés sur le nombre d'observations (valeurs manquantes comprises) et sont donc moins utiles.

Cette table de fréquences s'applique aussi dans le cas de variables non numériques, par exemple, si on désire compter le nombre d'étudiants de chaque sexe :



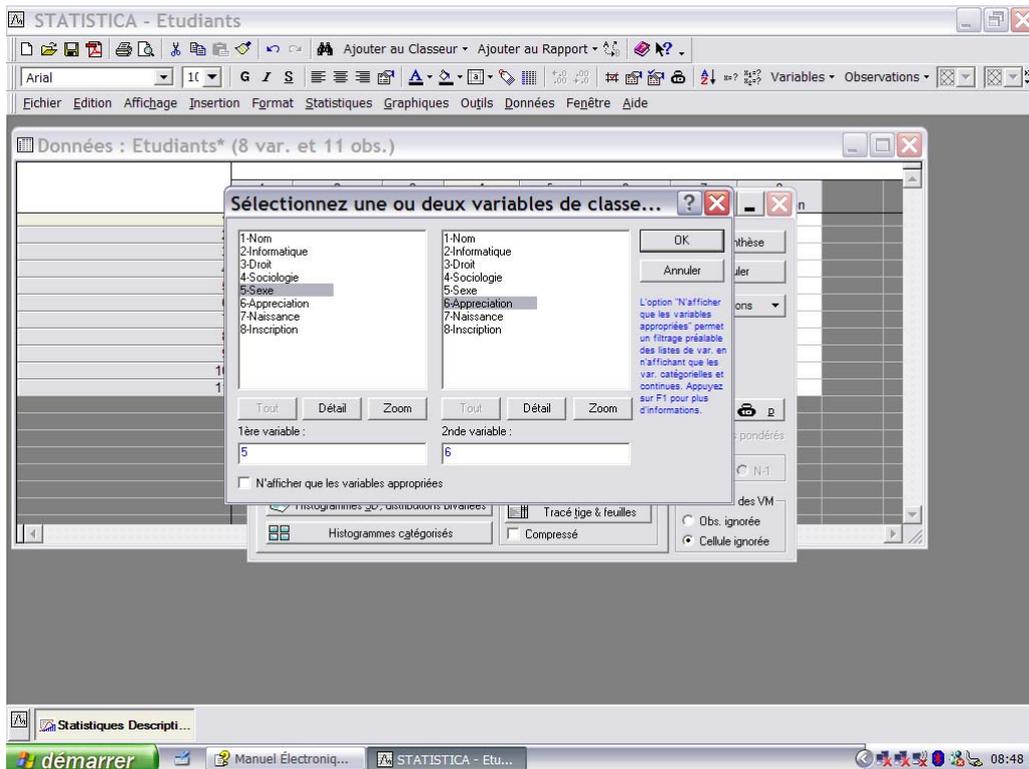
Représentation graphique des résultats

Enfin, on peut procéder à la représentation graphique de la distribution des valeurs via le bouton « Histogramme » :

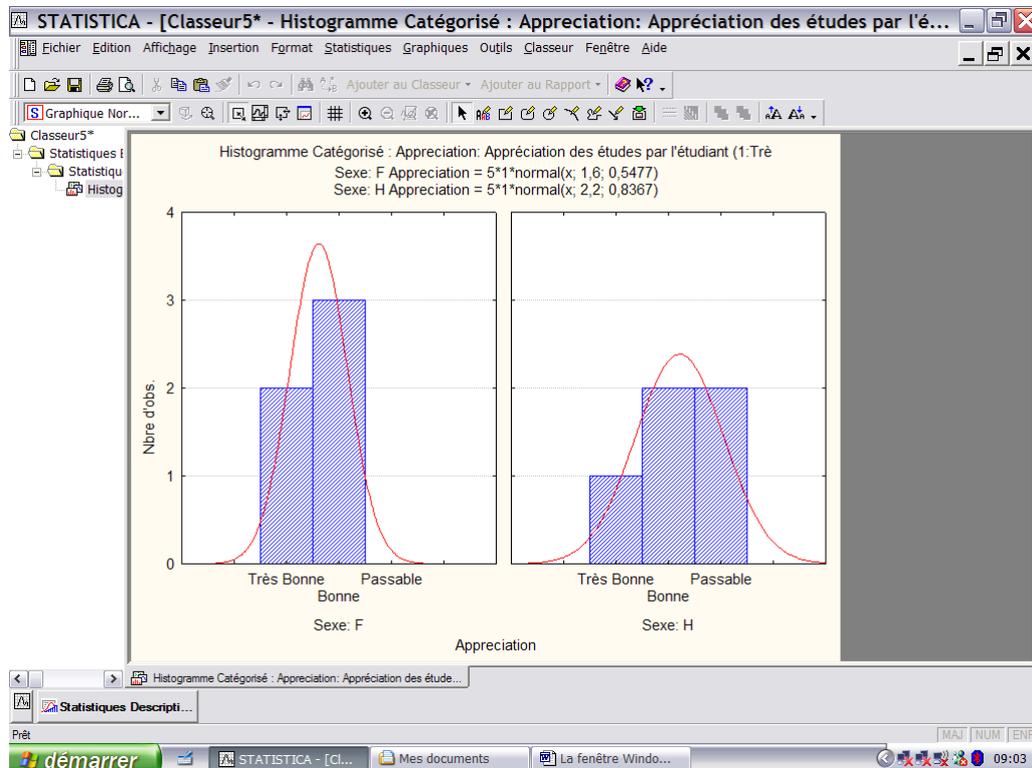


Ce graphique représente les effectifs de chaque catégorie de la distribution des fréquences en portant en ordonnées la fréquence observée et en abscisses les différentes catégories (dans ce cas, les notes qui ont été attribuées).

Les histogrammes peuvent être représentés en fonction de catégories à définir à partir des variables du tableau de données ; ainsi, il peut être utile de représenter graphiquement l'appréciation quant aux études en fonction du sexe afin de vérifier l'existence d'une éventuelle relation entre ces deux variables. On choisit d'abord la variable à catégoriser grâce au bouton « Variables » (dans ce cas, c'est la variable « Appreciation »). Le bouton « Histogrammes catégorisés » de l'onglet « Normalité » affiche l'écran suivant :



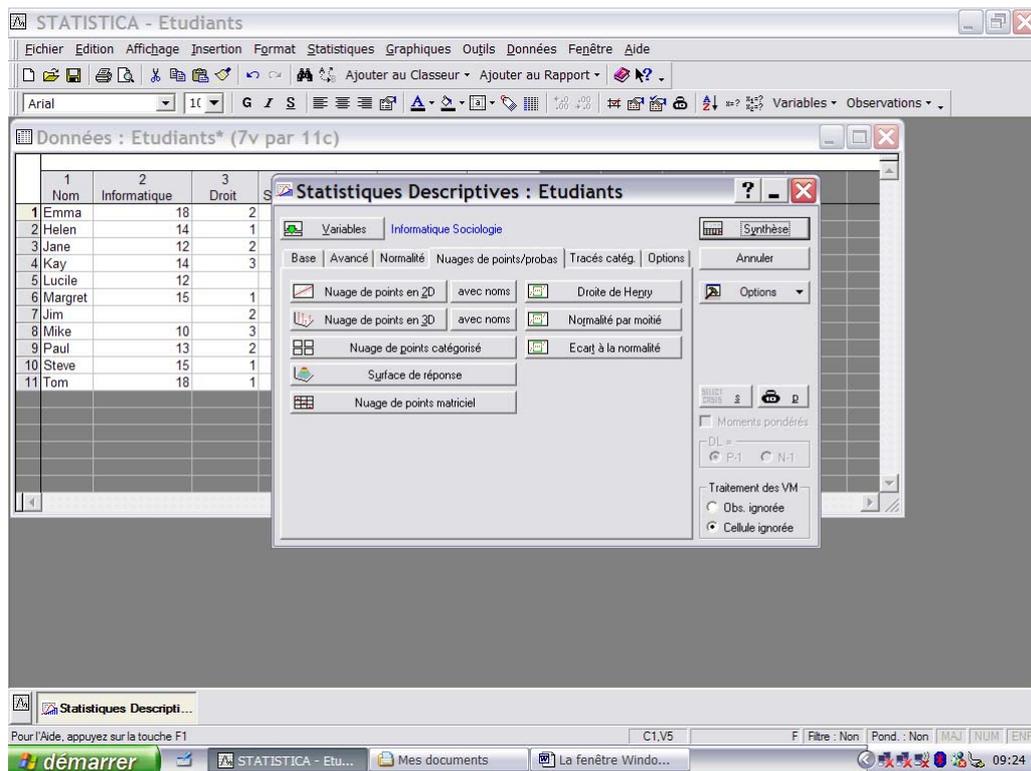
Si l'histogramme est catégorisé à partir de la variable « Sexe » seulement, on choisit le nom de cette variable dans la fenêtre de gauche, ensuite les valeurs des catégories pour cette variable, l'option « Tous » correspond aux codes « H » et « F » ; le résultat graphique est le suivant :



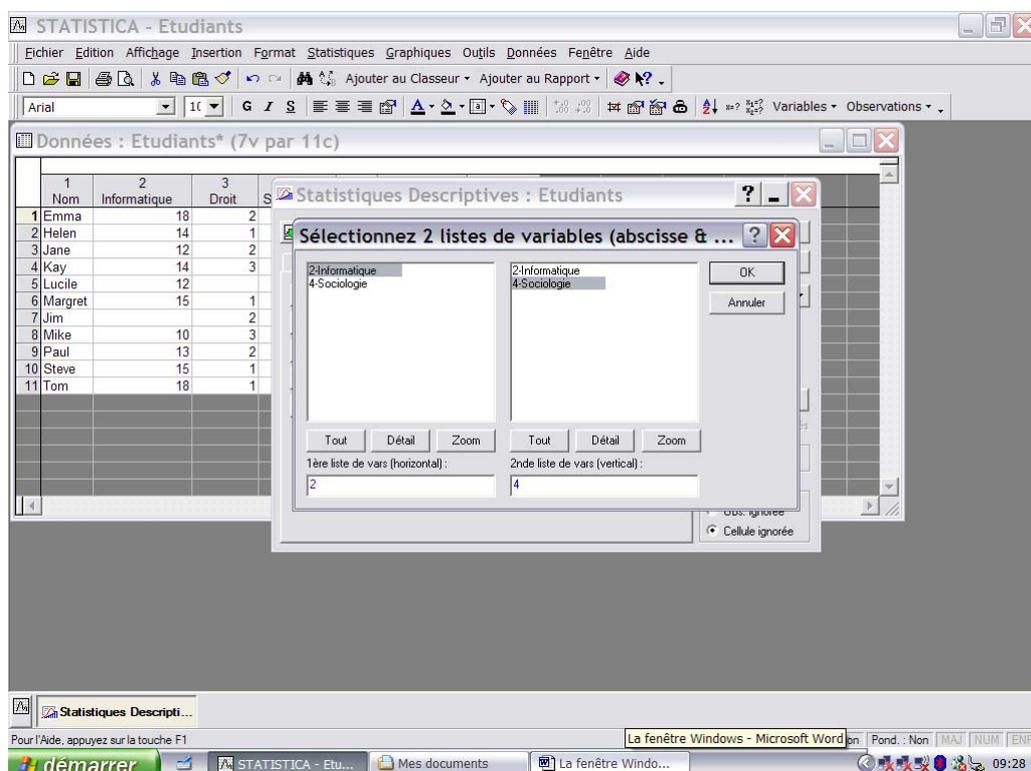
On reconnaît à gauche les différentes valeurs de la variable pour les Filles et à droite pour les Hommes ; un ajustement de type gaussien est automatiquement réalisé.

Les histogrammes peuvent être catégorisés à partir de deux variables et les graphiques sont affichés en fonction de tous les couples possibles de modalités des variables de classement.

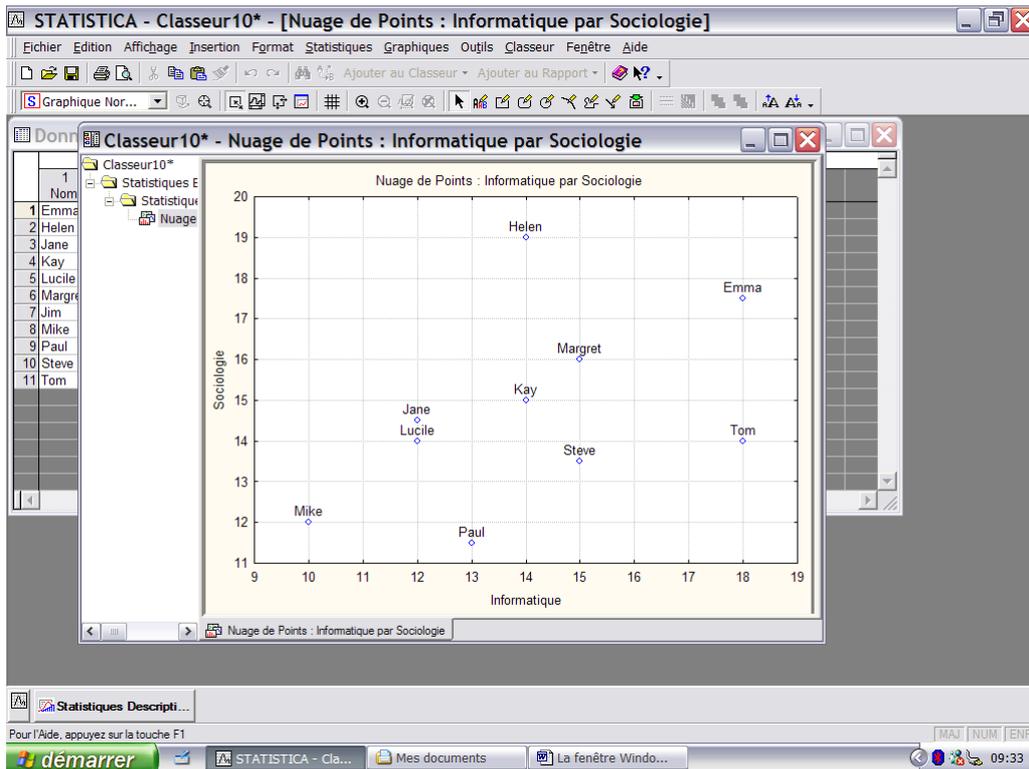
L'onglet « Nuage de points/probas » est utile, entre autres, pour représenter graphiquement les lignes à partir de différentes variables d'observations ; on désirerait par exemple positionner les différents étudiants à partir de leurs notes dans les examens d'Informatique et de Sociologie afin de vérifier l'hypothèse que celui qui obtient une note élevée à l'un d'eux réussit aussi bien dans l'autre. L'onglet affiche :



On commence l'analyse par « avec noms » à droite de « Nuage de points en 2D » ce qui produit l'écran suivant :

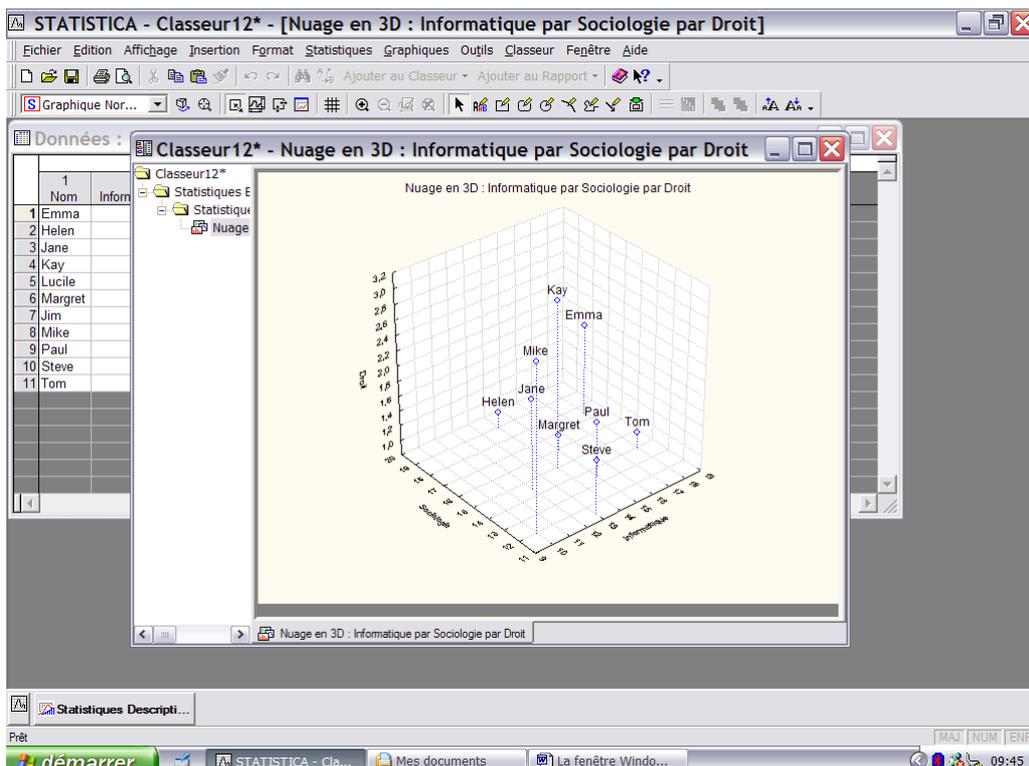


On choisit de placer les notes à l'examen d'Informatique sur l'axe horizontal (« 2 » dans « 1^{ère} liste de vars (horizontal) » et celles de Sociologie sur l'axe vertical (« 4 » dans « 2^{nde} liste de vars (vertical) »). Le tracé produit est :

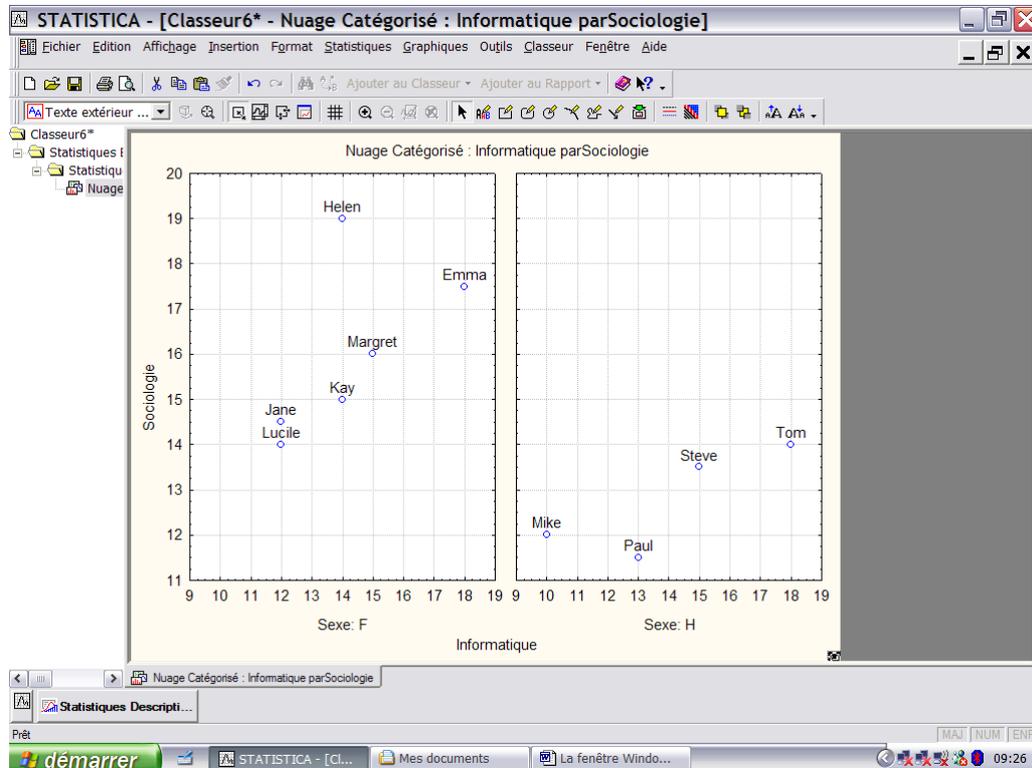


Les étudiants Helen et Tom semblent contredire notre hypothèse de départ.

L'analyse graphique du nuage de points peut être complétée par une « troisième dimension » ; par exemple, on peut étendre notre hypothèse concernant les notes obtenues pour les trois cours (y compris les notes à l'examen de Droit) en spécifiant « avec noms » à droite de « Nuage de points en 3D » et en positionnant ensuite les trois variables sur les axes (cf. ci-dessus pour le graphique en 2D) :

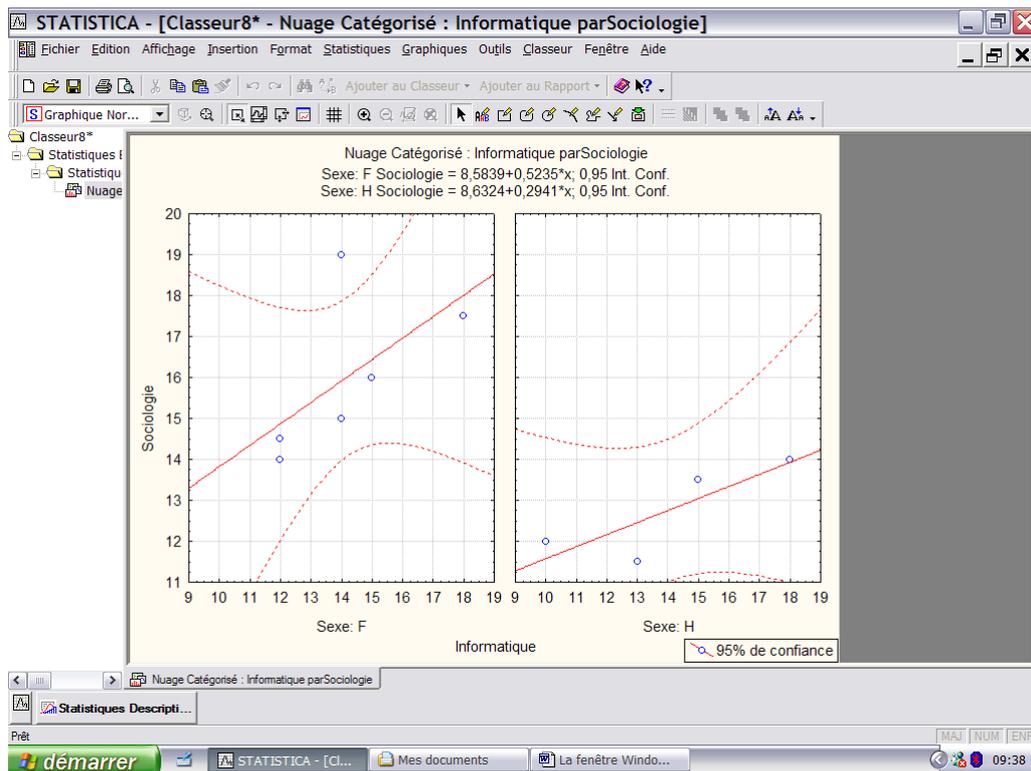


Enfin, le bouton « Nuage de points catégorisé » autorise la définition d'une ou deux variables de croisement pour la représentation du nuage des observations ; par exemple, si on représente les notes obtenues par les étudiants dans les cours d'Informatique et de Sociologie en fonction du sexe (cf. Histogrammes catégorisés ci-dessus pour les paramétrages des menus), on obtient (après quelques modifications graphiques) :

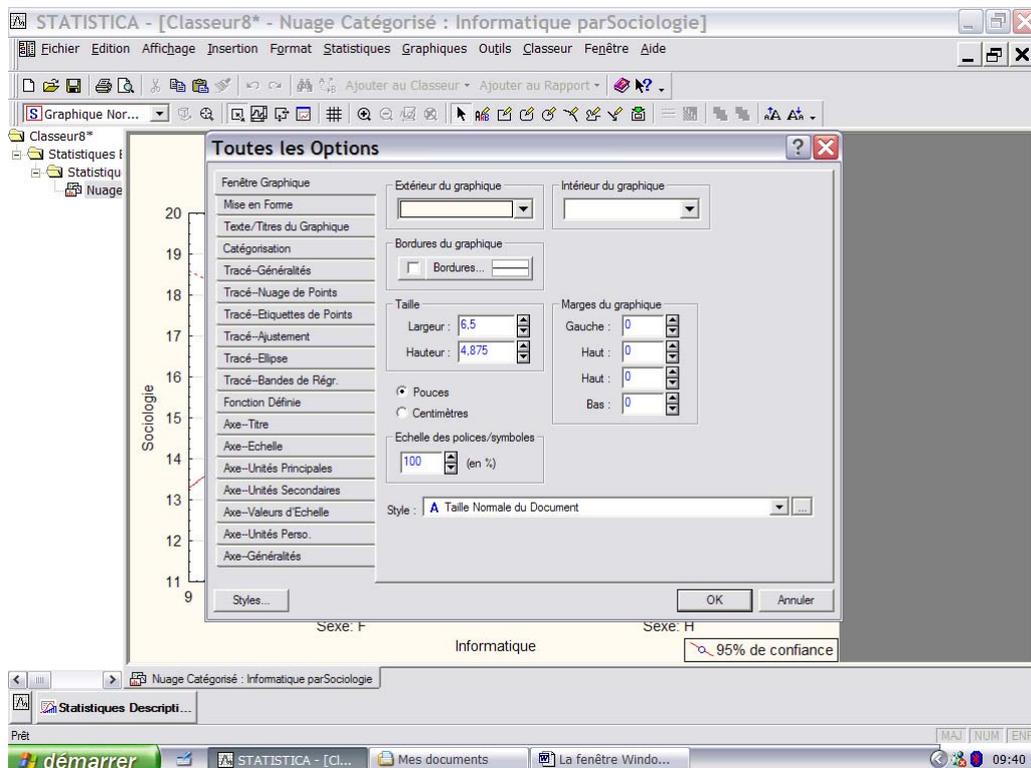


Commentaire général sur les graphiques en STATISTICA

Les nombreux résultats graphiques accessibles dans le logiciel STATISTICA ont une particularité commune : ils sont tous « éditables » c'est-à-dire modifiables. Par exemple, pour arriver à la représentation du graphique ci-dessus, il faut partir du résultat « par défaut » qui est en réalité le suivant :

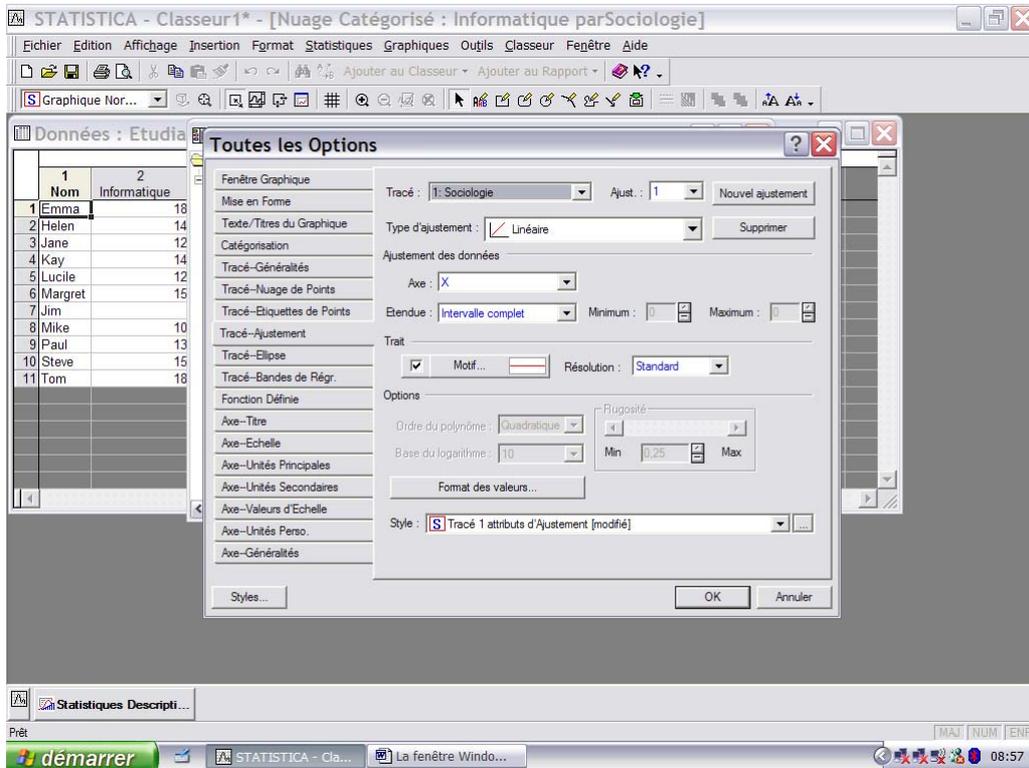


Pour modifier la présentation, on commence par un double clic souris dans la fenêtre du graphique (ou via le menu contextuel et l'option « Propriétés du Graphique ») afin de faire apparaître les outils d'édition sous la forme d'une liste d'onglets disposés verticalement :

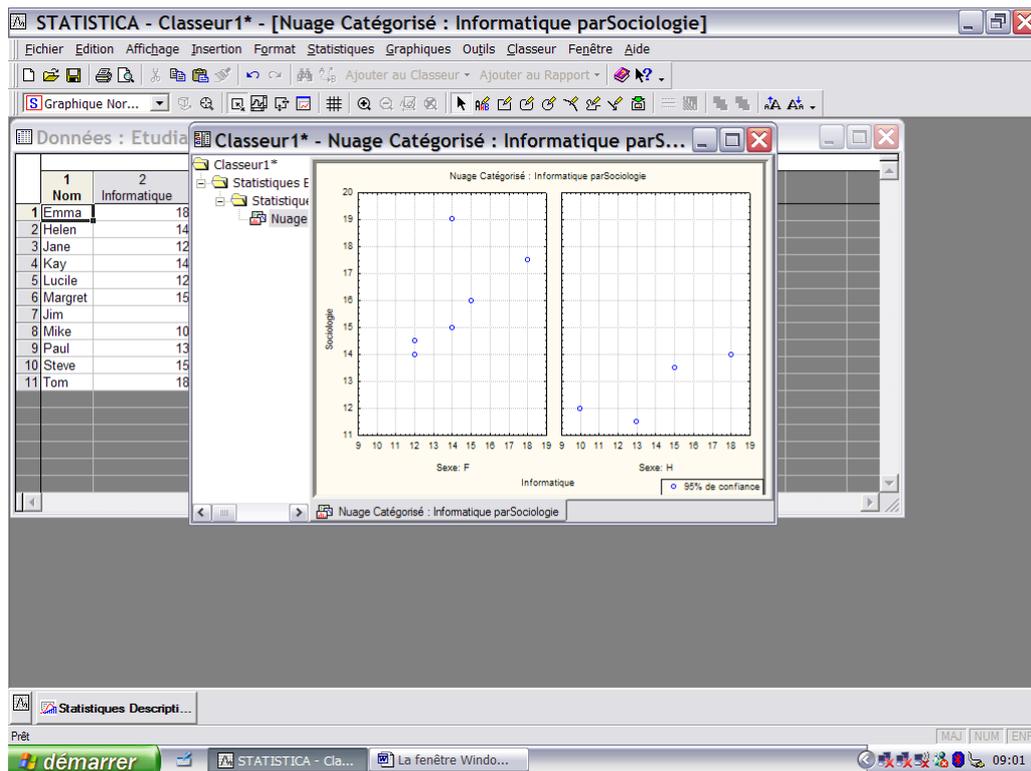


Les onglets permettent de modifier toutes les composantes aussi bien du graphique en lui-même que de sa fenêtre de présentation ; par défaut, c'est le premier onglet qui s'ouvre. LI permet de modifier les couleurs de la fenêtre, les bordures du graphique, sa taille, ses marges, etc...

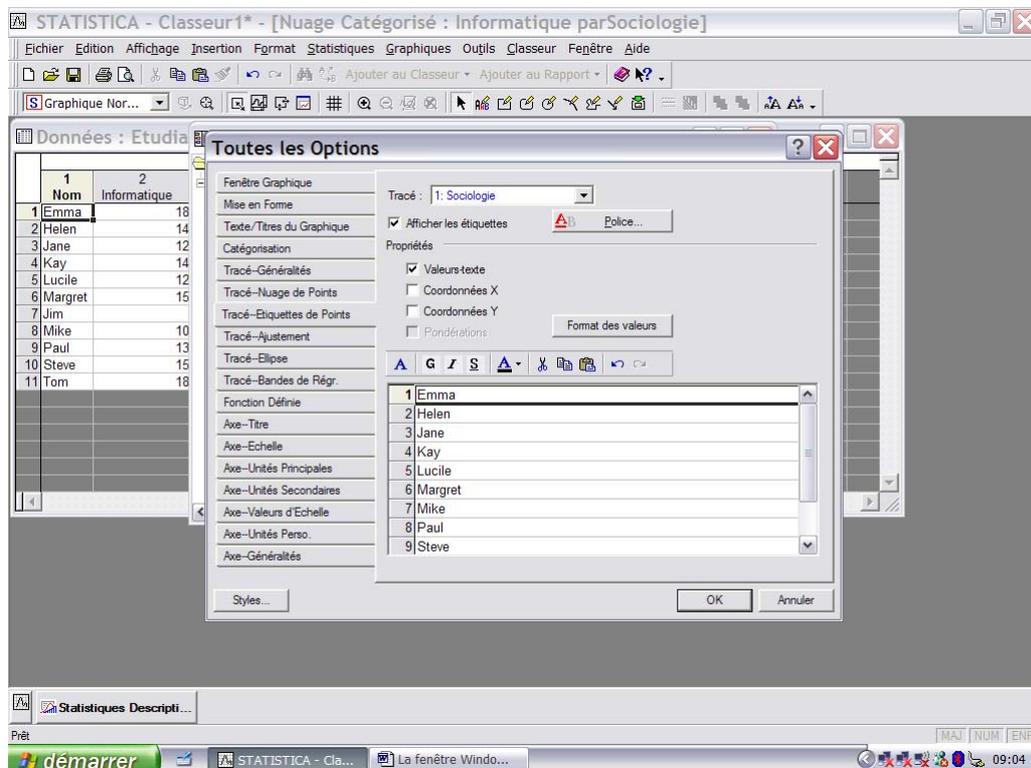
L'onglet « Tracé-Ajustement » nous permettra de supprimer du nuage de points les éléments qui ne sont pas indispensables à la représentation simple d'un nuage de points :



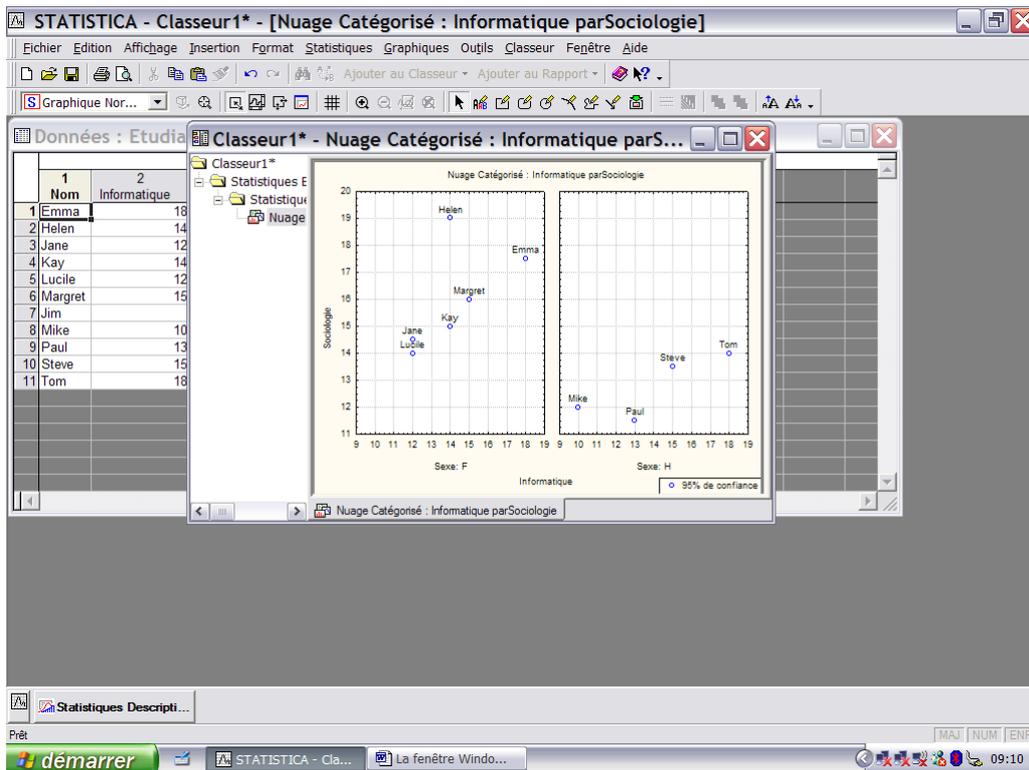
En utilisant le bouton « Supprimer » à droite de la fenêtre « Type d'ajustement », on commence par supprimer les droites présentes au départ ainsi que les courbes définissant des intervalles de confiance :



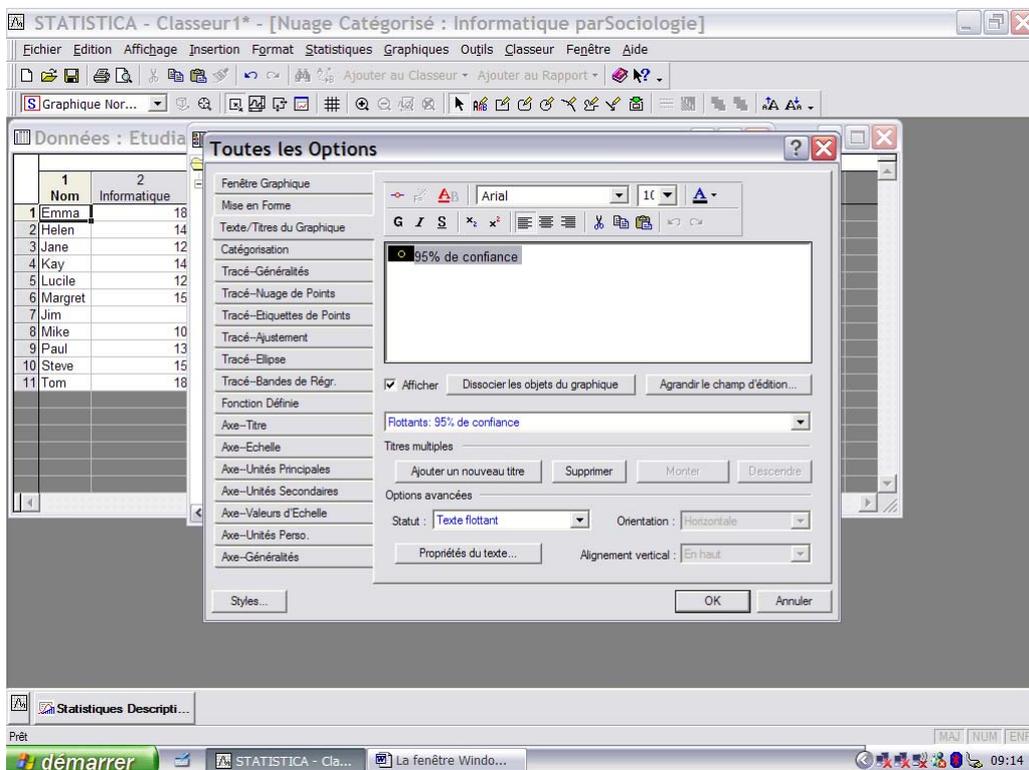
Il reste à « étiquetter » chaque point-observation par le nom de l'étudiant, on choisit pour cela l'onglet « Tracé-Etiquettes de points » et on coche l'option « Valeur-Texte », on peut modifier également les noms des étiquettes (noms d'étudiants) :



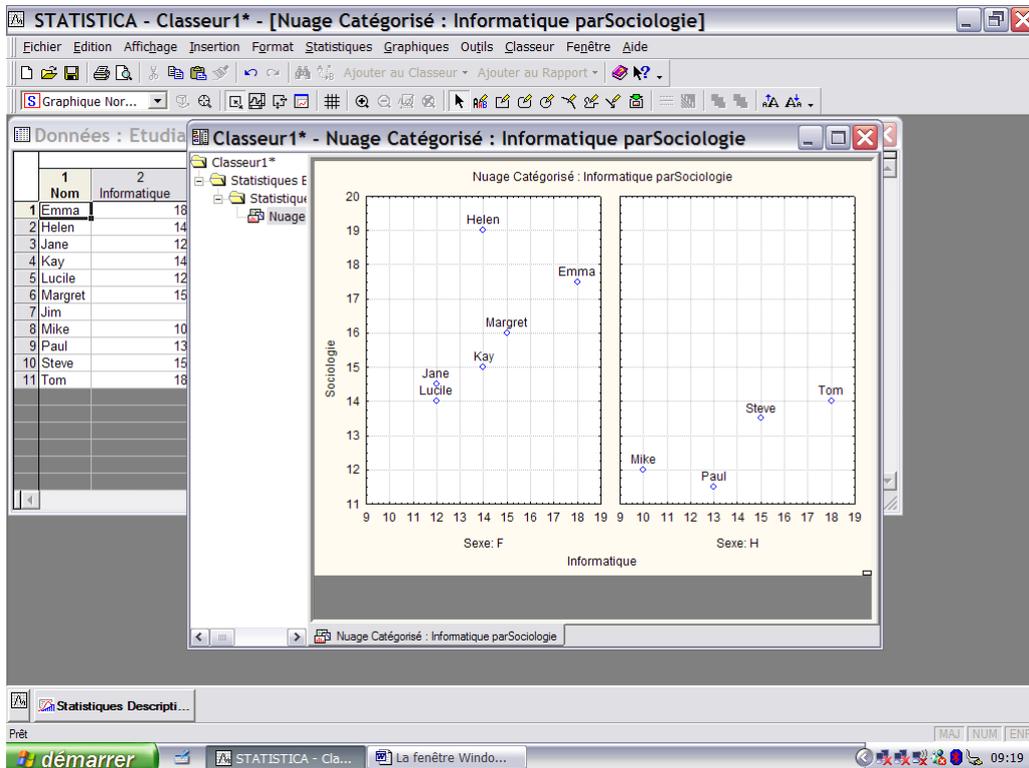
Cette étape affiche le graphique suivant :



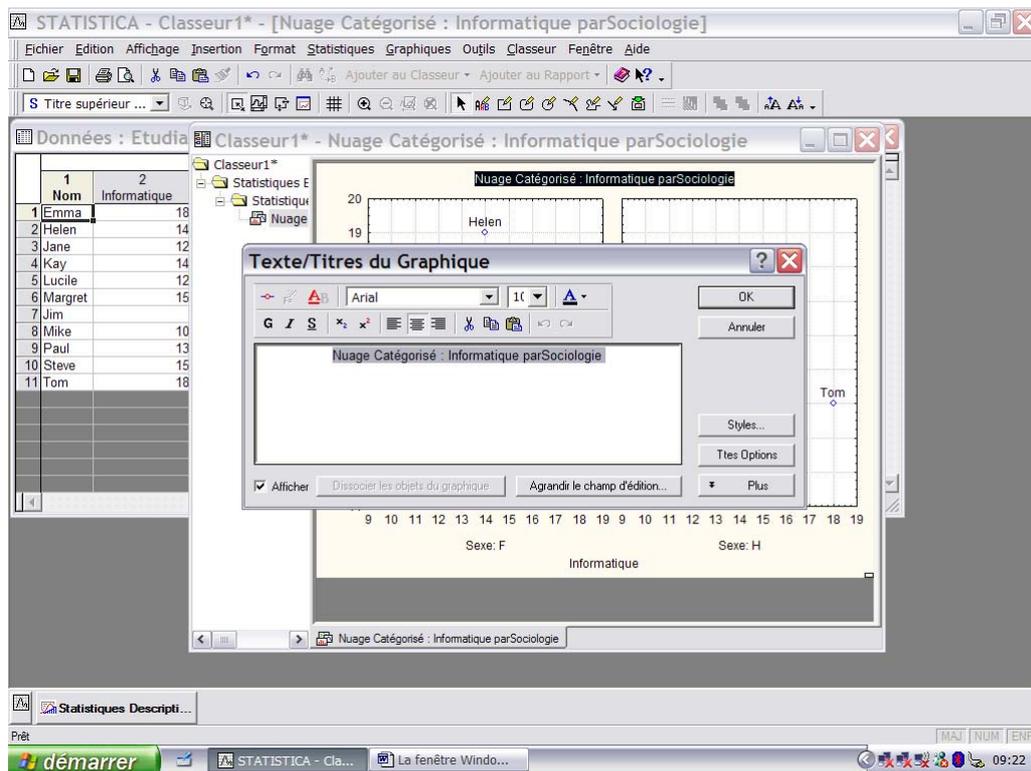
Il reste enfin à supprimer la légende présente en bas du graphique (« 95% de confiance ») par l'onglet « Texte/Titres du graphique » :



On supprime le titre dans la fenêtre centrale pour finalement reproduire la présentation recherchée :



n.b. : STATISTICA permet d'accéder à un onglet particulier sans être obligé de passer par la palette complète. Il suffit d'un double-clic sur l'élément du graphique à modifier pour afficher immédiatement les options de l'onglet correspondant. Par exemple, si on désire modifier le titre général « Nuage Catégorisé : Informatique par Sociologie », un double-clic sur cet élément provoque l'ouverture de l'onglet « Texte/Titres du Graphique » et rend accessibles les outils spécifiques :

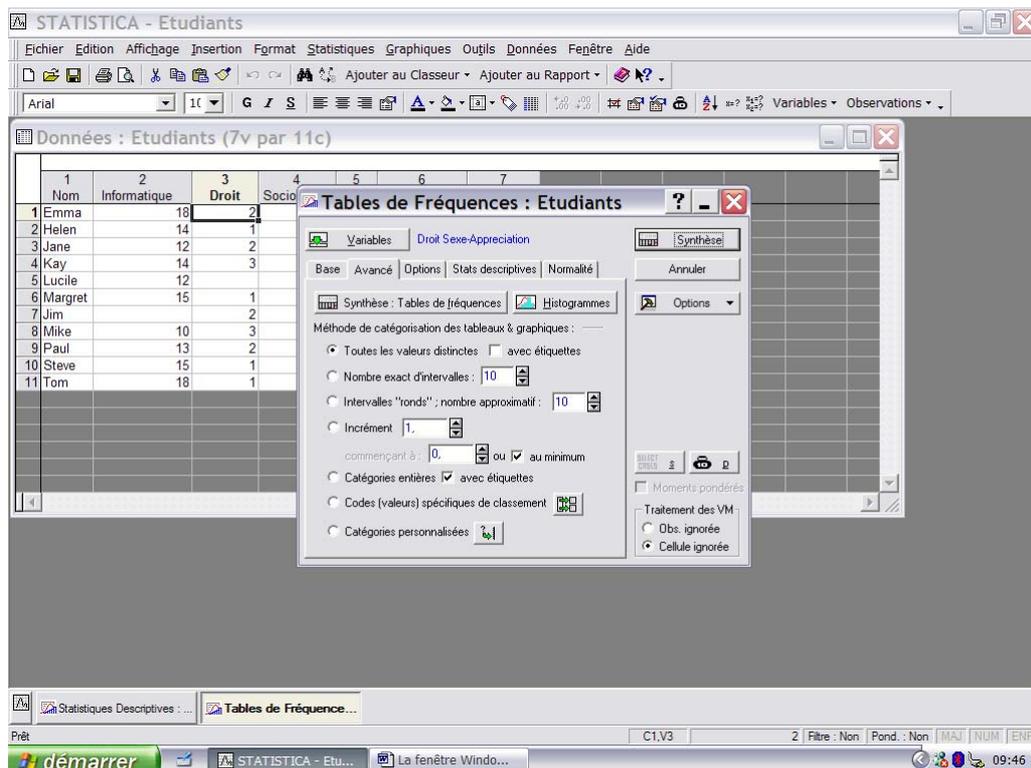


L'affichage de résultats sous une forme graphique est accessible par le menu « Graphiques » qui permet d'autres possibilités qui seront vues dans la suite. Dans le menu « Statistiques » et l'option « Statistiques élémentaires », deux outils retiennent encore l'attention, ce sont :

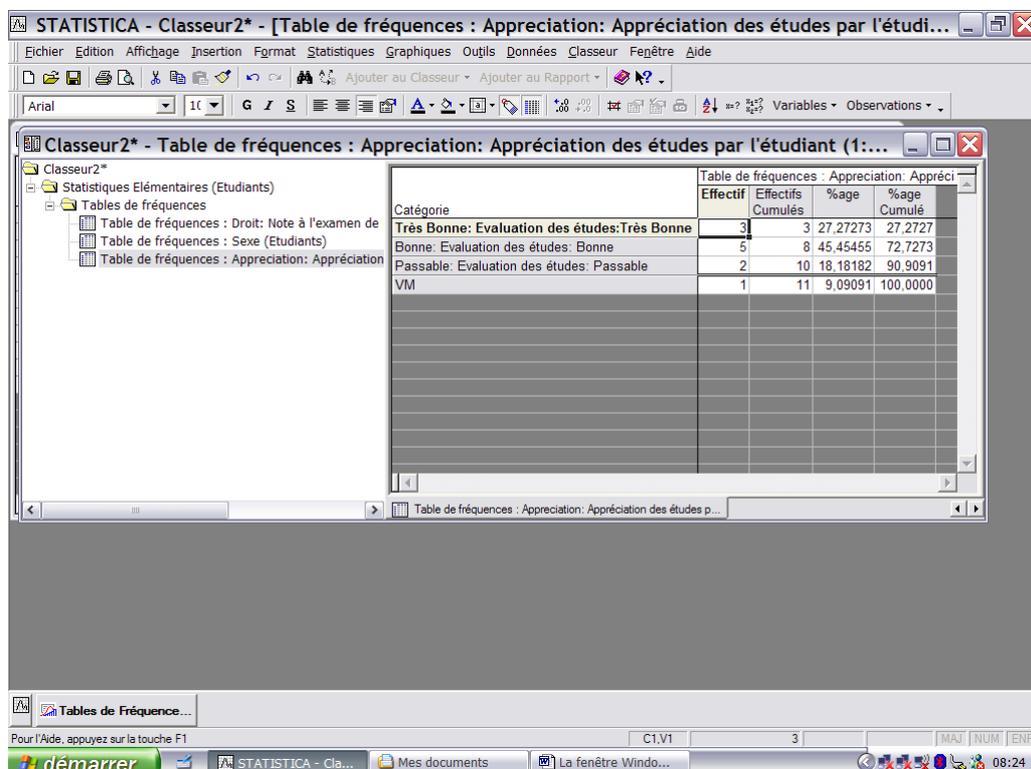
- Tables de fréquences
- Tableaux et Tris croisés

b. Tables de fréquences

L'outil « Tables de fréquences » permet le traitement des données pour des variables nominales et ordinales (dans notre exemple, les colonnes « Droit », « Sexe », « Appreciation ») :

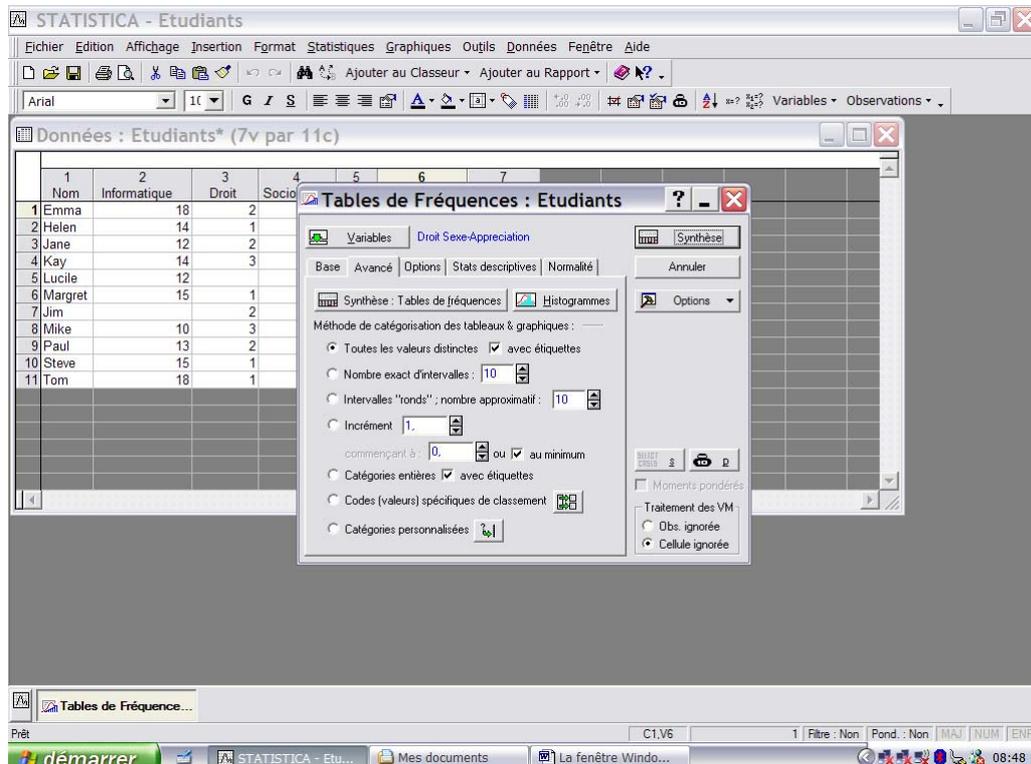
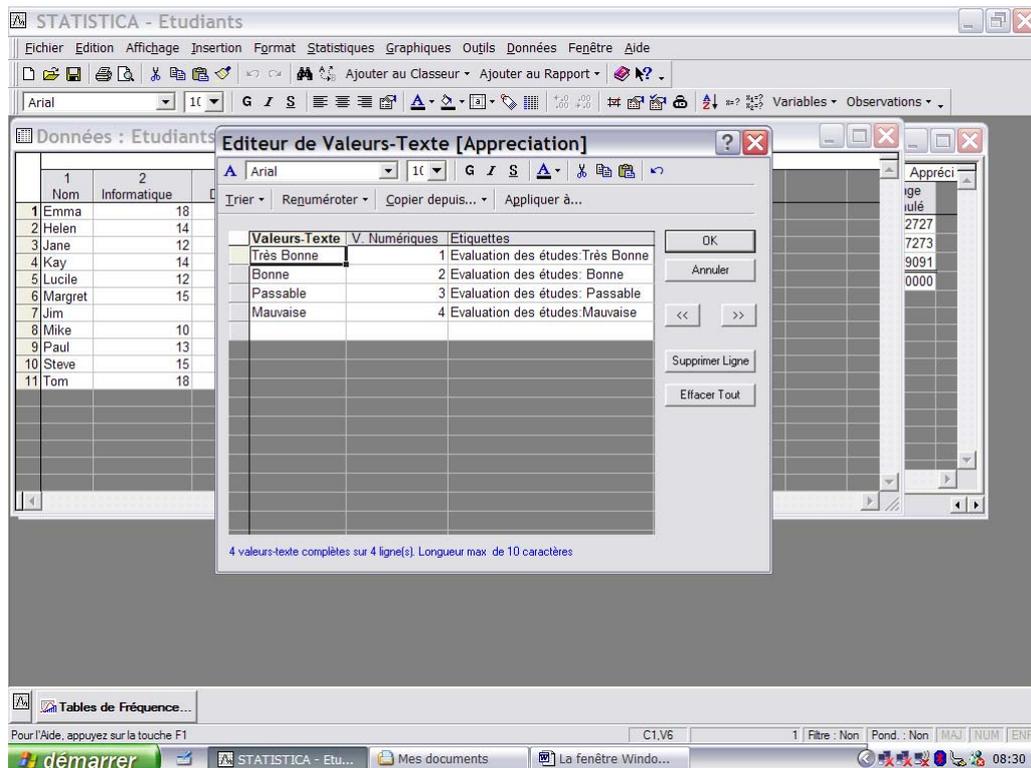


Le bouton « Synthèse : Tables de fréquences » édite les comptages des modalités de toutes variables retenues ; par défaut, les modalités sont les valeurs-texte et les étiquettes qui ont été encodées dans le menu « Données » « Editeur des valeurs-texte ».



Sans option particulière, les pourcentages de la table sont calculés par rapport au nombre total d'observations sans tenir compte du nombre de valeurs manquantes (VM).

Pour la variable « Appréciation », nous avons encodé les valeurs suivantes :



L'onglet « Avancé » permet de structurer les valeurs des variables analysées en modalités qui seront retenues dans la table des fréquences :

- « Toutes les valeurs distinctes » signifie qu'une valeur de la variable correspond à une modalité (pour une variable numérique décimale, p.ex. « Sociologie », les notes non entières sont des modalités)
- « Nombre d'intervalles » on indique pour les variables numériques le nombre exact d'intervalles
- « Intervalles ronds » : idem mais les limites des intervalles sont entières
- « Incrément » : on fixe la valeur de l'intervalle
- « Catégories entières » : on ne retient que les valeurs entières
- « Codes (valeurs) spécifiques de classement » : on fournit les modalités de la table qu'on désire retenir
- « Catégories personnalisées » : sur la base d'une expression, STATISTICA catégorise les valeurs d'une variable

Exemples :

1. Variable « Sociologie » - option « Toutes les valeurs distinctes »

The screenshot shows a window titled 'Classeur11* - Table de fréquences : Sociologie: N...'. The main table displays the following data:

Catégorie	Effectif	Effectifs Cumulés	%age	%age Cumulé
11,500000000	1	1	9,09091	9,0909
12	1	2	9,09091	18,1818
13,500000000	1	3	9,09091	27,2727
14	2	5	18,18182	45,4545
14,500000000	1	6	9,09091	54,5455
15	1	7	9,09091	63,6364
16	1	8	9,09091	72,7273
17,500000000	1	9	9,09091	81,8182
19	1	10	9,09091	90,9091
VM	1	11	9,09091	100,0000

2. Variable « Sociologie » - option « Nombre exact d'intervalles : 5 »

The screenshot shows a window titled 'Classeur11* - Table de fréquences : Sociologie: N...'. The main table displays the following data:

De... à...	Effectif	Effectifs Cumulés	%age	%age Cumulé
10,56250<x<=12,43750	2	2	18,18182	18,1818
12,43750<x<=14,31250	3	5	27,27273	45,4545
14,31250<x<=16,18750	3	8	27,27273	72,7273
16,18750<x<=18,06250	1	9	9,09091	81,8182
18,06250<x<=19,93750	1	10	9,09091	90,9091
VM	1	11	9,09091	100,0000

3. Variable « Sociologie » - option « Intervalles ronds – Nombre approximatif :5 »

De...	à...	Effectif	Effectifs Cumulés	%age	%age Cumulé
10,00000	<x<=12,00000	2	2	18,18182	18,1818
12,00000	<x<=14,00000	3	5	27,27273	45,4545
14,00000	<x<=16,00000	3	8	27,27273	72,7273
16,00000	<x<=18,00000	1	9	9,09091	81,8182
18,00000	<x<=20,00000	1	10	9,09091	90,9091
20,00000	<x<=22,00000	0	10	0,00000	90,9091
VM		1	11	9,09091	100,0000

4. Variable « Sociologie » - option « Incrément de 2 »

De...	à...	Effectif	Effectifs Cumulés	%age	%age Cumulé
11,50000	<x<=13,50000	2	2	18,18182	18,1818
13,50000	<x<=15,50000	5	7	45,45455	63,6364
15,50000	<x<=17,50000	1	8	9,09091	72,7273
17,50000	<x<=19,50000	2	10	18,18182	90,9091
19,50000	<x<=21,50000	0	10	0,00000	90,9091
VM		1	11	9,09091	100,0000

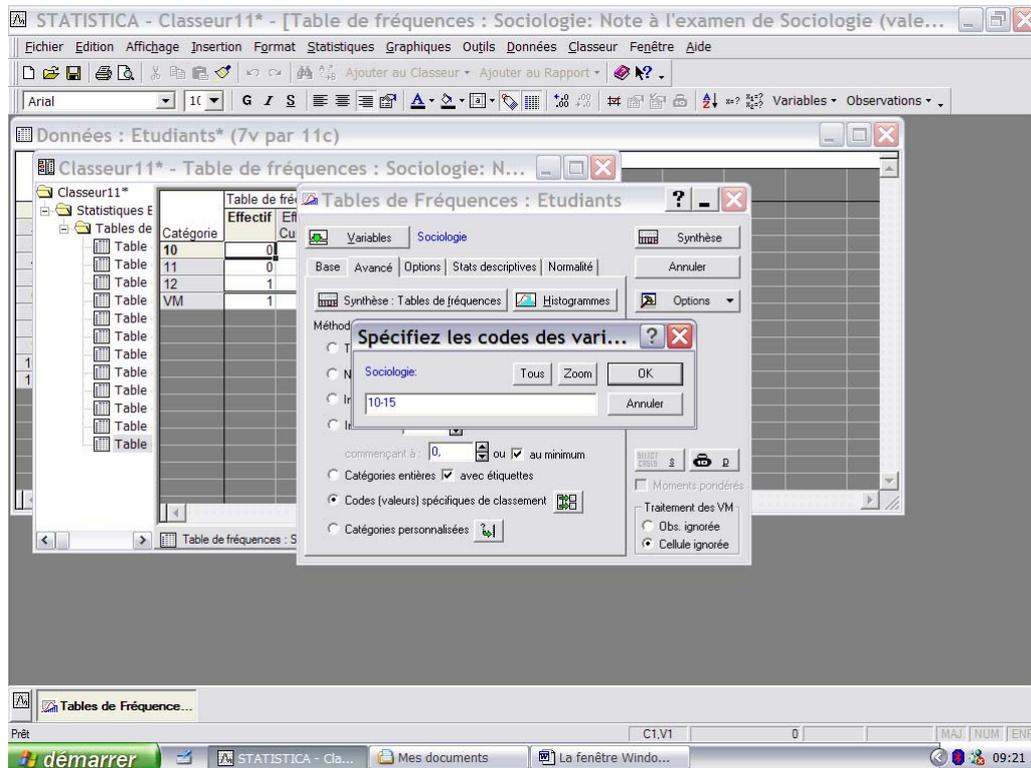
5. Variable « Sociologie » - option « Catégories entières »

Catégorie	Effectif	Effectifs Cumulés	%age	%age Cumulé
12	1	1	14,28571	14,2857
14	2	3	28,57143	42,8571
15	1	4	14,28571	57,1429
16	1	5	14,28571	71,4286
19	1	6	14,28571	85,7143
VM	1	7	14,28571	100,0000

6. Variable « Sociologie » - option « Codes spécifiques de classement »

Le choix de l'option affiche une fenêtre dans laquelle on fournit à STATISTICA les codes qu'on désire retenir ; ceux-ci peuvent être encodés individuellement séparés par un espace ou

sous la forme d'un intervalle de valeurs (dans notre exemple, « 10-15 » signifie toutes les notes comprises entre 10 et 15) :

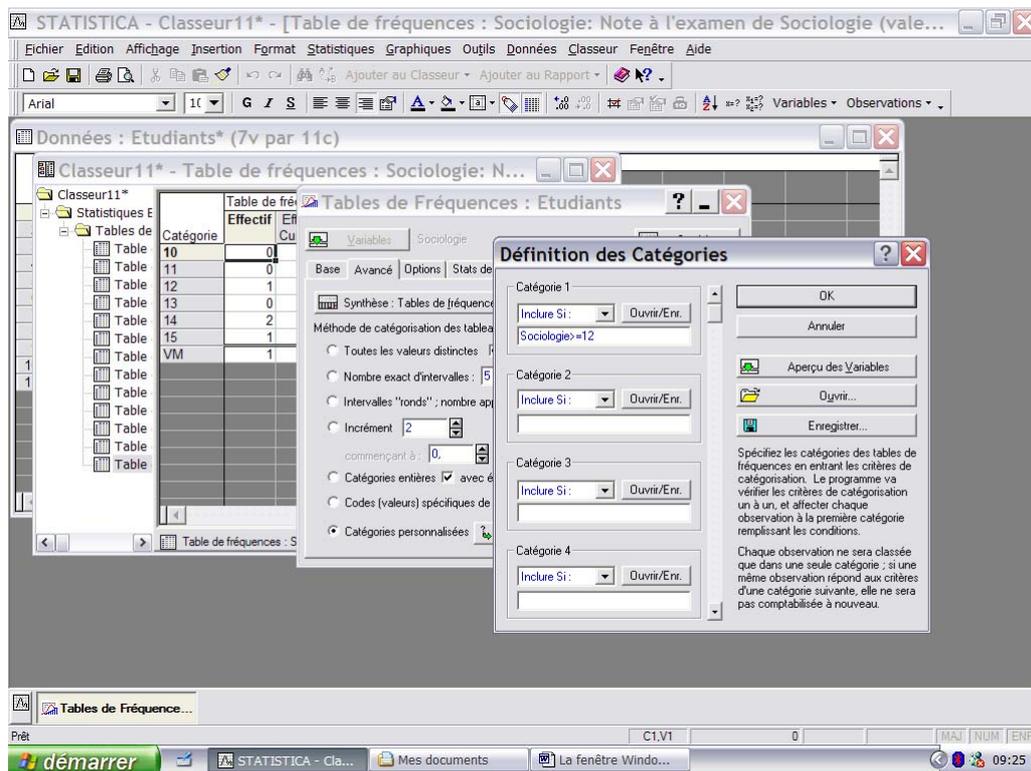


Voici le résultat :

Catégorie	Effectif	Effectifs Cumulés	%age	%age Cumulé
10	0	0	0.00000	0.0000
11	0	0	0.00000	0.0000
12	1	1	20.00000	20.0000
13	0	1	0.00000	20.0000
14	2	3	40.00000	60.0000
15	1	4	20.00000	80.0000
VM	1	5	20.00000	100.0000

7. Variable « Sociologie » - option « Catégories personnalisées »

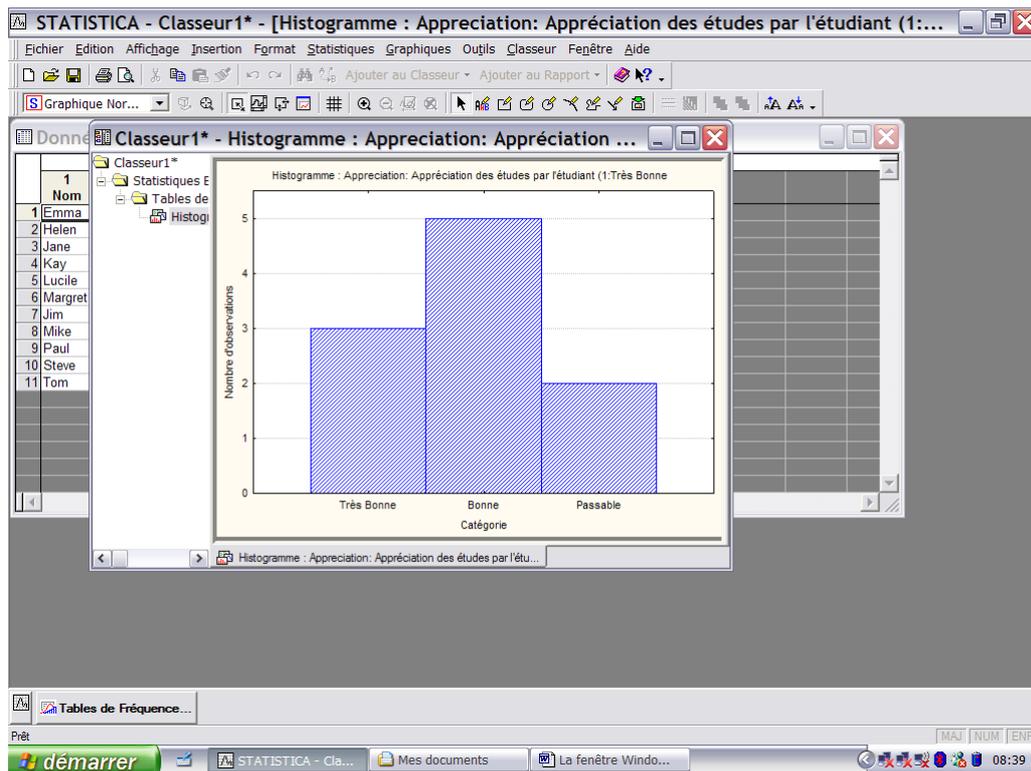
Cette option permet d'affecter les valeurs d'une variable à des catégories sur la base de l'évaluation d'une expression ; il faut fournir à STATISTICA cette expression. Nous y reviendrons plus longuement dans la suite :



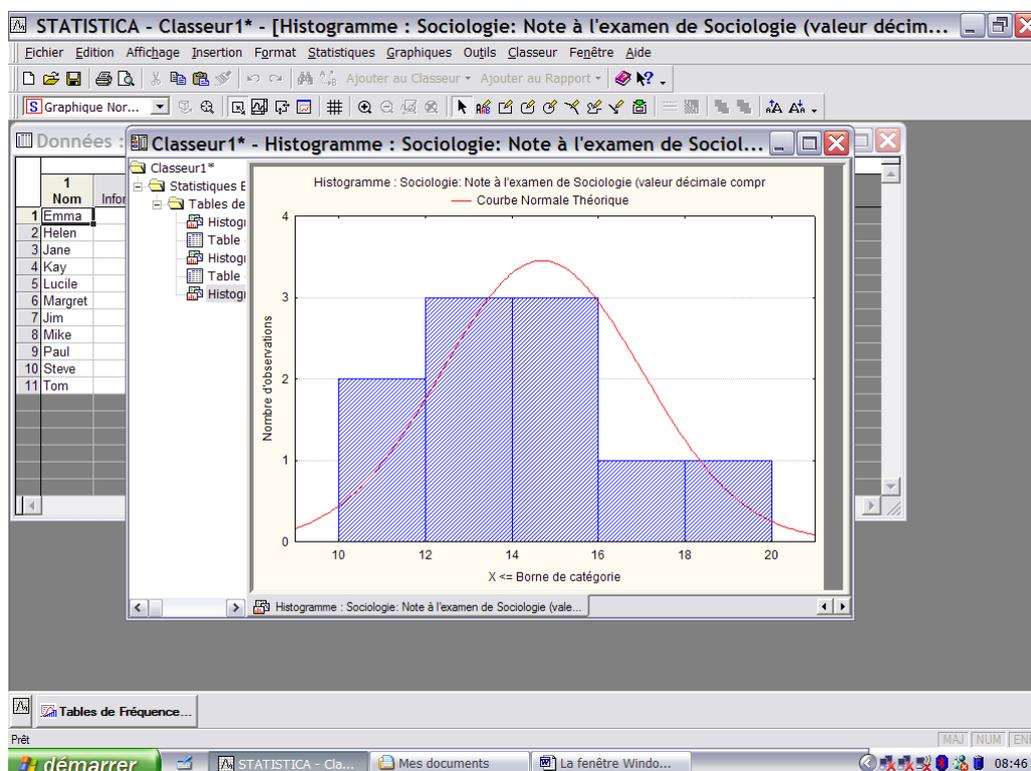
Dans l'écran ci-dessus, on définit une « Catégorie 1 » pour toutes les notes de Sociologie supérieures ou égales à la valeur 12, voici la table des fréquences :

Table de fréquences	Table de Fréquences po			
Catégorie	Effectif	Effectifs Cumulés	%age	%age Cumulé
Inclure Sociologie >= 12	9	9	81,81818	81,8182
Non sélect.	2	11	18,18182	100,0000

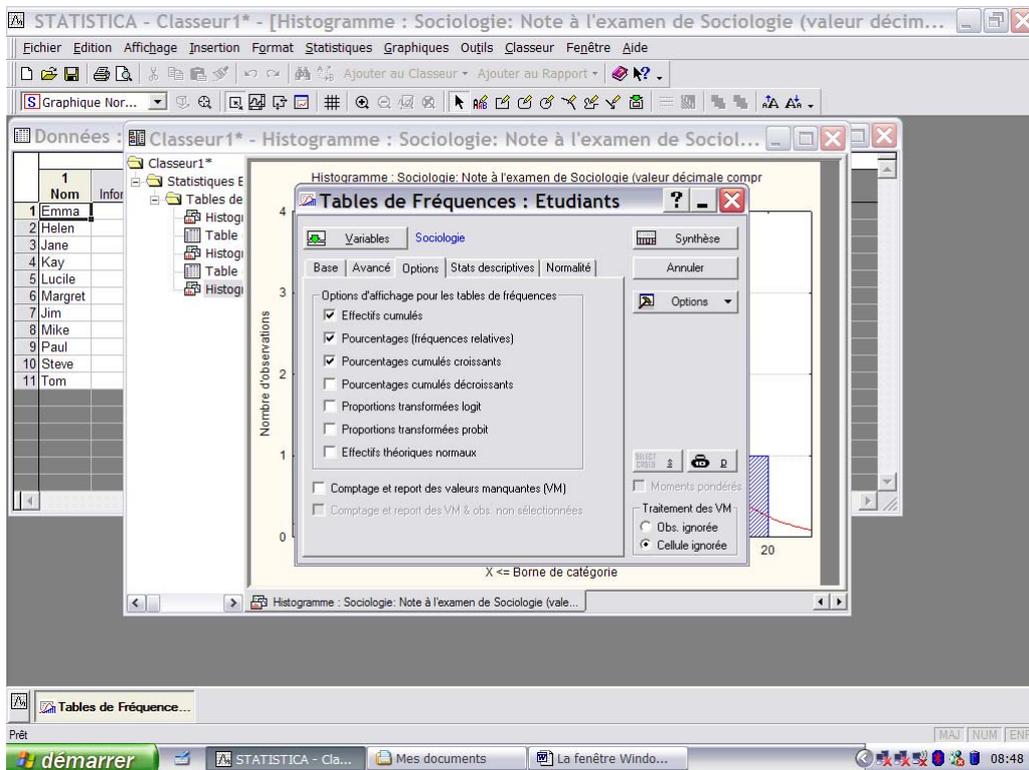
L'onglet «Avancé » offre enfin la possibilité d'afficher graphiquement les résultats du tableau des fréquences en choisissant le bouton «Histogrammes »; dans le cas de variables nominales et ordinales, on parlera plutôt de diagrammes en barres. Pour la variable « Appreciation », voici le diagramme :



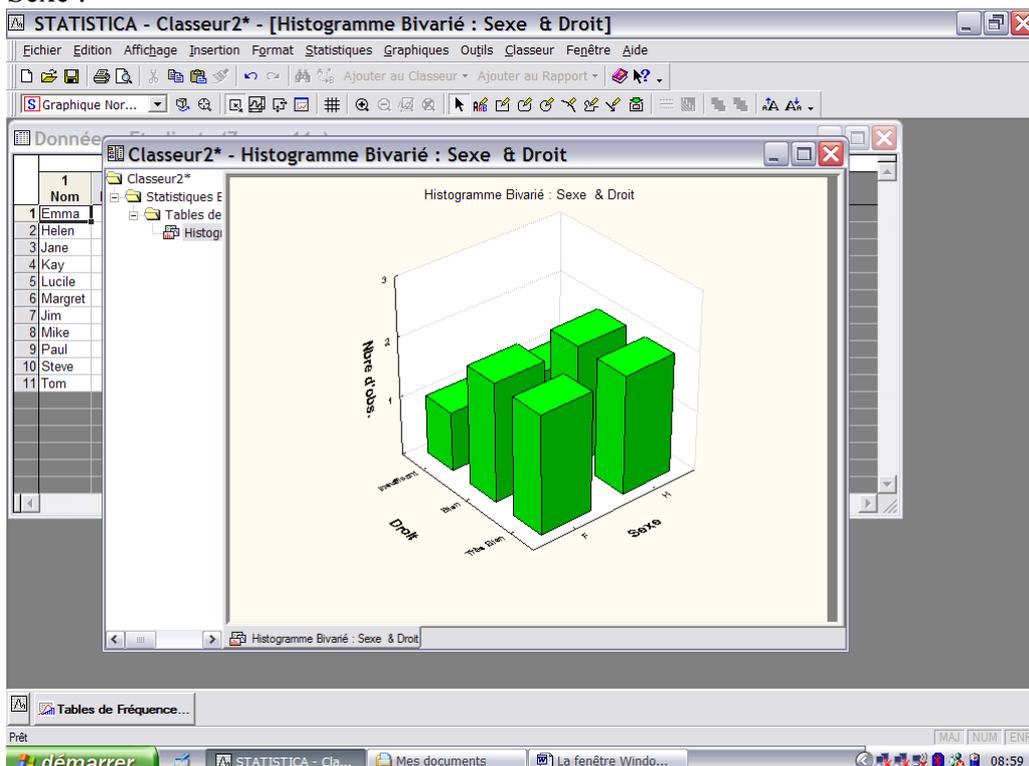
La représentation d'une variable numérique par un histogramme se fera toujours sur base d'intervalles de valeurs placés sur l'axe horizontal. Si on désire par exemple l'histogramme des notes à l'examen de « Sociologie », on commencera par spécifier l'option « Intervalles ronds – Nombre approximatifs : 8 » pour obtenir :



L'onglet « Options » permet de préciser les colonnes du tableau des fréquences ; dans cet onglet, on trouve aussi la possibilité d'éviter de tenir compte des valeurs manquantes dans les calculs de pourcentages :

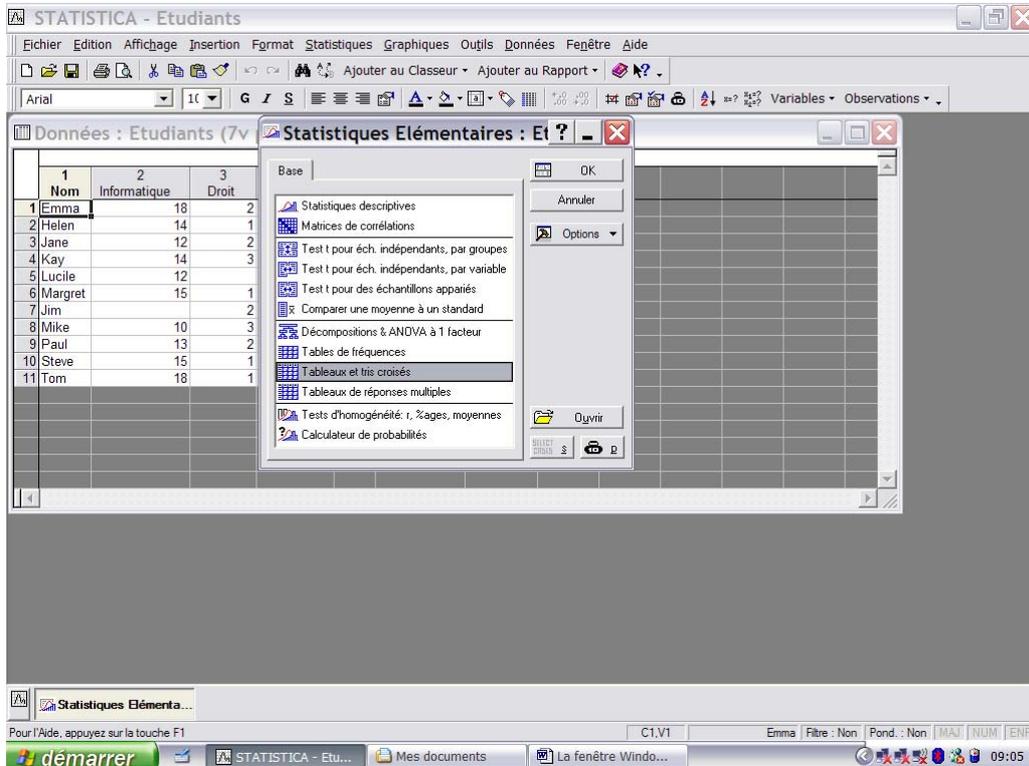


L'onglet « Stats Descriptives » permet, outre le calcul des indices classiques, une représentation d'un histogramme (ou d'un diagramme en barres) de deux variables Simultanées. Ci-après, le diagramme en barres des notes à l'examen de Droit en fonction du Sexe :



c. Tableaux à plusieurs entrées

On poursuit l'option « Statistiques élémentaires » du menu « Statistiques » par l'analyse des tableaux croisés :

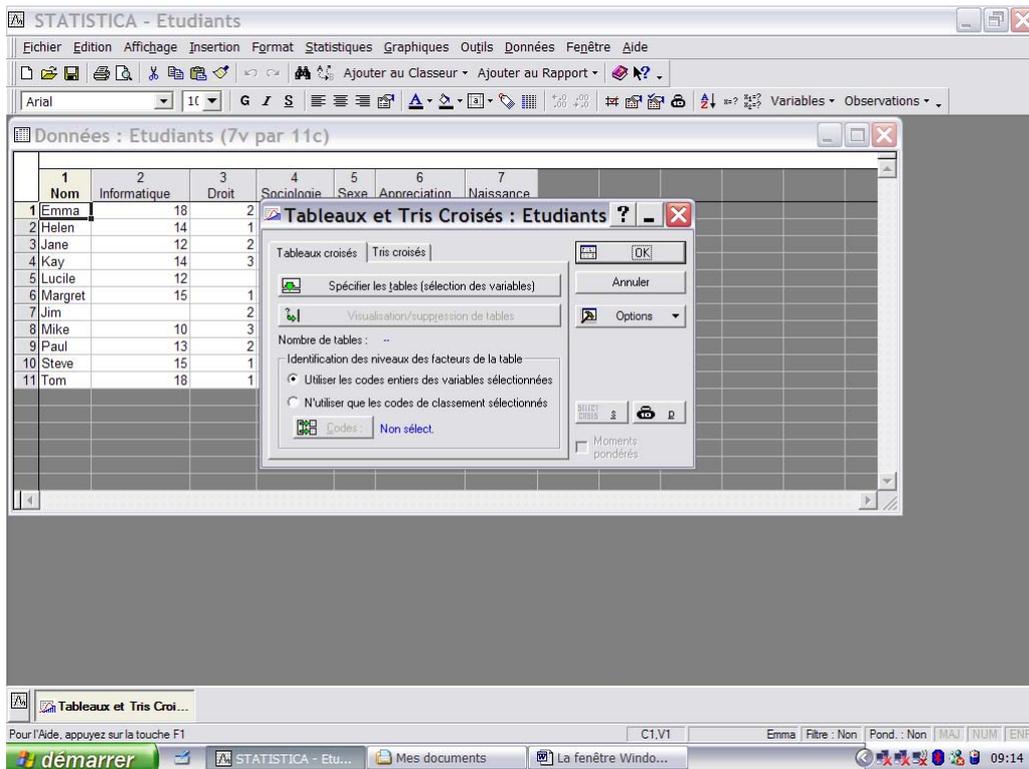


Ce type d'outil permet d'analyser les valeurs de deux variables, souvent des variables nominales et ordinales, mais aussi des variables numériques qui ont été recatégorisées sur des échelles nominales (cas le plus fréquent) pour découvrir si elles sont liées ; si deux variables apparaissent comme étant liées, leurs valeurs évoluent ensemble (mais pas nécessairement dans le même sens).

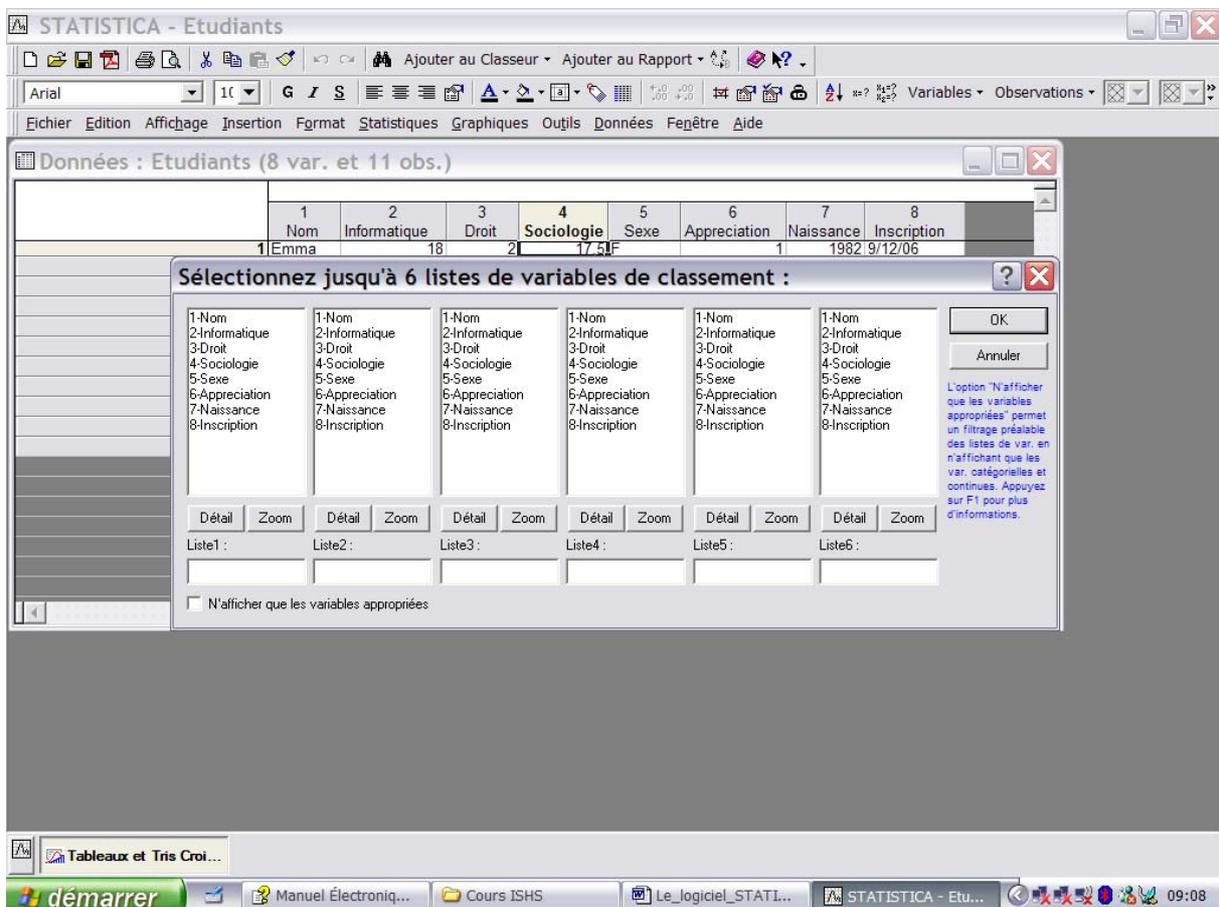
L'analyse des tableaux croisés se réalise en deux étapes :

- une première étape consiste à spécifier les tables (donc les variables) qu'on désire analyser
- une deuxième étape est l'analyse proprement dite

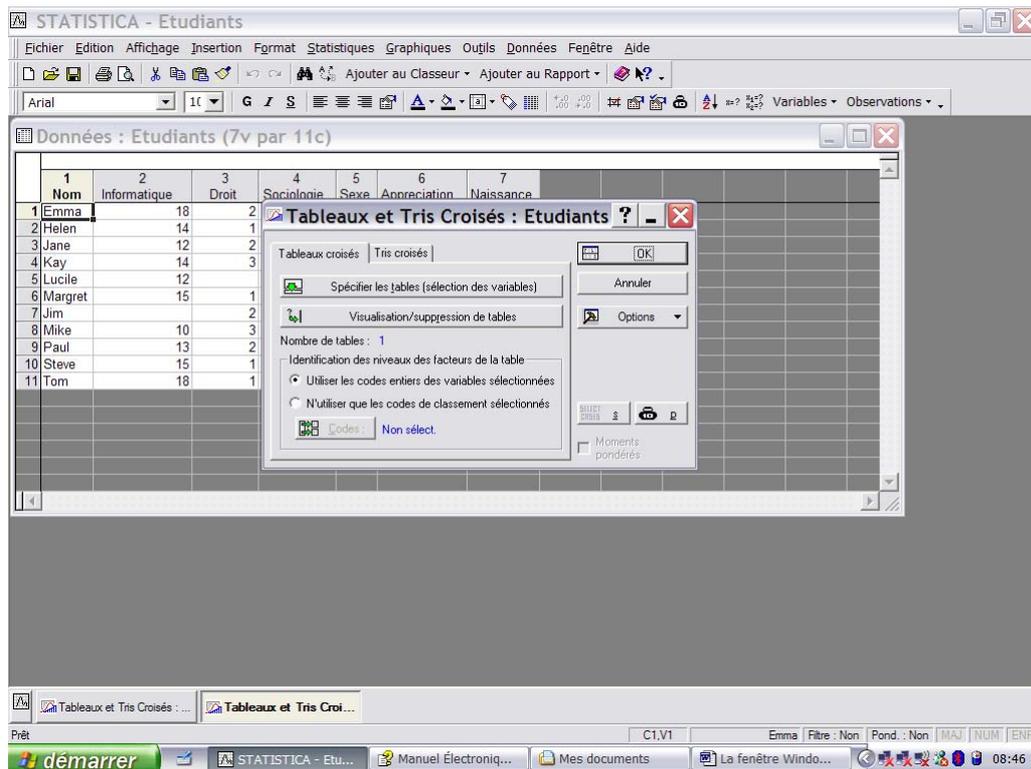
Première étape : Spécification des tables



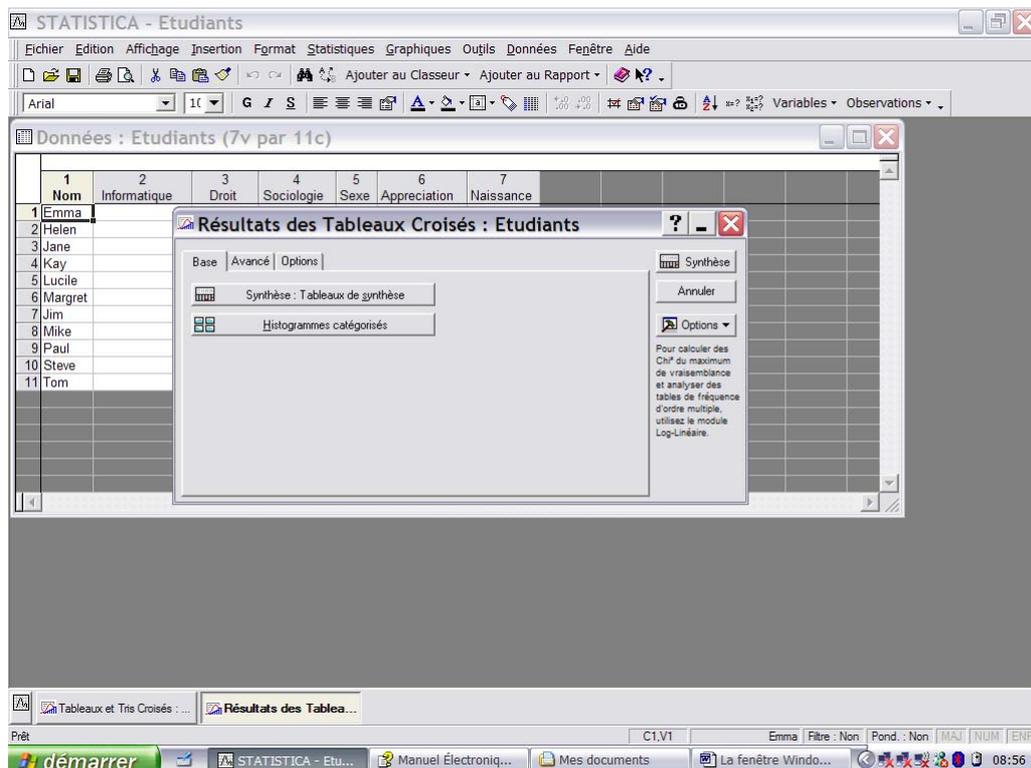
Le bouton « Spécifier les tables (sélection des variables) » permet de choisir les variables de la feuille de données qui seront croisées :



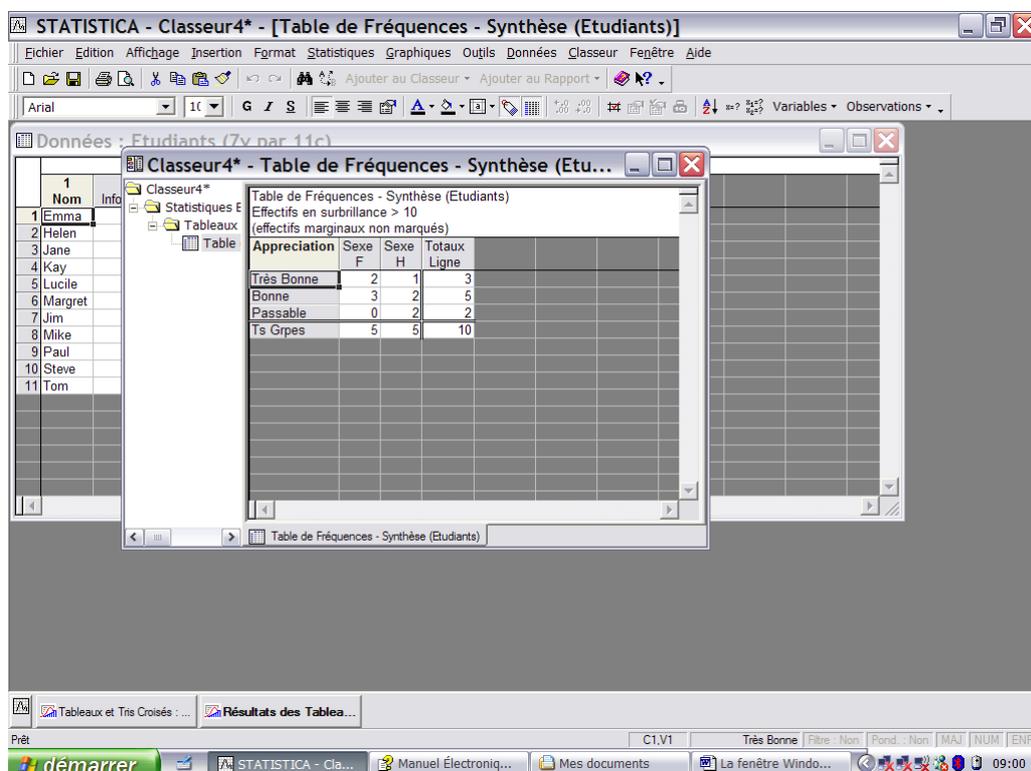
STATISTICA permet de construire des tables à partir de six variables de classement ; nous nous limiterons au cas (le plus fréquent) de deux variables. Le bouton « Détail » affiche la description détaillée des variables ; « Zoom » ouvre la fenêtre « Valeurs Stats » (cf. infra) qui donne une description détaillée de la variable et des valeurs. Dans le cas ci-dessus, nous désirons analyser l'évaluation des études faites par les étudiants en fonction de leur sexe ; on construit la table pour les variables « Appreciation » (variable 6) et « Sexe » (variable 5). Dès que la sélection est réalisée, l'écran de départ réapparaît en indiquant le nombre de tables sur lesquelles l'analyse portera et propose la création d'une nouvelle table (« Spécifier les tables ») ; on peut visualiser les tables déjà construites et éventuellement les supprimer (« Visualisation/Suppression de tables »). Par défaut, toutes les modalités des variables seront utilisées dans les croisements :



Dès que la dernière table a été spécifiée et validée (« Ok »), l'étape d'analyse (« Résultats des Tableaux croisés ») en débutant par l'onglet « Base » :

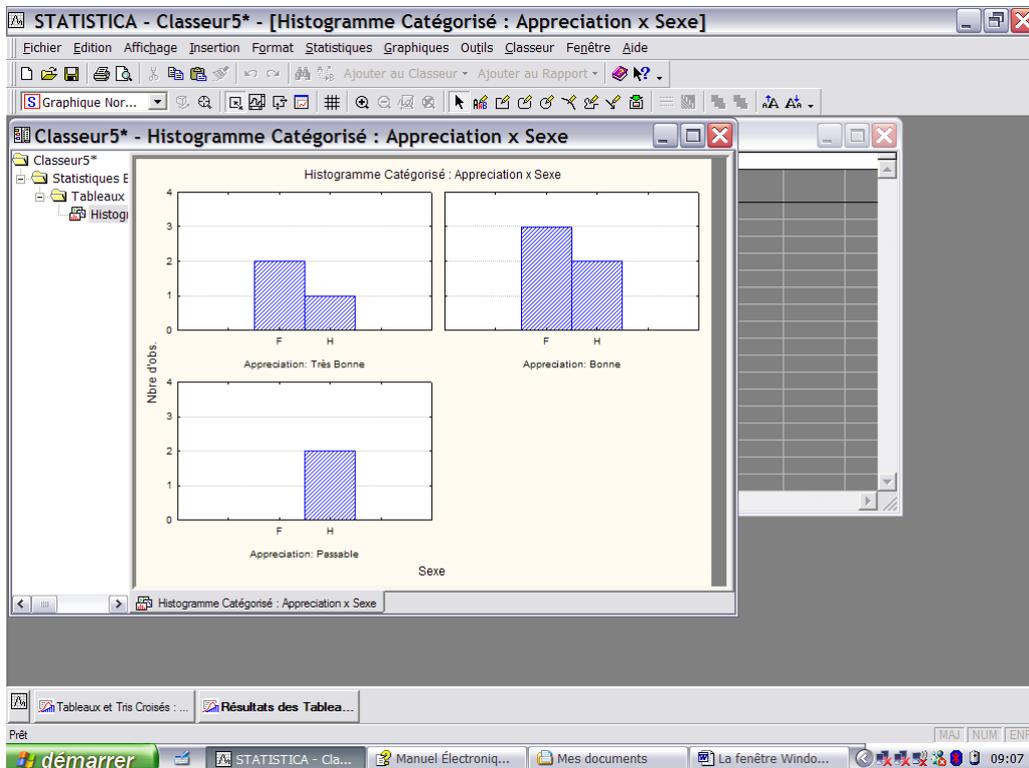


Les boutons « Synthèse : Tableaux de synthèse » et « Synthèse » afficheront la feuille des résultats :

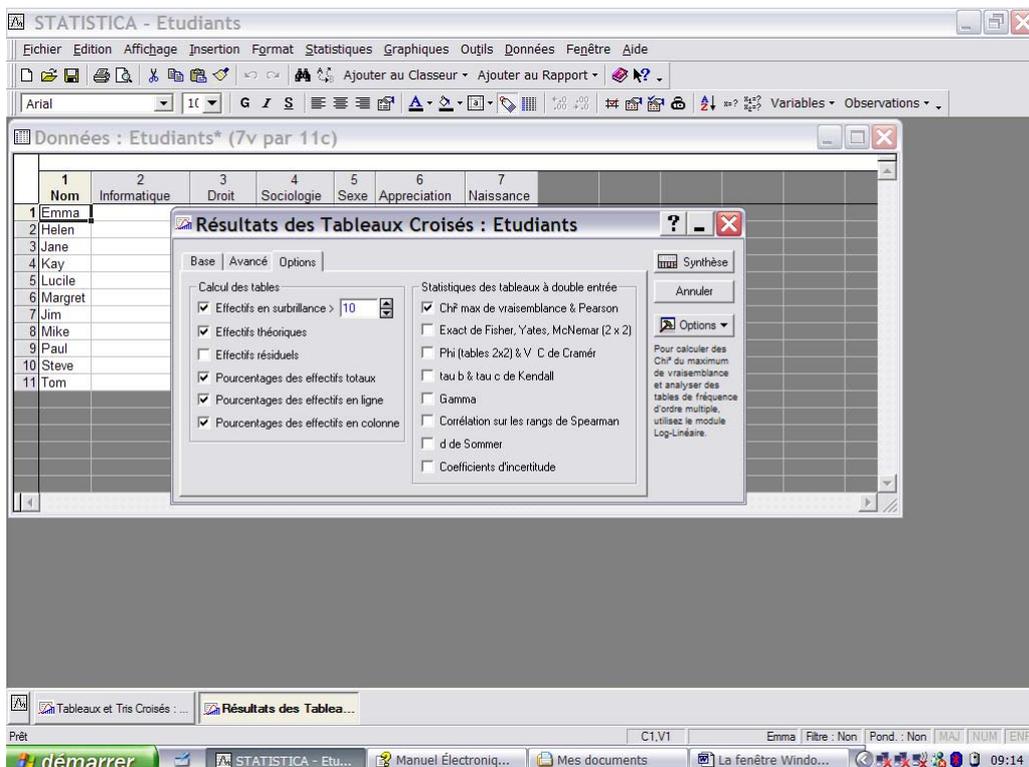


n.b : La variable « Appreciation » dont les différentes modalités apparaissent en ligne est la première des variables spécifiées dans la construction de la table.
Par défaut, les modalités de la table correspondent aux valeurs-texte qu'on a encodées.

« Histogrammes catégorisés » permet d'obtenir un diagramme en bâton des modalités de la seconde variable (« Sexe ») spécifiée pour chaque catégorie des valeurs de la première (« Appreciation ») :



L'onglet « Options » permet de préciser les informations que les tables de fréquences demandées devront fournir :

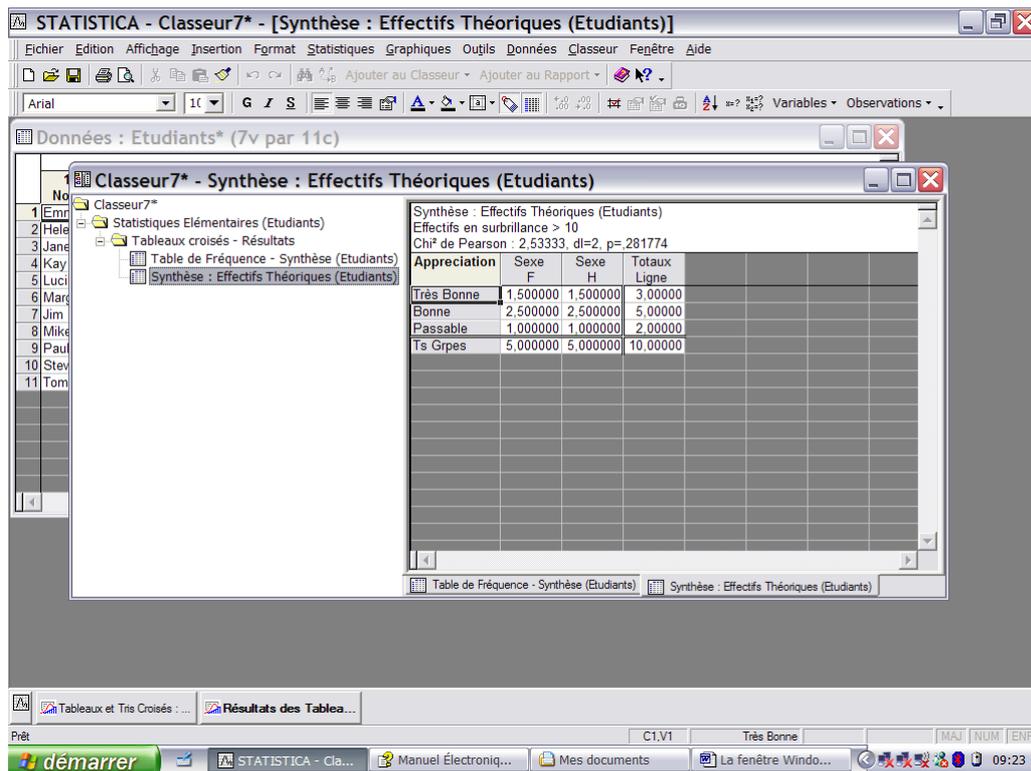


Dans l'exemple, on demande que les fréquences des tables soient exprimées en termes de différents pourcentages et qu'un test de chi-carré soit réalisé pour tester l'hypothèse d'indépendance ; on trouvera ci-dessous les résultats de ces options (l'arborescence du classeur indique que les résultats portent sur deux tables :

The screenshot shows the STATISTICA software interface. The main window displays a contingency table titled "Table de Fréquence - Synthèse (Etudiants)". The table is structured as follows:

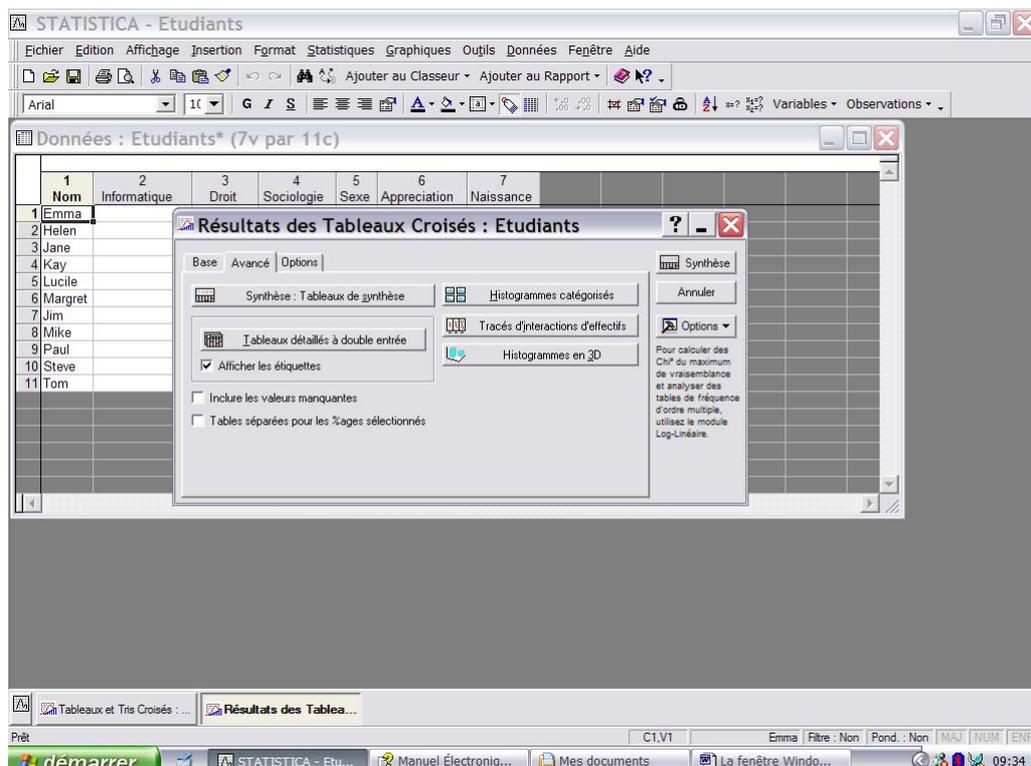
		Sexe		Totaux Ligne
		F	H	
Appreciation	Très Bonne	2	1	3
	%age Colonne	40,00%	20,00%	
	%age Ligne	66,67%	33,33%	
	%age Total	20,00%	10,00%	30,00%
Bonne	Effectif	3	2	5
	%age Colonne	60,00%	40,00%	
	%age Ligne	60,00%	40,00%	
	%age Total	30,00%	20,00%	50,00%
Passable	Effectif	0	2	2
	%age Colonne	0,00%	40,00%	
	%age Ligne	0,00%	100,00%	
	%age Total	0,00%	20,00%	20,00%
Ts Grpes	Effectif	5	5	10
	%age Colonne	50,00%	50,00%	
	%age Ligne			
	%age Total			

La première table des fréquences correspond aux comptages croisés des modalités exprimés suivant les précisions apportées dans l'onglet « Option », la deuxième table donne les comptages (« Effectifs Théoriques ») qui auraient dû être observés en cas d'indépendance des deux variables, de même que le chi-carré et le niveau de confiance en fonction des degrés de liberté :



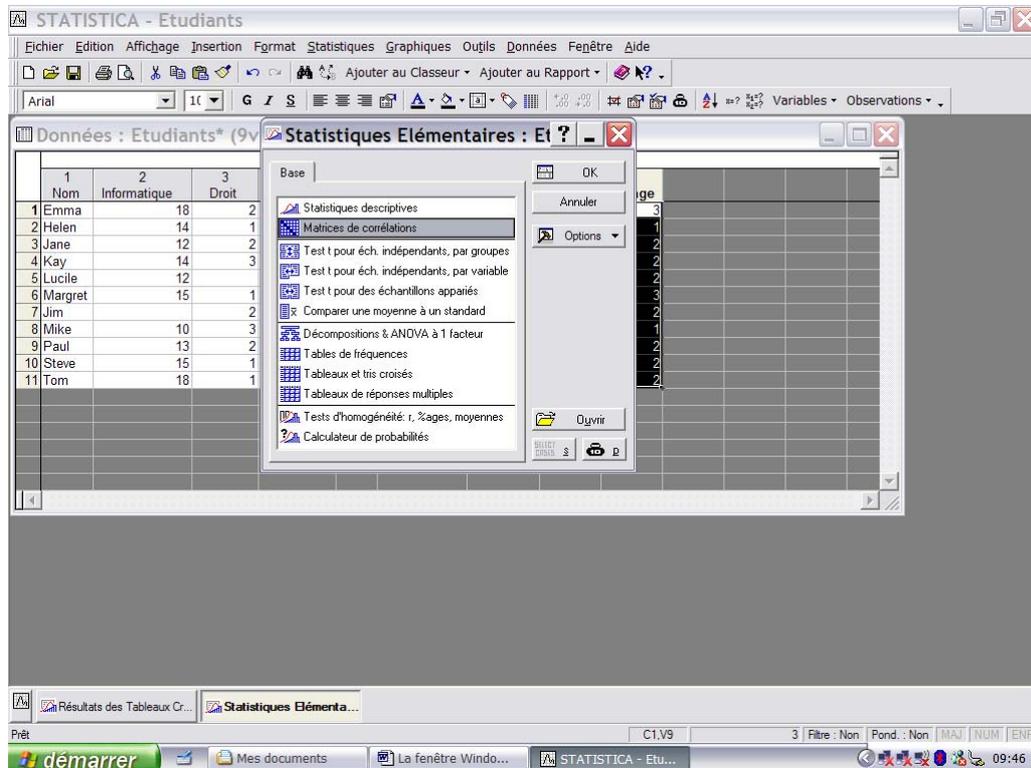
n.b. : D'autres paramètres (coefficients d'association) sont accessibles dans l'onglet « Options ».

Dans l'onglet « Avancé », on aura la possibilité d'afficher dans une feuille de résultats spécifiques tous les paramètres demandés au niveau d'« Options » (« Tableaux détaillés à double entrée »), les histogrammes catégorisés (cf. supra) et les histogrammes à 3 dimensions (idem) :

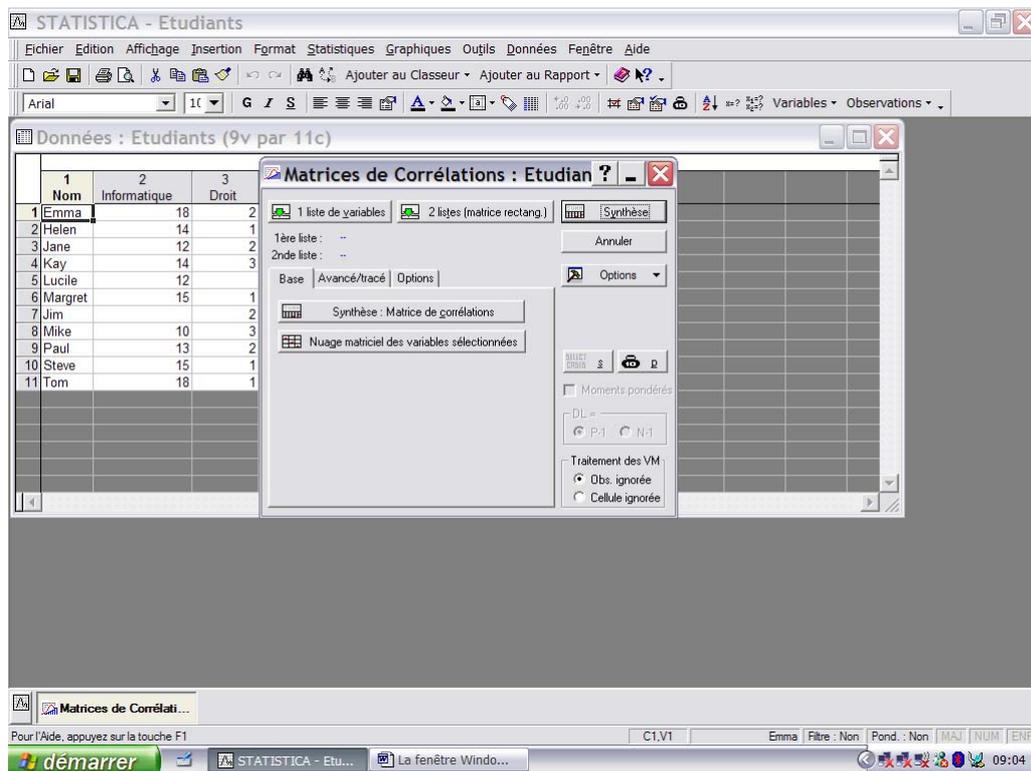


d. Corrélation

On termine l'option « Statistiques élémentaires » du menu « Statistiques » par « Matrice de Corrélations » :

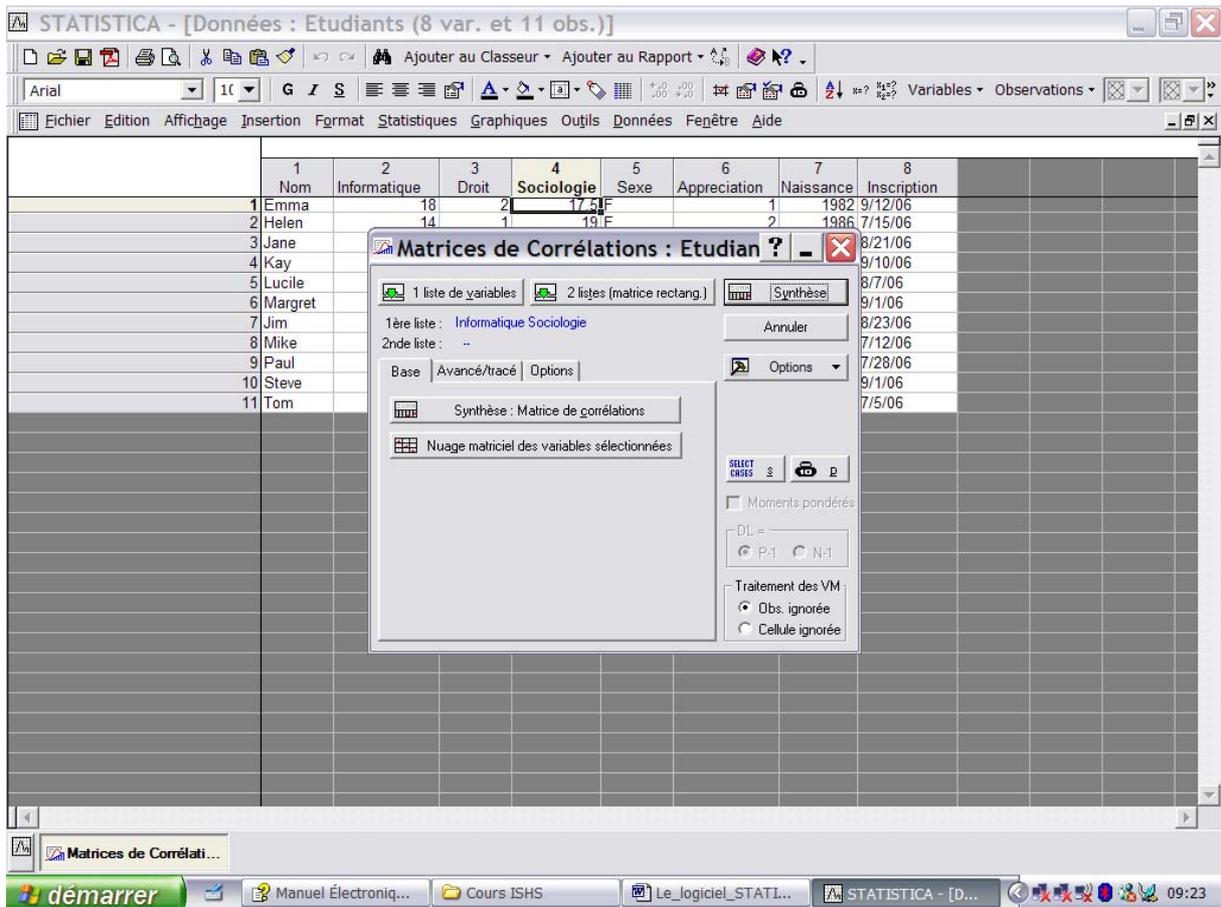


Le calcul de la corrélation linéaire entre deux variables de ratio est celui de Pearson ; les résultats sont donnés dans un tableau appelé « matrice des corrélations ». Il est composé de lignes et de colonnes qui correspondent aux noms des variables pour lesquelles la coefficient de corrélation a été demandé, pour cela, on remplit l'écran suivant :

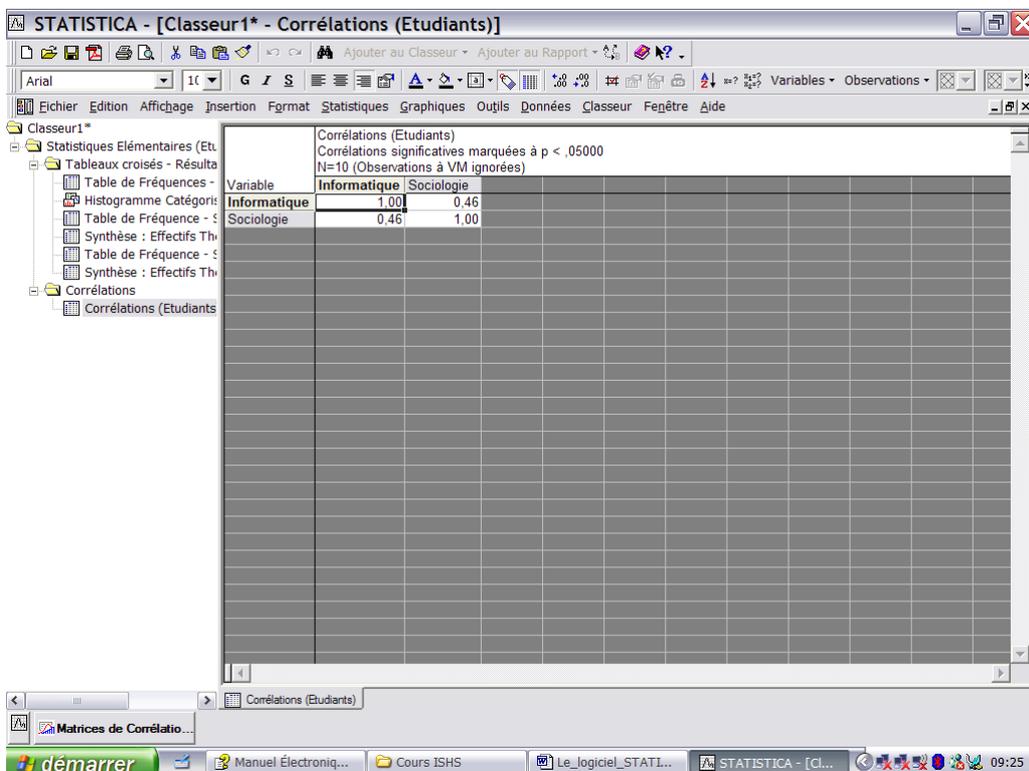


Le bouton « 1 liste de variables » permet de fournir la liste des variables d'une feuille de données entre lesquelles on veut calculer la corrélation 2 à 2.

Par exemple, si on calcule la corrélation entre « Informatique » et « Sociologie », en choisissant cette option :



Ce qui donne le résultat :

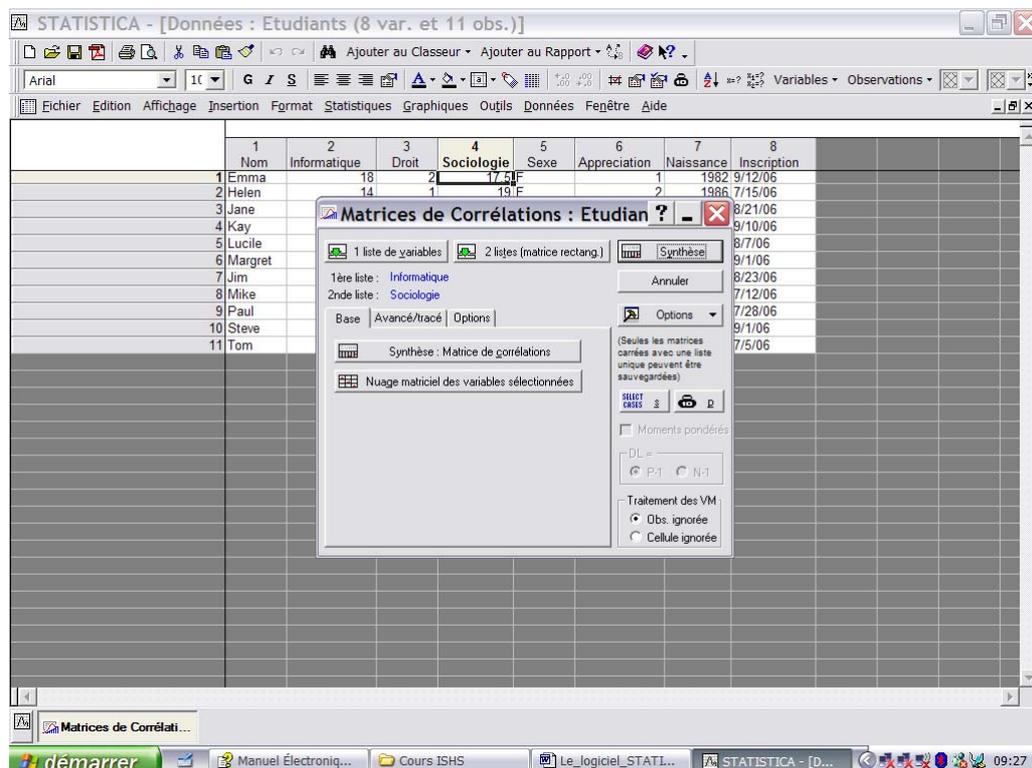


La corrélation d'une variable avec elle-même vaut évidemment 1 (diagonale de la matrice) ; le tableau est symétrique puisque la corrélation entre une variable X et une variable Y est égale à la corrélation entre Y et X.

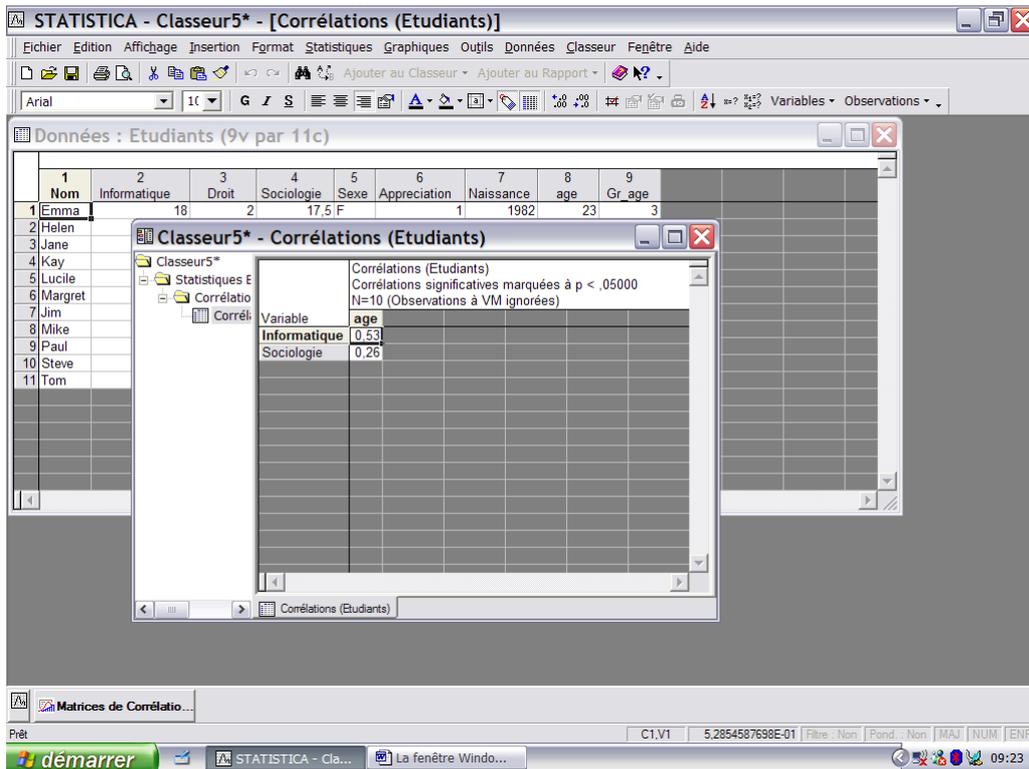
La corrélation entre les notes obtenues à l'examen d'informatique et l'âge des étudiants est positive et vaut 0,53 ; elle est supérieure à celle entre « Sociologie » et « age ».

Un test de vérifie que la corrélation trouvée dans un échantillon est significative pour la population.

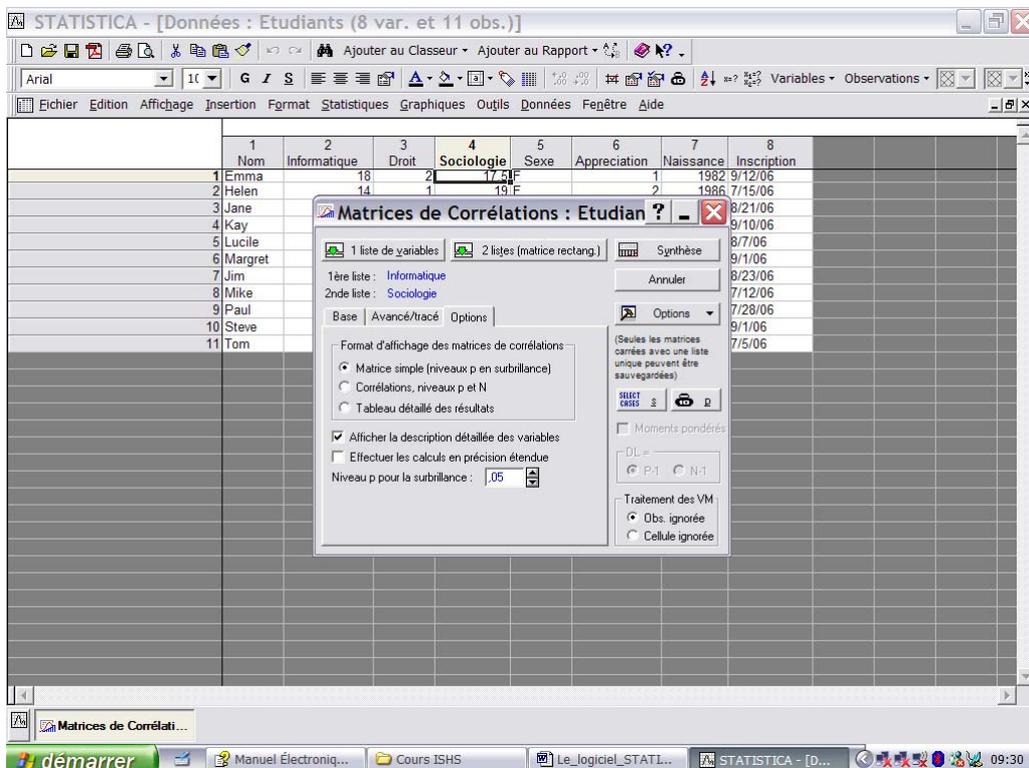
L'option « 2 listes » permet de calculer les coefficients de corrélation uniquement entre les variables fournies dans les listes et prises 2 à 2 ; dans ce cas, on évite d'afficher les corrélations unitaires (diagonale) et le valeurs symétriques :



Produit les résultats :



L'onglet « Options » permet d'obtenir tous les détails relatifs à la corrélation (test d'hypothèse compris) de même que les ajustements obtenus par la régression simple entre les variables de(s) la (les) liste(s) :



Donnera les résultats suivants :

The screenshot shows the STATISTICA software interface with a window titled 'Corrélations (Etudiants)'. The window displays a table of correlation results for two variables: 'Informatique: Note à l'examen d'Informatique (Valeur e)' and 'Sociologie: Note à l'examen de Sociologie (valeur dé)'. The table includes columns for 'Ec-Type', 'r(X,Y)', 'r²', 't', 'p', 'N', 'Constnte dép : Y', 'Pente dép : Y', 'Constnte dép : X', and 'Pente dép : X'. The correlation coefficient r(X,Y) is 0,456537, r² is 0,208426, t is 1,451361, and p is 0,184736.

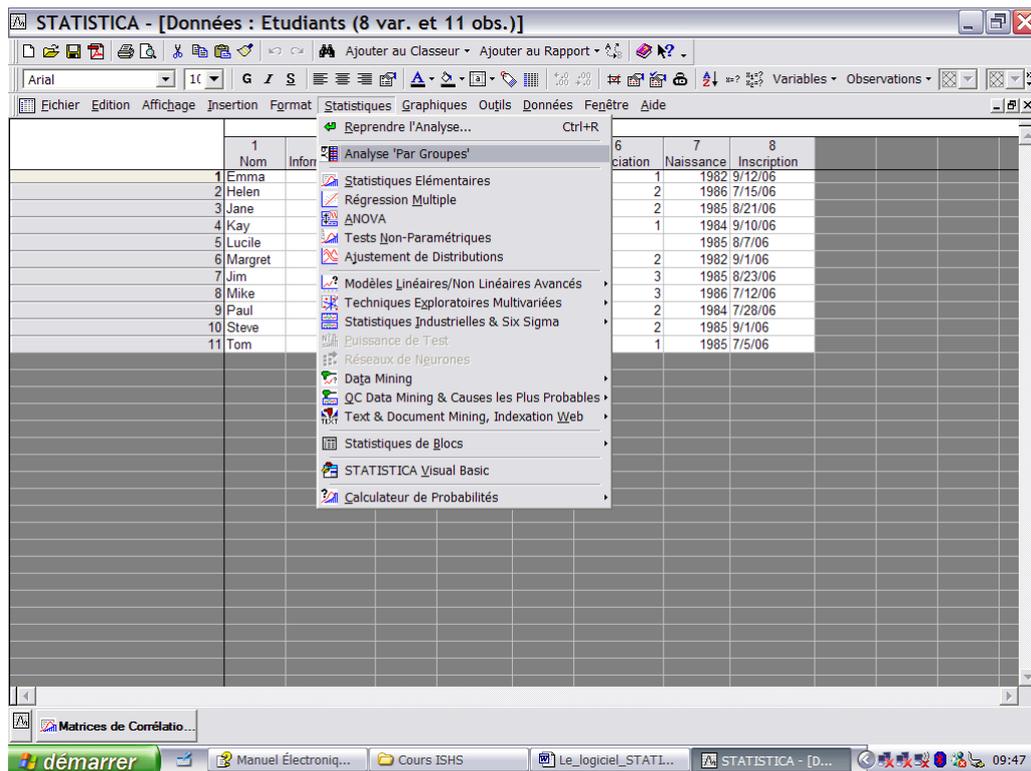
Var. X et Var. Y	Ec-Type	r(X,Y)	r ²	t	p	N	Constnte dép : Y	Pente dép : Y	Constnte dép : X	Pente dép : X
Informatique: Note à l'examen d'Informatique (Valeur e)	2,558211									
Sociologie: Note à l'examen de Sociologie (valeur dé)	2,311805	0,456537	0,208426	1,451361	0,184736	10	8,882852	0,412564	6,673597	0,505198

Les coefficients de corrélation se trouvent dans la colonne « r(X,Y) » : 0,456 est la corrélation entre « Informatique » et « Sociologie ». r^2 est le coefficient de détermination (le pourcentage de variation d'une variable expliqué par son ajustement linéaire par une autre variable). t est la valeur de la statistique de Student sous l'hypothèse nulle que la corrélation au niveau de la population est 0 : pour la corrélation entre « informatique » et « Sociologie », t vaut 1,451 et la probabilité d'une valeur égale ou supérieure est encore de 0,18 ce qui conduit à l'acceptation de l'hypothèse nulle.

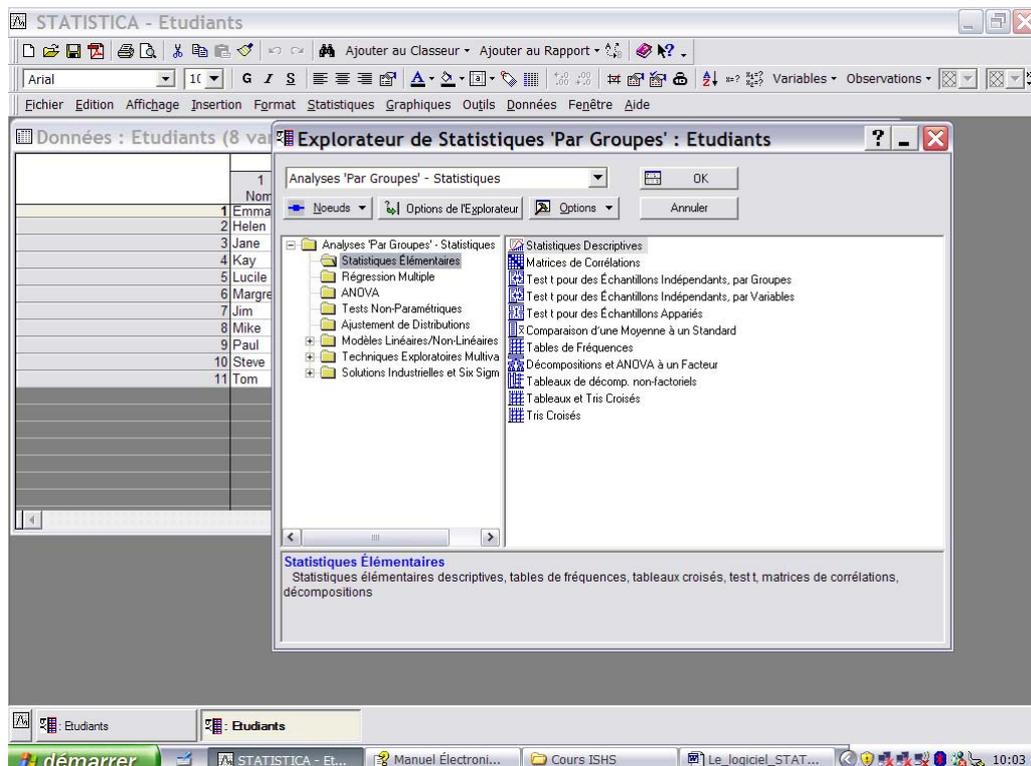
Les dernières colonnes donnent les résultats des coefficients de régression des variables.

4. Analyse par groupes d'observations

Il s'agit d'une fonctionnalité entièrement nouvelle accessible dans la version 7 ; elle permet de répéter automatiquement une analyse (et ses différentes options) sur des groupes d'observations prédéfinis. La constitution des groupes est réalisée à partir des valeurs prises par les variables qui les déterminent. Par exemple, dans le fichier Etudiants, les valeurs de la variable « Sexe » permet de constituer deux groupes : les Femmes et les Hommes à partir desquelles on désire procéder à des analyses distinctes. On commence par sélectionner le menu « Statistiques », ensuite l'option « Analyse 'Par Groupes' » :



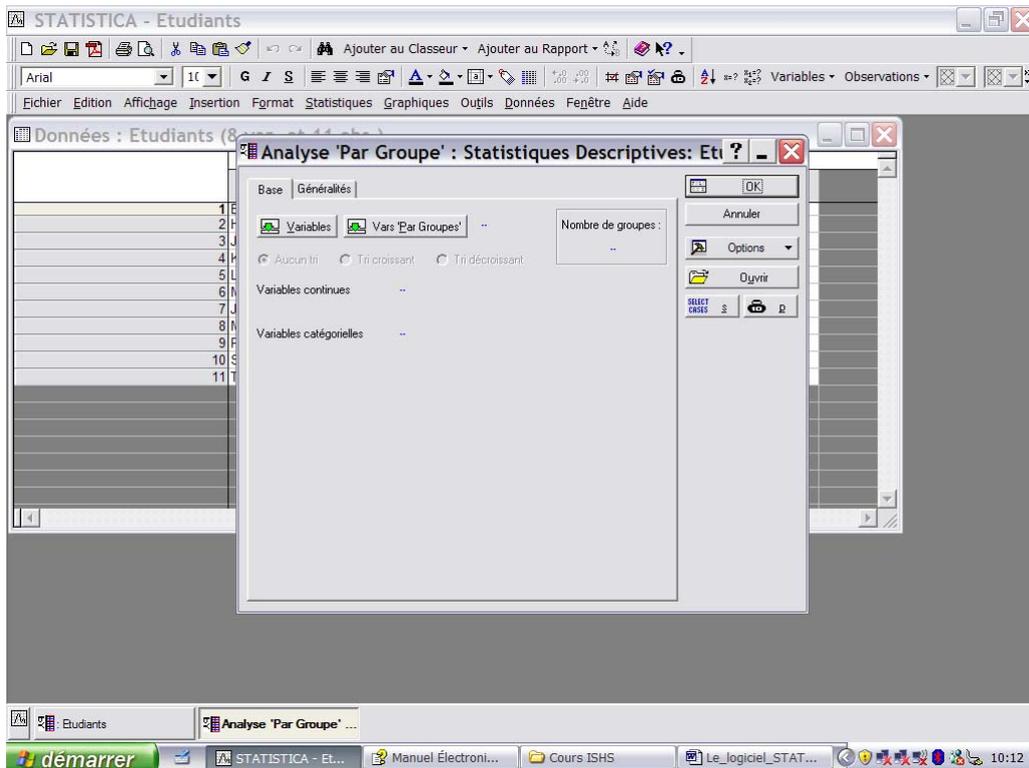
pour obtenir l'écran suivant :



dans lequel les différentes analyses possibles sous le menu « Statistiques » sont transposées sous la forme de lignes dans la partie gauche.

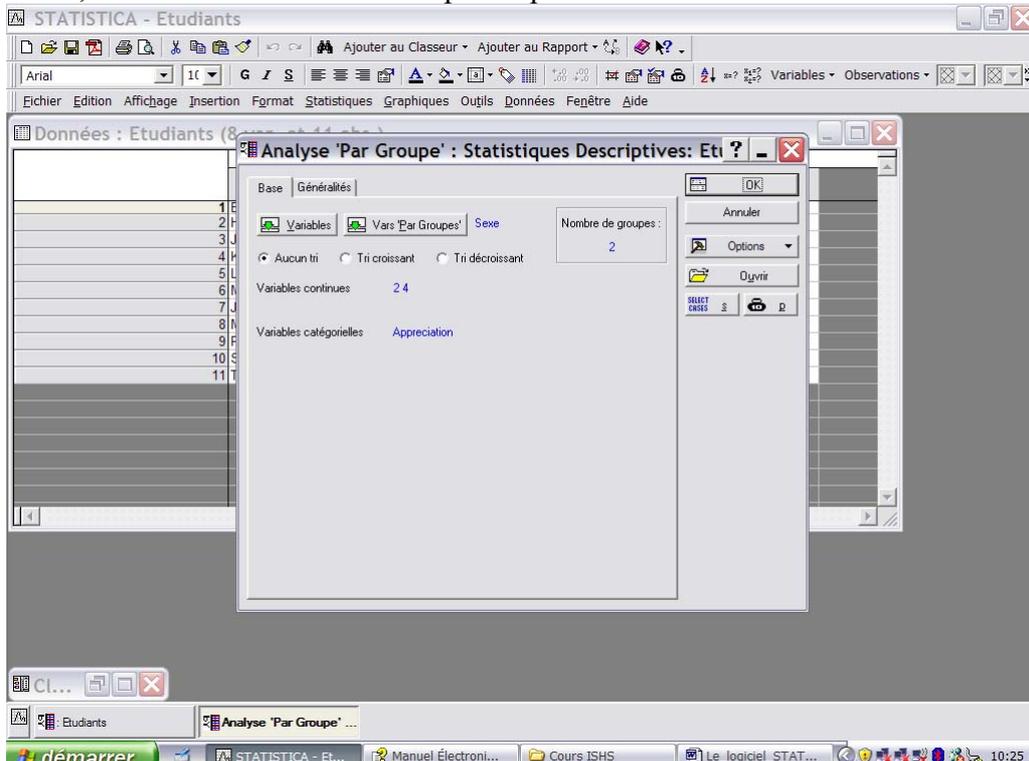
On désire obtenir les statistiques descriptives des variables « Sociologie », « Appréciation » et « Informatique » selon le Sexe ; le choix « Statistiques Descriptives » conduit à l'écran qui

permet de préciser quelle(s) variable(s) on désire analyser et quelle(s) variable(s) va (vont) constituer les groupes :



Le bouton « Variables » permet de sélectionner « Informatique » et « Sociologie » définies dans le fichier comme des variables qui prennent des valeurs entières ou décimales ; elles apparaissent comme variables continues (en réalité des variables de ratio). On définit ensuite « Appréciation » comme variable catégorielle (en réalité, une variable ordinale).

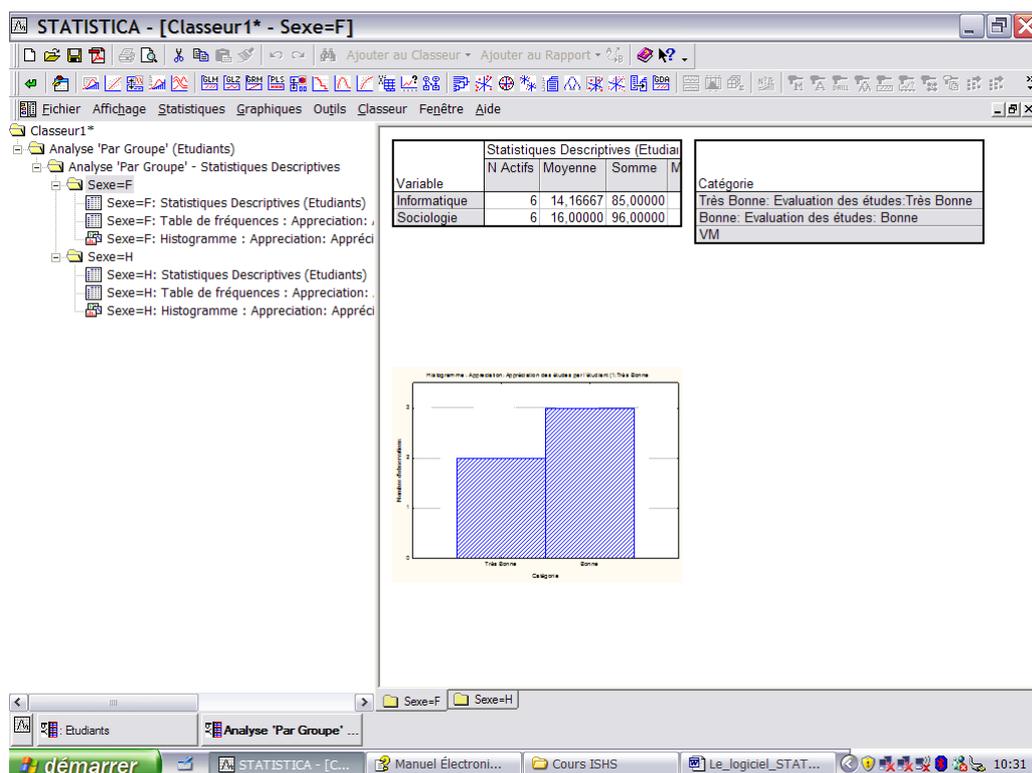
Enfin, le bouton « Vars 'Par Groupes' » permet de sélectionner la variable « Sexe » :



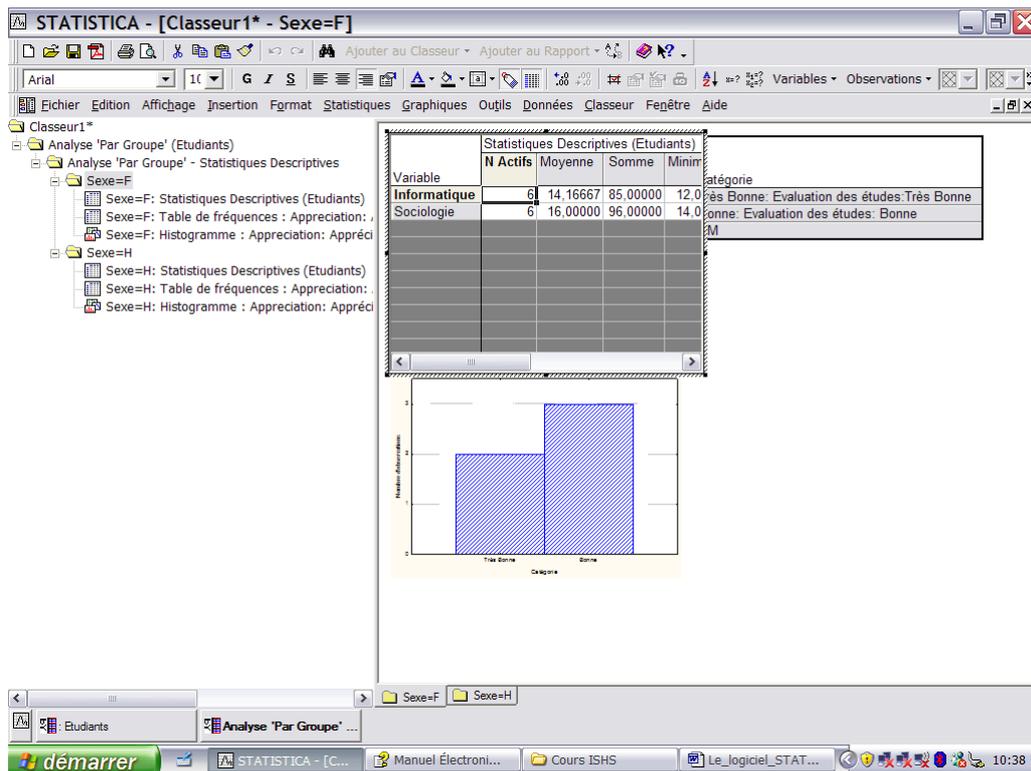
On peut présenter les résultats selon les valeurs retenues pour la constitution des groupes (Cases à cocher « Aucun Tri », « Tri croissant », « Tri décroissant »)

Le nombre de groupes est précisé selon le nombre de valeurs trouvées dans la variable qui permet de constituer les groupes (Nombre de groupes : 2). Lorsque plusieurs variables ont été sélectionnées pour la constitution des groupes, le nombre de ceux-ci est le produit du nombre de modalités des différentes variables.

L'exécution de l'Analyse par Groupes produit les résultats suivants :



Les statistiques descriptives sont affichées selon les groupes qui ont été constitués (voir l'arborescence des résultats dans la partie de gauche du Classeur), et au sein de chaque groupe, l'analyse spécifique au type de variable (continue – catégorielle). Chaque objet de la partie de droite est accessible par un double-clic qui permet d'obtenir l'entièreté des résultats, par exemple, pour le premier tableau en haut et à gauche :



Les onglets en bas de l'écran permettent de se déplacer dans les groupes.

C. TRAITEMENTS AVANCES SUR LES VARIABLES

1. Les expressions

L'analyse des variables de ratio (variables numériques) par les tableaux croisés est une technique utilisée très souvent en statistique ; par exemple, notre feuille de données contient l'année de naissance des étudiants et nous pourrions nous demander si l'appréciation qu'ils ont donnée sur les études ne peut pas être mise en relation avec leur âge. Plutôt que d'utiliser immédiatement des techniques de régression, on désirerait donc construire la table des fréquences qui croisent la variable « Appreciation » et des groupes d'âges qui nous intéressent.

La première difficulté que nous devons surmonter consiste à transformer l'année de naissance (« Naissance ») en âge ; il faut donc créer une variable qui est le résultat d'un calcul réalisé sur une autre variable.

A la page 8, lorsque nous avons décrit les spécifications des variables, nous avons vu comment indiquer une description détaillée en insérant un texte qui décrit la variable et nous avons précisé que si cette description commence par « = », la variable est en réalité le résultat d'une expression. C'est cette possibilité qui va être développée en détails.

Une expression est la combinaison d'opérateurs, d'opérandes et de fonctions qui aboutit à l'évaluation d'un résultat. Dans le logiciel STATISTICA, le résultat peut être de plusieurs types : numérique, texte (ou caractère) et logique (Vrai-Faux).

Parmi les opérateurs, on trouve les opérateurs arithmétiques (+,-,/,*,**) classiques, les opérateurs relationnels (>,>=,<>,<=,<=) et les logiques (NOT, AND,OR). Les opérandes sont des constantes ou des variables (les colonnes du fichier STATISSTICA). Les fonctions appartiennent à 5 catégories : Générales, Distributions, Mathématiques, Statistiques et Texte.

Chaque fonction porte un nom qui l'identifie et qui figure dans l'expression et un ou plusieurs arguments (des opérandes) sur lesquels elle va s'exécuter.

Exemples :

- LEN (var) : calcule la longueur du texte contenu dans var
- UCASE (var) : convertit le texte contenu dans var en majuscules
- MEAN (v3 ;v4) : calcule la moyenne arithmétique des variables v3 et v4

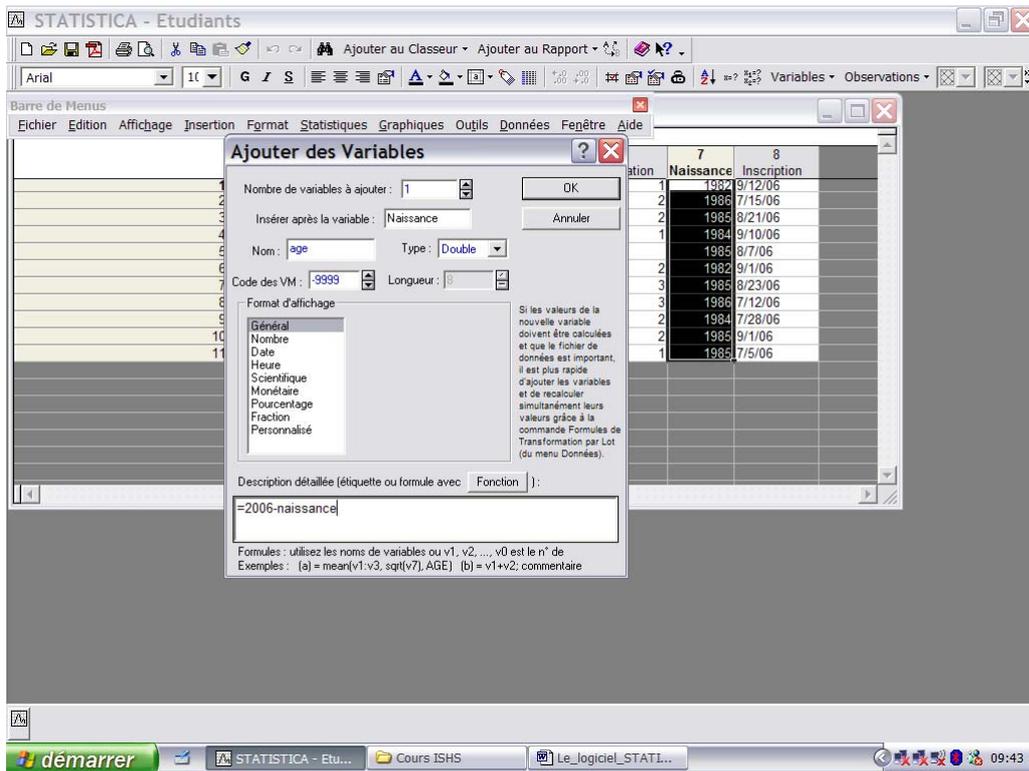
Dans les arguments d'une fonction, on peut trouver d'autres fonctions (fonctions de fonctions), dans ce cas, ce sont les fonctions les plus internes qui sont évaluées d'abord.

Exemple :

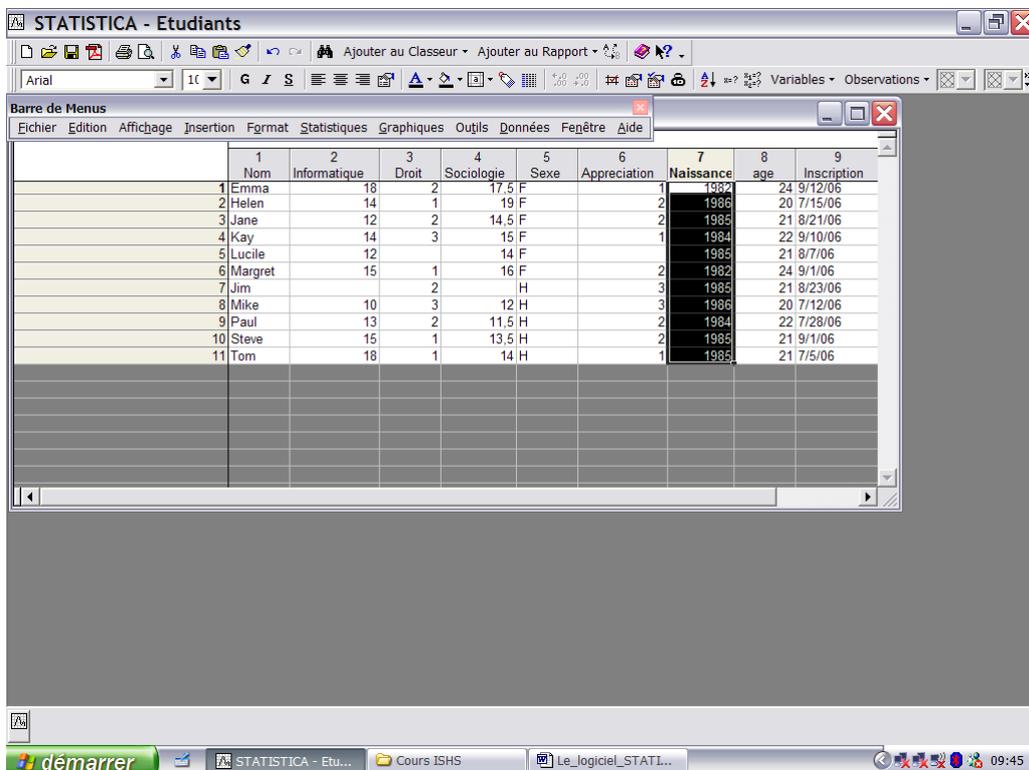
- UCASE(MID(v1 ;1 ;4))

MID extrait les 4 premiers caractères de la variable (texte) v1 qui sont convertis en majuscules par UCASE.

On commence par ajouter une colonne à la feuille de données via le menu « Insertion » et la ligne « Ajouter des variables » qui ouvre les spécifications d'une nouvelle variable que STATISTICA appelle par défaut Nv_Var, on lui donne le nom « Age » et on l'insère après la variable « Naissance », c'est une variable numérique (type « double ») et elle est égale à la différence entre l'année 2005 et l'année de naissance de l'étudiant (c'est-à-dire la variable « Naissance ») :



La validation de la fenêtre (« OK ») donne une nouvelle colonne qui est bien l'âge des étudiants :

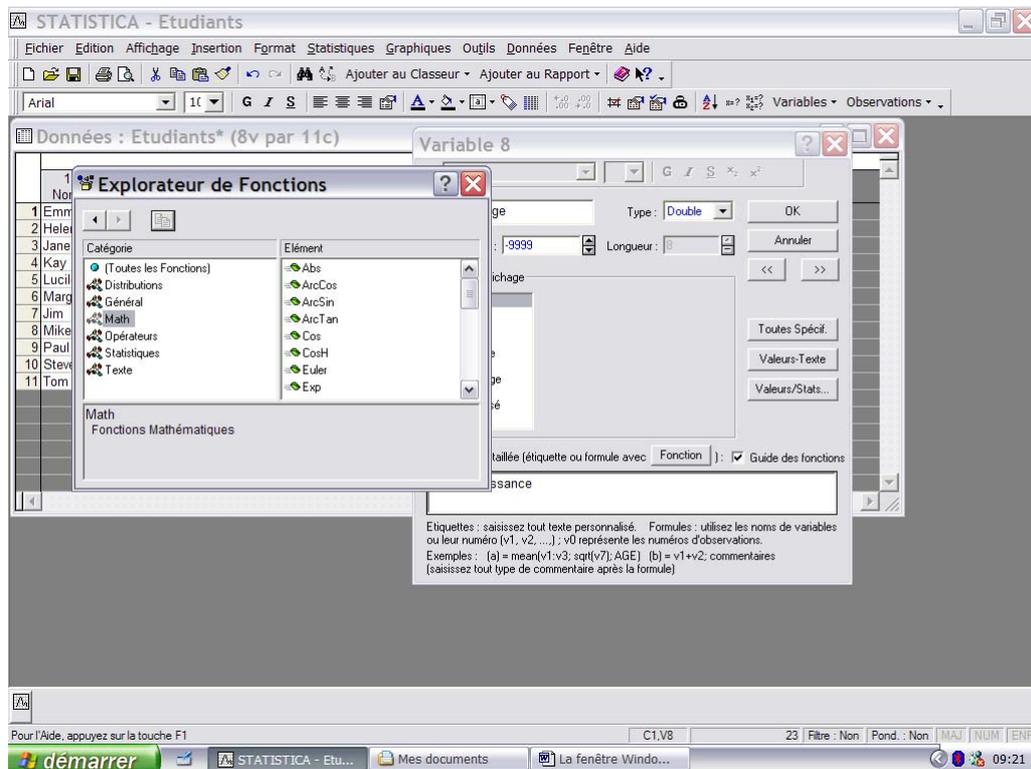


L'évaluation de l'expression qui permet d'obtenir l'âge a été réalisée pour toutes les lignes de la feuille de données. Dans ce cas, il s'agit d'une expression utilisant seulement l'opérateur de soustraction entre deux variables numériques ; en réalité, on peut écrire dans la zone de description détaillée d'une variable des expressions beaucoup plus complexes en utilisant

d'autres opérateurs que les opérateurs algébriques classiques (addition, soustraction, multiplication et division), en l'occurrence :

- les opérateurs relationnels : <, >, <=, >=, =
- les opérateurs logiques : NOT, AND, OR

Enfin, on peut y insérer les fonctions internes très nombreuses dont la documentation est obtenue en revenant à la fenêtre « Spécifications » (double clic sur un nom de variable de la feuille de données) et en utilisant le bouton « Fonctions » :



la fenêtre « Explorateur de Fonctions » indique à gauche les grandes catégories et à droite les fonctions utilisables (dans l'exemple, les opérateurs de la catégorie « Math »).

En présence d'une colonne de la feuille de données qui est le résultat du calcul opéré sur d'autres colonnes, on peut se demander si une modification des données a un impact sur le résultat. Par exemple, supposons que nous ayons fait une erreur d'encodage et qu'en réalité l'étudiant Tom soit né en 1987, on entre donc cette valeur dans la cellule :

STATISTICA - Etudiants

Echier Edition Affichage Insertion Fgmat Statistiques Graphiques Outils Données Fenêtre Aide

Ajouter au Classeur - Ajouter au Rapport

Données : Etudiants* (8v par 11c)

	1 Nom	2 Informatique	3 Droit	4 Sociologie	5 Sexe	6 Appreciation	7 Naissance	8 age
1	Emma	18	2	17,5	F		1982	23
2	Helen	14	1	19	F		1986	19
3	Jane	12	2	14,5	F		1985	20
4	Kay	14	3	15	F		1984	21
5	Lucile	12		14	F		1985	20
6	Margret	15	1	16	F		1982	23
7	Jim		2		H		1985	20
8	Mike	10	3	12	H		1986	19
9	Paul	13	2	11,5	H		1984	21
10	Steve	15	1	13,5	H		1985	20
11	Tom	18	1	14	H		1985	22

Prêt C1.V8 23 Filtre : Non Pond.: Non MAJ NUM ENR

démarrer STATISTICA - Etu... Mes documents La fenêtre Windo... 09:05

On peut constater que l'âge de l'étudiant n'a pas été modifié et qu'il y a donc une erreur dans la colonne « age ».

Pour que la mise à jour d'une variable « calculée » soit automatique, il est nécessaire de le spécifier dans les paramètres généraux de STATISTICA. Pour cela, on utilise le menu « Outils » dont on choisit « Options ».

L'onglet « Feuille de données » contient une option « Recalcul automatique des formules lorsque les données changent » qui doit être sélectionnée.

STATISTICA - Etudiants

Echier Edition Affichage Insertion Format Stati

Ajouter au Classeur - Ajouter au Rapport

Options

Barre de Menus

	1 Nom	Inform
1	Emma	
2	Helen	
3	Jane	
4	Kay	
5	Lucile	
6	Margret	
7	Jim	
8	Mike	
9	Paul	
10	Steve	
11	Tom	

Listes Personnalisées | Gestionnaire de Configurations | Programmes SVB (Macros) | Classeurs | Rapports | Graph. 1

Généralités | Analyses/Graphiques | Gestionnaire de Sorties

Graph. 2 | Feuilles de Données | Importation | Fenêtre du Navigateur

Déplacement dans la feuille de données après Entrée
 Aucun déplacement Vers le bas A droite

Action de la touche de tabulation
 Naviguer dans le doc. Insérer le caractère Tabul.

Largeur d'affichage maximale des colonnes : 20 cm
 Personnalisée : 20.32 cm

Recalcul auto. des formules si les données changent
 Inclure la couleur du fond en copiant les cellules
 Inclure la couleur du fond en copiant les entêtes
 Autoriser l'annulation des requêtes (déconseillé)

Extrapolation des données (glisser/déplacer)
 Ajouter des obs./variables supplémentaires si besoin
 Extrapoler dans les limites du fichier de données
 Toujours demander confirmation

Mise en Forme par Défaut de la Feuille de Données : Paramètres par Défaut du Système

Année de référence (début du siècle) pour les dates à 2 chiffres : 1930

Utiliser le format de date Julien de type Excel
 Afficher un message d'avertissement sur l'état des Filtres de Sélection de la feuille de données
 Afficher un message d'avertissement sur l'état des Pondérations de la feuille de données

Avertissement sur les performances lors du tri des données (formats spéciaux)
 Afficher un avertissement Inclure la mise en forme des cellules Supprimer la mise en forme des cellules

Commencer la numérotation automatique des valeurs-texte à partir de la valeur numérique : 101

Afficher le Guide de Fonctions contextuel (info-bulles) dans les éditeurs de formules
 Message d'avertissement si un fichier (avec nombreuses vars inutilisées dans l'analyse) dépasse : 50 Mo

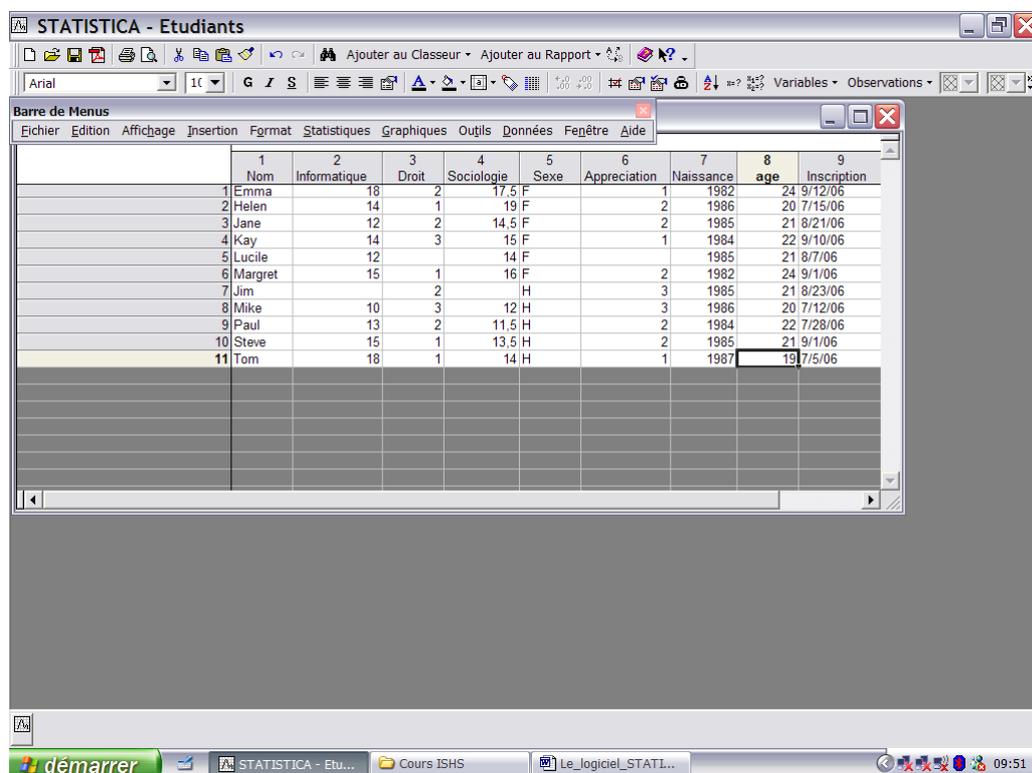
Dossier des fichiers temp. : défaut personnalisé : C:\Documents and Settings\hborn\Local Setti

Taille du cache : défaut personnalisé : 64 kilo-octets

OK Annuler

démarrer STATISTICA - Etu... Cours ISHS Le_logiciel_STATL... 09:52

On valide cette option et on revient à la feuille de données où l'on réintroduit pour Tom l'année de naissance « 1987 » :



L'âge de l'étudiant Tom a bien été modifié. La colonne « age » sera dorénavant enregistrée au même titre que les autres colonnes lorsque nous procéderons à une sauvegarde de la feuille de données.

2. Recodification des variables

Pour répondre à notre question de départ, il nous reste à catégoriser la nouvelle variable, chaque catégorie correspondant à un groupe d'âges ; supposons que nous sommes intéressés aux groupes d'âges suivants :

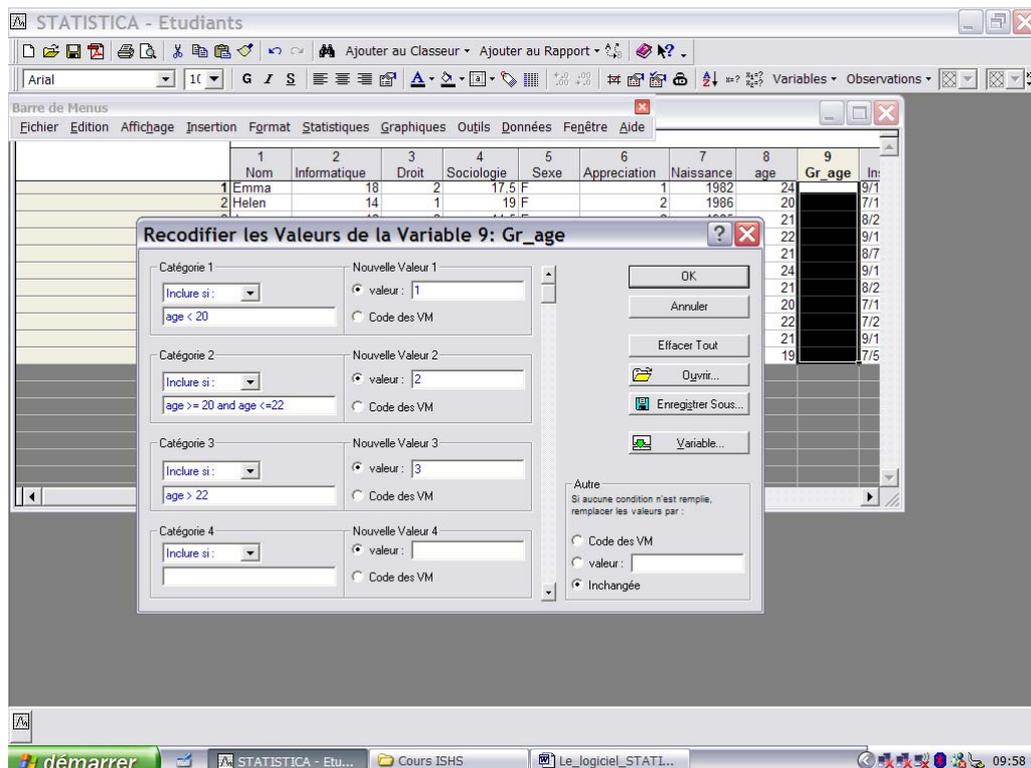
- étudiants de moins de 20 ans
- étudiants de 20 à 22
- étudiants de plus de 22 ans

On commence par créer une nouvelle variable dans le tableau dans laquelle on rangera les catégories d'âges, on appelle cette variable « Gr_ages ».

On décide que cette variable prendra 3 valeurs :

- 1 pour les étudiants de moins de 20 ans
- 2 pour le groupe 20 à 22 ans
- 3 pour les plus de 22 ans

On sélectionne la variable Gr_age (clic sur son nom) ; dans le menu « Données », on choisit « Recodifier » :

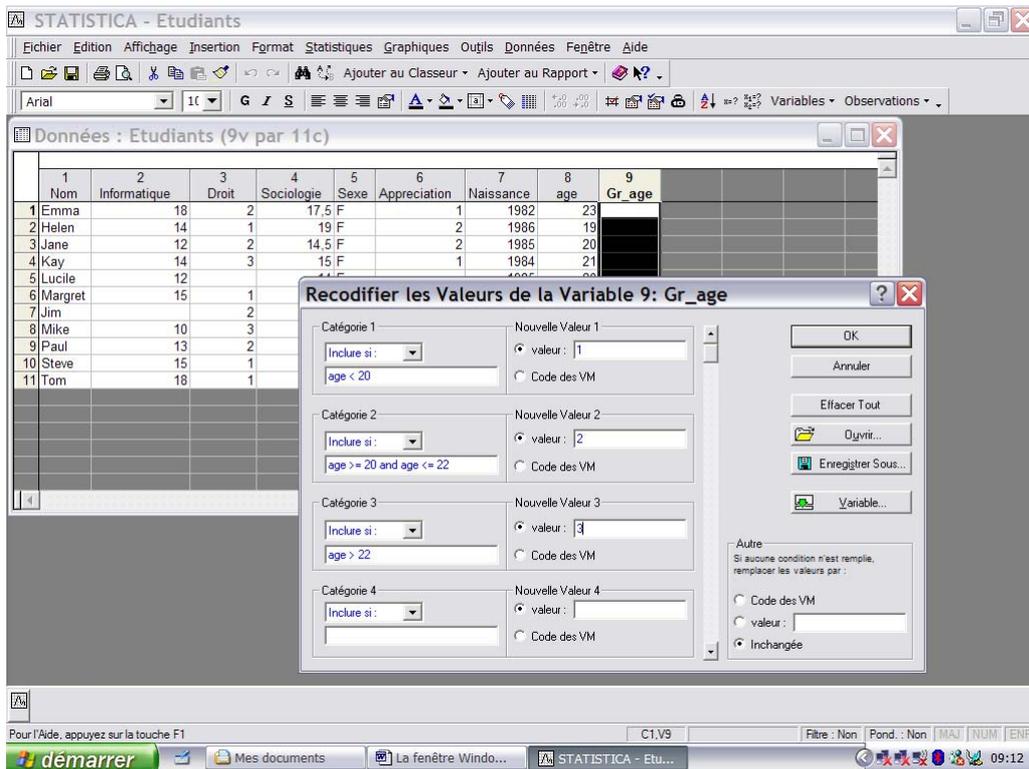


La fenêtre de recodification de la variable Gr_age permet d'écrire les expressions conditionnelles dont le résultat logique « Vrai » créeront les catégories dont les valeurs commencent à 1.

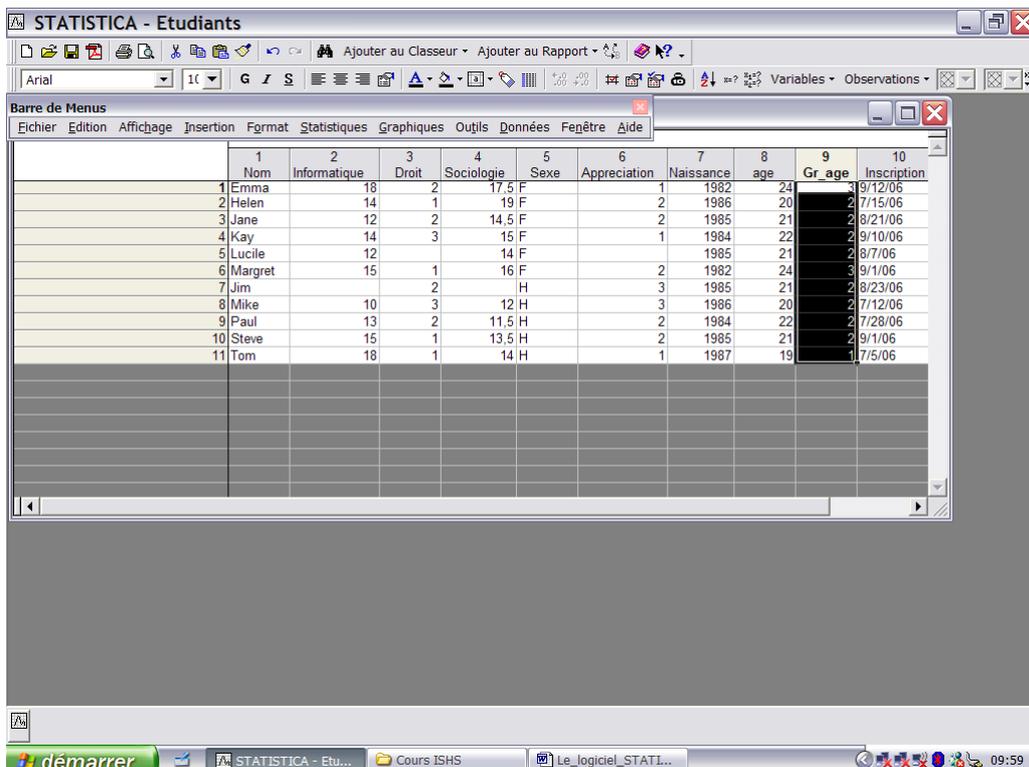
Exemple : Nous avons décidé que les étudiants de moins de 20 ans constitueront la catégorie « 1 ». Il faut donc écrire une condition qui teste l'âge du type :

« SI age < 20 alors Gr_age = 1 »

Cette écriture est obligatoire pour tous les groupes que nous avons décidé de retenir, en STATISTICA, voici le résultat :



Pour la catégorie 2, on a l'expression « age >= 20 and age <= 22 » qui définit bien l'intervalle des valeurs à l'intérieur duquel les étudiants devront se trouver pour constituer le deuxième groupe. Un espace au moins doit séparer les opérateurs logiques (AND, OR, NO), des espaces peuvent être insérés entre les autres opérateurs et les opérandes mais pas à l'intérieur des opérateurs eux-mêmes (>= est considéré comme UN opérateur).
Voici le résultat :



Le résultat du tableau de fréquences qui croise les variables « Appreciation » et « Gr_age » est finalement :

The screenshot shows the STATISTICA interface with a window titled 'Classeur1* - Table de Fréquences - Synthèse (Etudiants)'. The window displays a frequency table with the following data:

Appreciation	Gr_age 1	Gr_age 2	Gr_age 3	Totaux Ligne
Très Bonne	1	1	1	3
Bonne	0	4	1	5
Passable	0	2	0	2
Ts Grpes	1	7	2	10

The background shows a data table with columns: 1 Nom, 2 Informatique, 3 Droit, 4 Sociologie, 5 Sexe, 6 Appreciation, 7 Naissance, 8 age, 9 Gr_age, 10 Inscription. The 'Gr_age' column is highlighted in black.

n.b. : Pour recoder la variable Gr_age, nous aurions pu utiliser la possibilité de la définir comme le résultat d'une expression écrite à partir de la variable «age» à condition toutefois de posséder quelques rudiments à propos des langages informatiques. Voici une façon de faire :

The screenshot shows the 'Variable 9' dialog box in STATISTICA. The 'Nom' field is set to 'Gr_age' and the 'Type' is 'Double'. The 'Code des VM' is '-9999' and the 'Longueur' is '3'. The 'Format d'Affichage' section is set to 'Général'. The 'Description détaillée' field contains the formula:

$$=age < 20 * 1 + (age >= 20 \text{ and } age <= 22) * 2 + (age > 22) * 3$$

The dialog box also includes options for 'Toutes Spécif.', 'Valeurs-Texte', and 'Valeurs/Stats...'. The background shows a data table with columns: 1 Nom, 2 Informatique, 3 Droit, 4 Sociologie, 5 Sexe, 6 Appre.

Quelques explications à propos de l'expression:
 $(age < 20) * 1 + (age \geq 20 \text{ and } age \leq 22) * 2 + (age > 22) * 3$

- chaque parenthèse contient une condition dont le résultat est Vrai ou Faux
- ce résultat est une opérande d'une expression numérique (il est multiplié par une constante) ; le résultat Vrai vaut alors 1 et le résultat Faux vaut 0
- on fait ensuite la somme algébrique de ces trois expressions

Par exemple, s'il s'agit d'un étudiant qui est âgé de 22 ans, alors :

$(age < 20) * 1$ vaut 0

$(age \geq 20 \text{ and } age \leq 22) * 2$ vaut 2

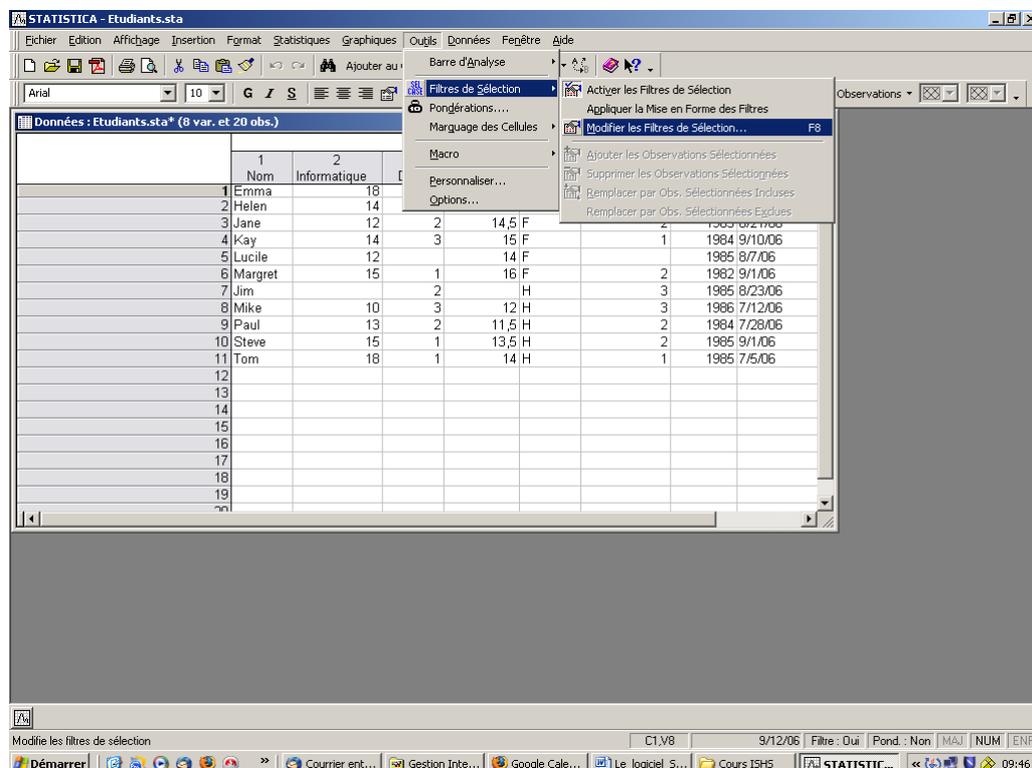
$(age > 22) * 3$ vaut 0

La somme vaut 2 ce qui est le résultat attendu.

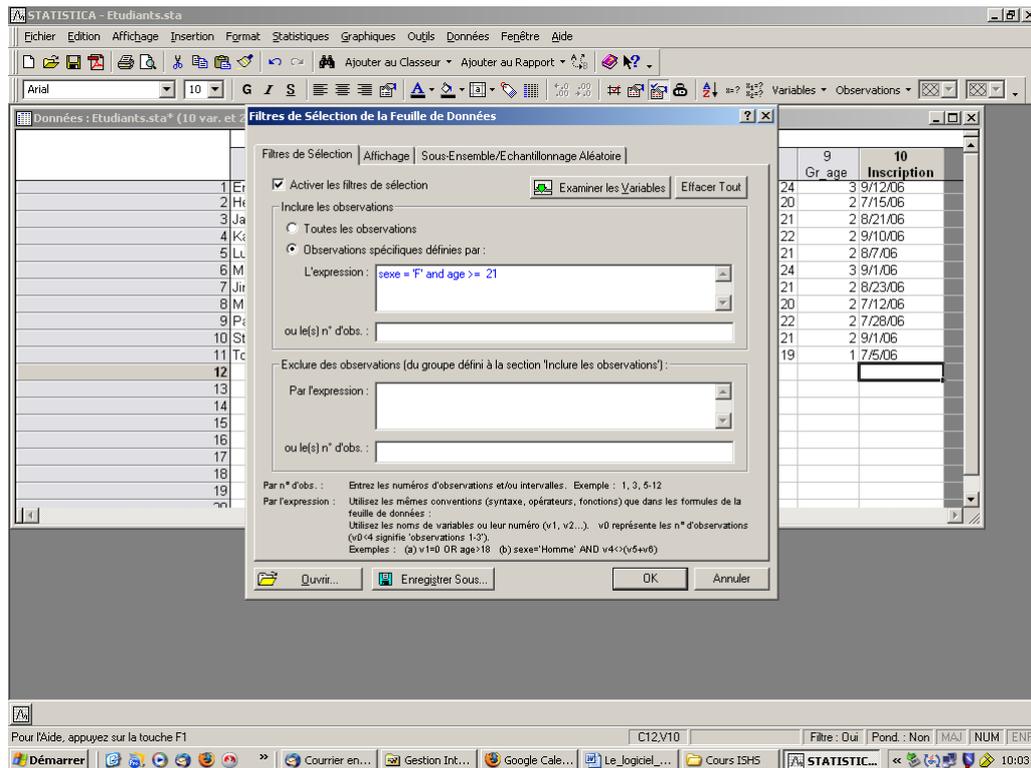
3. Les filtres de sélection

Les expressions conditionnelles qui conduisent à un résultat logique « Vrai » ou « Faux » permettent de sélectionner les lignes de la feuille de données aux quelles seront appliqués les analyses.

Supposons, par exemple, que nous désirions analyser les notes à l'examen d'Informatique pour les étudiants Filles âgées de 21 ans et plus ; le menu « Outils » et le choix « Filtre de Sélection » va permettre de rédiger la condition (le « Filtre ») qui permettra de sélectionner les étudiants visés :



La rédaction du filtre est accessible par «Modifier les filtres » :



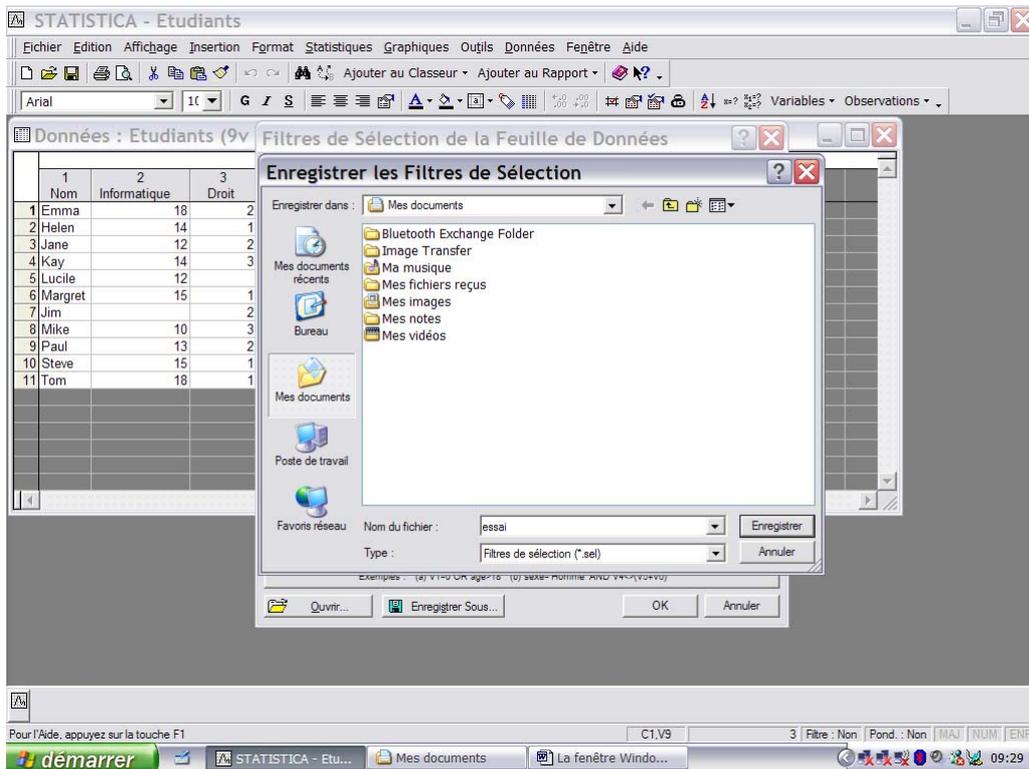
Dans cette fenêtre d'édition d'un filtre de sélection, on commence par « Activer le filtre » ; deux possibilités sont alors accessibles :

- i. soit on rédige l'expression conditionnelle d'inclusion (toutes les observations pour lesquelles l'expression est « Vrai » sont retenues)
- ii. soit on rédige l'expression conditionnelle d'exclusion (toutes les observations pour lesquelles l'expression est « Vrai » sont exclues).

Dans les deux cas, on peut remplacer l'expression conditionnelle directement par les numéros des observations à inclure ou à exclure.

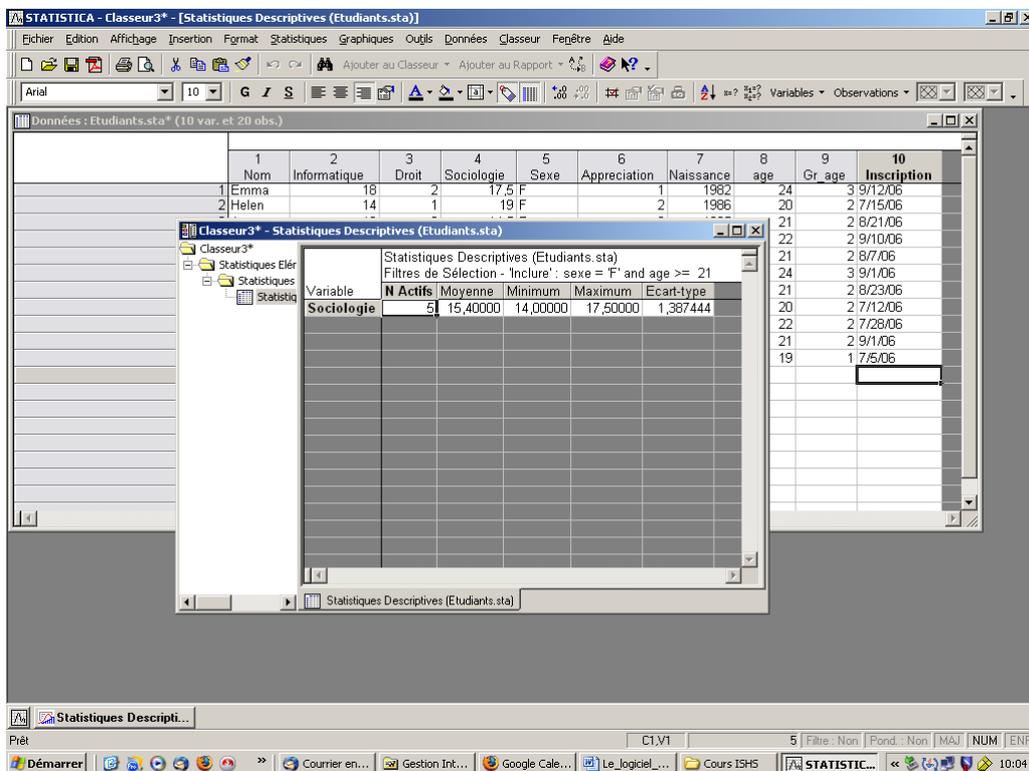
Dans l'exemple, on choisit l'expression d'inclusion « sexe='F' and age >=20 » qui définit bien le sous-ensemble des étudiants auquel on s'intéresse. On note que les constantes de type texte sont obligatoirement entourées d'une apostrophe.

L'expression peut être enregistrée pour une utilisation future dans un fichier à suffixe « sel ». Le filtre enregistré pourra être ouvert dans la fenêtre d'édition (bouton « Ouvrir »).



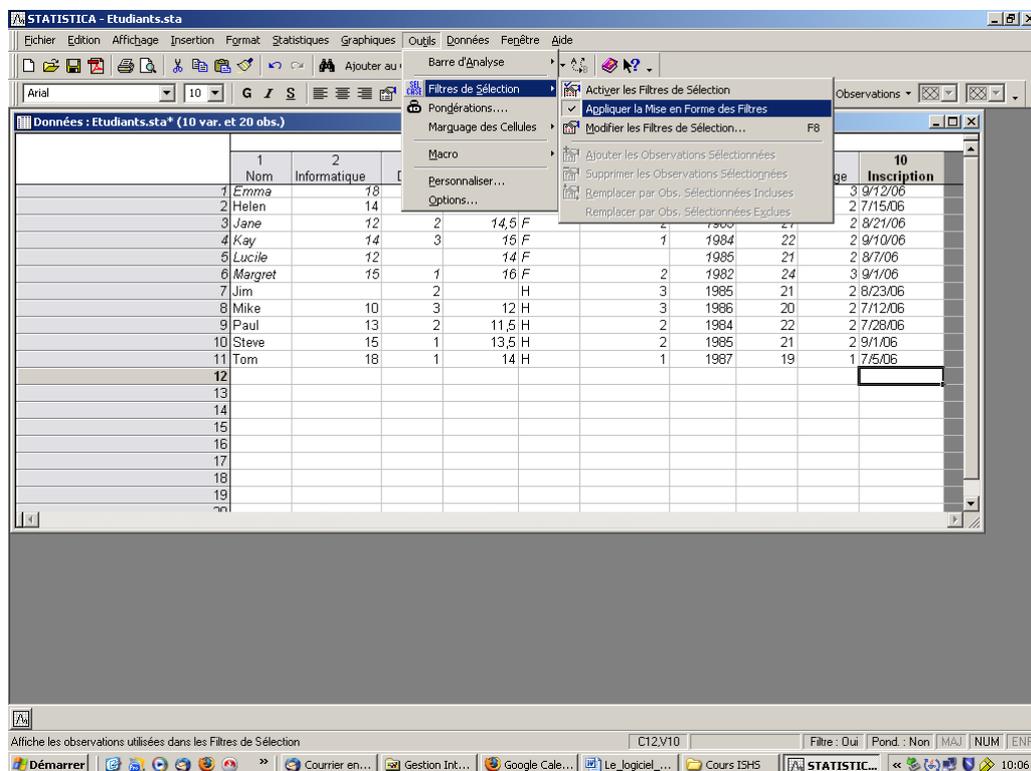
Dès que la fenêtre d'édition du filtre est activée (« OK »), la feuille de données réapparaît à l'écran et la ligne d'état affiche l'information « Filtre :Oui » indiquant qu'une sélection a été activée. Toute analyse qui sera réalisée se limitera aux lignes sélectionnées grâce à ce filtre.

Par exemple, une analyse descriptive des notes à l'examen de sociologie donne le résultat :



Le nombre « N Actifs » vaut 5 ce qui correspond aux filles âgées d'au moins 21 ans.

Pour visualiser les lignes de la feuille de données qui sont sélectionnées, on peut utiliser dans le menu « Données », « Filtres de sélection » l'option « Appliquer la mise en forme des Filtres » qui permet d'obtenir les lignes filtrées dans un format particulier (par défaut, en italiques, ce format est modifiable dans la fenêtre d'édition du filtre, onglet « Affichage ») :



La désactivation d'un filtre est accessible par le menu « Outils », « Filtres de Sélection » et « Activer » ; on vérifie ensuite que la ligne d'état de STATISTICA affiche « Filtre : Non ».