

Le systeme SICLA : manuel de l'utilisateur

Y. Ok

▶ **To cite this version:**

Y. Ok. Le systeme SICLA : manuel de l'utilisateur. RT-0092, INRIA. 1987, pp.66. inria-00070074

HAL Id: inria-00070074
https://inria.hal.science/inria-00070074

Submitted on 19 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INRIA

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports Techniques

N° 92

Le Système SICLA

MANUEL DE L'UTILISATEUR

Yvette OK-SAKUN

NOVEMBRE 1987

MANUEL DE L'UTILISATEUR

USER'S MANUAL

Yvette OK-SAKUN

I.N.R.I.A
Domaine de Voluceau Rocquencourt B.P. 105
78153 LE CHESNAY CEDEX

Abstract

Sicla is an interactive software developed at the INRIA by the automatic clustering and pattern recognition team.

In its present form it consists of a set of 66 independent commands categorized by subjects which, in the area of data analysis and especially in that of automatic clustering, permit the use of basic methods or the development of new ones. The purpose of this manual is to explain the use of the Sicla software. It also describes the terminology used in the dialogue between the user and the system and the methods of recording data as well as the various possibilities of data processing.

A reference manual and a programmer's manual are also available.

Résumé

SICLA est un logiciel interactif développé à l'INRIA par l'équipe de Classification Automatique et de Reconnaissance des Formes. Il est constitué, dans sa version actuelle, par un ensemble de 66 commandes indépendantes classées par thèmes qui permettent dans le domaine de l'analyse des données et en particulier de la classification automatique d'accéder aux méthodes de base ou de développer de nouvelles méthodes.

L'objet de ce manuel est de faciliter l'utilisation du logiciel Sicla. Il décrit aussi bien la terminologie utilisée dans le dialogue système-utilisateur que les modalités d'enregistrement des données ou encore les différentes possibilités de traitement.

Un manuel de référence et un manuel du programmeur sont également disponibles.

PAPIER RÉCUPÉRÉ ET RECYCLÉ

1. DESCRIPTION DE SICLA 1
1.1. GENERALITES 1
1.2. GUIDE PRATIQUE DU DIALOGUE 4
1.2.1. UTILISATION PAR COMMANDES DIRECTES 4
1.2.2. UTILISATION PAR MENU 5
2. ASSISTANCE 6
2.1. EXEMPLE D'UTILISATION SOUS MULTICS 6
2.2. ASSISTANCE SYSTEME SICLA 7
2.3. ASSISTANCE EXTERIEURE 7
2.4. EXEMPLE D'ENCHAINEMENT DES THEMES 8
3. TERMINOLOGIE UTILISEE 9
3.1. INFORMATIONS CONCERNANT LES VARIABLES 9
3.1.1. ENSEMBLE DES VARIABLES QUANTITATIVES 9
3.1.2. ENSEMBLE DES VARIABLES QUALITATIVES 10
3.2. INFORMATIONS CONCERNANT LES INDIVIDUS 11
4. COMMENT UTILISER SICLA 12
4.1. COMMENT ENREGISTRER 12
4.1.1. LES DIFFERENTES PHASES D'ENREGISTREMENT PAR SICLA 12
4.1.2. ENREGISTREMENT DU DICTIONNAIRE ABREGE 16
4.1.3. EXEMPLE ENREGISTREMENT PAR MENU ET DICTIONNAIRE 18
4.1.4. ENREGISTREMENT CONVERSATIONNEL PAR COMMANDE DIRECTE 23
4.1.5. CORRECTION DES ERREURS 36
4.1.6. LES COMMANDES D'ENREGISTREMENT 37
4.1.6.1 ENRDIS 37
4.1.6.2 ENRFIC 38
4.1.6.3 ENRI 38
4.1.6.4 ENRIT 39
4.1.6.5 ENRV 39
4.2. COMMENT GERER 42
4.2.1. COMMANDES DE MANIPULATION DES STRUCTURES DE DONNEES 43
4.2.2. COMMANDES DE GENERATION DE NOUVELLES VARIABLES 46
4.2.3. RECODIFICATION DES VARIABLES QUALITATIVES 48
4.2.4. COMMANDES DE GESTION DE PARTITIONS 48
4.3. COMMENT DECRIRE 49
4.3.1. IMPRESSION DE DESCRIPTIFS 49
4.3.2. CALCUL D'INDICATEURS 49
4.3.3. EXEMPLES DE DESCRIPTION 50
4.4. COMMENT ANALYSER 52
4.5. COMMENT INTERPRETER 52
4.6. COMMENT SAUVEGARDER ET UTILISER L'ARCHIVE 53
4.7. COMMENT EDITER 54
5. BIBLIOGRAPHIE 55
6. ANNEXES 56
6.1. LISTE ALPHABETIQUE DES COMMANDES 56
6.2. LISTE DES THEMES ET DES COMMANDES PAR THEME 59
6.2.1. LE THEME EDITION 60
6.2.2. LE THEME GESTION 61
6.2.3. LE THEME DESCRIPTION 62
6.2.4. LE THEME INTERPRETATION 62
6.2.5. LE THEME CLASSIFICATION 63
6.2.6. LE THEME FACTORIEL 63
6.2.7. LE THEME DISCRIMINATION 64
6.2.8. LE THEME HIERARCHIE 64
6.2.9. LE THEME CODAGE 65

1. DESCRIPTION DE SICLA

1.1. GENERALITES

SICLA est un Systéme Interactif de CLassification Automatique permettant la GESTION, la DESCRIPTION ELEMENTAIRE et le TRAITEMENT des données par les méthodes d'analyse muldimensionnelle.
Son objectif principal est la diffusion de méthodes issues de recherches récentes, principalement celles concernant la classification automatique et notamment les Nuées Dynamiques, mais il permet également d'accéder aux méthodes factorielles de la bibliothéque SPAD et aux méthodes de la bibliothéque de programmes MODULAD.

Les caractéristiques de SICLA sont :

- LA FACILITE D'ACCES

  SICLA nécessite peu de connaissances informatiques préalables (utilisation du terminal, exploitation de fichiers, connaissance des formats Fortran [11])

- LA SECURITE D'UTILISATION

  Le systéme contrôle l'adéquation entre les donneés, les paramétres et l'algorithme choisi.

- LA POSSIBILITE DE TRAITER DES GRANDS TABLEAUX

  En évitant de garder les données en mémoire centrale, on peut traiter des tableaux de grande taille, de l'ordre de 1000 à 10000 individus selon la méthode.

- L'EXTENSIBILITE

  Toute nouvelle méthode peut étre intégrée sans que cela nécessite un investissement informatique important.
  On trouvera dans la Documentation Sicla no 2 "Manuel du programmeur" [3], les informations nécessaires à l'intégration des programmes.

SICLA a été résolument orienté vers l'interactivité:

1- SICLA dialogue avec l'utilisateur pour la saisie des informations concernant les données et le mode de traitement souhaité. Ces informations sont contrôlées et validées avant la poursuite du traitement.

2- SICLA est constitué d'un ensemble de commandes réparties en plusieurs thèmes (enregistrement, gestion, analyse, .etc..).

3- SICLA peut conserver dans une archive des résultats intermédiaires (partitions, multi-partitions, tableaux de distances, facteurs, etc....) utilisables à l'entrée d'autres algorithmes.

4- SICLA permet d'enchaîner facilement au vu des résultats déjà acquis d'autres traitements.

5- SICLA présente des interfaces avec quatre logiciels : avec PEPIN [5], système de gestion de base de données relationnel permettant une gestion complexe de l'information ; avec CLAVECIN [6] système expert pour une assistance intelligente ; avec le logiciel SPAD [8] pour le traitement par les méthodes factorielles ; avec un logiciel GRAPHIQUE respectant les normes GKS [7].

L'architecture de SICLA peut être schématisée de la facon suivante :

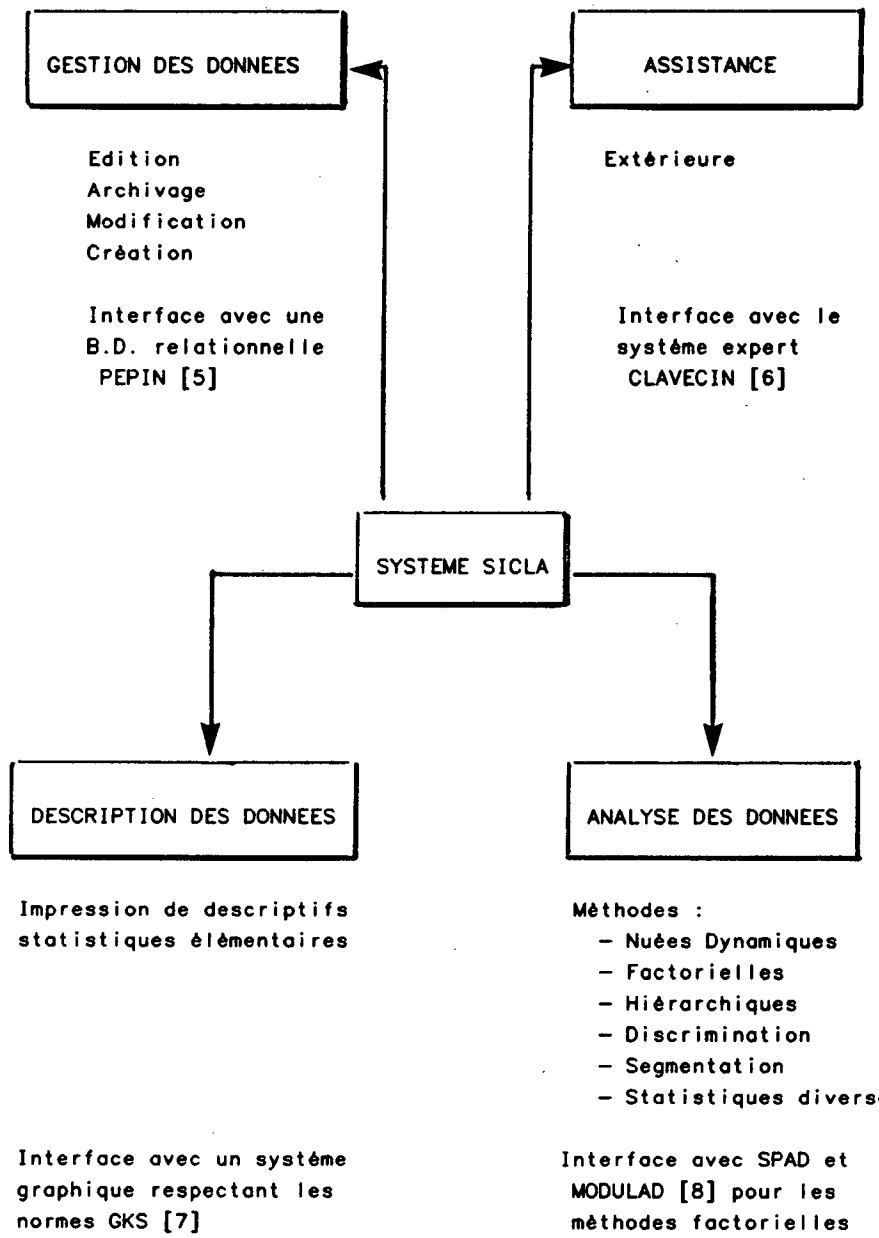


Figure 1 : ARCHITECTURE DE SICLA

1.2. GUIDE PRATIQUE DU DIALOGUE

Le logiciel SICLA est un systéme interactif auto-documenté qui dialogue avec l'utilisateur .

Le dialogue se déroule de la façon suivante :

Sur l'écran apparaît une séquence qui est une instruction á exécuter ou une question (un point d'interrogation s'imprime), Si l'utilisateur ne sait pas répondre, il peut demander l'aide du systéme.

Il faut taper :

" ? " pour obtenir des précisions sur les questions posées.

" / " pour obtenir l'arrêt de la session SICLA ou le retour á la question précédente selon le module.

" .. " pour transmettre une chaîne de caractéres au processeur de commandes sans sortir de SICLA.

Deux modes d'utilisation sont possibles :

- par COMMANDES DIRECTES .
- par MENU

(Cette dernière utilisation dépend du matériel utilisé)

1.2.1. UTILISATION PAR COMMANDES DIRECTES

Les commandes dont on trouvera en annexe la liste par théme et la liste par ordre alphabétique peuvent être appelées directement.
Par exemple on pourra enchaîner :

d'abord :

- ENRV (Enregistrement des informations sur les variables)
- ENRI (Enregistrement des données)
  Une structure de données (en abrégé s.d.) est alors créée.

puis :

- MODSDO (Modification de la s.d. en cas d'erreur)
- MNDQAN (Traitement des données par les Nuées Dynamiques)
- INPAQN (Interprétation de la partition obtenue)

etc.....

N.B. Lors de l'enchaînement des modules le fichier traité est la s.d. courante spécifiée lors de la première commande. Si l'on désire traiter une autre s.d. il faut utiliser la commande DONNEE pour se positionner sur une autre structure de données.

1.2.2. UTILISATION PAR MENU

Dans le cas d'utilisation de SICLA par MENU (cf paragraphe 4.1.3), il y a dialogue avec l'utilisateur pour :

1- ENREGISTRER LES INFORMATIONS concernant les données
2- CHOISIR LE TYPE DE TRAITEMENT.

Après l'enregistrement des informations sur les données (qui sera précisé dans le paragraphe COMMENT ENREGISTRER et qui n'est effectué qu'une seule fois), la poursuite du traitement peut être immédiate ou différée .

2. ASSISTANCE

2.1. EXEMPLE D'UTILISATION SOUS MULTICS

1. ajouter dans sa start_up

```
add_search_rule pathname_objets
add_search_path info pathname_info
```

où "pathname_objets" est le nom du directory contenant les objets sicla et "pathname_info" le nom du directory contenant les info de sicla.

2. crèer ou choisir une structure de données
Pour crèer une s.d. on utilisera les commandes ENRI ouENRV ; Les fichiers crèés sont :
<nom_des_données>.sdo et <nom_des_données>.sar

Pour choisir une s.d. on utilisera la commande DONNEE.

On peut également utiliser ENRFIC qui permet d'enregistrer des variables à partir d'un fichier, ENRIT qui enregistre des données transposées, ENRDIS pour un tableau de distances.(voir paragraphe 4.1.6)

3. appeler une commande
On en obtiendra la liste en faisant :

```
help sicla
```

4. recueillir les résultats en imprimant le fichier "listing" dont le numéro est indiqué par le systéme au dèbut.

```
**********************************************************************
* SICLA              Nom_utilisateur          Date : 05/06/85  *
* S.D :                          Archive :                    *
* Listing : lis16_10             Mode interactif              *
**********************************************************************
```

2.2. ASSISTANCE SYSTEME SICLA

Utilisation de SICLA par le " MENU " :

Taper " ? " pour obtenir de l'aide

Utilisation de SICLA par commandes directes :

à voir suivant l'ordinateur,sous MULTICS taper :

"help sicla"

pour obtenir la liste des commandes SICLA.

Assistance en ligne :

Il faut taper :

" ? " pour obtenir des précisions sur les questions posées.

" / " pour obtenir l'arrêt de la session SICLA ou le retour à la question précédente selon le module.

" .. " pour transmettre une chaîne de caractères au processeur de commandes sans sortir de SICLA.

2.3. ASSISTANCE EXTERIEURE

Un effort est fait pour aider et faciliter l'utilisation du système par un utilisateur peu formé en analyse des données.
Le système étant d'utilisation facile, l'assistance porte essentiellement sur les différents choix d'algorithmes, de codages ou de distances.
Actuellement ces différentes aides sont extérieures au système, ce qui permet de faciles extensions, suivant les remarques et réactions de l'utilisateur.
Le système expert CLAVECIN [6] peut être utilisé, Il aide au choix d'une stratégie de traitement des données.

2.4. EXEMPLE D'ENCHAINEMENT DES THEMES

SAISIE	GESTION
SELECTION	
DESCRIPTION ELEMENTAIRE	DECRIRE
CREATION DE NOUVELLES DONNEES	CODAGE
CHOIX DE METHODE	CLASSIFICATION | FACTORIELLE | HIERARCHIE | DISCRIMINATION
EXPLOITATION DES RESULTATS	INTERPRETATION
STOCKAGE DES RESULTATS	ARCHIVE
IMPRESSION DES RESULTATS	EDITION

3. TERMINOLOGIE UTILISEE

Pour dialoguer avec SICLA, l'utilisateur devra connaître le vocabulaire usuel de l'analyse des données.

Les données se présentent sous la forme d'un tableau contenant un ensemble d'observations (ou INDIVIDUS) sur lesquels ont été mesurés un certain nombre de paramètres (ou VARIABLES).

Le traitement des données par le logiciel nécessite la saisie, le contrôle et la validation d'informations concernant les VARIABLES (colonnes) et les INDIVIDUS (lignes) .

3.1. INFORMATIONS CONCERNANT LES VARIABLES

Les variables sont caractérisées par :

- Un ABREGE (identificateur abrégé) de 4 caractères maximum.

Le premier caractère doit être obligatoirement une lettre.

N.B. Il faut veiller à l'unicité des identificateurs abrégés aussi bien pour les variables et leurs modalités que pour les individus et se méfier des homonymies involontaires obtenues lorsque l'identificateur complet est tronqué à 4 caractères.

Lorsque le nombre de variables est important, une option permet la génération automatique des identificateurs sous la forme :

var1, var2,... , varN,.....

- Un LIBELLE (identificateur en clair) de 40 caractères maximum.

Ce libellé qui est facultatif, complète la description de la variable. Pour faciliter la lecture et l'interprétation des résultats, il est fortement conseillé de définir les libellés des variables.

- Un TYPE qui est un nombre compris entre 1 et 7 et qui précise un type de VARIABLE QUANTITATIVE (1,5,6,7) ou un type de VARIABLE QUALITATIVE (3,4)

3.1.1. ENSEMBLE DES VARIABLES QUANTITATIVES

Ce sont les variables de mesure (poids, âge, etc,...), de comptage (nombre de cadres supérieurs résidant dans une commune par exemple), ou de présence/absence (variable binaire en 0 et 1).

D'une manière générale, tout paramètre sur lequel existe une structure d'ordre peut être considéré comme une variable quantitative.

Les variables quantitatives peuvent comporter des valeurs manquantes qui doivent être repérées par un code (nécessairement entier) que l'on choisira différent des valeurs prises par les données et global (i.e. le même quelle que soit la variable).

On distingue quatre types de variables quantitatives :

a) Mesures (nombres réels) TYPE SICLA No 1
b) Notes (entiers compris entre deux bornes) TYPE SICLA No 5
c) Binaire (0 ou 1) TYPE SICLA No 6
d) Comptage (entiers positifs) TYPE SICLA No 7

Exemple de définition de VARIABLE QUANTITATIVE :

ABREGE	: math
LIBELLE	: note en mathématiques
TYPE	: 1
FOURCHETTE	: 0,20
CODE VALEUR MANQUANTE	: 999

3.1.2. ENSEMBLE DES VARIABLES QUALITATIVES

Une variable qualitative est une variable prenant un certain nombre d' ETATS ou MODALITES. Les modalités d'une variable qualitative peuvent être ordonnées ou non ordonnées.
On distingue trois types de variables qualitatives :

a) qualitatif non ordonné TYPE SICLA No 2
b) qualitatif ordonné TYPE SICLA No 4
c) texte TYPE SICLA No 3

N.B. Le TYPE SICLA no 3, celui des variables " TEXTE ", est en cours d'implantation. Les variables TEXTE sont des variables qualitatives dites ouvertes c.a.d. des variables pour lesquelles on ne connaît pas a priori l'ensemble des possibilités de réponses.

Exemples :

1) Variable de type 2 (qualitatif non ordonnè)

ABREGE : coul
LIBELLE : COULEURS
TYPE : 2
NOMBRE DE MODALITES : 4
LIBELLES DES MODALITES: BLANC NOIR ROUGE VERT

2) Variable de type 4 (qualitatif ordonnè)

ABREGE : telv
LIBELLE : regardez-vous la télévision?
TYPE : 4
NOMBRE DE MODALITES : 3
LIBELLE DES MODALITES : JAMAIS DE_TEMPS_EN_TEMPS SOUVENT
ORDRE DES MODALITES : JAMAIS < DE TEMPS EN TEMPS < SOUVENT

Pour traiter des variables qualitatives il faut connaître :

- le NOMBRE DE MODALITES dont le maximum est fixé à 99, mais certaines méthodes n'admettent pas plus de 26 modalités.

- les ABREGES, qui peuvent aussi être gènèrès automatiquement.

- le LIBELLE DE LA MODALITE : identificateur en clair de 40 caractéres maximum (ceci est facultatif).

3.2. INFORMATIONS CONCERNANT LES INDIVIDUS

Les individus sont caractèrisés par :

- un ABREGE de 4 caractéres maximum, ces identificateurs doivent être uniques, les homonymes ne sont pas admis. Les 4 premiers caractéres étant pris comme abrègès, il faut qu'ils soient différents pour chaque individu. De plus le premier caractére doit être alphanumérique.

Si les données ne comportent pas d'identificateurs, ceux-ci peuvent être gènèrès automatiquement sous la forme : in01, in02, in03, ..

- un LIBELLE de 40 caractéres (facultatif).

- les valeurs prises par les individus sur les variables. Ce sont LES DONNEES PROPREMENT DITES qui sont considèrées comme des rèels .

Le format de lecture pourra contenir des instructions FORTRAN [11] du type:

na1 : pour la lecture de l'identificateur alphanumérique de l'individu s'il existe ou pour la lecture des variables textes.
n doit être compris entre 1 et 40 ; si n > 4, les quatre premiers caractéres sont pris comme ABREGE.

e,f,g : pour la lecture des valeurs réelles.

Exemples : (4a1,10f4.0)
(12g12.5)

Mais il n'est pas indispensable de donner le format de lecture, c'est la cas de la lecture en format dit "libre" ou encore "variable" (voir paragraphes 4.1.1. et 4.1.6).

4. COMMENT UTILISER SICLA

Les commandes de SICLA sont regroupées en thémes (cf paragraphe 6.2) correspondant aux fonctions principales suivantes : ENREGISTRER, GERER, DECRIRE, ANALYSER, INTERPRETER, STOCKER, EDITER.

Rappel : Tout ce qui est fait par un utilisateur pendant une session est conservé dans un fichier "lishh_mm" dont le nom est donné en début de session.

Exemple :

```
**************************************************************
* SICLA              Nom_utilisateur      Date : 05/06/85  *
* S.D :                              Archive :             *
* Listing : lis16_10     Mode interactif                   *
**************************************************************
```

Il est trés conseillé de mettre frèquemment ce fichier à jour , en supprimant les rèsultats des essais infructueux, pour éviter qu'il ne dépasse les limites de taille permises.

4.1. COMMENT ENREGISTRER

4.1.1. LES DIFFERENTES PHASES D'ENREGISTREMENT PAR SICLA

L'enregistrement des données comporte DEUX PHASES distinctes :

PHASE 1 : Enregistrement des informations sur les variables

Il s'agit d'enregistrer le NOMBRE, les IDENTIFICATEURS, les LIBELLES et les TYPES des variables. Il existe sept types, trois qualitatifs, quatre quantitatifs. (cf. paragraphe 3.1.1 et 3.1.2.)

Les commandes à utiliser sont ENRV (Enregistrement des variables) ou ENRFIC (Enregistrement à partir d'un fichier).

Cette phase peut être réalisée :

- soit en CONVERSATIONNEL (ENRV) : le système Sicla interroge l' utilisateur qui doit fournir les différentes informations concernant les variables. Il faut noter que cette option risque d'être fastidieuse et sujette à erreur si le nombre d'informations à fournir est important, par exemple lorsque les données comportent beaucoup de variables.

- soit par COPIE DU DESCRIPTIF des variables d'une autre s.d. (ENRV).

- soit par la LECTURE D'UN DICTIONNAIRE (ENRFIC) qui est la transcription sous un format donné des différentes informations concernant les variables et qui doit être enregistré sur un fichier .

La réponse à la question "que désirez-vous faire ?" détermine le choix du mode de saisie du descriptif des variables.
Lorsque le nombre de variables est important (supérieur à 20 par exemple), il est fortement conseillé d'enregistrer, au moyen d'un éditeur de textes, les informations variables sous forme de dictionnaire (cf.paragraphe 4.1.2.).

N.B. A la fin de l'enregistrement du dictionnaire par ENRV, le système propose de générer un fichier dictionnaire au moyen de la commande ENRFIC. L'intérêt est de disposer d'un fichier modifiable par un éditeur de textes pour des traitements ultérieurs.

PHASE 2 : Enregistrement du tableau des données
(informations sur les individus)

La commande à utiliser est ENRI (Enregistrement des individus).

Les données se présentent comme un tableau rectangulaire (LIGNES x COLONNES) , les individus (LIGNES) pouvant comporter des identificateurs alphanumériques en tête et éventuellement un libellé de 40 caractères maximum.

Les données doivent être enregistrées sur un fichier et sont lues en réels.

Deux types de lecture sont possibles :

1. avec format "libre" ; on ne précisera pas le format à condition que les valeurs du tableau de données soient séparées par un blanc et que chaque ligne comporte au plus 80 caractères,

Exemple de format dit "libre" ou "variable" :

```
IND1 10 20 30
       40 50
IND2 20 20 60
       40 55
```

2. avec un format FORTRAN donné lorsque les conditions précédentes ne sont pas satisfaites.
Le format est alors nécessairement composé d'ordres FORTRAN du type :

f, g, e	pour les données proprement dites de type réel uniquement.
na1	pour les identificateurs éventuels des individus. (facultatif)

Exemples : (4a1,12f3.0)
(10a1,5f8.5)
(10e12.4)

Attention: n doit être compris entre 1 et 40 ; si n > 4 , les quatre premiers caractères sont pris comme ABREGE . Comme il a été signalé plus haut, il faut veiller à ce que cette troncature n'engendre pas des homonymies.

Deux autres commandes permettent l'enregistrement de données dans des cas particuliers :
ENRIT permet d'enregistrer des données mises sous forme transposée c.a.d. telles que les individus correspondent aux colonnes et les variables aux lignes du tableau. L'identificateur de la variable (4 caractères) peut être placé en tête pour contrôle, l'étape d'enregistrement des informations relatives aux variables (ENRV ou ENRFIC) doit évidemment être faite avant.
ENRDIS permet d'enregistrer un tableau de distances. Les données sont des valeurs réelles positives ou nulles recensées dans un tableau symétrique ayant n lignes et n colonnes ou n est le nombre d'individus.

Différents contrôles sont effectués (existence du fichier, adéquation du format, etc,...), lorsque la lecture peut se dérouler normalement les deux premiers enregistrements sont imprimés sur l'écran afin que l'utilisateur puisse en vérifier l'exactitude.

Remarque importante :

La phase 1 d'enregistrement du dictionnaire des variables n'est à faire qu'une seule fois ; en cas d'erreur dans la phase 2 de lecture des données, il n'est pas nécessaire de recommencer la phase 1 qui demeure valide.

En cas d'erreur dans la phase 1 il est possible de modifier, par la commande MODSDO, l'identificateur et le libellé.

Par contre MODSDO ne permet pas de modifier le type statistique des variables ou encore le nombre de modalités. Si l'enregistrement des données est déjà achevé on peut utiliser, les commandes CREQAN ou CREQAL qui crèent des variables de type quantitatif ou qualitatif.

La commande CREQAL permet,par exemple, de diminuer le nombre de modalités par regroupement.

Pour tous les autres cas , il faut refaire l'enregistrement des variables ou modifier le dictionnaire. (voir le résumé des différentes possibilites au paragraphe 4.1.5)

La commande GENFIC peut servir à générer le dictionnaire des variables sur un fichier séquentiel avec format. Un éditeur de textes permet de modifier un tel fichier et par suite des modifications du dictionnaire sont possibles.

La commande MODSDO permet également de corriger certaines erreurs de la phase 2.

A la fin de l'enregistrement des données est crée, sous le nom : <nom_des_données>.sdo, la structure de données standard (en abrégé s.d.) contenant toutes les informations, sur laquelle opéreront les différents modules du système SICLA .

4.1.2. ENREGISTREMENT DU DICTIONNAIRE ABREGE

Il s'agit d'enregistrer les informations sur les variables, au moyen d'un éditeur de textes, selon un format précisé ci-dessous. Le dictionnaire doit contenir, pour chaque variable, un enregistrement présenté comme suit :

```
Col  1 - 2 : code = 00
Col  3 - 3 : blanc
Col  4 - 7 : Identificateur abrégé en 4 caractères
Col  8 - 8 : blanc
Col  9 -48 : libellé de la variable en 40 caractères
Col 49 -49 : blanc
Col 50 -51 : type statistique (1 à 7)
Col 52 -52 : blanc
Col 53 -54 : Nombre p des modalités (1 à 99)
             Pour une variable qualitative p est supérieur ou
             égal à 2.
             Si  p = 1 , il s'agit d'une variable quantitative.
```

Pour les variables quantitatives on peut faire suivre le premier enregistrement (codé 00) par un deuxième facultatif (codé 01) sur laquelle seront enregistrées les bornes de validité des valeurs de cette variable.

Enregistrement des "bornes" (pour une variable quantitative)

```
Col 1 - 2 : code = 01
Col 3 - 3 : blanc
Col 4 - ..: 2 nombres REELS séparés par une VIRGULE
```

Exemple : Dans le fichier ci-dessous, sur la première ligne est enregistrée une variable quantitative var1 de type 1 ; sur la ligne suivante sont enregistrées les bornes de validité de cette variable à valeurs comprises entre 0.5 et 20.

```
    12 4567 9.....................................48.51.54
    ____________________________________________________
  /                                                    |
 /  00 var1 ********libelle_de_la_variable**********  1 01   |
|   01 0.5,20                                         |
|                                                     |
 ____________________________________________________
```

Pour les variables qualitatives, on peut faire suivre le premier enregistrement (codé 00) de p enregistrements relatifs aux p modalités. Ces enregistrements ne sont pas obligatoires, en cas d'absence ils sont générés automatiquement.

Enregistrement "modalités" (pour une variable qualitative)

Col 1 - 2 : numéro de la modalité
Col 3 - 3 : blanc
Col 4 - 7 : identificateur abrégé de la modalité
Col 8 - 8 : blanc
Col 9 -48 : libellé de la modalité en 40 caractères

Exemple :

Dans le fichier ci-dessous, le 1er enregistrement (00) représente une variable qualitative à trois modalités .
Les trois enregistrements suivants (01, 02, 03) sont relatifs à chacune de ces modalités.

```
    12 4567 9.....................................48.51.54
    ____________________________________________________
  /                                                    |
 /  00 var1 ********libelle_de_la_variable**********  2 03 |
|   01 mod1 ******libelle_de_la_modalite_1**********       |
|   02 mod2 ******libelle_de_la_modalite_2**********       |
|   03 mod3 ******libelle_de_la_modalite_3**********       |
|                                                          |
 ----------------------------------------------------------
```

Si les trois derniéres lignes sont absentes, le système génére automatiquement le nom des modalités en tronquant le nom de la variable à 3 caractéres et en écrivant 1, 2, 3, etc.. à droite de ces 3 caractéres.
Exemple : si le nom de la variable est AVOR, les identificateurs des modalités générés sont AVO1,AVO2,AVO3,...
Il est donc impossible dans ce cas d'utiliser des noms de variables comme var1, var2, etc..sinon il y aura homonymie entre variable et modalité de cette variable.

4.1.3. EXEMPLE ENREGISTREMENT PAR MENU ET DICTIONNAIRE

Il faut taper "SICLA", si l'enregistrement des données n'a pas été effectué, les commandes ENRV et ENRI se succédent, puis le traitement peut être enchaîné automatiquement. Dans l'exemple ci-dessous l'enregistrement des variables est fait à partir d'un dictionnaire généré au moyen d'un éditeur de textes (cf. paragraphe 4.1.2.) puis on enregistre les données et ensuite deux commandes de description des variables sont exécutées.

N.B. Dans les exemples les réponses de l'utilisateur sont signalées par le symbole " ► ".

```
► SICLA

   creation du fichier profil : sicla.profil

   ************************************************************************
   * SICLA                Ok                            Date : 21/10/86    *
   * S.D :                              Archive :                          *
   * Listing : lis15_55                 Mode interactif                    *
   ************************************************************************
   * SICLA              Commande SICLA                Version Multics 1.0 *

   entrez le nom de la structure de données choisie
   ------------------------------------------------

► micro

   ---------> fichier : micro.sdo

   ce sont des nouvelles données à enregistrer ?
   ----------------------------------------------

► oui
   ************************************************************************
   * données : micro.sdo      individus :   0  *  variables :   0          *
   * ---------                ----------        ----------                  *
   *                          quantitatives :  0   qualitatives :  0       *
   ************************************************************************

   que desirez-vous faire ?
   ------------------------

   1 : saisie du descriptif en conversationnel
    (déconseillé si nombre important de variables)

   2 : saisie du descriptif à partir d une autre s.d.

   3 : saisie du descriptif à partir d'un fichier dictionnaire
    (conseillé si nombre important de variables)
► 3
```

```
   entrez un titre pour les données :
   ----------------------------------
► Performances des micro-ordinateurs

   entrez le nom du fichier dictionnaire
   -------------------------------------
► micro.di

   entrez un majorant du nombre total de variables traitées :
   ----------------------------------------------------------
► 90

   entrez un majorant du nombre total de modalites traitées :
   ----------------------------------------------------------
► 90

   desirez-vous une impression du dictionnaire pour controle ?
   -----------------------------------------------------------
► oui

     1. variable : coul : moniteur couleur                    type : 2
        modalites =  3
    1.modalite :cou1  absent
    2.modalite :cou2  optionnel
    3.modalite :cou3  present

   ..........................................................................
   ..........................................................................
   ....................................

    10. variable : flop : nombre d'unites de disquettes      type : 1
        modalites =  1
   ..........................................................................
   ..........................................
    14. variable : mass : taille de la memoire de masse      type : 1
        modalites =  1


   nombre total de variables :   14
   --------------------------------


   variables quantitatives    :   5
   variables qualitatives     :   9
   nombre total de modalites  :  26


   fin de lecture du dictionnaire

   les variables quantitatives comportent-elles des valeurs manquantes ?
   ---------------------------------------------------------------------
► non
```

```
*****************************************************************
* nu  * id   * typs * mod *  libelles  des variables           *
*****************************************************************
*   1 * coul     2      3 * moniteur couleur                     *
*   2 * CP/M     2      3 * systeme d'exploitation CP/M          *
*   3 * MSDO     2      3 * systeme d'exploitation MS-DOS        *
*   4 * AUTR     2      3 * autres systemes d'exploitation       *
*   5 * proc     2      3 * type de microprocesseur              *
*   6 * para     2      3 * interface parallele                  *
*   7 * seri     2      3 * interface serie                      *
*   8 * IEEE     2      3 * interface IEEE 488                   *
*   9 * dur      2      2 * disque dur                           *
*  10 * flop     1      1 * nombre d'unites de disquettes        *
*  11 * prix     1      1 * prix du micro-ordinateur             *
*  12 * ram1     1      1 * taille de la memoire centrale        *
*  13 * ram2     1      1 * taille maximale de la memoire centrale*
*  14 * mass     1      1 * taille de la memoire de masse        *
*****************************************************************

fin de l enregistrement du descriptif des variables

que desirez-vous faire ?
------------------------

1. modifier le descriptif des variables
2. passer á la phase d'enregistrement des données relatives aux individus
► 2

nombre maximum d individus ?
----------------------------

► 200

nombre de caracteres de l identificateur des individus ?
---------------------------------------------------------

► 16

Type de format de lecture des donnees :
---------------------------------------

1 : lecture des donnees a l'aide d'un format FORTRAN
2 : lecture des donnees en format variable (faire ? pour plus de precisions)
► ?

Utiliser le format variable, si les données sont sèparées par des blancs ;
les données doivent être sur des enregistrements de longueur maximum
80 caractéres.
Une donnée peut comporter en dèbut de ligne l'identificateur ou libellé
de l'individu.

Type de format de lecture des donnees :
---------------------------------------

1 : lecture des donnees a l'aide d'un format FORTRAN
2 : lecture des donnees en format variable (faire ? pour plus de precisions)
```

```
► 1
   entrez le nom du fichier sequentiel contenant les données :
   ____________________________________________________

► micro.do

   image du premier enregistrement de votre fichier
   ____________________________________________________

  0        1         2         3         4         5         6         7
  12345678901234567890123456789012345678901234567890123456789012345678901234567890
  ..........................................................................
    1 PAP            1131222211 20000  192  512  720

   les ordres formats acceptables sont :
   (n)*a1 pour l identificateur de l individu
   f,g,e pour la lecture des valeurs réelles ,i est interdit

   entrez le format fortran de lecture (2 lignes maxi)
   ____________________________________________________

   (ex (4a1,10f3.0) )
► (4x,16a1,10f1.0,f6.0,3f5.0)

   impression de controle pour le premier individu :
   ____________________________________________________

    (format d impression g7.2)

   individu :  PAP
  1.0     1.0     3.0     1.0     2.0     2.0     2.0     2.0     1.0     1.0
   .20E+05 .19E+03 .51E+03 .72E+03

   individu :  AX2
  1.0     1.0     3.0     1.0     2.0     3.0     2.0     1.0     1.0     1.0
   .20E+05 .13E+03 .10E+04 .32E+03

    ——> ok ?
► oui

   ——>fin de lecture des données :     169 objets


   ***************************************************************
   * nu  * id  * typs * mod *  libelles  des variables           *
   ***************************************************************
   *   1 * coul    2      3 * moniteur couleur                    *
   *   2 * CP/M    2      3 * systeme d'exploitation CP/M         *
   *   3 * MSDO    2      3 * systeme d'exploitation MS-DOS       *
   *   4 * AUTR    2      3 * autres systemes d'exploitation      *
   *   5 * proc    2      3 * type de microprocesseur             *
   *   6 * para    2      3 * interface parallele                 *
   *   7 * seri    2      3 * interface serie                     *
   *   8 * IEEE    2      3 * interface IEEE 488                  *
```

```
*   9 * dur     2      2 * disque dur                                  *
*  10 * flop    1      1 * nombre d'unites de disquettes               *
*  11 * prix    1      1 * prix du micro-ordinateur                    *
*  12 * ram1    1      1 * taille de la memoire centrale               *
*  13 * ram2    1      1 * taille maximale de la memoire centrale      *
*  14 * mass    1      1 * taille de la memoire de masse               *
**********************************************************************

entrer une commande :                (aide : ?  fin : /)
```

► ?

```
 les commandes sont decrites dans les themes :

1. gestion          2. decrire      3. codage         4. manipulation
5. classification   6. factoriel    7. hierarchies    8. discrimination
9. interpretation  10. archive     11. edition       12. divers

rang du theme pour informations ? (fin= / liste= ?)
```

► 2

```
liste des commandes relatifs au theme : decrire


1.DESQAN description de variables qualitatives
2.DESQAL description de variables quantitatives
3.AKHI2  calcul de tableaux de contingence
4.HISTO  histogrammes
5.CORREL calcul de la matrice des corrélations
6.COURBE Edition simultanée de courbes
7.BIPLOT Calcul des croisements en variables quantitatives

rang du theme pour informations ? (fin= / liste= ?)
```

► /

```
entrer une commande :                (aide : ?  fin : /)
```

On peut exécuter à la suite de l'enregistrement les commandes DESQAN et DESQAL pour une description élémentaire des données.(voir exemple paragraphe 4.3.3)

4.1.4. ENREGISTREMENT CONVERSATIONNEL PAR COMMANDE DIRECTE

NdR : les astérisques (*),(**),(***), renvoient à une note concernant les erreurs en fin de paragraphe.

```
► ENRV

**********************************************************************
* SICLA                Ok                        Date : 07/10/86     *
* S.D :                          Archive :                           *
* Listing : lis17_31             Mode interactif                     *
**********************************************************************
* SICLA              Commande ENRV              Version 21/09/85 *

          enregistrement du descriptif des variables

   entrez le nom de la structure de données choisie :

► napoli

   ————————> fichier : napoli.sdo
   **********************************************************************
   * données : napoli.sdo   individus :   0  *  variables :   0       *
   * ————————               ————————          ————————              *
   *                      quantitatives :  0   qualitatives :  0      *
   **********************************************************************

   que desirez-vous faire ?

   1 : saisie du descriptif en conversationnel
    (déconseillé si nombre de variables important)

   2 : saisie du descriptif à partir d une autre s.d.
  moniteur
   3 : saisie du descriptif à partir d'un fichier dictionnaire
    (conseillé si nombre de variables important)
► 1


   ——> attention la reponse / à une question annulle la session

   ——> en cas d'erreur poursuivre la session jusqu'au bout, il sera toujours
       possible de rectifier les informations par un autre module(sauf si
       l'erreur porte sur le type ou le nombre de modalites des variables)

   entrez un titre pour les données :

► Données médicales (Naples 1990)
```

```
 nombre total exact de variables ?
 ---------------------------------
► 12

 desirez-vous definir des fourchettes de controles ?
 ---------------------------------------------------
► ?

 ce sont des valeurs min et max entre lesquelles doivent etre comprises
 les variables quantitatives

 desirez-vous definir des fourchettes de controles ?
 ---------------------------------------------------
► non

 %% identificateur, libelle, type de la variable numero : 1 ?
 ----------------------------------------------------------
► type,type à priori,2

 desirez-vous definir les libelles des modalites ?
 -------------------------------------------------
► oui

 nombre de modalites de la variable type ?
 -----------------------------------------
► 4

 libelle de la modalite numero : 1 identificateur genere:typ1
 ------------------------------  ---------------------------
► bronchite                                                (*)

 libelle de la modalite numero : 2 identificateur genere:typ2
 ------------------------------  ---------------------------
► ?                                                        (**)

 exemple : modal,   libelle de modalite
           I--->identificateur  I----->libelle

 libelle de la modalite numero : 2 identificateur genere:typ2
 ------------------------------  ---------------------------
► ENFI,enfiseme

 libelle de la modalite numero : 3 identificateur genere:typ3
 ------------------------------  ---------------------------
► ENbr,enfiseme dominante

 libelle de la modalite numero : 4 identificateur genere:typ4
 ------------------------------  ---------------------------
► BRen,bronchite dominante
```

```
   %% identificateur, libelle, type de la variable numero : 2 ?
   ------------------------------------------------------------
► gaw,resistance à l'inspiration1

   vous avez oublie la 2e virgule et le type

   %% identificateur, libelle, type de la variable numero : 2 ?
   ------------------------------------------------------------
► gaw,resistance à l'inspiration,1                          (***)

   ..............................................................

   ..........Enregistrement des variables 2 à 11................

   ..............................................................

   %% identificateur, libelle, type de la variable numero :12 ?
   ------------------------------------------------------------
► pco2,echange en gaz carbonique,1

   les variables quantitatives comportent-elles des valeurs manquantes ?
   ---------------------------------------------------------------------
► oui

   code entier des valeurs manquantes ?
   ------------------------------------
► 999
```

nu	id	typs	mod	libelles des variables
1	type	2	4	*type a priori
2	gaw	1	1	*resistance a l'inspiration
3	sgaw	1	1	*resistance a l'inspiration 2
4	vems	1	1	*capacite respiratoire 1
5	vemt	1	1	*capacite respiratoire 2
6	vecv	1	1	*capacite respiratoire 3
7	mef5	1	1	*capacite respiratoire 4
8	mef2	1	1	*cpacite respiratoire 5
9	tcl	1	1	*dilatation pulmonaire 1
10	vrtl	1	1	*dilatation pulmonaire 2
11	po2	1	1	*echange en oxygene
12	pco2	1	1	*echange en gaz carbonique

```
   fin de l enregistrement du descriptif des variables

   ——> executer la commande MODSDO si vous desirez modifier le descriptif
```

```
——> executer la commande ENRI pour enregistrer les données

=== Fin de ENRV          ===

STOP
```

Remarques :

L'utilisateur a fait quelques erreurs en cours d'enregistrement. Certaines de ces erreurs sont immédiatement détectées par le système et la correction est exigée pour pouvoir poursuivre ; par exemple l'oubli du type de la variable.
Par contre les erreurs de déclaration de nom et de libellé ne peuvent être remarquées que par l'utilisateur lui-même.
Par exemple :
(*) Croyant que l'identificateur est obligatoirement par défaut, il ne l'a pas donné. Sicla a généré un identificateur "typ1" et enregistré le libellé "bronchite".

(**) Ayant un doute, l'utilisateur s'informe et constate que les identificateurs des modalités ne sont pas nécessairement générés par le système. Souhaitant remplacer "typ1" par "BRON", il prévoit donc une correction pour la 1ère modalité.

(***) Erreur dans le libellé de la 2ème modalité.

Ces corrections vont être faites au moyen de MODSDO.

```
► MODSDO
   ************************************************************************
   * SICLA                Ok                          Date : 07/10/86     *
   * S.D : napoli.sdo                Archive :                            *
   * Listing : lis17_31              Mode interactif                      *
   ************************************************************************
   * SICLA              Commande MODSDO            Version 04/03/85       *
   ************************************************************************
   * données : napoli.sdo  individus :   0  *  variables :  12            *
   * ————————              ——————————          ——————————                *
   *                       quantitatives : 11   qualitatives :  1         *
   ************************************************************************

            modification de la structure de données
            ———————————————————————————————————————

    votre choix ?

   1.modifier les informations gènèrales relatives aux données
   2.modifier les informations relatives aux variables
   3.modifier les informations relatives aux individus
► 1

   ancien titre :
   ——————————————

► Données médicales (Naples 1990)

   nouveau titre ?
   ———————————————

► Données médicales (Naples 1980)

    votre choix ?

   1.modifier les informations gènèrales relatives aux données
   2.modifier les informations relatives aux variables
   3.modifier les informations relatives aux individus
► 2


   ************************************************************************
   * nu  * id  * typs * mod *  libelles  des variables                     *
   ************************************************************************
   *   1 * type    2      4 *type a priori                                 *
   *   2 * gaw     1      1 *resistance a l'inspiration                    *
   *   3 * sgaw    1      1 *resistance a l'inspiration 2                  *
   *   4 * vems    1      1 *capacite respiratoire 1                       *
   *   5 * vemt    1      1 *capacite respiratoire 2                       *
   *   6 * vecv    1      1 *capacite respiratoire 3                       *
```

```
*   7 * mef5    1      1 *capacite respiratoire 4              *
*   8 * mef2    1      1 *cpacite respiratoire 5               *
*   9 * tcl     1      1 *dilatation pulmonaire 1              *
*  10 * vrtl    1      1 *dilatation pulmonaire 2              *
*  11 * po2     1      1 *echange en oxygene                   *
*  12 * pco2    1      1 *echange en gaz carbonique            *
*****************************************************************

 rang ou identificateur de la variable à modifier ? (fin=/)
 ----------------------------------------------------------

► type

  1. changer l identificateur    = type
  2. changer le libelle          = type a priori
  3. pour changer le type statistique    2
  4. changer les identificateurs des modalites
  5. changer les libelles des modalites
  6. changer la valeur pour un individu
  7. pour changer le nombre de modalites :  4

  votre choix ? (fin = /)
► 4
   1.                                    typ1
   2.                                    ENFI
   3.                                    ENbr
   4.                                    BRen

   rang de l identificateur à changer ? (fin=/)
► 1

  entrez le nouvel identificateur :
► BRON
   1.                                    BRON
   2.                                    ENFI
   3.                                    ENbr
   4.                                    BRen

   rang de l identificateur à changer ? (fin=/)
► /

  1. changer l identificateur    = type
  2. changer le libelle          = type a priori
  3. pour changer le type statistique    2
  4. changer les identificateurs des modalites
  5. changer les libelles des modalites
  6. changer la valeur pour un individu
  7. pour changer le nombre de modalites :  4

  votre choix ? (fin = /)
► /
```

```
 rang ou identificateur de la variable á modifier ? (fin=/)
 ──────────────────────────────────────────────────────────
► 2

 1. changer l identificateur    = gaw
 2. changer le libelle          = resistance á l'inspiration
 3. pour changer le type statistique    1
 4. changer les identificateurs des modalites
 5. changer les libelles des modalites
 6. changer la valeur pour un individu
 7. pour changer le nombre de modalites :  1

 votre choix ? (fin = /)
► 2

 entrez le nouveau libelle :
resistance á l'inspiration 1

 1. changer l identificateur    = gaw
 2. changer le libelle          = resistance á l'inspiration 1
 3. pour changer le type statistique    1
 4. changer les identificateurs des modalites
 5. changer les libelles des modalites
 6. changer la valeur pour un individu
 7. pour changer le nombre de modalites :  1

 votre choix ? (fin = /)
► /

 rang ou identificateur de la variable á modifier ? (fin=/)
 ──────────────────────────────────────────────────────────
► /

 ═══ Fin de MODSDO      ═══

STOP fin normale
```

L'enregistrement des variables étant correct, on peut poursuivre en enchatnant la phase 2 d'enregistrement des données en tapant ENRI.

```
► ENRI
   ************************************************************************
   * SICLA                Ok                          Date : 07/10/86     *
   * S.D : napoli.sdo                Archive :                            *
   * Listing : lis17_31              Mode interactif                      *
   ************************************************************************
   * SICLA              Commande ENRI                Version 27/11/85     *

             enregistrement des données
             ──────────────────────────

   ************************************************************************
   * données : napoli.sdo  individus :   0  *  variables : 2              *
   * ────────              ──────────        ──────────                   *
   *                       quantitatives : 11   qualitatives :  1         *
   ************************************************************************

   nombre maximum d individus ?
   ────────────────────────────
► 1000

   nombre de caracteres de l identificateur des individus ?
   ────────────────────────────────────────────────────────
► 4
   Type de format de lecture des donnees :
   ───────────────────────────────────────

   1 : lecture des donnees a l'aide d'un format FORTRAN
   2 : lecture des donnees en format variable (faire ? pour plus de precisions)
► ?

   Utiliser le format variable, si les données sont séparées par des blancs ;
  les données doivent être sur des enregistrements de longueur maximum
  80 caractères.
   Une donnée peut comporter en début de ligne l'identificateur ou libellé
  de l'individu.

   Type de format de lecture des donnees :
   ───────────────────────────────────────

   1 : lecture des donnees a l'aide d'un format FORTRAN
   2 : lecture des donnees en format variable (faire ? pour plus de precisions)

► 1
   entrez le nom du fichier sequentiel contenant les données :
   ───────────────────────────────────────────────────────────
► >udd>Clorec>Sicla>d>napoli.do

   image du premier enregistrement de votre fichier
   ────────────────────────────────────────────────

  0       1         2         3         4         5         6         7
  12345678901234567890123456789012345678901234567890123456789012345678901234567890
```

```
.....
001 3. 0.185 0.035  1.45 60.80 61.70  0.80  0.25 145.00 65.30 80.00 40.00

 les ordres formats acceptables sont :
 (n)*a1 pour l identificateur de l individu
 f,g,e pour la lecture des valeurs réelles ,i est interdit

 entrez le format fortran de lecture (2 lignes maxi)
 ---------------------------------------------------

 (ex (4a1,10f3.0) )
► (4a1,f2.0,2f6.3,5f6.2,f7.2,3f6.2)

 impression de controle pour le premier individu :
 -------------------------------------------------

  (format d impression g7.2)

 individu : 001
3.0     .19     .35E-01 1.5     61.     62.     .80     .25     .15E+03 65.
80.     40.

 individu :   2
3.0     .19     .46E-01 1.2     61.     43.     .50     .20     .12E+03 53.
66.     35.

  ----> ok ?
► oui

 ---->fin de lecture des données :     211 objets

 ********************************************************************
 * nu  * id  * typs * mod *  libelles  des variables                 *
 ********************************************************************
 *   1 * type    2      4 *type a priori                              *
 *   2 * gaw     1      1 *resistance a l'inspiration 1               *
 *   3 * sgaw    1      1 *resistance a l'inspiration 2               *
 *   4 * vems    1      1 *capacite respiratoire 1                    *
 *   5 * vemt    1      1 *capacite respiratoire 2                    *
 *   6 * vecv    1      1 *capacite respiratoire 3                    *
 *   7 * mef5    1      1 *capacite respiratoire 4                    *
 *   8 * mef2    1      1 *cpacite respiratoire 5                     *
 *   9 * tcl     1      1 *dilatation pulmonaire 1                    *
 *  10 * vrtl    1      1 *dilatation pulmonaire 2                    *
 *  11 * po2     1      1 *echange en oxygene                         *
 *  12 * pco2    1      1 *echange en gaz carbonique                  *
 ********************************************************************

 fin de l'enregistrement des données

 ----> il vous est possible (et meme conseillé)
de faire une sauvegarde par la commande GENFIC
```

```
 ═══ Fin de ENRI          ═══
STOP fin normale
```

1ere remarque :

a) L'utilisateur peut choisir de lire les données en format "libre" ; mais les données n'étant pas enregistrées correctement pour ce type de lecture ,on aura alors la séquence suivante :

```
 Type de format de lecture des donnees :
 ---------------------------------------

 1 : lecture des donnees a l'aide d'un format FORTRAN
 2 : lecture des donnees en format variable (faire ? pour plus de precisions)

► 2
 Impression des individus pour controle
 --------------------------------------

  (format d'impression g7.2)

 ERREUR lors du traitement de la valeur No  1.

 La zone traitee est  :  PAP
 La derniere valeur correcte lue est la     0 ieme : .00E+00
 Caractere non numerique pour l' element No  1

 L'enregistrement traite est relatif a l'individu numero :     1

 ENREGISTREMENT des donnees INACHEVE : recommencez

  ——> reprendre l'etape enregistrement des informations des variables
      si l'erreur est due a une mauvaise definition
      du type ou du nombre de modalites de variable


 ********************************************************************************
 * SICLA                Commande SICLA                Version Multics 1.0 *

 ═══ Fin Anormale de SICLA       ═══

STOP
```

b) Par contre si l'enregistrement des données correspond bien au type de lecture format "libre" on aura la séquence suivante :

```
 Type de format de lecture des donnees :
 ---------------------------------------

 1 : lecture des donnees a l'aide d'un format FORTRAN
 2 : lecture des donnees en format libre (faire ? pour plus de precisions)
```

```
► 2
  entrez le nom du fichier sequentiel contenant les donnees :
  ___________________________________________________________

► vaa

  Impression des individus pour controle
  ______________________________________

   (format d'impression g7.2)

  individu :     1
  ________

  individu : aaaa
1.0     2.0     3.0     4.0

  individu :     2
  ________

  individu : bbbb
8.0     7.0     1.0     1.0

  ——->fin de lecture des donnees :       2 objets


*****************************************************************************
* nu  * id  * typs * mod *  libelles  des variables                         *
*****************************************************************************
*   1 * a       1      1 *a                                                  *
*   2 * b       1      1 *b                                                  *
*   3 * c       1      1 *c                                                  *
*   4 * d       1      1 *d                                                  *
*****************************************************************************

  fin de l'enregistrement des donnees

  ___________________________________________________________

  2 eme remarque :
  L'utilisateur décide de conserver le dictionnaire des variables
  et de réenregistrer le fichier des données en modifiant son
  format en appelant GENFIC.
  ___________________________________________________________
```

```
► GENFIC
   ************************************************************************
   * SICLA                 Ok                          Date : 07/10/86    *
   * S.D : napoli.sdo                Archive :                            *
   * Listing : lis17_31              Mode interactif                      *
   ************************************************************************
   * SICLA              Commande GENFIC              Version 16/05/85     *
   ************************************************************************
   * données : napoli.sdo      individus : 211  *  variables :  12        *
   * ————————                  ————————————     ————————————            *
   *                           quantitatives : 11   qualitatives : 1      *
   ************************************************************************

   entrez un nom pour le fichier dictionnaire
   ——————————————————————————————————————————

► napoli.di

   entrez un nom pour le fichier données
   —————————————————————————————————————

►napoli.do

   entrez un format d'ecriture type FORTRAN
   ————————————————————————————————————————

   exemple :     (a4,10f5.2)
►(a4,12f7.3)

   ═══ Fin de GENFIC       ═══

  STOP fin normale
   Le fichier crèé est édité à la page suivante.
```

Dictionnaire imprimé par GENFIC : fichier napoli.di

00	type	type a priori	2	4
1	BRON	bronchite		
2	ENFI	enfiseme		
3	ENbr	enfiseme dominante		
4	BRen	bronchite dominante		
00	gaw	resistance a l'inspiration 1	1	1
00	sgaw	resistance a l'inspiration 2	1	1
00	vems	capacite respiratoire 1	1	1
00	vemt	capacite respiratoire 2	1	1
00	vecv	capacite respiratoire 3	1	1
00	mef5	capacite respiratoire 4	1	1
00	mef2	capacite respiratoire 5	1	1
00	tcl	dilatation pulmonaire 1	1	1
00	vrtl	dilatation pulmonaire 2	1	1
00	po2	echange en oxygene	1	1
00	pco2	echange en gaz carbonique	1	1

4.1.5. CORRECTION DES ERREURS

1. Erreurs d'enregistrement du dictionnaire des variables

Le moyen le plus efficace est d'exécuter la commande GENFIC, après avoir terminé l'enregistrement ; puis de corriger, au moyen d'un éditeur de textes le fichier dictionnaire <nom>.di.
S'il s'agit d'erreurs de définition des différents identificateurs ou libellés des variables ou modalités, les rectifications peuvent se faire au moyen de MODSDO.
MODSDO ne permet pas de corriger une mauvaise définition du nombre de modalités d'une variable qualitative. Une telle erreur est récupérable de la manière suivante :

- nombre sous-évalué, exécuter GENFIC et modifier le dictionnaire au moyen d'un éditeur de textes.
- nombre surévalué, exécuter ELIMOD.

Si les bornes ont été mal définies, la correction se fait également par GENFIC.

2. Erreurs d'enregistrement des données

La lecture du tableau des données ne peut être exécutée dans les cas suivants :

- accès au fichier impossible ;
- fichier vide
- format inadéquat (caractères alphanumériques dans des données déclarées réelles par exemple)

Les remèdes sont évidents : vérification des accès, du fichier, du format.

Des erreurs de type logique peuvent également rendre la lecture des données impossible : les valeurs lues n'appartiennent pas au domaine de définition de la variable. Ce domaine étant déterminé par le type et les bornes déclarées, il faut vérifier les données d'une part et le dictionnaire des variables pour localiser l'erreur, puis effectuer les corrections .Lorsqu'il s'agit de corriger le dictionnaire on utilisera GENFIC (voir ci-dessus)

4.1.6. LES COMMANDES D'ENREGISTREMENT

4.1.6.1 ENRDIS

ENRDIS permet d'enregistrer un tableau de distances. Les données sont des valeurs réelles positives ou nulles recensées dans un tableau symétrique ayant n lignes et n colonnes, n étant le nombre d'individus.

Chaque individu peut être caractérisé par un identificateur abrégé et un libellé placé alors en tête de chaque enregistrement.

Deux types de lecture sont possibles :

1. avec un format FORTRAN
2. sans préciser le format à condition que les différentes valeurs soient séparées par des blancs et que les enregistrements ne dépassent pas 80 caractères.

Exemple de format "libre" :

```
IND1 0 20 30
IND2 20 0 40
IND3 30 40 0
.............
```

Les PARAMETRES d'utilisation de la commande sont :

- Le nombre réel d'individus qui correspond au nombre de valeurs réelles à lire par enregistrement

- Le nombre total de caractères permettant d'identifier un individu ; les quatre premiers caractères sont considérés comme l'identificateur abrégé.

- Le type de format de lecture :

  1 : fortran
  2 : "libre" avec contrôle de l'identificateur variable-ligne
  3 : "libre" sans contrôle

- Le nom du fichier séquentiel contenant les données

- Le format fortran éventuel

4.1.6.2 ENRFIC

ENRFIC permet d'enregistrer le dictionnaire des variables relatif à un tableau (individus x variables) enregistré sous une forme standard sur un fichier.

Les PARAMETRES d'utilisation de la commande sont :

- un nom pour la structure de données
- un titre pour les données
- le nom du fichier contenant le dictionnaire des variables
- un majorant du nombre des variables à saisir
- un majorant du nombre des modalités à saisir
- la demande éventuelle d'édition du dictionnaire pour contrôle
- l'indication de la présence ou non de données manquantes
- le code éventuel des données manquantes

S'il existe des variables ou des modalités ayant des identificateurs identiques (homonymies), l'enregistrement du dictionnaire n'est pas possible.

4.1.6.3 ENRI

ENRI permet d'enregistrer un tableau de données (Individus x Variables) ; les individus correspondent aux lignes et les variables aux colonnes du tableau. Chaque individu (ligne) peut posséder un identificateur de 4 caractères et un libellé de 40 caractères placés en tête de l'enregistrement. Les données sont lues ,soit selon un format fortran, soit sans spécifier le format si les valeurs sont séparées par des blancs (format dit "libre"). L'étape d'enregistrement des informations relatives aux variables (ENRV ou ENRFIC) doit évidemment être effectuée avant.

Les PARAMETRES d'utilisation de la commande sont :

- un majorant du nombre d'individus
- le nombre total de caractères de l'identificateur des individus
- le nom du fichier séquentiel contenant les données
- le type de format de lecture
  1 format fortran

2 format libre
- le format fortran éventuel.

Les principales ERREURS sont les suivantes :

- erreur dans l'accès au fichier
- erreur dans le format de lecture du fichier
- erreur dans les bornes définies pour les variables

On peut voir un exemple dans le paragraphe 4.1.4.

4.1.6.4 ENRIT

ENRIT permet d'enregistrer des données mises sous forme transposée, c.a.d. telles que les individus correspondent aux colonnes et les variables aux lignes du tableau. L'identificateur de la variable (4 caractéres) peut être placé en tête pour contrôle.

Deux types de lecture sont possibles avec format fortran ou format libre (voir paragraphes 4.1.1 et 4.1.6.1.).

L'étape d'enregistrement des informations relatives aux variables (ENRV ou ENRFIC) doit évidemment être effectuée avant.

Les PARAMETRES d'utilisation de la commande sont :

- le nombre réel d'individus qui correspond au nombre de valeurs réelles à lire par enregistrement
- le type de format de lecture :
  1 : format fortran
  2 : format libre avec contrôle de l'identificateur des lignes
  3 : format libre sans contrôle
- le format fortran éventuel
- le nom du fichier séquentiel contenant les données .

4.1.6.5 ENRV

ENRV permet d'enregistrer le dictionnaire des variables relatif à un tableau (Individus x Variables) en conversationnel.

Les PARAMETRES d'utilisation de la commande sont :

- le nom pour la s.d.
- le choix du mode de saisie du dictionnaire
  1 en conversationnel
  2 à partir du dictionnaire d'une autre s.d.
  3 à partir d'un dictionnaire stocké sur un fichier
- le titre pour les données.

- le nombre exact des variables à saisir
- les bornes de contrôle pour les variables (facultatif)
- pour chaque variable :
  l'identificateur abrégé
  le libellé
  le type statistique
- le nombre de modalités pour les variables qualitatives
- les identificateurs et libellés des modalités éventuellement
  Un identificateur des modalités est généré par défaut à partir de l'identificateur de la variable correspondante mais il peut être redéfini.Si deux variables ont des libellés de modalités identiques (oui/non par exemple), il suffit de faire référence à la première variable qui doit alors avoir un nombre de modalités supérieur ou égal à la seconde (voir l'exemple ci-dessous).
- le code entier repérant les données manquantes s'il y en a

Exemple :

identificateur, libelle, type de la variable numero : 1 ?

► avor, êtes vous pour l'avortement,2

nombre de modalités de la variable avor ?

► 2

libelle de la modalité numero : 1 identificateur généré:avo1

► avvr,oui (redéfinition de l'identificateur de la modalité)

libelle de la modalité numero : 2 identificateur généré:avo2

► non

identificateur, libelle, type de la variable numero : 2 ?

► nucl,pour les centrales nucleaires,2

nombre de modalités de la variable nucl ?

► 2

libelle de la modalité numero : 1 identificateur généré:nuc1

► =avor
(les libelles de modalités seront ceux de la variable avor)

On peut voir un exemple d'enregistrement par ENRV dans le paragraphe 4.1.4.

4.2. COMMENT GERER

Le logiciel SICLA permet une certaine transparence des données au sens suivant:

1.- L'utilisateur n'a plus à connaître l'emplacement physique de l'information ; il a "ACCES PAR NOM" aux individus et aux variables, la correspondance étant assurée par le système.

2.- Lorsqu'au vu des résultats obtenus, l'utilisateur décide de sélectionner des individus ou des variables selon des régles plus ou moins complexes, de créer de nouvelles variables au moyen d'opérateurs mathématiques, de créer de nouveaux tableaux comme les tableaux de contingence, de distances, de coordonnées factorielles, ... le système assure la GESTION (création, recherche, édition, ...) de ces constituants et évite à l'utilisateur cette tâche qui est en général la plus fastidieuse sans logiciel approprié, et prend 70 à 80% de son temps .

Le logiciel SICLA offre à l'utilisateur un ensemble de possibilités de gestion recouvrant la majorité des cas habituellement rencontrés en analyse des données :

- édition, modification des données
- sélection de variables
- création de variables quantitatives ou qualitatives
- recodification des variables qualitatives
- filtrage d'individus selon des régles logiques
- création de nouveaux tableaux

Les opérateurs utilisés pour SELECTIONNER ou CREER des variables et pour filtrer des individus sont regroupés dans le tableau ci-dessous.

TABLEAU DES OPERATEURS

Dans une expression logique ou arithmétique peuvent intervenir :

1. l'opérateur d'affectation	: :=
2. les opérateurs arithmétiques	: + , - , * , / , ** .
3. les fonctions	: abs, exp
	: log, log1 (log decimal)
	: sqrt
4. les opérateurs de comparaison	: =, < , > , <= , >= , <>
5. les opérateurs logiques	: ! (ou), & (et),#(non)
	: (sur multics \#(non))

Exemple : cout := qu * prix + 150
pour créer une nouvelle variable quantitative cout.

Exemple : Pour sélectionner, par la commande SELOGI, les individus appartenant aux catégories cadres moyens (CSP=10) et supérieurs (CSP=11), travaillant dans le secteur informatique (SECT=5) et âgés de plus de 30 ans, l'ordre de sélection introduit à la console sera le suivant :

((CSP=10)!(CSP=11))&(SECT=5)&(AGE>30)

Cet ordre sera analysé et validé et une structure de données comportant les individus vérifiant le critère de sélection sera générée.

Du point de vue de l'implémentation informatique, l'ensemble des objets se regroupent dans deux types de fichiers en accès direct :
a) La STRUCTURE DES DONNEES (en abrégé s.d.) : ensemble des données de base sur lequel porteront les divers traitements.
b) L'ARCHIVE dans laquelle seront regroupés les différents objets issus des traitements (partitions, coordonnées factorielles, etc,..)

4.2.1. COMMANDES DE MANIPULATION DES STRUCTURES DE DONNEES

Il existe une commande permettant de se positionner sur une s.d.

DONNEE : Changement de données, au cours d'une session cette commande permet de traiter une autre s.d. que la s.d. courante.

Les commandes suivantes (à l'exception de MODSDO) génèrent une nouvelle s.d. ; il s'agit de transferts à partir de l'archive ou de création de données.

CONIND : Fusion de données, cette option permet de fusionner des s.d. ayant le même ensemble de variables.

CONTMU : Création de tableaux de contingence croisant deux ou plusieurs variables qualitatives.
Tableau de Burt croisant l'ensemble des variables qualitatives.

CONVAR : Concaténation de données, permet de concaténer des s.d. ayant le même ensemble d'individus.
Exemple : soient deux s.d. DON1 et DON2, les variables de DON1 sont concaténées à celles de DON2.
Le fichier résultat est une nouvelle structure de données DON3.

CRDIQN : Génération d'un tableau de distances ; permet à partir d'une s.d. comportant des variables quantitatives la création d'une nouvelle structure de données constituée par le tableau de distances entre les individus. Cette commande est utile si l'on désire appliquer certaines méthodes d'analyse fonctionnant uniquement sur des tableaux de distances.

CRQLBI : Mise sous forme disjonctive complète de données qualitatives.
Il est parfois utile d'"éclater" les variables qualitatives en 0 et 1 pour certaines analyses. Cette commande permet de générer une nouvelle s.d. contenant les données "éclatées".

GENFIC : Génération de fichiers textes relatifs à une s.d. (voir exemple paragraphe 4.1.4.)

GRAPAR : Construction de données dont les individus sont les centres de gravité des classes d'une partition. Cette commande est utile pour ceux qui désirent étudier les classes d'une partition en considérant les centres de gravité. Les méthodes hiérarchiques peuvent par exemple être appliquées sur les données centre de gravité.

GROUPE : Génération de groupes de variables ou d'individus. Des ensembles de variables ou d'individus ayant un même statut doivent être définis avant l'éxécution de certaines méthodes. GROUPE permet de les définir.

MODSDO : Modification de la s.d. en cas d'erreur lors de l'enregistrement .
on peut :
1.modifier les informations générales relatives aux données
2.modifier les informations relatives aux variables
3.modifier les informations relatives aux individus

Puis, par exemple, en ce qui concerne les variables (modification 2) on peut :
1. changer l identificateur
2. changer le libellé
3. pour changer le type statistique
4. changer les identificateurs des modalités
5. changer les libelles des modalités
6. changer la valeur pour un individu
7. pour changer le nombre de modalités

Pour les points 3 et 7, le système renvoie à d'autres commandes : GENFIC, CREQAL ou CREQAN ou ELIMOD.

Pour les commandes suivantes, lorsque l'ensemble des individus est modifié, par sélection des individus ou des variables ou par transposition, une nouvelle s.d. est créée.

SELIND : Sélection d'un sous-ensemble d'individus ; cette sélection est faite selon les rangs des individus ou les identificateurs.
Exemple : L'ordre (1a10,20a$) supprimera les individus 11 à 19.

SELOGI : Sélection d'un sous-ensemble d'individus selon un critère logique portant sur les variables et utilisant des opérateurs logiques usuels (cf tableau des oppérateurs au paragraphe 4.2.)
Exemple : L'ordre " ((age<30)&(sect=5)&(csp=10)) " sélectionnera les individus de moins de trente ans, cadres moyens (csp=10),travaillant dans le secteur informatique (sect=5).
Rappel : L'accès aux variables qualitatives se fait par le numéro de modalité.

SELVAR : Sélection des variables.
Il est possible de sélectionner en supprimant certaines variables et de créer en même temps une nouvelle structure de données. La sélection est faite soit par identificateurs, soit selon les rangs en utilisant les opérateurs : " k , n a p , q a $ "
où k,n,p,q désignent des rangs et $ le rang de la dernière variable.
Exemple : Si les données comportent 16 variables, le critère de sélection " 1 a 3,13,15 a $ " sélectionnera les variables : var1, var2, var3, va13, va15, va16

TRANSP : Création d'une s.d. transposée pour l'éxécution de certaines méthodes.

4.2.2. COMMANDES DE GENERATION DE NOUVELLES VARIABLES

Remarques

1. Une nouvelle variable s'exprime comme une expression arithmétique ou logique des variables existantes.

2. En ce qui concerne la création de variables, les différentes commandes sont réalisées directement sur la s.d. Les variables sont ajoutées ou supprimées sur la s.d. et il n'y a pas de création de nouvelles données.

CREQAN : Création de variables quantitatives (en abrégé QN).

Exemple 1 : Création des variables QN qua1,qua2
(au moyen d'une expression arithmétique)

```
qua1:=(var1+var2)/3;
qua2:=log(var3);
```

Exemple 2 : Création de la variable QN v1
(au moyen d'une expression logique)

```
v1:=v3*(v4>v5)+v2*(v4<v5)
signifie :
si v4>v5 alors v1=v3, sinon v1=v2
```

CREQAL : Création de variables qualitatives (en abrégé QL).
Ces variables sont créées à partir des variables de la s.d. et en utilisant les opérateurs arithmétiques ou logiques de la liste donnée au paragraphe 4.2.
Deux types de codages sont possible :

1 : un codage classique dit en disjonctif complet. L'exclusion mutuelle et l'existence d'une expression logique ayant une valeur vraie sont controlées par le système.

exemple :

```
mod1:=(age<35)&(csp=10)
mod2:=(age>=35)&(csp<>10)
```

2 : un codage dit hiérarchique. On n'impose pas que les expressions logiques soient exclusives. La première vérifiée par un individu est choisie pour le codage. L'expression logique de la dernière modalité est définie comme étant toujours vraie. Ainsi la dernière modalité regroupera les individus ne vérifiant pas les expressions logiques des autres modalités.

exemple :
mod1:=(var1=1)
mod2:=(var1=1) ! (var2=3)
pour créer une variable à 3 modalités ; le système génère automatiquement la modalité 3 dont le nom est demandé.

Cette possibilité est intéressante pour créer un variable qualitative ayant deux modalités, il suffit de donner la condition logique relative à la première modalité.

CRPAQL : Création d'une partition sur l'archive à partir d'une variable qualitative.

CRQLPA : Création d'une variable qualitative à partir d'une partition de l'archive.

CODQAN : Codage de variables quantitatives en qualitatives par découpage en classes. Les découpages proposés sont :
- le découpage à intervalles égaux
- le découpage par lecture des bornes
- le découpage à effectifs égaux
- le découpage optimal au sens de la variance

Il existe également une option qui permet d'analyser simultanément tous ces découpages.

METRIQ : Choix d'une métrique pour un tableau de données (variables QN).
Le choix d'une métrique particulière permet de pondérer l'influence des différentes variables dans le calcul des distances entre les individus. On peut toujours se ramener à une métrique euclidienne en transformant d'une manière adéquate les données initiales.
Les transformations possibles sont : centrage ,centrage réduction , calcul des profils ,calcul des profils et normalisation du chi-deux.

Divers contrôles de cohérence sont effectués : validité de calculs des fonctions (log, sqrt, /) ; validité des expressions logiques ; exclusion des modalités ; présence des modalités (modalité representée par au moins un individu.).

En cas d'erreur, les corrections peuvent être faites en conversationnel, en appelant MODSDO, ELIMOD.

4.2.3. RECODIFICATION DES VARIABLES QUALITATIVES

Lors du traitement de données d'enquête ou de données comportant des variables qualitatives, on est souvent amené à recoder les variables, par exemple lorsque peu ou pas d'individus ont répondu à une question.

Les options suivantes permettent de recoder une variable qualitative .

ELIMOD : Elimination des modalités d'effectif nul

Cette option permet de détecter les modalités d'effectif nul et de les éliminer. Le système n'élimine cependant pas les variables qualitatives ayant un nombre insuffisant (inférieur à 2) de modalités.

GROMOD : Regroupement de modalités d'une même variable.

Il est parfois conseillé de regrouper les modalités de faible effectif afin d'assurer la robustesse de certaines analyses.

4.2.4. COMMANDES DE GESTION DE PARTITIONS

Diverses options permettent la gestion de partitions ; toutes les commandes ci-dessous sont décrites de manière détaillée dans le Manuel de référence [4].

CRMUMU : Regroupement de multi-partitions.

CRMUPA : Regroupement de partitions en multi-partitions.

CRPAQL : Création d'une partition sur l'archive à partir d'une variable qualitative

CRQLPA : Création d'une variable qualitative à partir d'une partition.

EDMUAR : Edition d'une multi-partition.

EDPAR : Edition d'une partition archivée.

EXPAMU : Extraction d'une partition d'une multi-partition.

GRAPAR : Création d'une structure de données constituée par les centres de gravité des classes d'une partition.

4.3. COMMENT DECRIRE

Une fois l'enregistrement et le contrôle des données effectués, il est d'usage de faire une première description des données avant d'aborder les méthodes d'analyse de données proprement dites.

4.3.1. IMPRESSION DE DESCRIPTIFS

Une simple édition constitue une description élémentaire : les commandes EDSDO et EDDIC impriment respectivement le descriptif des données et le dictionnaire des variables.

4.3.2. CALCUL D'INDICATEURS

Il est possible de calculer divers indicateurs qui sont, suivant les types de variables :

- Minimums,maximums,moyennes,etc,...
- Histogrammes
- Tris à plats,tris croisés,etc,...

Ces indicateurs constituent une première approche des données .

Les indicateurs de description unidimensionnelle des variables quantitatives sont calculés par :

AKHI2 : Calcul du tableau de contingence entre deux variables.

BIPLOT : Représentation plane d'un ensemble de deux variables quantitatives

BOX : Représentation d'une variable binaire par une boite à moustaches ; permet de fournir les quantiles empiriques d'ordre p notés Qp(n).
Si les n individus sont rangés par ordre croissant alors Qp(n) correspond à la valeur prise par le [np] + 1 éme individu.
La médiane correspond à la valeur p=1/2, le premier quartile correspond à la valeur p=1/4 et le dernier à p=3/4.

CORREL : Calcul et impression de la matrice de corrélation

COURBE : Edition simultanée de courbes.
Cette commande trace un ensemble de courbes. elle est applicable sur des variables quantitatives (types 1,5,6,7)

DESQAL : Description des variables qualitatives ; Calcul des effectifs et pourcentages des modalités par variable.

DESQAN : Calcul pour chaque variable quantitative du MINIMUM, du MAXIMUM, de la MOYENNE, et de l'ECART-TYPE.

HISTO : Calcul d'HISTOGRAMMES, le nombre de classes étant imposé par l'utilisateur et devant être compris entre 2 et 25 .
Trois sortes d'histogrammes sont imprimés :
-histogramme horizontal
-histogramme vertical
-histogramme vertical avec noms des individus.

Il existe aussi un INTERFACE avec un système graphique respectant les normes GKS.

4.3.3. EXEMPLES DE DESCRIPTION

```
► DESQAN
 ************************************************************************
 * SICLA                Commande DESQAN              Version 15/03/85   *
 ************************************************************************
 * données : micro.sdo      individus : 169  *  variables :  14         *
 * ————————                 ————————————                 ————————————  *
 *                          quantitatives :  5   qualitatives :  9      *
 ************************************************************************

         description des variables quantitatives
         ———————————————————————————————————————

 ************************************************************************
 * variables * minimum      *   maximum   *   moyenne   * ecart-type  *
 ************************************************************************
 *    flop   *  1.0000      *  4.0000     *  1.7278     *  .48333     *
 *    prix   *  8500.0      *  69900.     *  40055.     *  12652.     *
 *    ram1   *  32.000      *  640.00     *  168.83     *  106.70     *
 *    ram2   *  64.000      *  16000.     *  699.80     *  1320.6     *
 *    mass   * 0.00000E+00  *  31200.     *  3635.1     *  4762.6     *
 ************************************************************************
 ═══ Fin de DESQAN      ═══
 entrer une commande :            (aide : ?  fin : /)
 ———————————————————

► DESQAL
```

```
► DESQAL
   ************************************************************************
   * SICLA                Commande DESQAL              Version 15/03/85   *
   ************************************************************************
   * données : micro.sdo      individus : 169   * variables :  14         *
   * ————————                 ————————          ————————                 *
   *                          quantitatives :  5   qualitatives :  9      *
   ************************************************************************
          description des variables qualitatives
          ——————————————————————————————————————

       1. variable :coul   moniteur couleur
          ————————

               cou1 :  absent
               cou2 :  optionnel
               cou3 :  present

    **********************
    * modl * nbre *   %  *
    **********************
    * cou1 *  100 *   59 *   ****************************************************
    * cou2 *   60 *   36 *   ************************************
    * cou3 *    9 *    5 *   *****
    **********************

       2. variable :CP/M   systeme d'exploitation CP/M
          ————————

               CP/1 :  absent
               CP/2 :  optionnel
               CP/3 :  present

    **********************
    * modl * nbre *   %  *
    **********************
    * CP/1 *   62 *   37 *   ***************************************
    * CP/2 *   25 *   15 *   ***************
    * CP/3 *   82 *   49 *   ************************************************
    **********************

   encore ?
► non
   ═══ Fin de DESQAL      ═══
   entrer une commande :          (aide : ?  fin : /)
   ————————————————————
  /

   ************************************************************************
   * SICLA                Commande SICLA               Version Multics 1.0 *
   ═══ Fin de SICLA       ═══
```

4.4. COMMENT ANALYSER

Après les différentes mises en forme préalables, l'utilisateur choisit, selon la nature de ses données et le but poursuivi, entre les diverses techniques d'analyse des données :
ANALYSES FACTORIELLES, CLASSIFICATIONS, ET HIERARCHIES

Les conventions utilisées pour l'enregistrement des paramètres sont :

/ pour retourner à la question précédente ou arrêter.
? pour obtenir des indications sur les réponses possibles.
.. pour exécuter une commande du système.

Les résultats seront enregistrés dans un fichier "lis.heure_mn",(cf paragraphe 4.) et les structures-résultats dans le fichier archive crée ou choisi en début de session.

Les méthodes proposées sont principalement celles relatives aux Nuées Dynamiques et leurs variantes. Cependant il est possible d'avoir accès à d'autres méthodes d'analyse des données.

Les méthodes actuellement disponibles sont précisées dans les paragraphes suivants :

a) Les méthodes de Classification Automatique ; voir la liste des commandes dans le thème "Classification".(paragraphe 6.2.5)

b) les méthodes factorielles ; voir la liste des commandes dans le thème "Factoriel".(paragraphe 6.2.6)

c) les méthodes de Discrimination ; voir la liste des commandes dans le thème "Discrimination".(paragraphe 6.2.7)

d) les méthodes hiérarchiques ; voir la liste des commandes dans le thème "Hierarchie".(paragraphe 6.2.8)

e) des méthodes statistiques diverses ; voir la liste des commandes dans le thème "Divers".(paragraphe 6.2.11)

Toutes ces méthodes sont décrites de manière détaillée dans le manuel de référence [4].

4.5. COMMENT INTERPRETER

Cette partie concerne actuellement essentiellement la structure partition.
Les algorithmes des nuées dynamiques fournissent des partitions localement optimales qui peuvent être :

- soit décrites d'une manière interne au niveau des classes (recherche d'individus représentatifs, paramètres discriminants, etc, ..) ou externe (comparaisons des paramètres des différentes classes).
- soit validées, les partitions optimales obtenues à chaque tirage dépendent de la configuration initiale, par conséquent il est nécessaire de valider les formes obtenues pour poursuivre la recherche et la réduction des formes fortes.

Les modules spécifiques disponibles pour l'interprétation des partitions sont :

INPAQN : Calcul d'indices d'interprétation basés sur la décomposition de l'inertie totale en inertie intra-classe et inter-classe, décrivant les rôles respectifs des classes et des variables. Ce module d'aide à l'interprétation concerne les variables quantitatives.

INPAQL : Interprétation d'une partition à partir de variables qualitatives. ayant ou non participé à la construction de la partition et recherche des modalités significatives pour chaque classe.

FF : Calcul et édition des formes fortes. [9]

4.6. COMMENT SAUVEGARDER ET UTILISER L'ARCHIVE

Les différentes analyses effectuées fournissent des structures résultats : partitions, tableaux de distances, coordonnées factorielles, hiérarchies,...dont Le système SICLA assure la sauvegarde dans un fichier ARCHIVE sous une forme normalisée.
Ce fichier est crée automatiquement, sous le nom :
<nom_de_la_s_d.sar>.
Divers modules permettent de réaliser les fonctions suivantes :

EDAR : impression des constituants de l'archive sur listing. A chaque constituant est associé un ensemble d'informations caractéristiques,(type,identificateur et divers autres renseignements) qui sont imprimées.

DELAR : destruction de constituants .
L'utilisateur peut détruire des constituants ne présentant plus d'intérêt.

EDPAR : Edition d'une partition contenue dans l'archive courante.

EDHIAR : Edition d'une hiérarchie contenue dans l'archive courante.

4.7. COMMENT EDITER

Les éditions suivantes sont possibles :

EDDIC : Edition en clair des informations variables
Le dictionnaire des variables est imprimé sur listing ou sur l'écran sous forme d'un tableau récapitulatif de leurs caractéristiques.

EDFAC : Edition et Analyse d'un tableau de facteurs
Cette commande permet d'éditer divers indicateurs : corrélations, contributions,.etc.., liés à un tableau de facteurs et les plans factoriels relatifs aux individus ou variables.

EDSDO : Edition des données
Les données complétes ou un sous-ensemble des donneés peuvent être édités sur écran ou listing afin de les contrôler ou pour assurer l'interface avec d'autres logiciels.

EDIND : Edition sur l'écran du vecteur représentant un individu.

5. BIBLIOGRAPHIE

[1] Y. Lechevallier, H. Ralambondrainy, G. Govaert
Le système Sicla : principes et architecture.
Rapport de recherche No 500, INRIA (Mars 1986).

[2] Documentation Sicla no 1
Présentation du Système.

[3] Documentation Sicla no 2
Manuel du programmeur.

[4] Documentation Sicla no 4
Manuel de référence.

[5] G. Jomier, O. Kezouit, H. Ralambondrainy
Contribution aux bases de données statistiques : Le système PEPIN_SICLA
4èmes journées internationales d'analyse des données
Versailles Octobre 1985.

[6] E. Demonchaux, J. Quinqueton, H. Ralambondrainy
Un système expert en analyse des données : CLAVECIN
Rapport de recherche INRIA No 431, Juillet 1985.

[7] Y. Rakotomalala
Graphique en analyse des données
Mémoire d'ingénieur IIE, 1984.

[8] SPAD
Système Portable d'Analyse de Données
CESIA 82 rue de Sèvres 75007 Paris.

[9] E.Diday et collaborateurs
Optimisation en Classification automatique ,INRIA 1979.

[10] E. Diday, J. Lemaire, J. Pouget, F. Testu
Eléments d'analyse des données, DUNOD (1982)

[11] P. Lignelet
Fortran 77, Chapitre 10,Masson 1984

6. ANNEXES

6.1. LISTE ALPHABETIQUE DES COMMANDES

Les noms des commandes sont en majuscules, les suffixes ne sont ajoutés que pour rappeler le thème de la commande, ils ne doivent pas être utilisés pour exécuter la commande.

Les codes utilisés pour les noms de thème sont les trois premières lettres du mot qui les désignent.

arc = ARCHIVE
cla = CLASSIFICATION
cod = GESTION : CODAGE
dec = DECRIRE
dsc = DISCRMINATION
div = METHODES DIVERSES
edi = EDITION
fac = FACTORIEL
ges = GESTION : COMMANDES GENERALES
hie = HIERARCHIE
int = INTERPRETATION
man = GESTION : MANIPULATION de DONNEES

Les abréviations suivantes sont utilisées :
QL pour (variables) QuaLitatives ; QN pour (variables) QuaNtitatives ; s.d. pour structure de données.

AFTD -fac Analyse factorielle d'un tableau de distances
AKHI2 -dec Analyse d'un tableau de contingence entre deux variables QL
ANAVAR-div Ajustement de modèles linéaires par les moindres carrés
BIPLOT-dec Représentation plane de deux ensembles de variables QN
BOX -dec Représentation d'une variable binaire par une boite à moustaches
CLADIS-cla Classification à partir d'un tableau de distances
CODQAN-cod Découpage en classes d'un ensemble de variables quantitatives
CONIND-man Concaténation de deux ensembles d'individus
CONTMU-man Création d'un tableau de contingence simple ou multiple
CONVAR-man Concaténation de deux ensembles de variables
CORREL-dec Impression du tableau des corrélations
COURBE-dec Edition simultanée de courbes
CRDIQN-man Génération d'un tableau de distances à partir de variables QN
CREQAL-cod Création de variables qualitatives
CREQAN-cod Création de variables quantitatives
CRMUMU-arc Regroupement de multipartitions
CRMUPA-arc Regroupement de partitions en multipartitions
CROBIN-cla Classification croisée d'un tableau binaire
CROEUC-cla Classification croisée d'un tableau individus x variables QN

CROKI2-cla	Classification croisée d'un tableau de contingence
CROMUL-cla	Classification croisée d'un tableau individus x variables QL
CRPAQL-cod	Création d'une partition à partir d'une variable QL
CRQLBI-man	Génération de données binaires
CRQLPA-cod	Création d'une variable QL à partir d'une partition
DELAR -arc	Destruction d'un constituant de l archive
DESQAL-dec	Description de variables qualitatives
DESQAN-dec	Description de variables quantitatives
DISC -dsc	Méthode de discrimination linéaire
DNP -dsc	Segmentation non paramétrique
DONNEE-ges	Choix des paramètres pour une session SICLA
EDAR -arc	Impression des constituants de l'archive
EDDIC -edi	Edition du dictionnaire d'une s.d.
EDFAC -edi	Edition et analyse d'un tableau de facteurs
EDHIAR-arc	Edition d'un arbre hierarchique
EDIND -edi	Edition des valeurs prises par un individu
EDMUAR-arc	Edition d'une multipartition
EDPAR -arc	Edition d'une partition de l archive
EDSDO -edi	Edition des données
ELIMOD-cod	Elimination des modalités absentes
ENRDIS-ges	Enregistrement d'un tableau de distances
ENRFIC-ges	Enregistrement des variables à partir d'un fichier
ENRI -ges	Enregistrement des individus
ENRIT -ges	Enregistrement d'un tableau de données
ENRV -ges	Enregistrement des variables
EXPAMU-arc	Extraction d'une partition d'une multipartition
FACDSC-dsc	Analyse factorielle discriminante
FASPAD-fac	Création d'une structure de données relative aux facteurs
FF -int	Analyse de plusieurs partitions d'une multi-partition
GENFIC-man	Création des fichiers textes relatifs à une s.d.
GRAPAR-man	Création d'une s.d. centres de gravité d'une partition
GROMOD-cod	Regroupement de modalités d'une variable QL
GROUPE-man	Génération de groupes de variables et d'individus
HISTO -dec	Tracés d'histogrammes
INPAQL-int	Interprétation d'une partition par des variables QL
INPAQN-int	Interprétation d'une partition par des variables QN
METRIQ-cod	Choix d'une metrique
MNDQAL-cla	Méthode des Nuées Dynamiques sur variables QL
MNDQAN-cla	Méthode des Nuées Dynamiques sur variables QN
	Classification de grands tableaux
MODSDO-man	Modification de la structure des données
SAUT -hie	Hiérarchie du saut minimum ou maximum
SELDSC-dsc	Sélection de variables en discrimination
SELIND-man	Sélection d'individus par les rangs ou identificateurs
SELOGI-man	Sélection d'individus suivant un critère logique
SELVAR-man	Sélection d'un ensemble de variables
SEM -div	Reconnaissance d'un mélange gaussien unidimensionnel
SEMMUL-div	Reconnaissance d'un mélange gaussien multidimensionnel
SISPAD-fac	Interface avec le logiciel SPAD
TRANSF-cla	optimisation ou création de partitions
TRANSP-man	Création d'une s.d. transposée

VOISIN-dsc Discrimination par la méthode des K plus proches voisins
WARD -hie Hiérarchie avec l'indice de l'accroissement de l'inertie

6.2. LISTE DES THEMES ET DES COMMANDES PAR THEME

Les commandes de SICLA sont réparties en thèmes :

1 EDITION

2 GESTION (ENREGISTREMENT ET MANIPULATION DES DONNEES)

3 DESCRIPTION

4 INTERPRETATION

5 CLASSIFICATION (contient INTERPRETATION DES PARTITIONS)

6 FACTORIEL

7 DISCRIMINATION

8 HIERARCHIE

9 CODAGE

10 ARCHIVE

11 METHODES DIVERSES

6.2.1. LE THEME EDITION

Fonction : Editer (écran) ou Imprimer (listing)

Commandes :

EDAR : Impression des constituants de l'archive.
EDDIC : Impression du dictionnaire d'une s.d.
EDFAC : Edition et analyse d'un tableau de facteurs
EDHIAR : Edition d'un arbre hiérarchique
EDIND : Edition sur l'écran d'un vecteur-objet.
EDMUAR : Edition d'une multipartition
EDPAR : Edition d'une partition de l'archive
EDSDO : Impression des données d'une s.d.

6.2.2. LE THEME GESTION

Fonctions : Enregistrement des données
Manipulation des Structures de données
Génération de variables

Commandes pour l'enregistrement des données

DONNEE : Changement de structure de données courante
ENRDIS : Enregistrement d'un tableau de distances
ENRFIC : Enregistrement des variables à partir d'un fichier
ENRI : Enregistrement des données
ENRIT : Enregistrement d'un tableau de données
ENRV : Enregistrement du dictionnaire des variables

Commandes pour la manipulation des Structures de données

CONIND : Fusion de deux structures de données
(concaténation sur les individus)
CONTMU : Création d'un tableau de contingence multiple
CONVAR : Fusion de deux structures de données ;
Concaténation sur les variables
CRDIQN : Génération d'un tableau de distances
(à partir de variables quantitatives)
CRQLBI : Création d'un tableau disjonctif complet
GENFIC : generation des fichiers dictionnaires relatifs à une s.d.
GRAPAR : Création d'une s.d. des centres de gravité
GROUPE : Mise en forme des données pour une méthode ;
Définition de groupes de variables et d'individus
MODSDO : Modification de la structure de données
SELIND : Sélection d'individus par rang ou identificateur
SELOGI : Sélection d'individus selon un critère logique
SELVAR : Sélection de variables
TRANSP : Création d'une s.d. transposée pour un accès rapide
par colonnes

6.2.3. LE THEME DESCRIPTION

Fonction : Décrire les données

Commandes :

AKHI2 : calcul de tableaux de contingence et analyse par khi2 locaux.
BIPLOT : Représentation planede deux ensembles de variables QN.
BOX : Représentation d'une variable binaire par une boite à moustaches
CORREL : calcul et impression de la matrice de corrélation.
COURBE : Edition simultanée de courbes
DESQAL : description des variables qualitatives.
DESQAN : description des variables quantitatives.
HISTO : calcul et édition des histogrammes.

6.2.4. LE THEME INTERPRETATION

Fonction : Interpréter les résultats

Commandes :

FF : Construction et analyse du tableau des formes fortes
INPAQL : Interprétation d'une partition par des variables QL
INPAQN : Interprétation d'une partition par des variables QN

6.2.5. LE THEME CLASSIFICATION

Fonction : Analyser par les méthodes des "NUEES DYNAMIQUES" [9] [10]

Commandes :

CLADIS : Classification d'un tableau de distances et choix de la meilleure classification entre 2 et KMAX classes.

CROBIN : Classification simultanée de lignes et de colonnes d'un tableau binaire

CROEUC : Classification simultanée de lignes et de colonnes d'un tableau de mesures

CROKI2 : Classification simultanée de lignes et de colonnes d'un tableau de contingence (critère du Khi2)

CROMUL : Classification simultanée de lignes et de colonnes d'un tableau de modalités (critère du Khi2)

MNDQAL : Méthode des Nuées Dynamiques sur variables qualitatives

MNDQAN : Méthode des Nuées Dynamiques sur variables quantitatives

TRANSF : Optimisation ou création de partitions

6.2.6. LE THEME FACTORIEL

Fonction : Analyser par des méthodes factorielles directement ou au moyen d'un interface avec le logiciel SPAD

Commandes :

AFTD : Analyse factorielle d'un tableau de distances.

EDFAC : Edition et analyse d'un tableau de facteurs

SISPAD : Cette commande permet d'éviter les deux modules de SPAD DONNE et LILEX, on a directement accès aux modules de traitement.

FASPAD : Cette commande permet de "récupérer" les facteurs et crée une nouvelle s.d. SICLA.

6.2.7. LE THEME DISCRIMINATION

Fonction : Construction de régles de décision

Commandes :

DISC : Discrimination linéaire.
DNP : Segmentation non paramétrique (var.quantitatives).
FACDSC : Analyse factorielle discriminante
SELDSC : Sélection de variables discriminantes
VOISIN : Méthode des K plus proches voisins.

6.2.8. LE THEME HIERARCHIE

Fonction : Analyser par les méthodes hiérarchiques

Commandes :

EDHIAR : Interprétation d'une hiérarchie.
Edition de l'arbre et choix de la meilleure coupure
SAUT : Hiérarchie avec l'indice d'agrégation du saut minimum
ou du saut maximum ,Algorithme des voisins réciproques
avec la formule de Lance et Williams.
WARD : Hiérarchie avec l'indice de l'accroissement d'inertie.
Algorithme des voisins réciproques.

6.2.9. LE THEME CODAGE

Fonction : Codage

Commandes :

CODQAN : Codage de variables quantitatives en variables qualitatives (effectifs égaux).
CREQAN : Création de variables quantitatives par des opérateurs arithmétiques ou logiques
CREQAL : Création de variables qualitatives par des opérateurs arithmétiques ou logiques ; Regroupement de modalités
CRPAQL : Création d'une partition à partir d'une variable qualitative
CRQLPA : Création d'une variable qualitative à partir d'une partition
ELIMOD : Elimination des modalités absentes
GROMOD : Regroupement des modalités d'une même variable
METRIQ : Choix d'une métrique pour les variables quantitatives

6.2.10. LE THEME ARCHIVE

Fonction : Gestion de l'archive
Edition de segments de l'archive

Commandes :

CRMUMU: Regroupement de multipartitions
CRMUPA: Regroupement de partitions en multipartitions
DELAR : Destruction logique d'un constituant de l'archive
EDAR : Edition des constituants de l'archive
EDHIAR: Edition d'un arbre hiérarchique
EDMUAR: Edition d'une multipartition
EDPAR : Edition d'une partition de l'archive
EXPAMU: Extraction d'une partition ou d'une multipartition

6.2.11. LE THEME METHODES DIVERSES

Fonction : Analyser par d'autres méthodes.

Commandes :

ANAVAR : Ajustement de modéles linéaires par les moindres carrés
SEM : Reconnaissance de mélanges gaussiens unidimensionnels
Estimation des paramétres
SEMMUL : Reconnaissance de mélanges gaussiens multidimensionnels

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique