

~~Annex 1/BAC/14/2/2/1/2/3/4/5/6/7/8/9/10/11/12/13/14/15/16/17/18/19/20/21/22/23/24/25/26/27/28/29/30/31/32/33/34/35/36/37/38/39/40/41/42/43/44/45/46/47/48/49/50/51/52/53/54/55/56/57/58/59/60/61/62/63/64/65/66/67/68/69/70/71/72/73/74/75/76/77/78/79/80/81/82/83/84/85/86/87/88/89/90/91/92/93/94/95/96/97/98/99/100~~
2003
27/5

ECA/STAT/TW/84/2



NATIONS UNIES

Commission économique pour l'Afrique

**ATELIER DE FORMATION SUR L'EXECUTION
DES RECENSEMENTS ET L'UTILISATION DE
PROGICIELS STATISTIQUES POUR LES PAYS
AFRICAINS FRANCOPHONES**

YAOUNDE : 1^{er} - 26 OCTOBRE 1984

RESUME DES TRAVAUX

**DEUXIEME PARTIE :
(16-26 Octobre 1984)**

UTILISATION DE PROGICIELS STATISTIQUES

MAI 1985

COMMISSION ECONOMIQUE POUR L'AFRIQUE

ECA/STAT/TW/84 -2

Atelier de formation sur l'exécution
des recensements et l'utilisation de
progiciels statistiques pour les pays
africains francophones

Yaoundé, 1er-26 octobre 1984

RESUME DES TRAVAUX

Deuxième partie :
(16-26 octobre 1984)

UTILISATION DE PROGICIELS STATISTIQUES

MAI 1985

ECA/BO2/003

PREAMBULE

Cette publication est le résumé des travaux relatifs à la deuxième partie de l'Atelier de formation sur l'exécution des recensements de population et l'utilisation de progiciels statistiques pour les pays africains francophones, organisé du 1er au 26 octobre 1984 à Yaoundé (République du Cameroun).

Elle porte sur l'utilisation de progiciels statistiques particulièrement dans le cadre du traitement des données d'un recensement de la population.

Un premier document sur la première partie du dit Atelier, à savoir les différentes phases d'exécution d'un recensement (travaux préparatoires, dénombrement, évaluation et analyse des résultats), à l'exclusion du traitement des données, a été publié sous le n° de référence ECA/STAT/TW/1.

Pourquoi deux publications ? Les raisons qui ont amené la CEA à procéder ainsi sont à la fois de fond et de forme.

En effet, afin de permettre d'accorder à chacune des parties l'importance qu'elle mérite et tirer le maximum d'enseignements des expériences vécues dans les différents pays, il avait été initialement prévu la tenue de deux ateliers distincts : l'un sur l'exécution des recensements et l'autre sur l'utilisation des progiciels pour le traitement des données issues de ces opérations.

Mais pour des raisons budgétaires, il a été décidé d'organiser ces deux ateliers l'un à la suite de l'autre. Bien que complémentaires, ces deux parties ont été traitées séparément et ont été animées par des instructeurs différents. De plus, les participants n'ont pas exactement été les mêmes pour les deux parties.

Une note succincte relative aux principales observations et conclusions sur le déroulement de l'atelier figure en annexe I.

Puisse les responsables nationaux chargés du traitement des données des recensements des différents pays concernés, trouver dans ce document les éléments pouvant les aider dans leurs tâches en vue d'une obtention rapide et efficace des résultats des futures opérations.

TABLE DES MATIERES

	Page
Préambule	
Introduction	1
1- Généralités sur les ordinateurs	4
A. Concepts , définitions , terminologie	4
B. Fonctionnement des ordinateurs	6
2- Application aux recensements et enquêtes	12
2-1 Introduction	12
2-2 Le questionnaire	13
2-3 Le traitement des données	15
A. Le traitement manuel	15
B- La saisie des données	19
C. Le redressement par ordinateur	20
E. La mise en tableaux	22
3- Les progiciels statistiques	23
3-1 Progiciels de contrôle et de correction des données	23
A. COBOL CONCOR	23
B. UNEDIT	32
3-2 Progiciels de tabulation	37
A. CENTS-4	37
B. XTALLY	47
C. COXTALLY	64
4- Résumé des discussions	98
Annexe I : observations des participants sur l'organisation et l'efficacité de l'atelier	
Annexe II : liste des participants	
Annexe III : liste des animateurs	

RESUME DES EXPOSES

I N T R O D U C T I O N

L'Atelier de formation sur l'exécution des recensements et l'utilisation des progiciels statistiques pour les pays africains francophones, s'est déroulé du 1er au 26 octobre 1984 à Yaoundé (Cameroun) dans les locaux de l'Institut de statistique, de planification et d'économie appliquée (ISPEA), centre qui participe au Programme de formation statistique pour l'Afrique (PFSA).

Cet Atelier a été financé par le Fonds des Nations Unies pour les activités en matière de population (FNUAP). L'organisation et l'exécution étaient assurées par la Division de la statistique de la Commission économique des Nations Unies pour l'Afrique (CEA). Le Bureau de recensement des Etats Unis, le Bureau de statistique des Nations Unies à New York et l'Institut de formation et de recherche démographiques (IFORD), ont apporté leur appui technique.

Il s'inscrit dans le cadre de la préparation de l'exécution des recensements de la décennie 1980-1989 en Afrique, après le séminaire organisé par le Bureau de recensement des Etats Unis en mars-avril 1980 à Yaoundé, et poursuit les mêmes objectifs que celui qui s'est tenu au Ghana en 1978 à l'intention des pays africains anglophones.

Quinze pays africains francophones ont participé à cet Atelier. Ce sont : Algérie, Bénin, Burkina Faso, Cameroun, Comores, Congo, Djibouti, Gabon, Guinée, Madagascar, Mali, Rwanda, Sénégal, Togo et Zaïre (voir la liste des participants en annexe II).

L'Union douanière et économique des Etats de l'Afrique centrale (U.D.E.A.C) avait également envoyé un représentant.

Les objectifs de l'Atelier peuvent être ainsi définis :

- appuyer les pays concernés dans leurs efforts pour la préparation et l'exécution des recensements de la décennie 1980-1989 par un approfondissement des techniques.
- favoriser le dialogue entre démographes, informaticiens et statisticiens en vue d'une meilleure coordination des différentes phases du recensement.
- créer un cadre d'échanges d'expériences entre les pays de la région sous la supervision d'instructeurs qualifiés et expérimentés.

Au cours de l'Atelier de formation, toutes les opérations des différentes phases de l'exécution d'un recensement ont été abordées, notamment les définitions et concepts, la préparation du questionnaire et des manuels d'instructions, la formation, la cartographie, le dénombrement, l'exploitation et le traitement (utilisation des progiciels statistiques), l'analyse et la publication des données.

Les instructeurs de l'Atelier (voir annexe III) étaient constitués, pour la plupart de conseillers régionaux de la Division de la statistique et de la Division de la population de la CEA, d'experts du Bureau de recensement des Etats Unis, du Bureau de statistique des Nations Unies à New York et de professeurs de l'Institut de formation et de recherche démographiques (IFORD) à Yaoundé.

L'allocution d'ouverture a été prononcée par Monsieur Meka-Meka René, représentant du Ministre d'Etat chargé du plan et de l'aménagement du territoire de la République du Cameroun.

Après avoir souhaité la bienvenue à tous les participants aux noms du Président de la République et de son Gouvernement et de tout le peuple camerounais, Monsieur Meka-Meka a indiqué que cet Atelier de formation s'insère dans le cadre de la recherche permanente du perfectionnement et de l'adaptation de nouvelles méthodes et techniques relatives à la collecte, au traitement et à l'analyse des données de base sur la population. Il a ensuite souligné le fait que les données démographiques nécessaires à une formulation judicieuse des politiques des pays africains en matière de population ou en matière de développement tout court font encore cruellement défaut. L'une des principales raisons de cette situation est le manque d'informations fiables permettant aux pays africains d'établir de manière claire l'étroite relation qui lie les tendances démographiques et le développement économique et social. Une autre raison est que les pays ne disposent ni de moyens ni de personnel qualifié nécessaires pour collecter les données démographiques de base, les analyser et les utiliser comme support essentiel des décisions à prendre. L'Atelier de formation répond donc aux préoccupations des pays africains de disposer de cadres capables et aptes à maîtriser la science démographique qui donne à l'homme sa valeur dans les différents mécanismes du développement.

Monsieur Nkounkou-Louya, assurant l'interim de la Direction du MULPOC de Yaoundé, a d'abord mis l'accent sur le fait que le traitement informatique des données en vue d'une utilisation rapide et efficace a toujours constitué un goulot d'étranglement dans une opération aussi onéreuse et aussi importante que le recensement général de la population et de l'habitation. C'est pourquoi l'utilisation des progiciels statistiques, qui sont des programmes déjà conçus de traitement et de tabulation des données, tient une place privilégiée dans cet Atelier de formation. Monsieur Nkounkou-Louya a ensuite exprimé, aux noms des participants et de la CEA, sa profonde gratitude au Gouvernement du pays hôte pour avoir bien voulu accueillir cet Atelier de formation à Yaoundé, ville située au coeur

de l'Afrique, et mettre à la disposition des organisateurs toutes les facilités nécessaires. Il a enfin remercié le FNUAP pour l'appui financier qu'il a apporté pour l'organisation de l'Atelier de formation, ainsi que tous les organismes qui ont fourni un support technique pour la tenue de la réunion.

1. GENERALITES SUR LES ORDINATEURS

A. CONCEPTS, DEFINITIONS, TERMINOLOGIE

- i) L'informatique est la science du traitement rationnel de l'information (communiqué du 6 avril 1967 de l'Académie française), ce traitement pouvant être effectué par l'intermédiaire de machines automatiques généralement connues sous le nom d'ordinateur (mot lancé en 1951 par IBM).

En informatique, le mot information désigne un facteur qualitatif réperant la position d'un système et qui est éventuellement transmis par ce système à un autre.

Une information se présente toujours sous deux aspects : un aspect syntaxique (ou forme) et un aspect sémantique (ou sens). Les ordinateurs actuels ne sont capables d'opérer des traitements d'information qu'au niveau de la syntaxe. Parmi les principaux traitements susceptibles d'être assurés automatiquement, citons : le tri, la reproduction, la saisie, la ventilation, le calcul, le classement, la transmission.

ii) Intérêt de l'utilisation d'un ordinateur

Il faut bien prendre conscience qu'un ordinateur n'est pas "intelligent" en ce sens qu'il n'exécute que ce qui a été prévu lors de l'écriture du programme (succession prédéterminée d'opérations élémentaires correspondant au traitement). Toute éventualité non prévue ne pourra pas donner lieu à un traitement improvisé par la machine, même si pour le concepteur le traitement à effectuer apparaît évident compte tenu du contexte.

Par ailleurs, un ordinateur ne peut effectuer que des opérations très élémentaires et les méthodes de résolution sur ordinateur diffèrent sensiblement de celles utilisées par l'homme. Par exemple le calcul d'une intégrale définie se fait pas à pas. Mais la machine, grâce à sa rapidité, obtient le résultat beaucoup plus rapidement que l'homme.

En règle générale, l'intervention d'un ordinateur est souhaitable et même indispensable soit quand une même opération est à répéter un grand nombre de fois, soit lorsqu'une opération peut se décomposer en un grand nombre d'opérations identiques.

Vu sous cet angle l'utilisation de l'ordinateur pour l'exploitation d'un recensement est tout à fait justifiée.

iii) Evolution historique du traitement de l'information

De tout temps l'homme a cherché à faire effectuer par des machines les travaux qui visent à modifier le monde qui l'entoure. Les travaux de traitement de l'information n'échappent pas à cette règle.

Les premiers éléments d'une machine à traiter l'information remontent au boulier japonais. Mais il faut attendre PASCAL pour voir apparaître une machine à calculer digne de ce nom.

Les premiers concepts d'une machine automatique sont apparus au début du 19^e siècle avec l'anglais BABBAGE.

Il fallut attendre un siècle et le développement de l'électronique pour que les machines prennent leur forme actuelle.

1944 : apparition de MARK I qui pouvait effectuer 10 additions par seconde

1947 : la firme américaine UNIVAC commercialisait une machine à tubes électroniques.

Il faut mentionner, en parallèle, les importants travaux de HOLLERITH (fin 19^e siècle), fondateur de IBM, qui, pour dépouiller une enquête, eut l'idée de représenter l'information sous forme de perforations : Code HOLLERITH. Ainsi apparut la carte perforée et un certain nombre de machines électro-mécaniques (perforatrice, trieuse, interclasseuse).

Les machines électroniques de traitement de l'information peuvent être classées suivant leur période d'apparition en quatre générations :

- Première génération : machine à tubes électroniques jusque vers 1958.
- Deuxième génération : machines à transistors 1958 - 1964 qui se caractérisent par une réduction de la taille de l'unité centrale et une augmentation de la vitesse de fonctionnement. C'est la prise de conscience de la partie SOFTWARE.

En effet, les constructeurs fournissaient : compilateurs, programmes utilitaires, système d'exploitation.

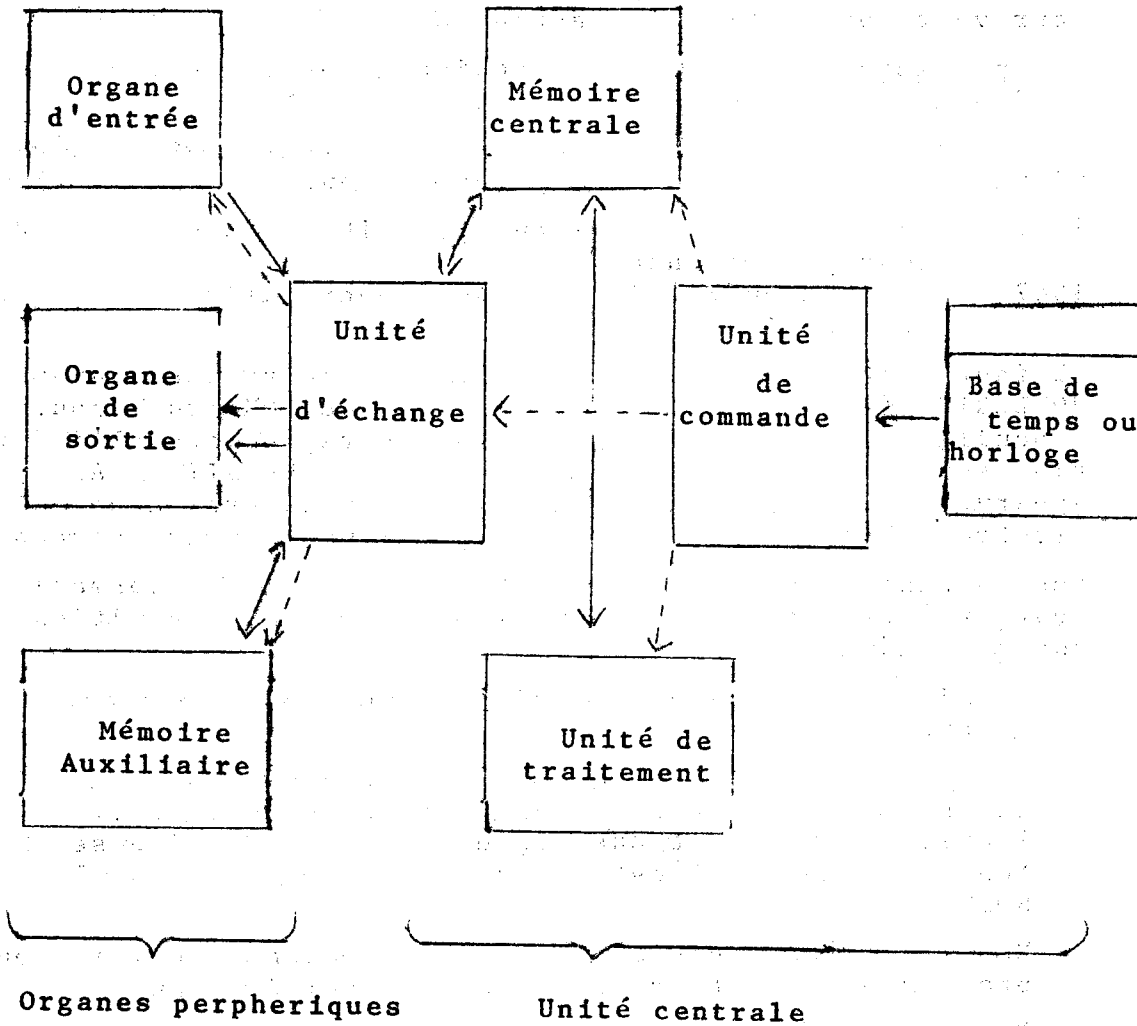
- Troisième génération : à partir de 1964, ordinateurs actuels dont les hautes performances sont dues à :
 - l'utilisation de circuits intégrés;
 - l'existence d'un "système d'interruption" qui permet l'utilisation de la machine en multiprogrammation;
 - l'existence de systèmes d'exploitation évolués.

A partir de 1970 tous les constructeurs ont proposé une nouvelle série de machines (IBM 370, CIT IRIS 80, CDC 7600 etc...)

- Actuellement, en raison de performances accrues tant du côté hardware que du côté software, certains estiment que les machines des années 80 constituent la quatrième génération d'ordinateurs.

B. FONCTIONNEMENT DES ORDINATEURS

1) Schéma général d'un ordinateur → cheminement des données --> cheminement des commandes



ii) Unité de mémoire

La mémoire centrale contient deux types d'information :

- les instructions du programme que la machine devra exécuter;
- les données sur lesquelles la machine effectue les traitements (sous forme binaire).

Une mémoire est caractérisée par :

- sa capacité qui varie de quelques milliers à quelques milliards d'éléments binaires;
- sa vitesse de fonctionnement caractérisée par le temps d'accès et le temps de cycle;
- son mode de fonctionnement.

Temps d'accès : c'est l'intervalle de temps qui sépare le déclenchement du signal d'action et l'instant où l'information est disponible.

Temps de cycle : c'est le temps minimal séparant deux accès à la mémoire.

iii) Unité de commande :

Elle préside au fonctionnement de la machine et gère l'enchaînement de différentes étapes de l'exécution des instructions.

iv) Unité de traitement :

C'est dans ce bloc que s'effectue le traitement d'une opération spécifiée par l'unité de commande; le résultat est disponible à la sortie du bloc.

v) Unité d'échange (ou canal) :

Elle gère les transferts d'informations entre l'unité centrale et les organes périphériques. Il n'y a pas de modification du contenu des informations au cours de ces échanges mais seulement transcodage, la représentation de l'information étant, pour des raisons économiques, rarement la même dans l'unité centrale et dans les organes périphériques.

vi) Unités périphériques :

Il y a deux grandes classes :

- Les unités de communication (lecteur de cartes, de disquettes, imprimante, unité de visualisation oscilloscopique, etc...) qui permettent le dialogue avec l'extérieur
- Les mémoires auxiliaires (disques, bandes magnétiques) qui servent à soulager temporairement l'unité centrale ou à stocker pendant un temps relativement long des données ou des programmes. Leur capacité est généralement supérieure à celle de la mémoire centrale, mais leur temps d'accès plus long.

Disques magnétiques :

L'enregistrement se fait sur une couche magnétique déposée recto-verso sur un ou plusieurs disques solidaires du même axe. L'axe est animé d'un mouvement de rotation uniforme.

Les disques peuvent être fixes ou amovibles. Chaque face est divisée en pistes concentriques. Le nombre de pistes peut varier entre 100 et 1000.

Capacité : jusqu'à peu près de 450 millions d'éléments binaires.

Bandes magnétiques :

La couche magnétique oxyde est déposée sur un film de nylon. Une bobine a une longueur d'environ 730 m et d'une bande large de 1,27 cm (0,5 pouce). On trouve le long de la bande 7 ou 9 pistes accessibles par une rangée de tête de lecture et d'écriture.

Capacité : jusqu'à peu près de 18 millions de caractères.

Disquettes :

C'est un disque de 18 cm environ en matière plastique souple revêtu sur une surface d'oxyde magnétique et placé à demeure dans une enveloppe en carton.

- 77 pistes environ
- vitesse 360 tours/mn
- densité : 3200 bits au pouce
- temps d'accès moyen : 300 ms

Imprimantes :

Ce vocable désigne l'ensemble des organes connectées à un ordinateur et permettant l'impression directe sur papier des caractères alphanumériques issus de l'ordinateur sous forme codée.

On distingue les imprimantes séries et les imprimantes parallèles.

- . Imprimantes séries : l'impression se fait caractère par caractère.

- . Imprimantes parallèles : l'impression se fait ligne par ligne.

Traceur de courbes :

Le principe consiste à déplacer un stylet sur une feuille de papier. Le déplacement continu du stylet est assuré grâce à une conversion numérique analogique des informations numériques issues du calculateur.

Consoles de visualisation :

Elles permettent d'afficher sur l'écran des points, des vecteurs, des courbes obtenues par déplacement du spot commandé par l'ordinateur.

Unités à réponse vocale :

Encore peu développées, elles offrent la possibilité de sortie d'informations sous forme vocale. Le dispositif d'entrée est généralement un clavier alphanumérique car il n'a pas encore été possible de faire reconnaître par l'ordinateur la parole humaine avec suffisamment de fiabilité.

Autres périphériques :

Système de transmission de données : ce système permet d'échanger à distance des informations avec un ordinateur. Les supports de transmission utilisés sont les moyens classiques de télécommunications : voies téléphoniques, télégraphiques, câbles et faisceaux herziens.

Le modulateur sert à convertir les données en un signal modulé capable d'être transmis par le support de transmission.

Le démodulateur assure la conversion inverse.

vii) Base de temps (horloge)

Pour assurer le rôle d'ordonnement temporel, l'unité de commande reçoit des impulsions délivrées par l'horloge.

viii) Registres et bus :

Les registres sont des organes de mémorisation à lecture et écriture très rapides. Un registre comprend des positions binaires. Tous les organes principaux de la machine comprennent un ou plusieurs registres.

Les échanges entre ces organes se font via les registres par l'intermédiaire de lignes appelées bus. Ce bus se présente physiquement comme un ensemble de fils, à raison soit d'un fil par bit, soit 2 fils par bit, l'un portant la grandeur vraie, l'autre la grandeur fausse.

ix) Représentation des nombres

Dans les ordinateurs, les nombres sont représentés soit à partir de la numération binaire pure, soit à partir de la numération décimale en utilisant une représentation codée de chaque chiffre décimal. Dans tous les cas se pose le problème de la représentation des nombres négatifs, ainsi que celui de la représentation des nombres fractionnaires.

- Nombres entiers binaires :

Exemple de représentation par valeur absolue et signe : on représente le signe et la valeur séparément. Il suffit d'un seul bit (chiffre binaire) pour représenter le signe, par exemple 0 pour + et 1 pour -. Si on dispose de n éléments binaires $n - 1$ seront réservés à la valeur absolue. Ainsi on pourra représenter de $(2^{n-1} - 1)$ à $+(2^{n-1} - 1)$ zéro a 2 représentations possibles : +0 et -0.

- Nombres entiers décimaux :

Certaines machines calculent dans le système décimal. Les chiffres constituant le nombre décimal sont alors codés séparément ainsi que les signes.

Pour coder les 10 chiffres décimaux, 4 bits sont nécessaires. Mais 4 bits permettent de coder 16 informations différentes, qui correspondent au système de numération hexadécimal.

x) Représentation des caractères alphanumériques :

Il est indispensable de pouvoir représenter, en plus des chiffres, des caractères alphanumériques, les ordinateurs ayant à travailler sur des informations autres que des nombres.

De nombreux codes sont utilisés :

- le code HOLLERITH

C'est le code utilisé depuis longtemps sur les cartes perforées.

- le code EBCDIC (Extended Binary Coded Decimal Interchange code);

Est utilisé comme code interne, code de représentation sur les bandes magnétiques à 9 pistes ou comme code d'échange avec les organes d'entrée-sortie.

xi) Domaines d'utilisation des ordinateurs :

La coutume veut que l'on sépare le domaine d'emploi des ordinateurs en trois ou quatre classes. Il fut, en effet, un temps où les constructeurs proposaient une machine adaptée à chaque classe. La tendance actuelle est aux ordinateurs universels.

a) Problèmes de gestion

Ils présentent 2 caractéristiques :

- volume important de données ;
- simplicité des opérations de traitement.

b) Problèmes scientifiques : ils se caractérisent par :

- un volume plus ou moins important de données
- la complexité des calculs à effectuer.

L'exploitation des recensements et enquêtes statistiques entre dans cette classe parmi d'autres applications (analyse numérique, recherche opérationnelle, simulation etc...)

c) Problèmes en temps réel.

Un ordinateur travaille en temps réel lorsqu'il existe un lien imposé de l'extérieur entre l'instant où les données à traiter sont fournies et l'instant où le résultat doit être disponible.

d) Problèmes divers.

De par ses caractéristiques, un ordinateur peut intervenir dans des applications très variées telles que l'enseignement, la documentation, la traduction, la reconnaissance des formes.

xii) Systèmes d'exploitation

La machine doit être munie d'un certain nombre de programmes auxiliaires si l'on veut assurer l'exécution convenable des programmes des utilisateurs.

Pour améliorer le rendement de l'ordinateur, on est amené à rendre automatique l'enchaînement des travaux.

Aussi, les tâches d'un système d'exploitation peuvent se résumer aux suivantes :

- Gestion de l'enchaînement des travaux ;
- Gestion des entrées et des sorties ;
- Protection du système lui-même vis-à-vis des erreurs faites par les programmeurs ;
- Gestion des mémoires centrales et auxiliaires ;
- Prise en charge des erreurs détectées ;
- Tenue à jour de la comptabilité des temps d'utilisation.

2. APPLICATION AUX RECENSEMENTS ET ENQUETES

2.1. INTRODUCTION

S'agissant, par définition, de relever des informations sur chaque ménage et les caractéristiques individuelles de chaque habitant d'un pays, le recensement général de la population est certainement l'opération la plus vaste, la plus coûteuse et la plus complexe que devra entreprendre un service national de statistique.

Par ailleurs, le temps qui s'écoule entre deux recensements successifs, généralement de l'ordre de 10 ans, est trop long pour pouvoir pleinement bénéficier des expériences précédentes dans le domaine du traitement des données en raison des progrès technologiques qui s'opèrent entre-temps entraînant une évolution du matériel et des méthodes.

On entend par traitement des données la série complète d'opérations qui transforme l'information recueillie au cours d'un recensement en tableaux statistiques.

C'est un des aspects les plus importants de cette opération d'envergure. C'est pourquoi il est essentiel que le personnel informatique soit représenté dès le départ, dans l'équipe chargée de définir les objectifs et la méthodologie de cette opération.

Le calendrier des opérations d'un recensement doit tenir compte de tous les éléments de la phase de traitement.

Les contraintes budgétaires, les ressources humaines à engager et les différentes tâches à accomplir doivent être clairement définies. Le plan doit comprendre une description générale du système, une indication de la façon dont il doit être mis en place en utilisant les équipements et le personnel existants, et une estimation du temps d'analyse et de programmation nécessaires.

Dans ce document, nous essaieront de considérer toutes les étapes du traitement des données, depuis le dessin du questionnaire jusqu'à la phase finale de mise en tableaux et la constitution de la base de données statistiques.

2.2. - Questionnaire

Les documents de collecte de données sur le terrain sont des éléments-clés d'un recensement et doivent, de ce fait, être conçus avec le plus grand soin. Ils auront une incidence importante sur le coût et la qualité de la collecte, du contrôle manuel, du codage, de la saisie et du redressement des données.

Le questionnaire doit répondre à trois objectifs principaux :

- la collecte d'informations sur le terrain
- le chiffrement
- la saisie (si l'on exclut l'entrée directe de données au moyen de la lecture optique).

On évitera certainement bien des problèmes en associant les informaticiens à la conception du questionnaire, à la confection des codes et nomenclatures, à la mise au point des spécifications de contrôle et à l'élaboration du plan de tabulation.

- La forme du questionnaire et un choix judicieux de codes et nomenclatures peuvent améliorer de façon notable les conditions de travail et la qualité des tâches informatiques, sans pour autant affecter les opérations de collecte. Lorsque l'on met au point un questionnaire, il est important de savoir sur quel type de machine s'effectuera la saisie, de façon à tirer davantage des possibilités offertes par les équipements envisagés et pour éviter des situations auxquelles il sera difficile ou impossible de faire face.

Le questionnaire doit être discuté et approuvé par toutes les catégories de personnel qui auront à l'utiliser à travers les différentes étapes du traitement : agents recenseurs, chiffreurs et personnel de saisie. Au cas où il y aurait un litige la priorité devrait être donnée à la collecte sur le terrain. Nous pensons cependant qu'il est possible de confectionner un modèle de questionnaire qui satisfasse tous ceux qui interviennent au long de la chaîne de travail.

Le questionnaire de recensement est souvent conçu sous forme de "feuille de ménage", de façon à recueillir sur une première partie les renseignements communs à tout le ménage sur une autre partie les caractéristiques individuelles de (chacun) des membres du ménage, puis sur une dernière partie des indications sur les événements survenus dans le ménage au cours des 12 derniers mois précédant la date de dénombrement.

Des bulletins individuels, c'est-à-dire un formulaire séparé pour chaque personne, ont été utilisés dans certains pays, en France, en République Fédérale d'Allemagne et en Suisse par exemple, la République Fédérale d'Allemagne et la Suisse passant du questionnaire directement à la machine au moyen d'un lecteur optique de marques (OMR) ou de caractères (OCR).

L'avantage d'un questionnaire par ménage est qu'il permet de réduire le volume de documents à produire et à manipuler, réduisant ainsi les coûts sur ce chapitre et facilitant par la même occasion le stockage, le traitement et la vérification.

En général, sur les "feuilles de ménage", on prévoit un espace pour dix personnes. Pour les ménages comportant plus de dix personnes, on utilisera des questionnaires supplémentaires en prenant soin d'indiquer qu'il s'agit de feuilles supplémentaires et en reportant sur chacun l'identifiant du ménage.

Il existe plusieurs méthodes de collecte d'informations sur le terrain. Dans certains cas, les réponses sont notées en clair sur le questionnaire pour être ensuite codées en salle, soit sur le questionnaire même si des emplacements sont prévus, soit sur une grille intermédiaire de chiffrage.

Dans d'autres cas, on a recours à la technique dite de pré-codage, qui consiste à ce que les réponses possibles soient imprimées sur le questionnaire avec le code correspondant qu'il s'agit d'entourer ou de cocher.

La méthode de plus en plus utilisée, consiste à combiner les deux précédentes : on utilise la technique de pré-codage pour les questions "fermées" et on note "en clair" les réponses aux questions "ouvertes", telles que la profession, la branche d'activité économique, ou le lieu géographique.

L'avantage du pré-codage est le gain de temps de codification qui peut être considérable. L'inconvénient majeur est que ce procédé exige beaucoup de place, pour peu que les nomenclatures utilisées comportent plusieurs modalités. Par ailleurs, un questionnaire pré-codé exige plus d'attention de la part du personnel d'entrée de données réduisant ainsi la vitesse de saisie en augmentant le risque d'erreurs de report des données.

Le précodage tel que décrit ici ne doit pas être confondu avec le chiffrement sur le terrain qui consiste à faire retranscrire en chiffre par l'agent recenseur la réponse donnée par la personne interrogée. Cette façon de procéder est à éviter, étant trop souvent une source d'erreurs. (1)

(1) Des expériences dans ce domaine auraient montré que des agents recenseurs commettaient en moyenne deux fois et demi plus d'erreurs que des codificateurs en salle, sur des modalités aussi simples que le sexe; il s'agissait d'indiquer 1 ou 2 pour masculin ou féminin au lieu de M ou F qui sont des codes mnémoniques.

Certains responsables de recensements et enquêtes statistiques répugnent d'ailleurs à faire coder des informations par des agents recenseurs ou des enquêteurs, en raison du peu de contrôle qu'on peut avoir sur le personnel de terrain qui est en général nombreux et temporaire et dont la durée de formation est limitée. Lorsque le chiffrement est entièrement prévu en salle, il faut éviter de recourir à des grilles intermédiaires de chiffrement.

Le chiffrement doit, dans ce cas s'effectuer sur le questionnaire même, de préférence immédiatement en dessous de l'information à coder. Pour faciliter la tâche au personnel de saisie, les grilles doivent être groupées, et si possible teintées, pour pouvoir guider l'oeil de l'opérateur de saisie. Lorsque les grilles de chiffrement sont séparées du questionnaire, le travail se complique et il y a de grands risques d'omission ou de double chiffrement. Par ailleurs, le contrôle devient plus difficile, deux ou plusieurs documents étant en jeu.

2.3. Le traitement des données

Le travail informatique proprement dit, consiste en les opérations suivantes :

- réception et vérification des questionnaires parvenant du terrain afin de s'assurer qu'il n'en manque pas;
- vérification manuelle et redressement des erreurs les plus évidentes ;
- chiffrement ;
- saisie des données sur cartes perforées, disquettes, cassette, disques ou bandes magnétiques;
- constitution de fichiers bruts sur support informatique triés selon le découpage administratif ;
- contrôle de validité des données et de cohérence des informations par ordinateur et redressement des données erronées ;
- constitution de la base de données, mise en tableaux et préparation des publications.

- Centralisation ou décentralisation

En raison de l'interférence de toutes les activités informatiques, il est préférable de les centraliser en un seul endroit pour éviter les problèmes de logistique et autres complications. Par ailleurs, la centralisation permet de garantir l'uniformité des méthodes de travail et d'assurer un bon système de communication. Si, pour des raisons d'espace par exemple, la décentralisation devient nécessaire, la séparation devra s'effectuer par type d'activité.

A. LE TRAITEMENT MANUEL.

Le traitement manuel est la partie de ces opérations qui ne nécessite pas le recours à un matériel électronique de traitement des données.

Il peut être utilisé comme étape préparatoire au traitement des données par ordinateur ou même constituer le seul moyen de traitement comme cela a été le cas pour les premiers tableaux du recensement de la population de l'Inde de 1971 qui ont été obtenus à partir d'un traitement entièrement manuel.

i) La réception et le contrôle

Dès la réception des questionnaires à l'Office central de statistique, on devra procéder à un premier contrôle pour s'assurer que tous les questionnaires sont renseignés pour toutes les zones de dénombrement. On contrôlera l'exhaustivité en se référant à la liste des zones de dénombrement communiquée par le service de cartographie. Les questionnaires seront ensuite classés temporairement selon des critères géographiques dans des étagères soigneusement étiquetées.

ii) Le redressement manuel et le codage

Le contrôle manuel du contenu des questionnaires consistera à s'assurer qu'ils sont complets et correctement remplis. Dans tout recensement ou enquête statistique il y a inévitablement des erreurs dans les informations recueillies. S'il y a généralement peu de falsifications délibérées de la part de la personne interrogée ou de l'agent recenseur, il peut par contre y avoir des biais systématiques ou des erreurs introduites par inadvertance lors de la collecte. Le rôle du redressement est de détecter ces erreurs et de les corriger.

Dans le cas de redressement manuel, on se concentre habituellement sur les erreurs les plus sérieuses, car une tentative de redressement de toutes les erreurs risquerait d'être une opération longue et dans la plupart des cas, pas très efficace.

Le codage est l'opération qui consiste à attribuer un code à une information relevée en clair sur le terrain. Un questionnaire n'est souvent jamais entièrement précodé et il y aura toujours des questions pour lesquelles il existe une grande variété de réponses. Dans ce cas, la consultation de nomenclatures est nécessaire pour l'attribution de codes. C'est le cas par exemple pour les codes géographiques, les professions et les branches d'activité économique.

Des erreurs peuvent être introduites lors de la phase de codification. Malgré la formation intensive qu'il a pu recevoir, le chiffreur peut attribuer un code erroné à une information. S'il s'agit d'un dépassement de valeur de code, c'est-à-dire si le code inscrit est en deça ou au-delà de la fourchette des valeurs permises pour la variable en question, cette erreur sera détectée par l'ordinateur lors du passage du programme de contrôle exhaustif de validité et de cohérence, qu'il faudra prévoir de toutes les façons. Par contre, si le code erroné est dans la tranche de valeurs permises, l'erreur ne sera pas détectée, à moins que cette valeur ne soit en contradiction avec une autre variable concernant la même unité statistique (ménage ou individu) et qu'un contrôle de cohérence soit effectué sur les deux variables considérées simultanément.

Si ces erreurs ne sont pas repérées, elles se retrouveront dans les tableaux qui seront publiés. Tous les systèmes de traitement des données ont donc leur système intégré de contrôle de qualité.

Le principal objectif du contrôle de la qualité statistique dans ce contexte est d'assurer qu'un nombre minimal d'erreurs apparaissent dans le produit final (les tableaux publiés) pour un coût déterminé, ou bien de déterminer le coût minimum qu'entraîne le niveau de qualité souhaité du produit final.

Les codeurs doivent être entraînés à travailler d'après les instructions reçues, selon des normes précises de production et de qualité. Les erreurs systématiques de codage peuvent être détectées au moyen de contrôle par sondage. Les codeurs qui n'arrivent pas à maintenir le niveau de qualité exigé, après un certain nombre de questionnaires, devraient être relevés de la section.

Les codes basés sur des calculs ou ceux qui sont une synthèse de plusieurs codes, tels que celui de la catégorie socio-professionnelle (basé sur la profession, la qualification et la branche d'activité économique) seront mieux attribués par ordinateur, qui peut effectuer cette opération sur instruction programmée de façon précise, uniforme et beaucoup plus rapide.

Dans le cadre des méthodes traditionnelles de codage et de vérification du codage, deux techniques sont utilisées pour détecter et corriger les erreurs des codeurs:

- La "vérification dépendante". Elle consiste à faire examiner par un vérificateur, un échantillon du travail pour détecter les erreurs. Si le nombre d'erreurs dépasse un niveau déterminé, l'ensemble du travail effectué par le codificateur devra être vérifié intégralement.
- Dans la "vérification indépendante", le vérificateur va coder les informations sur une grille de chiffrement. Le codificateur ne dispose pas de ces informations et doit chiffrer le questionnaire indépendamment. Après codage, le vérificateur compare ses codes avec ceux du codificateur.

Sur la base du nombre d'erreurs constatées, une décision est prise qui consiste à

- refaire le travail du codificateur
- donner un complément de formation au codificateur
- affecter le codificateur à d'autres tâches
- licencier le codificateur dans les cas extrêmes

Les contrôles et les redressements des données erronées doivent s'effectuer à tous les niveaux:

- sur le terrain, par l'agent recenseur lui-même qui devra examiner son travail à la fin de chaque interview pour s'assurer qu'il est complet et correct.
- par le contrôleur local, avant la transmission des documents à l'Office central.

L'avantage des contrôles sur le terrain est qu'il est possible de revenir au ménage en cas d'erreurs ou de doute.

- à l'Office central, selon des instructions précises.

A ce niveau il n'est plus possible de retourner sur le terrain et les critères de redressement se fondent avant tout sur le bon sens.

- le redressement par ordinateur.

Ce type de redressement est à privilégier dans un recensement en raison du volume à traiter, et nous reviendrons en détail sur cette importante méthode.

B. LA SAISIE DES DONNEES

Après le codage, l'étape suivante est la saisie des données sur support informatique.

La saisie sur carte perforée, utilisée dans le passé comme moyen privilégié de saisie, est tombée en désuétude et n'a presque plus cours de nos jours. Elle a été remplacée dans plusieurs pays par la saisie sur disquette, cassette ou directement sur disque ou bande magnétique. Il y a plusieurs types de ces machines (enregistreurs, systèmes multiclaviers sur mini-ordinateurs) dont les spécifications peuvent être très variables : la longueur des enregistrements qu'ils peuvent produire, les possibilités de contrôle en ligne (simultanément à la saisie) et le nombre de postes de travail (clavier) qu'on peut connecter à une même unité centrale. Bien qu'il soit possible d'effectuer sur certains types de machines des contrôles sophistiqués intra et inter-enregistrements au moment de la saisie, il n'est pas recommandé de multiplier ce genre de contrôles.

Le contrôle "en ligne" peut réduire la vitesse de saisie. De plus, si une erreur est détectée et si elle n'est pas une erreur de saisie (auquel cas la correction est évidente), l'opérateur n'est pas, dans la plupart des cas, en mesure de la rectifier lui-même. Il vaut mieux, dans ce cas, marquer l'enregistrement pour une correction ultérieure par une équipe spécialisée qui étudiera le problème à tête reposée.

Lorsque les opérations de codage et de saisie se chevauchent ce qui est souvent le cas, le contrôle effectué simultanément à la saisie est un bon moyen de déceler des erreurs systématiques de codage, toujours possibles surtout au début des opérations, il peut ainsi donner une idée de la qualité du chiffrement et permettre de prendre s'il le faut, les dispositions qui s'imposent pour rectifier le tir.

Ceci dit, le contrôle "en ligne" doit être conçu surtout pour détecter et corriger les erreurs de saisie et doit être limité aux contrôles de validité des codes et de présence de variables-clés.

La lecture optique est une autre technique d'entrée de données. Bien qu'elle soit utilisée dans certains pays lors de recensement comme nous l'avons vu, la lecture optique est généralement considérée comme mal adaptée à ce type d'application. Les contraintes imposées sont nombreuses (précisions des traces, taille et forme du papier, encrage spécial, qualité du papier). Les risques de rejet ou de substitution de caractère sont importants. De plus, la nécessité de passer par une étape de codification dans le cas d'un recensement, enlève toute sa raison d'être à la lecture optique qui est de passer directement de la source à la machine.

L'un des problèmes souvent rencontrés en saisie des données est la sous-estimation du volume du travail. Les estimations réalistes doivent tenir compte des indications suivantes :

- nombre de machines disponibles pour le travail donné
- nombre de vacations
- productivité des agents en termes de frappes/heures
- nombre de questionnaires et nombre moyen de caractères par questionnaire
- taux de vérification à effectuer.

D'autres facteurs tels que pannes machines, taux d'erreurs, absentéisme, doivent également être pris en considération et les estimations doivent être basées sur les conditions locales.

C. LE REDRESSEMENT PAR ORDINATEUR

Après le transfert des données sur support informatique, les fichiers bruts ainsi constitués doivent encore subir sur ordinateur un contrôle exhaustif de validité des données et de cohérence des informations en vue de leur redressement.

L'avantage du redressement par ordinateur est le degré de précision et d'uniformité qui ne peut être obtenu manuellement.

Contrôler manuellement de grandes quantités de documents est une tâche ardue et qui prend du temps. Certains pays ont mis plusieurs années à redresser manuellement les données de leur recensement en prenant le risque d'introduire d'autres erreurs.

Par ordinateur, il est possible de contrôler non seulement la validité d'une donnée, mais aussi d'effectuer des tests de cohérence sur plusieurs variables liées logiquement entre elles et étudiées simultanément.

Dans le cas d'un recensement les tests à effectuer sont :

- les tests de structure (cohérence inter-enregistrements)
- les tests de validité
- les tests de cohérence interne.

Les spécifications pour le programme de redressement doivent être mis au point par les démographes et statisticiens expérimentés, en collaboration avec les informaticiens responsables de la rédaction du programme. Afin d'éviter tout malentendu et des surprises désagréables, le programme de redressement devra être testé sur un "jeu d'essai" avant de le faire passer sur les fichiers réels.

Il est fortement recommandé pour un recensement de ne recourir à la correction manuelle que pour très peu de cas tels que les erreurs sur le code géographique et les erreurs de structure qui peuvent affecter les effectifs et leur répartition géographique. Une fois ces erreurs redressées, les dépassements de valeurs de codes et les cohérences internes doivent être redressées automatiquement, pour éviter de s'enliser dans des cycles itératifs saisie - traitement rejets - correction - saisie etc...

Jusqu'à une période récente, les programmes de redressement devaient être rédigés sur mesure. Il existe maintenant des logiciels de redressement très performants. COBOL-CONCOR par exemple, mis au point à l'origine par CELADE en langage Assembleur - IBM, a été réécrit en langage COBOL par le Bureau de recensement des Etats Unis qui l'a revu et amélioré par la même occasion en introduisant de nouvelles fonctions. COBOL-CONCOR permet d'effectuer :

- des tests de structure (contrôle de présence et multiplicité des enregistrements)
- des contrôles de fourchettes (contrôle des dépassements de codes)
- des tests de cohérence
- des remplacements (les valeurs erronées peuvent être remplacées automatiquement ou manuellement).

Pour les corrections automatiques, les techniques d'imputation peuvent être basées soit sur la méthode "Hot-Deck" soit sur la méthode de "Cold-Deck".

Dans la méthode de "Hot-Deck" l'information erronée est remplacée par une information plus vraisemblable. Les valeurs imputées proviennent des enregistrements précédents dont les autres caractéristiques sont similaires à l'enregistrement en cours. Les valeurs de remplacement sont ainsi changées constamment, au fur et à mesure du déroulement du fichier. Des statistiques sur les tests entrepris et les corrections effectuées sont produites en fin de traitement de sorte à donner à l'utilisateur une idée de la qualité des données sur lesquelles il aura à travailler.

Parmi les autres logiciels de redressement on peut citer encore :

- CAN-EDIT, mis au point par le service de statistique du Canada
- LEDA, produit par l'INSEE
- UNEDIT, mis au point par le bureau de statistique des Nations Unies.

UNEDIT ne permet pas de corrections par Hot-Deck. Certains de ces logiciels ne passent que sur un seul type de système.

D. RECODIFICATION ET CREATION DE SOUS-FICHIERS

Une fois un fichier "propre" obtenu, il sera peut-être nécessaire de procéder à une étape de recodification avant la phase de tabulation. C'est une opération qui consiste à créer de nouvelles variables en groupant des codes ou en combinant d'autres variables. La réduction d'une longue distribution de valeurs en un nombre plus restreint de modalités est nécessaire pour des variables telle que la branche d'activité, lorsque les codes d'origine sont détaillés. Durant cette étape, les données peuvent être réorganisées en sous-fichiers pour faciliter la mise en tableaux par thème d'étude.

E. LA MISE EN TABLEAUX

La phase finale du traitement des données est la mise en tableaux pour la publication. Il s'agit pour la majorité des tableaux du croisement de deux ou plusieurs variables.

Il existe des logiciels de tabulation qui réduisent considérablement le temps de programmation. Le plan de tabulation doit tenir compte des contraintes imposées par le logiciel, tels que la taille des tableaux, et le nombre de variables pouvant être croisées simultanément. Des détails sur ces logiciels peuvent être obtenus dans les publications suivantes :

- A Comparative Review of Statistical Software, Exhibition of Statistical program Packages, New-Dehli 1977, publié par Ivor Francis pour l'I.A.S.C.
- Statistical Software : A Comparative Review for Developers and Users, par Ivor Francis et Laurence Wood, El Sevier North Holland, New-York, 1980

Ces publications évaluent ces logiciels d'un point de vue statistique. Elles indiquent dans quelle mesure un statisticien peut les utiliser avec peu ou sans l'assistance d'informaticiens.

En voici quelques uns :

RGSP (Rothamsted General Survey Program), distribué par le Secrétariat de RGSP, Computer and Statistics Departments, Rothamsted Experimental Station, HARPENDEN, Angleterre.

Il permet de manipuler et imprimer des tableaux à partir de données d'enquête.

TPL (Table Producing Language), ce logiciel a été conçu et est distribué par le Bureau de statistique de travail des Etats-Unis. Pour l'Europe, la distribution est assurée par le centre informatique international à Genève (Suisse).

Ce logiciel produit des tableaux, génère de nouvelles variables par groupement de codes, suppression et réarrangement. Il est en mesure de prendre en charge des fichiers hiérarchisés et complexes.

XTALLY mis au point par le Bureau de statistique des Nations Unies, produit des tableaux à plusieurs dimensions, calcule des pourcentages, rapports, moyennes etc. Il peut pondérer les totaux à tous les niveaux de tabulation. Il s'adapte sur tout ordinateur équipé de logiciel GAP II (RPG II).

TAB - 68 : développé par le Bureau central de statistique de Suède.

LEDA : mis au point par l'INSEE (France)

CONCENTS ET CENTS : mis au point par le Bureau de recensement des Etats-Unis, sont des logiciels de tabulation très utilisés pour le dépouillement des recensements de la série 1970 en Afrique.

BIBLIOGRAPHIE

1. POPSTAN : analyse d'un exemple type pour les recensements de la population et des logements de 1980. US Department of Commerce, Bureau of the Census.

2. Traitement informatique des enquêtes statistiques

Groupe de démographie africaine
GDA, (IDP - INED - INSEE - MICOOP - ORSTOM) 1978

3. Recensements africains - monographies méthodologiques

Paris novembre 1980, GDA

4. National Household Survey, Comptability Programme : Survey data processing

United Nations, New-York 1982

5. Traitement manuel des données

Document préparé par la CEA pour le séminaire commun CEA/ENPC sur les statistiques des transports, Addis-Abeba, 11 - 22 juin 1984

3. LES PROGICIELS STATISTIQUES

- 3.1 Exemples de progiciels de contrôle et de correction des données

A. INTRODUCTION AU SYSTEME COBOL CONCOR (CONSISTENCY AND CORRECTION)

Depuis plus de 10 ans, le centre des programmes statistiques internationaux (ISPC) du bureau de recensement des Etats-Unis met au point et distribue des logiciels à l'intention des organisations statistiques des pays en développement. Dans le but d'améliorer la capacité informatique qui menaçait la bonne marche des programmes de recensement mondial, l'ISPC a mis au point un système de tabulation à entrées multiples (le CENTS) qui a été perfectionné d'année en année et est actuellement utilisé dans plus de 90 centres informatiques dans le monde.

Afin de satisfaire aux besoins informatiques analogues du tiers-monde dans le cadre des programmes de recensement mondial de 1980, l'ISPC a élaboré une nouvelle version du CONCOR, programme spécialement conçu pour le redressement et la correction automatique des données de recensement et d'enquête. Le CONCOR offre un langage de commande puissant et perfectionné qui permet au programmeur ou au spécialiste d'élaborer des spécifications de redressement exécutables au moyen de techniques comme l'examen inter-enregistrement, l'imputation "hot deck" et le contrôle de tolérance.

Le COBOL CONCOR version 2.0 a été offert au début de 1980. La version 2.2, qui comprend un certain nombre d'améliorations sur le plan de l'efficacité date de 1981. Le CONCOR 2.2 a été largement distribué parmi les organismes statistiques nationaux et utilisé pour redresser les données de plusieurs recensements de population nationaux. La dernière version en date du CONCOR, la version 2.3, présente de nouvelles fonctions d'application pratique suggérées par les utilisateurs du système. Le présent document examine ces différentes innovations. Le financement du CONCOR 2.3 a été assuré par l'Agence des Etats-Unis pour le développement international (USAID).

Le présent document définit les principales possibilités de redressement du CONCOR, décrit le langage et les restrictions du système CONCOR et explique les impératifs d'ordre informatique préalables à son installation. Il est également possible d'obtenir des informations plus détaillées concernant le CONCOR 2.3, notamment:

- un manuel d'emploi
- un manuel de référence du système
- un guide des messages de diagnostic
- un manuel d'explication du système
- un guide d'installation.

Pour obtenir cette documentation détaillée ainsi que des informations sur la disponibilité du CONCOR, prière de s'adresser au:

International Statistical Programs Center
U.S. Bureau of the Census
Washington, D.C. 20233

1) Principales fonctions de redressement

La liste ci-dessous donne un aperçu des possibilités du CONCOR en matière de redressement et de correction de données:

Contrôles de structures

Ces contrôles permettent de déterminer que tous les enregistrements requis pour un type de formulaire sont fournis et qu'aucun enregistrement superflu n'est inclus.

Contrôles d'application

Ces contrôles sont destinés à déterminer si la valeur d'une variable se trouve dans les limites valables ou en dehors. Souvent appelés "validation des données", ces contrôles traitent une variable à la fois.

Contrôles de compatibilité

Ces contrôles consistent à comparer deux ou plusieurs variables d'un questionnaire et à vérifier la compatibilité de leur valeur. Ces variables peuvent se trouver dans le même enregistrement ou dans différents enregistrements.

Correction automatique

Lors d'un recensement ou de toute autre enquête à grande échelle, il est fastidieux de faire les corrections à la main. Le CONCOR permet non seulement de repérer les données incorrectes ou contradictoires mais également de procéder aux corrections selon les spécifications de l'utilisateur.

Fichier de sortie sans erreur

Le CONCOR permet de créer un fichier de sortie auquel toutes les corrections conformes aux spécifications de l'utilisateur ont été apportées. Ce fichier possède un format identique au fichier d'entrée. En fait, le CONCOR peut traiter ce fichier de sortie et vérifier si certaines contradictions ont été introduites dans le passage machine CONCOR initial.

Création de fichier

Le CONCOR permet à l'utilisateur de créer un autre dossier de sortie qui combine des données d'entrée ou de nouvelles variables. Cette fonction rend possible la création de fichiers de dérivation contenant des données recodées ou reformatées.

Rapports d'erreur

Le CONCOR élabore des statistiques complètes sur les tests de redressement réalisés sur les données d'entrée. Les rapports de ces statistiques sont produits à la demande de l'utilisateur.

Le CONCOR permet de définir jusqu'à 50 types différents d'enregistrement, chacun d'entre eux pouvant avoir un maximum de 999 occurrences dans chaque questionnaire. Par ailleurs, il est possible de définir 700 identificateurs uniques à l'utilisateur.

ii) Le langage CONCOR

Le CONCOR utilise un langage de programmation simple qui offre à l'utilisateur la possibilité d'écrire des programmes structurés. Une analyse complète du programme de l'utilisateur est réalisée afin d'identifier et d'expliquer les erreurs et d'identifier les problèmes avant que le programme ne serve au traitement en série du fichier de données. Faciles à mettre à jour, les programmes CONCOR sont à la fois à la portée des programmeurs et des experts des sujets à étudier.

Au moyen du langage CONCOR, l'utilisateur décrit le fichier de données d'entrée, définit les spécifications de redressement et désigne le type de statistiques à produire. Le langage de l'utilisateur se compose de trois volets :

Le volet dictionnaire

Ce volet comprend les commandes qui décrivent et définissent les fichiers d'entrée et de sortie du système, les commandes qui définissent les rubriques de données d'entrée à redresser et celles qui définissent les nouvelles variables ou éventails de données. Ces commandes délimitent le champ de redressement de l'utilisateur. A l'instar de nombreux autres langages de programmation, chaque rubrique est assortie d'un nom. Toute rubrique - nouvelle ou calculée - requise pour le programme CONCOR devra également figurer dans le dictionnaire.

Le volet exécution

Ce volet comprend les commandes destinées au redressement effectif des données. Ici aussi, l'utilisateur doit spécifier le ou les types de statistiques erronées que le CONCOR est censé compiler et mémoriser en cours de traitement des données. Le volet exécution est de loin le volet le plus complexe et le plus puissant des trois.

Le volet rapport

Le programme de redressement du CONCOR, appelé EDITOR, génère un fichier de statistiques d'erreur. Les instructions de commande du volet rapport servent uniquement à permettre à l'utilisateur de produire des rapports sur les erreurs décelées et les corrections faites sur la base des informations continues dans ce fichier d'erreur.

Les informations contenues dans les trois volets du programme CONCOR constituent des entrées au système CONCOR. Le CONCOR analyse les commandes de l'utilisateur afin de s'assurer de la justesse syntaxique. Une fois que tout est correct, le CONCOR va générer un programme source en COBOL. Ce programme est compilé et mis en format exécutable. On peut alors le tester au moyen d'un échantillon de données de façon à s'assurer que sa logique est correcte et que les résultats obtenus sont ceux attendus. La procédure de redressement, de correction et d'essai sera répétée jusqu'à ce qu'un programme EDITOR correct soit mis au point. La version finale du programme EDITOR pourra ensuite être utilisée avec les données de recensement ou d'enquête de l'utilisateur. Le programme EDITOR lit les données non redressées, les redresse selon les spécifications données par l'utilisateur et produit un fichier de données "au net". L'EDITOR produit également un fichier de statistiques de redressement qui sera imprimé par le biais du volet rapport du CONCOR.

iii) CONCOR Program Outline

DICTIONARY-DIVISION:

. <DICTIONARY-ATTRIBUTES-SECTION:>

DICTIONARY-NAME:

. <FILE-SECTION:>

INPUT-FILE:
OUTPUT-FILE:
WRITE-FILE:
SAVE-ARRAY-FILE:
ERROR-FILE:

. <IDENTIFICATION-CONTROL-SECTION:>

AREA-CONTROL:
QUESTIONNAIRE-CONTROL:
RECORD-CONTROL:

. <INPUT-RECORD-SECTION:>

DEFINE-RECORD:

. <WORKING-DATE-SECTION:>

NEW-DATE:
ARRAY-DATE:

END<-DIVISION>;

EXECUTION-DIVISION:

< RUN-CONTROL-SECTION:>

DIVIDE<-CHECK>;
FIELD-TRUNCATION<-CHECK>;
NUMERIC-CLASS<-CHECK>;
SPECIAL-SYMBOL<-CHECK>;

< REPORT-CONTROL-SECTION:>

GENERATE<-EDIT-STATISTICS>;
COUNT-IMPUTES;
COLLECT-FREQUENCIES;

< EDIT-SPECIFICATION-SECTION:>

PROLOG<-ROUTINE>:

ALLOCATE; (ALLOC)
ASSERT; (ASRT)
CALL;
DRECODE; (DRCD)
EXIT;
GRECODE; (GRCD)
IF;
LET;
LIST;
OUTPUT
RANGE; (RNG)
RETURN;
STOP;
SUBSCRIPT<-CHECK;
SUBROUTINE : (SUBR)
UNTIL;
UPDATE; (UPD)
WRITE; (WRT)

FILTER<-ROUTINE> (name-of-record-type):

ALLOCATE; (ALLOC)
ASSERT; (ASRT)
.
.
.
WRITE; (WRT)

EPILOG<-ROUTINE>:

ALLOCATE; (ALLOC)
ASSERT; (ASRT)
.
.
.
WRITE; (WRT)

END<-DIVISION>;

REPORT-DIVISION

```
.    < DISPLAY-CONTROL<-SECTION>:>

        DISPLAY<-EDIT-STATISTICS>;

.    < TOLERANCE-CONTROL<-SECTION>:>

        ERROR-RATE<-CHECK>;
        REJECT<-FILE>;
```

END<-DIVISION>;

iv) Derniers perfectionnements du CONCOR

Ci-dessous figure une liste des nouvelles fonctions prévues dans la dernière version du CONCOR, la version 2.3 :

- Modification du sous-système d'analyse du langage de commande de manière à ce que les appels de procédure en provenance des trois volets du langage CONCOR puissent être lus comme un fichier enchaîné simple. Le CONCOR 2.2 nécessitait 3 fichiers séparés, un par volet.
- Possibilité de traiter des champs de données et des variables alphanumériques d'une manière compatible au moyen de données numériques. Le CONCOR 2.2 avait une capacité limitée de traitement des données alphanumériques.
- Possibilité de sauver le contenu de séries "hot deck" d'un passage machine CONCOR à l'autre. Ce perfectionnement est particulièrement utile lors du traitement par lots.
- Variables globales supplémentaires, aussi appelées identificateurs internes, permettant à l'utilisateur de vérifier le nombre total de questionnaires et/ou d'enregistrements traités.
- Exécution d'un sous-programme. Cette fonction permet à l'utilisateur de définir et de faire appel à des sous-programmes de sorte que les instructions à exécuter depuis différents points du programme CONCOR ne doivent pas être répétées.
- Possibilité d'entrer un code COBOL utilisateur dans le programme CONCOR: le programme EDITOR peut ainsi exécuter des fonctions difficiles à réaliser avec le langage CONCOR.

- Execution de commandes permettant à l'utilisateur de couper certains contrôles internes du programme EDITOR et d'ainsi améliorer la vitesse d'exécution de l'EDITOR et de réduire son ampleur.
- Possibilité de capture sélective et d'indication des distributions de fréquences des valeurs imputées. Un nouveau rapport de statistiques est à présent disponible qui montre le nombre d'entrées pour chaque valeur ou série de valeurs imputées à partir d'une série.
- Exécution de contrôles internes permettant à l'utilisateur d'éviter toute séquence erronée dans l'exécution des modules de programmes CONCOR.
- Modification des routines d'analyse de commandes de manière à uniformiser les règles de ponctuation pour les trois volets langage. Dans les versions précédentes du CONCOR, les règles de ponctuation du volet dictionnaire différaient de celles des autres volets.
- Possibilité d'obtenir des identificateurs internes CONCOR dans la liste de valeurs de la commande RANGE (série). Auparavant, l'utilisateur devait inclure un code spécial lorsqu'une rubrique de données destinée à une série valide de valeurs pouvait à juste titre être laissée en blanc.
- Modification du processus de génération du programme EDITOR de manière à modulariser le contrôle des suffixes de série. Ce perfectionnement permet de réduire l'ampleur du programme EDITOR et d'augmenter son efficacité.

v) Restrictions applicables à l'entrée des données

Lorsqu'on crée un fichier qui sera redressé par le CONCOR, il importe de tenir compte des restrictions suivantes :

- Tous les enregistrements doivent avoir la même longueur.
- Tous les enregistrements relatifs à un questionnaire en particulier doivent être physiquement contigus sur le fichier.
- Un code unique d'identification de questionnaire doit figurer au même endroit sur chaque enregistrement. Tous les enregistrements appartenant au même questionnaire doivent avoir le même code. On peut satisfaire à cette condition en ayant une série de champs qui donnent des informations de codage géographique et un numéro de ménage unique au niveau géographique le plus bas.

- Chaque enregistrement doit avoir un code d'enregistrement type d'une longueur déterminée et à un endroit fixe.
- Si l'utilisateur désire que le CONCOR compile des statistiques de redressement séparément pour chaque région géographique unique, tous les questionnaires pour une région particulière de contrôle géographique doivent être contigus sur le fichier.

vi) Informations sur le système

Le CONCOR se compose de plus de 20 programmes COBOL. Le plus grand module de programme requérant un redressement de liens dans le système CONCOR nécessite une capacité de mémoire de 128 K octets, cette dernière comprend tous les modules de communication entrée/sortie requis par le programme. Cette capacité de mémorisation est basée sur un système informatique IBM 370 OS modèle 168. Lors du redressement des liens, le programme EDITOR COBOL peut nécessiter une capacité de mémoire supérieure à celle des autres programmes CONCOR. Son ampleur dépend du nombre et de la complexité des instructions CONCOR codées par l'utilisateur.

Une capacité minimale de mémorisation sur disque de 5 mégaoctets est nécessaire pour la bibliothèque du programme qui comprend la source et l'objet des programmes COBOL, un espace de travail pour les fichiers intermédiaires et la bibliothèque de chargement des modules de programmes requérant un redressement de liens. Le matériel nécessaire comprend encore un poste d'entrée standard (bande, disque, lecteur de cartes), une imprimante directe et un poste de sortie standard (bande, disque, perforatrice).

vii) Matériel compatible

Le CONCOR 2.3 est actuellement disponible pour les ordinateurs suivants :

IBM - OS (370,4300,303X)

IBM SYSTEM/38

UNIVAC 1100 SERIES

Le CONCOR 2.2 est disponible pour les ordinateurs suivants:

IBM - OS (370, 4300,303X)

IBM - DOS/VSE

NEC 500 (ACOS-4)

HONEYWELL BULL

PERKIN-ELMER 3250

WANG -VS80

UNIVAC 1100 SERIES

B. UNEDIT

1) Introduction

- Le système UNEDIT est un ensemble portatif de logiciel généralisé qui a été mis au point pour répondre aux besoins communs d'apurement correspondant à la présentation en tableau des données de recensement et d'enquête. L'ensemble UNEDIT utilise la langue RPG-II et peut être passé sur les petits ordinateurs ayant un minimum de mémoire interne de 32 K-octets.
- Il n'est pas exigé que les utilisateurs de l'UNEDIT connaissent bien les techniques de programmation informatisée; il est préférable qu'ils soient des statisticiens au courant des détails du questionnaire, des règles de codage des données et des tableaux définitifs présentant les données d'un recensement ou d'une enquête.
- La conception du système UNEDIT a donné une place de choix à la facilité avec laquelle l'utilisateur peut apprendre à connaître les possibilités offertes par le système, les spécifications exigées et s'en servir sur de petits ordinateurs. Les utilisateurs n'ont qu'à remplir les feuilles à mise en page préalable des spécifications d'édition et n'ont pas besoin d'écrire des informations en langue-machine.
- L'UNEDIT commence par examiner les spécifications de l'utilisateur et s'il y a des erreurs de syntaxe, imprime les messages de diagnostic. Au moment de l'exécution, il effectue les opérations d'édition en fonction de règles découlant des spécifications de l'utilisateur. Durant l'exécution, il imprime les statistiques d'erreurs par catégorie d'erreur et par zone de données.
- Bien que l'UNEDIT soit conçu essentiellement pour la présentation des tableaux de recensement ou d'enquête sur la population, on peut s'en servir pour l'édition de données d'autres tableaux statistiques; en effet, l'UNEDIT peut traiter un fichier comportant diverses catégories d'enregistrements et exécuter une tâche d'édition quantitative et qualitative. De plus, étant donné que l'UNEDIT peut facilement réaliser l'édition conformément aux instructions du statisticien, on peut s'en servir pour vérifier et modifier les règles d'édition au cours des diverses étapes de la planification et de la préparation des tableaux, depuis l'enquête-pilote jusqu'à la décision définitive concernant les règles d'édition des tableaux.

11) Caractères et possibilités de l'ensemble UNEDIT

- Inscriptions n'utilisant pas la langue-machine faciles à apprendre, faciles à utiliser et faciles à évaluer.

Afin d'employer l'ensemble UNEDIT, l'utilisateur n'a qu'à remplir les feuilles à mise en page préalable pour indiquer les spécifications. Ces formulaires sont les suivants :

- Formulaire 1. Formulaire des spécifications de description du fichier,
- Formulaire 2. Formulaire des spécifications d'enregistrement d'entrée,
- Formulaire 3. Formulaire des spécifications des codes valides,
- Formulaire 4. Formulaire de calcul avant l'édition et
- Formulaire 5. Formulaire des spécifications de vérification d'incompatibilité.

Les spécifications concernant les règles d'apurement, telles que la vérification de validité et la vérification d'incompatibilité, sont portées sur le formulaire des spécifications de telle manière qu'un statisticien puisse rédiger les règles de présentation en utilisant ses propres vocables; ces règles sont communiquées au personnel de programmation au lieu de laisser ce dernier les mettre dans l'ordinateur en langage de programmation.

- Souplesse de la structure du fichier des données d'entrée-Diverses catégories d'enregistrements sont acceptées.

L'UNEDIT peut traiter un fichier d'entrée comportant diverses catégories d'enregistrements. L'utilisateur précise dans le paramètre "mode" du formulaire 1 si son fichier d'entrée est un fichier à enregistrement unique ou un fichier à enregistrements multiples. Dans ce dernier cas, il précise si une vérification entre les enregistrements est exigée ou non. En fait, il existe quatre catégories de traitement de fichier appelées "mode", que l'utilisateur peut choisir :

- Modell: Fichier à enregistrement unique
- Mode 2: Fichier à enregistrements multiples - vérification entre enregistrements non exigée
- Mode 3: Fichier à enregistrements multiples - vérification entre enregistrements exigée

Mode 4: Fichier de la catégorie questionnaire des ménages (fichier à deux enregistrements) dans lequel un enregistrement pour le ménage (ou l'unité de logement) est suivi d'enregistrement pour une ou plusieurs personnes.

- Dépistage des erreurs - vérification de validité, vérification d'incompatibilité et vérification structurale.

- Vérification de validité

La vérification de validité doit vérifier si un code ou une valeur de la zone des données est un code valide ou se trouve dans la gamme acceptable en le comparant à une liste de codes ou valeurs que fournit l'utilisateur. L'utilisateur se borne à dresser la liste des codes valides ou d'une gamme acceptable de valeurs par zone de données dans le formulaire 2, formulaire des spécifications des codes valides. Quand on dépiste un code ou une valeur invalide, un message d'erreur est imprimé et un indicateur de code de non validité (attribué par l'utilisateur) est placé dans la zone de donnée erronée. L'utilisateur peut préciser soit des codes de validité individuels soit une gamme acceptable de valeurs (une limite inférieure et une limite supérieure de la gamme).

- Vérification d'incompatibilité

La vérification d'incompatibilité permet de voir si une suite de codes ou les valeurs de plusieurs zones de données sont compatibles ou non lorsqu'on les prend dans leur ensemble, bien que chaque élément puisse être en soi valide et acceptable. L'utilisateur précise les conditions d'incompatibilité dans le formulaire 5, formulaire des spécifications de vérification d'incompatibilité.

- Vérification d'incompatibilité à l'intérieur d'un enregistrement

La vérification d'incompatibilité à l'intérieur d'un enregistrement examine les zones de données d'un même enregistrement. L'UNEDIT exécute la vérification à l'intérieur d'un enregistrement dans tous les "modes".

. Vérification d'incompatibilité entre les enregistrements.

La vérification d'incompatibilité entre les enregistrements examine des zones de données contenues dans plus d'un enregistrement. L'UNEDIT exécute la vérification entre les enregistrements lorsque le "mode" du traitement du fichier est 3 ou 4, c'est-à-dire lorsque l'ordinateur traite un fichier à enregistrements multiple ou un fichier de questionnaires de ménage.

. Vérification quantitative et vérification qualitative

Les zones de données qui font l'objet de la vérification d'incompatibilité peuvent être quantitatives ou qualitatives.

. Calcul arithmétique et comparaison des zones de données

Dans la vérification d'incompatibilité, il est parfois nécessaire de calculer des valeurs de zones ou de comparer les valeurs de deux zones afin de déterminer la condition d'incompatibilité. UNEDIT permet de l'effectuer en se servant du formulaire 4, formulaire de calcul avant l'édition.

- Vérification structurelle

La vérification structurelle permet de voir s'il y a un enregistrement absent, alors qu'il doit être présent dans le fichier.

. Comptage à l'entrée par secteur

UNEDIT imprime automatiquement le nombre d'enregistrements d'entrée correspondant au secteur désigné par l'utilisateur, par exemple la zone de dénombrement.

. Dépistage d'enregistrement absent

Lorsqu'on traite un fichier à enregistrements multiples, tout enregistrement absent sera automatiquement dépisté et un message sera imprimé.

. Comptage des enregistrements de personnes dans le fichier en mode 4

Dans le fichier en mode 4, où l'on traite un questionnaire de ménage, le nombre d'enregistrements de personne faisant partie d'un ménage est automatiquement compté et mis à la disposition de l'utilisateur aux fins d'essais dans le formulaire 5: formulaire des spécifications de vérification d'incompatibilité.

II- Indication d'erreur

L'UNEDIT imprime les messages d'erreur et les statistiques d'erreur suivants :

. Message d'erreur sur enregistrement individuel refusé

Message de code invalide : nom de zone, position de zone et code invalide

Incompatibilité : numéro de la condition d'incompatibilité

Statistiques d'erreur

A la fin de chaque secteur :

- (1) Nombre total d'enregistrements d'entrée
- (2) Nombre d'enregistrements refusés (total)
- (3) Nombre d'enregistrements refusés ayant un code invalide
- (4) Nombre d'enregistrements refusés ayant des codes incompatibles

A la fin du fichier :

- (1) Nombre total d'enregistrements d'entrée
- (2) Nombre d'enregistrements refusés (total)
- (3) Nombre d'enregistrements refusés ayant des codes invalides
- (4) Nombre d'enregistrements refusés ayant des codes incompatibles
- (5) Nombre de codes invalides par rubriques de données
- (6) Nombre d'enregistrements incompatibles par condition d'incompatibilité.

- Correction automatique (= affectation)

L'utilisateur d'UNEDIT peut affecter un code ou une valeur à la zone particulière de données quand un ensemble de conditions indiquées sur le formulaire 5 est rempli. Il peut le faire en inscrivant le nom de la zone de données auquel est affecté un nouveau code à la fin de l'ensemble de conditions du formulaire 5.

La zone en blanc ou la zone de code d'invalidité, dépistée durant la vérification du code de validité et munie d'un d'un indicateur d'erreur, peut également recevoir un nouveau code à ce stade.

- Options de sortie

L'utilisateur peut choisir les options de sortie suivantes :

- (1) Messages d'erreur et statistiques d'erreur seulement
- (2) (1) plus le fichier corrigé

- Messages de diagnostic concernant les spécifications de l'utilisateur :

L'ensemble UNEDIT vérifie en détail les spécifications de l'utilisateur et imprime le message de diagnostic concernant les erreurs de syntaxe ainsi que les redondances et les contradictions que renferment les ensembles de conditions d'incompatibilité.

3.2. EXEMPLES DE PROGICIELS DE TABULATION

A. CENTS 4

1) Introduction au CENTS 4.

(Census Tabulation System Version 4)

Depuis plus de dix ans, le centre des programmes statistiques internationaux du Bureau de recensement des Etats-Unis, patroné par l'Agence des Etats-Unis pour le développement international (AID), s'attache à mettre au point des logiciels informatiques généralisés à l'intention des organismes de statistiques des pays en voie de développement. Face aux besoins critiques d'amélioration de la capacité informatique constatés durant les programmes de recensement mondial de 1970, l'ISPC a élaboré un système général de tabulation à entrées multiples pour ordinateurs IBM (le CENTS) qui est perfectionné d'année en année. Le COCENTS, un autre système général de tabulation, a ensuite été mis au point en COBOL afin de pouvoir élargir la gamme d'ordinateurs compatibles. Les deux systèmes utilisent un langage de commande similaire pour produire le même type de tabulation. Il est apparu, sur la base des informations fournies par les utilisateurs du CENTS et du COCENTS, que des améliorations s'imposaient en termes de flexibilité et de compatibilité.

En 1979, un accord a été signé avec le Département des sciences et de la technologie au Bureau de soutien au développement de l'AID avec pour but de planifier la conception d'un système de tabulation généralisé amélioré, le CENTS 4. Cette version du système CENTS constitue la première approche unifiée au système de tabulation CENTS/COCENTS. Outre qu'il uniformise le langage utilisé, le CENTS 4 apporte également des améliorations importantes par rapport aux versions précédentes du système. Parmi les innovations les plus remarquables figurent une plus grande flexibilité de codage ainsi que des nouvelles fonctions et capacités dans le langage utilisateur qui permettent de réaliser des rapports sous forme de tableaux prêts à l'impression.

ii) Restrictions applicables à l'entrée de données primaires

Le CENTS 4 peut traiter efficacement de très grands fichiers de données. Aucune restriction n'existe quant au nombre maximum de cas traités.

Il est également possible de traiter un fichier possédant des formats de données multiples. Le CENTS 4 traite les entrées en format binaire, décimal condensé ou sous forme de caractères. Les données binaires peuvent être des demi-mots ou des mots entiers. Il n'est donc pas nécessaire de reformater la plupart des fichiers de données. Toutefois, le CENTS 4 n'a pas été conçu pour traiter des enregistrements de données à longueur variable ou des rubriques de données d'entrée à format libre (code de source et valeur). Le CENTS 4 peut reconnaître plus d'un cas par enregistrement et traiter une structure hiérarchique de fichier du type habituellement utilisé dans les fichiers sur le logement et la population. Dans la mesure où des rubriques de données d'autres enregistrements sont requises pour un tableau, seules les rubriques nécessaires au tableau doivent être mémorisées. On peut ainsi réduire la capacité de mémoire requise pour traiter les fichiers à types d'enregistrements multiples par observation. Les fichiers de données créés par d'autres logiciels statistiques comme le système de redressement des données et d'imputation CONCOR peuvent servir d'entrées au CENTS 4.

iii) Possibilités de tabulation

Le CENTS 4 permet d'obtenir un large éventail de tableaux fondés sur de nombreux types de données. Par exemple, il est possible de réaliser des tableaux en utilisant des données censitaires, économiques, agricoles, démographiques et même des données d'enquêtes sur la fécondité, les ménages, la nutrition, l'énergie, l'éducation et la main-d'oeuvre.

Les tableaux multi-dimensionnels ayant jusqu'à six dimensions sont réalisables avec le CENTS 4. Le langage est conçu de telle manière que les plans de géocodage sont facilement manipulés et permettent d'obtenir des tableaux de totaux et d'agrégations spéciales.

Un tableau ne doit pas nécessairement représenter un simple croisement de deux rubriques d'information. Une rubrique peut également être tabulée en utilisant différentes agrégations ou décompositions dans un tableau simple. De même, les rubriques multiples peuvent être tabulées dans le même tableau. Le CENTS 4 est capable de générer des tableaux multiples en un passage machine simple du fichier de données. La taille du programme augmente à mesure que l'utilisateur ajoute des tableaux et des commandes au passage machine. Aussi, la seule limite imposée est la capacité de l'ordinateur (mémoire primaire disponible).

Le langage du CENTS 4 permet à l'utilisateur de recoder les rubriques de données d'entrée. Ce faisant, il est possible de grouper des valeurs ou de créer des variables d'index. Toute rubrique recodée peut être mémorisée pour tous les tableaux qui requièrent cette nouvelle variable.

Aucune restriction ne s'applique à l'inclusion d'une observation dans l'univers d'un tableau. La définition de l'univers peut varier d'un tableau à l'autre en cours de passage machine et dépend totalement du contrôle de l'utilisateur.

En ce qui concerne les fichiers de données d'enquête, il est possible de définir un facteur de pondération ou d'échantillonnage pour l'ensemble du fichier de données ou pour chaque cas en particulier. Les facteurs de pondération fractionnés sont plus difficiles à traiter que les nombres entiers. Le CENTS 4 traite les données par nombre entier. En conséquence, si des facteurs de pondération fractionnés sont utilisés, il faudra modifier les tableaux avant l'impression.

Parmi les nouvelles fonctions du CENTS 4 figure la manipulation des séries. L'utilisateur peut définir une série ayant jusqu'à deux dimensions et mentionner une série dans la plupart des commandes de tabulation. Le CENTS 4 permet également de définir des sous-programmes et de procéder à des appels conditionnels de sous-programmes. Il permet également d'établir des références variables de tableaux et de série. La fin du traitement des fichiers a été simplifiée et de nouveaux identificateurs réservés ont été ajoutés afin d'offrir à l'utilisateur un meilleur contrôle du programme.

iv) Présentation des tableaux

Le CENTS 4 possède un sous-système séparé destiné à préparer les tableaux pour l'impression. Il est également possible d'effectuer des mesures statistiques au cours de cette phase. La présentation des tableaux peut être simple ou aussi détaillée que l'utilisateur le souhaite. Le CENTS 4 offre une plus grande variété de formats de tableaux que le COCENTS et les versions précédentes du CENTS.

La fonction qui consiste à définir le texte des régions géographiques constitue un sous-système séparé du CENTS 4. Elle permet à l'utilisateur de créer un fichier généralisé des noms de régions susceptible d'être utilisé dans de nombreuses applications du CENTS 4. Un bon exemple d'application serait la création en une seule opération d'un plan de géocodage national informatisé.

Etant donné que les représentations internes du tableau sont mémorisées entre le traitement des données d'entrée et les phases de présentation des tableaux, il est possible de réimprimer les tableaux avec différents en-têtes ou notes en bas de page sans devoir retraiter la totalité du fichier des données d'entrée. Cette fonction est particulièrement importante si les tableaux sont élaborés à partir de données de recensement et que des erreurs surviennent dans la dénomination des tableaux. La même fonction permet également de produire rapidement des copies multiples des tableaux finaux.

Le travail détaillé requis pour le codage de tableaux complexes assorties de labels compliqués constituait l'inconvénient majeur du COCENTS et des versions précédentes du CENTS. Le CENTS 4 rend cette tâche beaucoup moins ardue du fait de la simplification du langage d'interface de l'utilisateur. Les instructions de définition de texte permettent un codage et des modifications très faciles à exécuter. Le contrôle de l'utilisateur sur le formatage des tableaux a été grandement amélioré et simplifié. L'utilisateur peut soit définir le format des tableaux dans le détail ou prendre les défauts de système. Une option d'impression en arabe a été ajoutée et l'on prévoit d'offrir d'autres jeux de caractères.

Un autre inconvénient des progiciels de tabulation précédents était qu'ils ne permettaient d'imprimer qu'un seul chiffre après la virgule. Le CENTS 4 prévoit 3 chiffres après la virgule.

v) Mesures statistiques

Le CENTS 4 génère des mesures statistiques telles que sommes, moyennes, médianes et distributions de pourcentage. L'utilisateur décide lui-même de l'ampleur de l'utilisation et de l'univers de ces mesures.

Pour le calcul des médianes, on utilise une technique d'approximation linéaire qui ne sera pas toujours aussi précise que des techniques plus compliquées. Toutefois, le CENTS 4 n'est pas en mesure de produire des mesures statistiques complexes telles que l'analyse de variance, le chi carré, la dissymétrie, etc.

vi) Caractéristiques de langage

Le langage de commande utilisé sur le CENTS 4 constitue une amélioration importante par rapport au COCENTS et aux versions précédentes du CENTS. Les noms de commande et les identificateurs ont au plus 8 caractères et les composantes de la commande ont un format libre, c'est-à-dire qu'elles ne dépendent pas de colonnes. Les commentaires peuvent être entremêlés avec les instructions de commande du CENTS 4 soit sur une ligne séparée soit sur la même ligne à la droite de l'instruction. Le langage du CENTS 4 peut être appris moyennant un cours de deux à trois semaines.

vii) La procédure CENTS 4

Les quatre sous-systèmes de base du CENTS 4 sont le CENTAL, le CENCON, le CENAREA et le CENPREP. Chaque sous-système se compose essentiellement de deux étapes. La première étape, aussi appelée démarrage ou phase 1, consiste à lire les instructions de commande source de l'utilisateur, analyser leur exactitude et les imprimer avec des messages de diagnostic. Dans la mesure où l'on ne découvre aucun problème de la phase 1, la seconde étape, appelée exécution ou phase 2, consiste à réaliser les opérations spécifiées dans les instructions de commande de l'utilisateur.

Le sous-système CENTAL élabore les matrices initiales de tableaux. Sur la base des instructions fournies par l'utilisateur, les enregistrements des données d'entrée sont sélectionnées pour la tabulation et les processus de traitement des données sont exécutés. Des matrices PROVINCE sont incorporées pour créer des matrices REGION.

Le sous-système CENCON n'est utilisé que lorsqu'il est nécessaire de produire des tableaux hiérarchiques, par région géographique par exemple.

Les matrices initiales de tableaux sont créées par le sous-système CENTAL en fonction de la plus petite région géographique définie. Le sous-système CENCON regroupe les matrices en totaux uniques afin de pouvoir les transmettre au CENPREP. Ainsi, si l'utilisateur désire créer un ensemble de tableaux par REGION, PROVINCE et DISTRICT, il lui suffit de définir cette hiérarchie géographique au CENTAL et ce dernier générera des matrices pour chaque combinaison de codes REGION, PROVINCE et DISTRICT. Le sous-système CENCON créera une matrice pour chaque province mentionnée dans les données en rassemblant toutes les matrices de district pour cette province.

Le but du sous-système CENAREA est de permettre à l'utilisateur d'assigner des noms de régions à imprimer par le CENPREP. Le texte du nom de région est imprimé au début de chaque tableau et sert à identifier l'information du tableau avec une entité géographique. Le fichier de noms de régions produit par le CENAREA peut être mémorisé dans la mesure où l'on prévoit plusieurs passages machine utilisant le même ensemble de valeurs de régions.

Les matrices établies au cours des phases CENTAL et CENCON sont imprimées par le CENPREP. Les en-têtes, intitulés, légendes et notes en bas de page sont imprimés s'ils ont été définis par l'utilisateur. Le format d'impression des rangées, colonnes et cellules peut également être spécifié. Aussi, avant d'imprimer les tableaux, on procède à la définition des matrices finales telles que les totaux par rangée et colonne, les moyennes et les pourcentages en fonction des spécifications de l'utilisateur.

Le tableau ci-dessous montre les temps de cycle machine (en nanosecondes) pour différents ordinateurs IBM. Ces données ont été publiées dans la revue Computerworld des 13 septembre 1983, 30 juin 1980 et 8 janvier 1979.

S Y S T E M E

IBM 3081D	IBM 3031	IBM 370/168-3	IBM 4331	IBM 4341
26	115	80	900	150-300
(per 4 bytes)				

viii) Informations sur le système

Le CENTS 4 regroupe 15 programmes COBOL. Le plus grand module de programme requérant un redressement de liens dans le système CENTS 4 nécessite une capacité de mémoire de 128 K octets, cette dernière comprend tous les modules de communication entrée/sortie requis par le programme. Cette capacité de mémoire est basée sur un système informatique IBM IBM 370 OS modèle 168. Lors du redressement des liens, le programme de tabulation COBOL généré peut nécessiter une capacité de mémoire supérieure à celle des autres programmes CENTS 4. Son ampleur dépend du nombre et de la complexité des instructions CENTAL codées par l'utilisateur.

Une capacité minimale de mémoire sur disque de 4 méga-octets est nécessaire pour la bibliothèque du programme qui comprend la source et l'objet des programmes COBOL, un espace de travail pour les fichiers intermédiaires et la bibliothèque de chargement des modules de programmes requérant un redressement de liens. Le matériel nécessaire comprend encore un poste d'entrée standard (bande, disque, lecteur de cartes), une imprimante directe et un poste de sortie standard (bande, disque, perforatrice).

Du point de vue du logiciel, le système requiert un compilateur ANS COBOL et une fonction de tri. Pour cette version du CENTS 4, qui a été réalisée sur un système IBM 370 OS, on a utilisé le compilateur ci-dessous :

COBL - ANS X3.23-1974

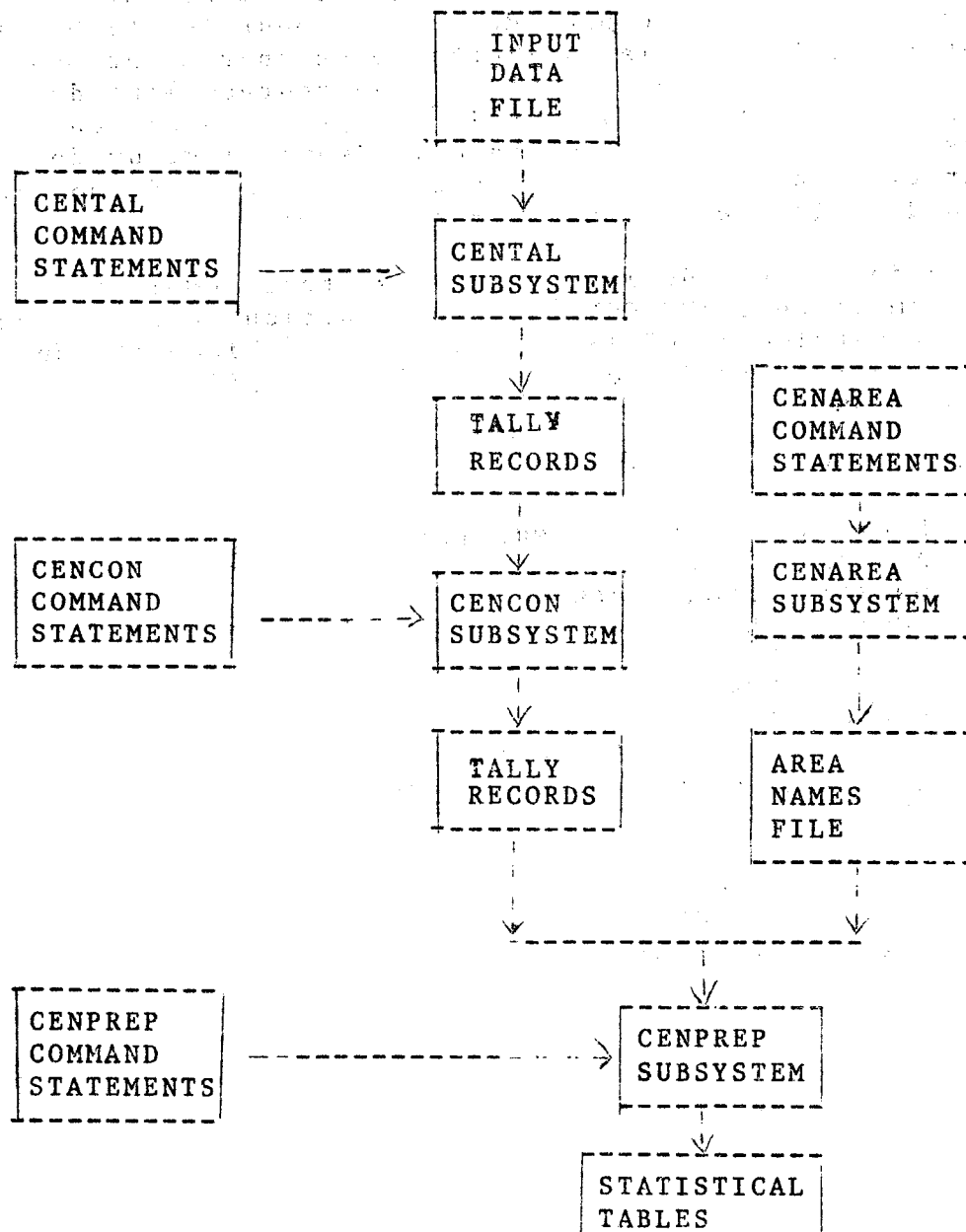
- a) NUCLEUS, LEVEL 1 (1 NUC 1,2)
- AVOIDED USING :
- ON SIZE ERROR PHRASE
 - ACCEPT VERB
 - ALTER VERB
 - INSPECT VERB
 - REMAINDER VERB
 - MOVE TO IDENTIFIER-LIST
 - ADD LITERAL TO IDENTIFIER-LIST
 - 77 LEVEL DECLARATIONS
- ADDED TO NUCLEUS :
- PERFORM ... UNTIL
 - PERFORM ... VARYING

b) TABLE HANDLING, LEVEL 1 (1 TBL 1,2)
RESTRICTED TO :
NO INDEXING
NO USE OF SET
MAXIMUM OF TWO (2) SUBSCRIPT LEVELS

c) SEQUENTIAL I-D, LEVEL 1 (1 SEQ 1,2)
ADDED :
RESERVE NO ALTERNATE AREAS
WRITE BEFORE IDENTIFIER LINES

d) RELATIVE/INDEXED I-D
THIS RELEASE OF CENTS 4 REQUIRES DIRECT ACCESS/
RELATIVE ORGANIZATION, BUT THE SYSTEM CAN EASILY BE
MODIFIED TO USE RANDOM OR INDEXED SEQUENTIAL ACCESS
METHODS.

THE CENTS 4 PROCESS



Matériel pour lequel CENTS 4 est disponible :

IBM - OS (370, 4300, 303X)
IBM - DOS/VSE
HDNEYWELL BULL
PERKIN-ELMER 3250
WANG - VS80
UNIVAC 1100 SERIES
IBM SYSTEM/38
NCR - 8500

Documentation disponible

SYSTEM REFERENCE MANUAL
DIAGNOSTIC MESSAGE GUIDE
SYSTEM INTERNALS MANUAL

Pour de plus amples informations :

TELEX : 892360
PHONE : (301) 763-2750

Ecrire à l'adresse suivante :

Assistant Chief for Data
Processing Activities
International Statistical
Programs Center
Bureau of the Census
Washington D.C. 20233

CENTS 4 SAMPLE PROGRAM

TABLE 2 : RELATIONSHIP TO HOUSEHOLD BY AGE AND SEX, URBAN AND RURAL

DATE OF RUN 02/06/84

AGE AND GENDER	TOTAL POPULATION	HEAD OF HOUSEHOLD	SPOUSE OF HEAD	CHILD OF HEAD	OTHER RELATIVE	NON RELATIVE	AGE AND GENDER
GOLD COAST REGION							
BOTH SEXES							
TOTAL	872	151	122	492	101	6	TOTAL
UNDER 5 YEARS	174	-	-	159	15	-	UNDER 5 YEARS
5 TO 14 YEARS	213	-	3	198	12	-	5 TO 14 YEARS
15 TO 24 YEARS	177	24	31	108	12	2	15 TO 24 YEARS
25 TO 34 YEARS	93	35	35	20	3	-	25 TO 34 YEARS
35 TO 44 YEARS	90	50	33	5	-	2	35 TO 44 YEARS
45 TO 54 YEARS	40	25	12	2	1	-	45 TO 54 YEARS
55 TO 64 YEARS	26	7	4	-	15	-	55 TO 64 YEARS
65 YEARS OR OLDER	59	10	4	-	43	2	65 YEARS OR OLDER
MEDIAN AGE	17.8	38.3	32.8	9.4	60.0	40.0	MEDIAN AGE
MALES							
TOTAL	450	135	-	261	50	4	TOTAL
UNDER 5 YEARS	84	-	-	75	9	-	UNDER 5 YEARS
5 TO 14 YEARS	115	-	-	105	10	-	5 TO 14 YEARS
15 TO 24 YEARS	101	24	-	70	5	2	15 TO 24 YEARS
25 TO 34 YEARS	49	35	-	11	3	-	25 TO 34 YEARS
35 TO 44 YEARS	47	45	-	-	-	2	35 TO 44 YEARS
45 TO 54 YEARS	22	22	-	-	-	2	45 TO 54 YEARS
55 TO 64 YEARS	9	4	-	-	5	-	55 TO 64 YEARS
65 YEARS OR OLDER	23	5	-	-	18	-	65 YEARS OR OLDER
MEDIAN AGE	17.6	36.9	-	10.3	28.4	30.0	MEDIAN AGE
FEMALES							
TOTAL	422	16	122	231	51	2	TOTAL
UNDER 5 YEARS	90	-	-	84	6	-	UNDER 5 YEARS
5 TO 14 YEARS	98	-	3	93	2	-	5 TO 14 YEARS
15 TO 24 YEARS	76	-	31	38	7	-	15 TO 24 YEARS
25 TO 34 YEARS	44	-	35	9	-	-	25 TO 34 YEARS
35 TO 44 YEARS	43	5	33	5	-	-	35 TO 44 YEARS
45 TO 54 YEARS	18	3	12	2	1	-	45 TO 54 YEARS
55 TO 64 YEARS	17	3	4	-	10	-	55 TO 64 YEARS
65 YEARS OR OLDER	36	5	4	-	25	2	65 YEARS OR OLDER
MEDIAN AGE	18.1	55.0	32.8	8.4	64.5	82.6	MEDIAN AGE

B. XTALLY et COXTALLY

1) XTALLY

- Introduction

XTALLY est un ensemble de logiciel portatif permettant d'élaborer des matrices ayant un titre et comportant jusqu'à 7 dimensions et 100.000 cellules, dont chacune résume 1 ou 2 variables. Les totaux partiels sont aussi présentés. XTALLY fournit des pourcentages du total général ou de n'importe quel niveau de total partiel et donne une inflation/déflation automatique des valeurs en fonction d'un ou deux totaux généraux indiqués à l'avance. Il est également possible de faire des sommes, des différences, des produits ou des quotients des deux variables résumés à multiplicateurs des cellules.

XTALLY a été employé pour présenter sous forme de tableau des données de recensement et d'enquête ou des données administratives à l'occasion d'un nombre de projets de coopération technique réalisés dans des pays en développement avec l'appui du Bureau des statistiques de l'Organisation des Nations Unies. Le système n'exige pas une compilation sur place et tous les programmes sont disponibles en forme "interprétative". Les instructions d'utilisation sont conservées sur le disque, la disquette ou la bande magnétique XTALLY, ce qui permet d'obtenir sur place les instructions d'utilisation chaque fois qu'on a besoin d'un exemplaire. Comme l'emploi d'XTALLY n'entraîne ni compilation ni tri et n'utilise que 4 types d'instruction et seulement 3 procédures principales, les utilisateurs peuvent apprendre à se servir d'XTALLY en quelques heures seulement. XTALLY est conçu expressément pour être employé par des non programmeurs de façon que des statisticiens ou d'autres techniciens puissent organiser directement leurs données en tableau.

L'utilisation d'XTALLY se fait en trois grandes étapes :

- Etape 1 : Définir la teneur et la présentation de l'enregistrement des données à la source et définir les ensembles de catégories - groupes de valeurs dont se servent les matrices.
- Etape 2 : Indiquer une matrice particulière en fournissant
- Les noms, par ordre hiérarchique, des ensembles de catégories de la colonne 1 à la colonne 3
 - Les noms, par ordre hiérarchique, des ensembles de catégorie de la rangée 1 à la rangée 4
 - Les noms de deux rubriques de données quantitatives dont les valeurs doivent être résumées dans la matrice comportant 2 à 7 dimensions.

- d. Il est possible, le cas échéant, de spécifier des totaux généraux arbitraires pour l'une ou l'autre des deux rubriques quantitatives de façon à obtenir une expansion proportionnelle de tous les totaux secondaires.

Etape 3 : Indiquer, le cas échéant, une fonction supplémentaire particulière du tableau en fournissant :

- a. Le nom de la fonction (A, B, A+B, A-B, A/B, A*B, B-A, B/A, %A, %B)
- d. Des multiplicateurs arbitraires pour 1 ou 2 variables fonctionnelles A ou B
- c. Si l'on souhaite une fonction en pourcentage, indiquer les ensembles de catégories qui forment la base des pourcentages.

Dans le cas de systèmes d'ordinateur à dialogue, tels que IBM S/32 ou S/34, l'utilisateur d'XTALLY peut réaliser toutes les étapes précitées en se servant du pupitre ou du terminal de l'ordinateur. Dans le cas de systèmes à traitement par lot, tel que IBM S/3, l'utilisateur établit des spécifications sous forme de cartes perforées, de disquette magnétique ou de cassette. Dans les deux cas, l'étape 1 qui fournit le dictionnaire de données et d'ensembles de catégories ne doit être réalisée qu'une seule fois pour préparer n'importe quel nombre de matrices. L'étape 2, élaboration de matrice, fournit le tableau résumé sur le disque magnétique de façon à ce qu'on puisse élaborer rapidement, si on le souhaite, un certain nombre d'exemplaires imprimés. L'étape 3, production d'une fonction du tableau résumé, peut donc être répétée aussi souvent qu'on le veut pour obtenir diverses fonctions et pourcentages qu'on pourrait souhaiter, sans avoir à répéter l'étape 2, élaboration du tableau résumé.

- Définition des enregistrements de données et des ensembles de catégorie

Définition des données.

Il s'agit d'identifier les rubriques de l'enregistrement de données qui serviront d'une façon ou d'une autre à élaborer les tableaux. Les rubriques qui ne doivent pas être utilisées n'ont pas besoins d'être définies; par exemple, le nom de famille n'est pas employé dans la présentation des données de recensement et sa position dans l'enregistrement ou les valeurs qu'il prend n'ont pas besoin d'être établis pour préparer les tableaux.

Les rubriques qui doivent être nommées et dont les positions de départ et fin dans l'enregistrement doivent être identifiées sont celles qui servent dans l'une des deux modalités suivantes :

- (a) variables d'accumulation : rubriques quantitatives par exemple, nombre d'enfants nés, somme gagnée ou dépensée, nombre de jours de travail, etc.) qui peuvent être résumées
- (b) variables de catégorie : variables quantitatives ou qualitatives dont les valeurs individuelles (ou ensemble de valeurs) identifient les catégories dans lesquelles on pourrait accumuler des sommes ou des comptages.

Une rubrique unique, telle que l'âge, les dépenses, les revenus, ou le nombre d'enfants nés, peut servir à la fois de variable d'accumulation et de variable de catégorie. Une rubrique de données peut faire partie d'une variable à deux chiffres, tel que l'âge en années, peut être une rubrique de données tandis que les deux chiffres peuvent représenter une autre rubrique de données.

- Présentation de définition des données.

Chaque rubrique servant de variable d'accumulation ou de variable de catégorie doit recevoir un nom qui lui est propre. Ce nom doit comporter trois caractères alphabétiques. La présentation permettant d'affecter des noms de rubrique de données et d'indiquer les positions de départ et de fin de l'enregistrement est la suivante :

<u>Position</u>	<u>Teneur</u>
1-3	nom de variable à 3 caractères
4-5	position de départ dans lenregistrement (01-99)
6-7	position de fin dans l'enregistrement (01-99)
8	* (indiquer s'il s'agit d'un enregistrement de définition de données)

Règles :

- a) Ne pas employer les caractères '*' ou '/' dans un nom de rubrique de données.
- b) La longueur des rubriques de données employées pour identifier les limites de catégorie ne doit pas dépasser 4 caractères; les zones condensées ou binaires ne sont pas permises.
- c) La longueur des rubriques quantitatives employées pour les accumulations de tableau ne peut pas dépasser 6 chiffres. (8 chiffres significatifs sont imprimés pour chaque variable d'accumulation dans chaque cellule; on élimine les zéros de tête. Les valeurs zéros ne sont pas imprimées.)
- d) Nommer jusqu'à 31 rubriques de données. Il n'est pas nécessaire de nommer les rubriques dont on ne se sert pas, mais il faut être sûr d'identifier les rubriques susceptibles de servir à définir des catégories ou à accumuler des totaux.
- e) Les emplacements de plusieurs rubriques de données peuvent chevaucher. Par exemple, les tableaux utilisant des codes d'industrie comme catégories peuvent employer un code à 1 chiffre pour certains tableaux et un code à 3 chiffres pour d'autres, en pareil cas, les codes à 1 et 3 chiffres doivent être définis comme deux rubriques séparées de données.

Définition d'un ensemble de catégories.

Une catégorie unique est constituée par une valeur de classification (ou ensemble de valeurs) de l'une des variables de catégorie. Par exemple, une catégorie peut comporter les valeurs 00, 01, 02, 03 et 04 pour une rubrique de données de deux chiffres appelée AGE; une autre catégorie peut être composée de la valeur AAMB pour une rubrique de données de 4 caractères appelée COD. Un ensemble de catégories est un ensemble de catégories séparées et distinctes qui, dans leur totalité, représentent toutes les valeurs possibles de l'une des variables de catégorie de telle façon que n'importe quelle valeur particulière de la variable se trouve dans une catégorie de l'ensemble et seulement dans une catégorie. Par exemple, les 20 catégories quinquennales d'âge 00-04, 05-09, 10-14 ... 95-99 peuvent être un ensemble de catégorie de la variable AGE. Un autre ensemble de catégorie de la variable AGE peut être les cents catégories à valeur unique 00, 01, 02, ... 99. Un troisième ensemble de catégories de la même variable AGE pourrait être -0-13, 14, 15, 16, 17, ..., 39, 40-43, 46-99.

Un seul ensemble de catégories classe les données sur une dimension. Deux ensembles de catégories classent les données sur deux dimensions de telle façon qu'un ensemble de 20 catégories d'âge, par exemple, et un ensemble de 4 catégories de situation de famille donnent 80 matrices ou cellules. Si l'on emploie XTALLY, on peut utiliser jusqu'à 7 ensembles différents de catégories pour établir les matrices d'un tableau quelconque qui comporte jusqu'à 99 . 999 cellules; chaque cellule peut contenir des sommes de valeurs correspondant à un ou deux variables d'accumulation, à une variable d'accumulation et à un comptage d'enregistrement, ou au seul comptage d'enregistrement.

D'ordinaire, il est souhaitable d'inclure des totaux partiels dans les matrices. Par exemple, une matrice faisant apparaître le nombre de personnes par sexe, âge et situation de famille comporte d'ordinaire non seulement les sommes d'hommes jamais mariés de chaque tranche d'âge. C'est pourquoi chaque ensemble de catégories XTALLY est automatiquement prolongé de façon à inclure une catégorie représentant un "total". Tous les totaux partiels et les totaux généraux sont automatiquement élaborés; pour chaque ensemble de catégorie utilisé dans un tableau les sommes accumulées de chaque catégorie individuelle sont additionnées de manière à obtenir le "total" de l'ensemble des catégories.

- Présentation de la définition d'un ensemble de catégorie

Une définition d'un ensemble de catégories exige un enregistrement pour chaque intervalle de valeur appartenant à l'une des catégories de l'ensemble, du fait que plusieurs intervalles différents peuvent être groupés dans une seule catégorie. Il est possible d'inclure dans un ensemble de catégories jusqu'à 126 intervalles qui sont chacun à distinguer par leur valeur supérieure. Les enregistrements de spécification doivent être organisés en fonction des valeurs croissantes des valeurs supérieures et doivent être présentés de la manière suivante:

POSITION

TENEUR

1-3	Nom de la variable de catégorie
4-5	Identificateur de 2 caractères (alpha-numérique) de l'ensemble de catégories (de cette manière les positions 1-5 constituent un seul nom pour l'ensemble de catégorie).
6-9	Valeur supérieure de l'intervalle (cadrée à droite).
10-13	Position relative de la rangée/colonne de cette catégorie (cadrées à droite) de gauche à droite ou de haut en bas.
14-17	Titre de 4 caractères à imprimer au sommet de la colonne ou à gauche de la rangée de cette catégorie.

Règles :

- a. Les catégories doivent s'exclure mutuellement et doivent épuiser toutes les valeurs possibles de la rubrique de données. Par exemple, un ensemble des catégories d'âge correspondant à une rubrique d'âge de 2 chiffres doit correspondre à toutes les valeurs possibles d'âge, 00-99;
- b. Si la séquence de valeurs appartenant à une catégorie est interrompue par des valeurs appartenant à une autre catégorie, faire pour chaque sous-séquence interrompue une inscription faisant apparaître leur caractère commun en indiquant leur position et titre identiques dans le tableau.

Par exemple :

XXX01	A	1	ODD
XXX01	B	2	EVEN
XXX01	C	1	ODD
XXX01	D	2	EVEN

XXX01 est l'ensemble de catégories

A, B, C, et D sont les valeurs possibles de la rubrique de données XXX.

A et C appartiennent à la catégorie dont le titre est 'ODD' et B et D appartiennent à 'EVEN'.

'ODD' précède 'EVEN' dans la colonne ou la rangée.

- c. Si certaines valeurs doivent être exclues de l'ensemble de catégories (par exemple, des âges inférieurs à quatorze ans), la position relative occupée dans les colonnes 10-13 doit être laissée en blanc ou fixée à 0;

Par exemple :

AGE 15	13	0
AGE 15	13	1 - 14

- d. Définir jusqu'à 126 intervalles de valeurs pour un ensemble quelconque de catégories.
- e. Définir jusqu'à 63 ensembles de catégories. Un nombre quelconque de ces ensembles de catégories peut correspondre à une seule rubrique de données. Par exemple, AGE peut rentrer dans les tranches d'âge de 5 ans ou dans les années particulières en utilisant des ensembles de catégorie dénommés AGE05 et AGE01; les mois de l'année peuvent être des catégories individuelles ou être groupés en trimestres par des ensembles de catégories appelés MON12 et MONQR.

- Procédure informatisée pour placer les définitions de données et d'ensemble de catégories dans le dictionnaire XTALLY.

Cette procédure varie selon le système de fonctionnement de l'ordinateur et les dispositifs d'entrée dont dispose l'utilisateur. Dans le cas du IBM S/3, la procédure est la suivante si l'on utilise le lecteur de carte d'entrée.

// CALL DEDATA, R1

// RUN

Enregistrement de définition de la rubrique de données.

Enregistrement de définition de l'ensemble de catégories; tous les enregistrements de chaque ensemble de catégories doivent être réunis et en séquence en fonction des valeurs des positions 6-9.

/*

Cette procédure fournit également une liste des définitions d'ensemble de catégories. La liste est ensuite vérifiée afin de s'assurer que les noms et les définitions sont corrects. S'il y a des erreurs ou des modifications, on répète la procédure. La liste sert ensuite de référence lorsqu'on veut identifier les ensembles de catégories et les rubriques de données pour établir les matrices décrites ci-dessous.

- Matrices

Une fois les définitions des données et des ensembles de catégories placées dans le dictionnaire XTALLY, on indique les matrices en désignant les ensembles de catégories et les variables particulières dont on doit se servir pour l'élaboration du tableau dans une instruction de contrôle et fournir un titre du tableau dans la deuxième instruction de contrôle. La rangée hiérarchique et la composition des colonnes du tableau sont indiquées par les positions relatives où se trouvent les noms des ensembles de catégories. De même, la position d'impression (haut ou bas dans la cellule) pour les variables d'accumulation ou le comptage est indiquée par la position de la carte du contrôle où figure le nom de la variable.

Présentation des spécifications de matrice La carte de contrôle se présente de la manière suivante :

<u>Position</u>	<u>Teneur</u>
1-5	nom de l'ensemble de catégories de colonne supérieure, le cas échéant,
6	' , '
7-11	nom du deuxième ensemble de catégories de colonne, le cas échéant
12	' , '
13-17	nom de l'ensemble le plus bas ou du seul ensemble de catégories (ne doit pas être laissé en blanc)
26-28	nom de la variable d'accumulation A ou blanc en cas de comptage
30-32	nom de la deuxième variable d'accumulation, le cas échéant, ou répétition du nom de la variable A (ou blanc) s'il y a une seule variable d'accumulation
33-37	nom de l'ensemble de catégories de la rangée supérieure, le cas échéant
38	' , '
39-43	nom de l'ensemble de catégories en troisième position depuis le bas, le cas échéant
44	' , '
45-49	nom de l'ensemble de catégories en deuxième position depuis le bas, le cas échéant
50	' , '
51-55	nom de l'ensemble de catégories de la rangée ou de la seule rangée
65-72	total général arbitraire des sommes d'inflation/déflation de la variable d'accumulation A, le cas échéant
73-80	total général arbitraire des sommes d'inflation/déflation de la variable d'accumulation B, le cas échéant (laisser un blanc s'il n'y a qu'une seule variable d'accumulation.

L'enregistrement du titre peut contenir n'importe quel titre alphanumérique de 96 caractères, mais ne doit pas renfermer ', dans les positions 6 ou 12.

- a) Le nombre total de cellules d'un tableau matriciel ne doit pas dépasser 99,999. Le nombre total de cellules du tableau est égal au produit des dimensions d'ensembles de catégories qui figurent sur la carte du contrôle. Il faut noter que la catégorie totale est automatiquement incluse dans les catégories individuelles que définit la procédure DEDATA. Par exemple, une matrice des données du recensement de la population pourrait être SEXOL (homme + femme + total = 3) par année individuelle d'âge (100 âges + total = 101) par situation de famille (jamais mariées; marié; veuf; divorcé ou séparé; + total x 4 catégories de réalisation d'éducation + total donne un tableau de $101 \times 101 \times 5 = 51,005$ cellules.
- b) La somme du nombre de catégories d'ensembles de catégories sur trois colonnes et des ensembles de catégories des 3 rangées supérieures ne doit pas dépasser 118 et le nombre de catégories de la catégorie de la rangée inférieure ne doit pas dépasser 126.

Indiquer jusqu'à deux variables d'accumulation dans les positions 26-28 et 30-32, séparées par une virgule.

- a) Si l'on doit accumuler les valeurs d'une seule rubrique de données, indiquer la rubrique de données à la fois dans 26-28 et 30-32.
- b) S'il faut accumuler seulement une variable mais s'il faut également faire le total du nombre d'enregistrements, laisser en blanc soit 26-28 soit 30-32.
- c) S'il faut uniquement faire le total du nombre d'enregistrements, laisser en blanc à la fois 26-28 et 30-32.

Si les données sont jugées représenter un échantillon d'après lequel il faut calculer des valeurs estimatives, indiquer la valeur totale générale estimée de la variable d'accumulation A (colonnes 26-28) dans les positions 65-72 et le total estimatif de la variable B (colonnes 30-32) dans les positions 73-80. On fait ainsi élaborer et imprimer un tableau des valeurs estimatives pondérées qui suit le tableau original et le tableau des pourcentages, qui sont toujours élaborés. Par exemple, une enquête par échantillon sur la population pourrait fournir 20,000 enregistrements de données dont chacun contient les données concernant une seule personne. Si l'on place la valeur 00400000 en position 65-72 et si on laisse en blanc les variables d'accumulation 26-28 et 30-32, tous les totaux de l'échantillon du premier tableau de sortie seront majorés en proportion pour donner les totaux estimés dans un tableau additionnel qui est imprimé après le tableau des pourcentages.

Si l'on se sert de moins de 7 ensembles de catégorie, laisser les paramètres inutilisés en blanc, mais veiller à indiquer la catégorie de la colonne la plus basse dans les colonnes 13-17 sans quoi il se produira un H1 HALT; veiller à ajouter les virgules dans les colonnes 6, 12, 38, 44 et 50 de la carte. Des erreurs d'orthographe ou de position des noms d'ensemble de catégories provoquent un H2 HALT.

Normalement deux arrêts seulement sont possibles, H1 et H2, tous deux provoqués par des spécifications incorrectes de l'enregistrement de contrôle. S'il se produit un arrêt H1 ou H2, annuler l'opération et la recommencer après avoir corrigé l'enregistrement de contrôle.

- a) H1 indique qu'il n'y a pas d'ensemble de catégories valable dans la colonne inférieure du 13-17.
- b) H2 indique qu'un ou plusieurs noms des autres ensembles de catégories indiquées ne sont pas valables, (ne se trouvent pas dans le dictionnaire XTALLY).
- c) Corriger les erreurs de présentation ou d'orthographe et procéder à un nouveau passage. S'il le faut, recommencer la procédure pour placer les définitions de données d'ensemble de catégories dans le dictionnaire XTALLY afin de désigner les nouveaux ensembles de catégories ou définir les nouvelles variables.

- Procédure informatisée produisant les matrices.

La procédure varie selon le dispositif d'entrée utilisé et le système d'exploitation. Dans le cas du IBM S/3, on se sert de la procédure suivante :

```
// CALL XTALLY , R1
```

```
// FILE NAME-INDATA, UNIT-XX, PACK-YYYYYY, LABEL-ZZZZZZ  
      (fournir le cas échéant, les para-  
      mètres X,Y,Z pour le fichier de  
      données)
```

```
1. CONTROL RECORD
```

```
2. TITLE RECORD FOR TABLE (jusqu'à 96 caractères de  
/*                           longueur)
```

Cette procédure fournit la matrice indiquée pour tous les enregistrements qui sont contenus dans le fichier des données particulières. Le tableau est présenté sous forme hiérarchique; les titres des ensembles de catégories du niveau le plus bas sont répétés dans les titres des ensembles de la catégorie du niveau supérieur pour les colonnes et les rangées jusqu'à ce qu'elles soient toutes équisées. Tous les totaux partiels sont imprimés pour les hiérarchies de rangées et de colonnes. Le total partiel de n'importe quel ensemble de catégories se trouve à droite pour les colonnes et au bas pour les rangées.

Les lignes continuent sur le plan vertical jusqu'à ce qu'elles soient toutes imprimées sans tenir compte d'une limite quelconque concernant le nombre des lignes par page d'imprimante. Les lignes du "total" correspondant aux membres des catégories de la rangée principale sont précédées d'un espace et suivies d'une ligne horizontale de "..."; les lignes du total de l'ensemble catégories de rangée secondaire sont précédées d'un espace et suivies d'une ligne de "---". Si l'on demande deux variables d'accumulation, le titre de la rangée est répété pour chacune des deux lignes imprimées pour chaque cellule.

Lorsqu'un tableau a plus de 15 colonnes de largeur, des pages supplémentaires sont automatiquement élaborées afin de renfermer ces colonnes supplémentaires. Les têtes de colonne continuent d'une page à l'autre et les têtes des rangées et les noms des variables d'accumulation sont répétés sur chaque page. Un tableau de 60 colonnes et de 1000 rangées avec deux variables d'accumulation aurait pour largeur 4 pages de papier d'imprimante et aurait en longueur d'au moins 20 pages de papier d'imprimante.

Après le tableau d'accumulation (sommés), l'ordinateur fournit automatiquement un tableau de présentation identique contenant les pourcentages du total général. Si la carte de contrôle fournit les valeurs estimatives de l'une ou de l'autre variable d'accumulation, un tableau de présentation identique des valeurs estimatives est fourni après le tableau des pourcentages.

- Fonction de la matrice

Après qu'une matrice ait été élaborée et imprimée, le tableau reste sur le disque XTALLY jusqu'à ce qu'un autre tableau soit élaboré. Il est donc possible d'imprimer des exemplaires supplémentaires du tableau ou d'obtenir des fonctions du tableau, si on le souhaite par exemple, divers pourcentages quotients, ou moyennes de cellules.

- Des exemplaires supplémentaires du tableau sont imprimés utilisant une procédure spéciale qui n'exige aucune carte de contrôle; les pourcentages et autres fonctions sont imprimés sous une présentation identique à celle du premier tableau avec les mêmes titres de colonne et de rangée, mais avec un titre qui est fourni accompagné de l'enregistrement de contrôle de la spécification de fonction.

- Présentation des spécifications de fonction.

Les pourcentages ou autres fonctions sont précisés en se servant de la présentation suivante de l'enregistrement de contrôle :

<u>Position</u>	<u>Teneur</u>
1-3	XFX code d'identification de l'enregistrement
4-12	un multiplicateur à 8 chiffres de la variable supérieure appelé "A" contenant 3 chiffres de fraction.
13-21	multiplicateur de 3 chiffres pour la variable inférieure appelée "B" contenant 3 chiffres de fraction.
22-24	un code de 3 caractères choisi parmi les suivants : (A+B, A*B, A/B, A-B, B-A, B/A, A, B, %, %A, %B)
31-65	Si l'on demande une fonction de pourcentage, les noms des ensembles de catégories qui identifient le dénominateur à utiliser (les noms d'ensemble de catégories, doivent être séparés des noms suivants par un trait ou une virgule.

Pourcentages.

Après avoir élaboré une matrice multi-dimensionnelle, il peut être souhaitable d'élaborer un ou plusieurs tableaux de pourcentages dans lesquels le contenu de chaque cellule est converti en pourcentage du total partiel approprié ou du total général. Cette conversion de sommes en pourcentages est facile à faire et n'exige que l'identification du niveau des totaux partiels ou des totaux généraux en employant les noms des ensembles de catégories qui indiquaient le tableau original qui réside sur le disque.

On peut se faire des pourcentages XTALLY pour n'importe quel niveau de la matrice multi-dimensionnelle. Le nombre d'ensembles de pourcentages disponibles pour une matrice donnée est 2 à la puissance 'N', dans laquelle 'N' = le nombre de dimension. Par exemple, il existe 4 ensembles de pourcentages pour un tableau bi-dimensionnel; 8 ensembles existent pour un tableau tri-dimensionnel; 128 ensembles différents sont disponibles pour un tableau à 7 dimensions.

Un pourcentage est une fraction inférieure ou égale à '100'. Le numérateur de la fraction est un membre d'un ensemble de quantités dont la somme générale est égale au dénominateur. Il est possible d'élaborer un grand nombre de tableaux de pourcentages à partir d'une matrice. Dans chaque tableau de pourcentages le numérateur est toujours le nombre contenu originalement dans la cellule, mais le dénominateur est la colonne ou la rangée appropriée du total partiel

ou du total général. (Comme une matrice XTALLY contient toujours tous les totaux partiels et les totaux généraux, chaque total partiel ou général sera également converti en pourcentage dans le tableau de pourcentages; les totaux partiels ou généraux employés pour calculer les pourcentages figurent bien sûr sous forme de 100%).

Un tableau particulier de pourcentages se distingue par l'ensemble de dénominateurs qui sert à calculer les pourcentages. Par exemple, si un tableau tri-dimensionnel a été élaboré dans lequel le seul ensemble de catégories des colonnes est AGE05 et les ensembles hiérarchiques de catégorie des rangées sont MARST et SEXMF, dans cette ordre, les 8 ensembles de pourcentages possibles sont identifiés par les ensembles de catégories combinés de manière à indiquer les dénominateurs :

Ensemble	Ensembles de catégories Indiquant les dénominateurs	Observations
1	Aucun	Le total général (total AGE05 x total, MARST x total SEXMF) sert de dénominateurs commun à chaque cellule; chaque cellule contient % du total général.
2	AGE05	Les dénominateurs sont des valeurs individuelles de AGE05 x total MARST x total SEXMF. Total MARST x total SEXMF = 100% pour chaque catégorie de AGE05; chaque cellule contient % du total de la colonne.
3	AGE05, MARST	Les dénominateurs sont des valeurs individuelles de AGE05 x valeurs individuelles de MARST x total SEXMF. Total SEXMF = 100% pour chaque catégorie de AGE05; chaque cellule contient % du total partiel dans la colonne.
4	MARST	Les dénominateurs sont des valeurs individuelles de MARST x total SEXMF x total AGE05. Total SEXMF x total AGE05 = 100% pour chaque catégorie de MARST; chaque cellule contient % de totaux partiels colonne/rangée.

Ensemble	Ensembles de catégorie indiquant les dénominateurs	Observation
5	SEXMF	Les dénominateurs sont des valeurs individuelles de SEXMF x total MARST x total AGE05. Total MARST x total AGE05 = 100% pour chaque catégorie de SEXMF; chaque cellule contient % de totaux partiels colonne/rangée.
6	MARST, SEXMF	Les dénominateurs sont des valeurs individuelles de MARST x valeurs individuelles de SEXMF x total AGE05. Total AGE05 = 100% pour chaque catégorie de MARST et chaque catégorie de SEXMF; chaque cellule contient % des totaux de rangée.
7	AGE05, SEXMF	Les dénominateurs sont les valeurs individuelles de AGE05 x valeurs individuelles de SEXMF x total MARST. Total MARST = 100% pour chaque catégorie de AGE05 et chaque catégorie de SEXMF; chaque cellule contient % des totaux partiels de colonne.
8	AGE05, MARST, SEXMF	Le cas simple. Les dénominateurs sont des valeurs individuelles de AGE05 x valeurs individuelles de MARST x valeurs individuelles de SEXMF, ainsi chaque dénominateur est identique à la valeur portée dans la cellule. Chaque cellule (non-vide) contient 100%.

- Planification et préparation des tableaux XTALLY.

Les tableaux que l'on souhaite obtenir d'un fichier de données sont souvent spécifiés par des descriptions proforma ou narratives. Ces spécifications servent facilement de base pour remplir les formules XTALLY concernant la définition d'ensemble de catégories ou la spécification du tableau, mais elles ne sont pas nécessaires. Les mesures suivantes concernant la planification et la préparation des tableaux peuvent contenir des références à des spécifications traditionnelles telles que le proforma mais un diagramme de préparation de tableaux peut aussi être élaboré afin d'assurer la conformité avec les possibilités et limitations de XTALLY.

Préparation d'un diagramme de tableau.

Pour planifier et préparer l'emploi d'XTALLY, il est utile d'exprimer les tableaux souhaités sous forme de diagramme comme suit :

1. Etiqueter les colonnes du diagramme avec les numéros d'identification ou les codes de divers tableaux de façon que chaque colonne soit identifiée à un tableau.
2. Etiqueter les rangées du diagramme avec les noms des divers ensembles de catégories suivis des noms des diverses variables d'accumulation. A côté du nom de chaque catégorie, indiquer le nombre (comptage) de diverses catégories de l'ensemble plus 1 (total).
3. Spécifier la présentation et la composition de chaque tableau :
 - a. indiquer les intersections de la colonne du tableau et des rangées étiquetées avec les ensembles de catégories qui figurent dans le tableau et les variables d'accumulation dont la somme figure au tableau;
 - b. pour chaque ensemble de catégories utilisé dans le tableau, indiquer par R ou C son emploi comme titre de rangée ou comme ensemble de catégories titre de colonne, et indiquer sa position hiérarchique dans un titre de rangée ou de colonne par 1, 2, 3 ou 4 après R ou C;
 - c. pour chaque variable d'accumulation dont le tableau présente la somme, indiquer sa position relative imprimée dans les cellules du tableau par 1 (haut) ou 2 (bas).
4. Calculer le nombre total de colonnes du tableau, qui est le produit des nombres (consignés en 2 ci-dessus) à côté des noms des ensembles de catégorie de colonne (C1, C2 ou C3).
5. Calculer le nombre total de rangées du tableau qui est le produit des nombres placés à côté des noms des ensembles de catégories de rangée (R1, R2, R3 ou R4).
6. Calculer le nombre total de cellules du tableau qui est le produit du total de colonnes x le total de rangées. (Ce chiffre ne doit pas dépasser 99.999 et un HALT apparaît s'il en est ainsi).

Exemple : Supposons que AGE01 est un ensemble de catégories de 100 années individuelles d'âge, et AGE05 est un ensemble de catégories de 20 tranches d'âge de 5 ans; SEX01 est un ensemble de catégorie de deux catégories de sexe; MAR01 est un ensemble de catégories de 5 codes de situation de famille, EDU01 est un ensemble de catégories de quatre codes de niveau d'éducation; OCC01 est un ensemble de catégories de 20 groupes de code de profession; IND01 est un ensemble de catégories de 60 groupes de code industriel; STA01 est un ensemble de catégories de quatre groupes de status d'activité. Supposons que DAW est une variable d'accumulation mesurant le nombre de jours travaillés pendant la dernière semaine et CEB est une variable d'accumulation exprimant le dernière semaine et CEB est une variable d'accumulation exprimant le nombre d'enfants nés. Un diagramme pour un ensemble de tableaux utilisant ces ensembles de catégories et ces variables d'accumulation pourrait se présenter comme suit :

Ensemble de catégories	Tableau 1	Tableau 2	Tableau 3	Tableau 4	Tableau 5
AGE01 (101)	R2				
AGE05 (21)				C1	C1
SEX01 (3)	C1	C2		C2	
MAR01 (6)	C2				
EDU01 (5)	R1		R2		
OCC01 (21)			R1	R1	R1
IND01 (61)		R1	C1		
STA01 (5)		C1			R2
Variables acc.					
DAW		1	2		2
CEB		2			
Personne (compagne enregistrements)	1		1	1	1
Total colonnes	18	15	61	63	21
Total rangées	505	61	405	81	405
Total cellules	9.090	915	24.705	5.103	8.505

- Les instructions de contrôle XTALLY utilisées pour élaborer les tableaux seraient les suivantes :

	COLONNES (1-17)	RANGÉES (33-55)	VARIABLES ACC. (26-32)	REMARQUES
Tableau 1	SEX01,....., MAR01	EDU01,....., AGE01,		- les cellules contiendront le comptage d'enregistrements
Tableau 2	STA01,....., SEX01, OCC01,....., DAW, CEB		- chaque cellule contiendra deux sommes : jours de travail et enfants nés
Tableau 3,....., IND01	OCC01,....., EDU01, ... , DAW		- chaque cellule contiendra deux sommes, comptage d'enregistrements et total jours de travail
Tableau 4	AGE05,....., SEX01,....., OCC01,		- les cellules contiendront le comptage d'enregistrements
Tableau 5,....., SEX01	OCC01,....., STA01 ... , DAW		- chaque cellule contiendra deux sommes, comptage d'enregistrements et nombre de jours de travail.

- Pour les tableaux 3 et 5, il peut être souhaitable d'obtenir des coefficients faisant apparaître le nombre de jours de travail par personne; dans les deux cas une instruction de contrôle indiquant qu'on souhaite le quotient DA/... (... représente nombre d'enregistrements) provoquera l'élaboration du tableau contenant ces coefficients. L'instruction de contrôle prendrait la forme suivante :

Position	Teneur	Remarques
1- 3	XFX	code d'identification
4-12	--	aucun multiplicateur nécessaire
13-21	--	aucun multiplicateur nécessaire
22-24	B/A	quotient de variable B et variable A (moyenne de cellule) est souhaité
31-55	--	utilisé seulement pour les fonctions de pourcentage

- Il est possible qu'on souhaite avoir les pourcentages des totaux de rangées pour le tableau 4. L'instruction de contrôle contiendrait les éléments :

Position	Teneur	Remarques
1- 3	XFX	code d'identification
4-12	--	aucun multiplicateur nécessaire
13-21	--	aucun multiplicateur nécessaire
22-24	%	une variable seulement (comptage) est totalisée dans le tableau. Par conséquent, il n'est pas nécessaire d'indiquer ni %A ni %B.
31-55	OCC01	base de % est l'ensemble de catégories d'occupation qui correspond aux totaux de rangée (OCC01 doit être placé en position 31-35)

C. COXTALLY

Introduction :

COXTALLY est la version COBOL du logiciel XTALLY. XTALLY est un logiciel généralisé qui permet la production de tableaux statistiques à partir de paramètres simples, sur tous les modèles d'ordinateurs, des plus petits aux plus gros.

Les programmes formant le logiciel XTALLY sont rédigés en RPG-II. L'ensemble de ce logiciel a été conçu et développé par Monsieur Michael R. Lackner de l'Office de la statistique des Nations Unies à New - York.

Le package XTALLY produit des tableaux statistiques au moyen de quatre (4) cartes paramètres.

COXTALLY a été rédigé en COBOL en n'utilisant que les instructions admises par les standards CODASYL, ce qui rend cette nouvelle version des plus compatibles sur tous les types d'ordinateurs.

Lorsque nous avons décidé de réécrire XTALLY en COBOL, les principes directeurs étaient :

- a) l'utilisation des paramètres de XTALLY
- b) l'augmentation des possibilités du logiciel
- c) de rendre l'impression des tableaux statistiques plus flexible.

Ces trois points majeurs ont été entièrement satisfaits.

Le logiciel COXTALLY emploie donc les mêmes paramètres que le logiciel XTALLY pour la description des variables de classification, pour la description des catégories et finalement pour la description des tableaux à produire. Il est bien évident que pour augmenter les possibilités du système il nous a fallu ajouter de nouvelles cartes paramètres.

De ce fait, trois types de cartes paramètres ont été ajoutées en ce qui concerne l'impression des tableaux, et une lors de la sélection des tableaux statistiques à produire.

Possibilités du logiciel COXTALLY :

- aucun tri nécessaire
- description d'un maximum de 99 variables de classification
- description d'un maximum de 2 variables d'addition par tableau statistique
- description d'un maximum de 999 catégories par variables de classification
- description d'un très grand nombre d'intervalles par catégories, le maximum étant de 99.999 pour l'ensemble
- description d'un maximum de 65.535 tableaux statistiques
- production d'un maximum de 99 tableaux statistiques par passe sur ordinateur
- jusqu'à un maximum de 12 conditions de rejet par tableau produit pour la sélection des enregistrements.
- jusqu'à un maximum de quatre caractères pour les variables de classification
- jusqu'à un maximum de huit (8) caractères pour les variables d'addition
- jusqu'à un maximum de 99.999 compteurs par tableau produit
- jusqu'à un maximum de 7 variables de classification par tableau produit (7 dimensions)
- option de génération automatique des titres des colonnes ou des lignes, ou insertion de ces titres au moment de l'impression
- aucune limitation quant au nombre de pages de largeur du tableau statistique produit
- impression des totaux suivis de détail
- lecture des données de l'utilisateur au moyen d'une routine de lecture.

Toutes les étapes du logiciel COXTALLY sont nommées de la façon XTALLYnn ou 'nn' est le numéro séquentiel de l'étape exécutée.

XTALLY01 : Cette étape est la première du logiciel COXTALLY et permet de décrire les variables de classification ainsi que les catégories nécessaires à la production de tableaux statistiques.

NOTE: Cette étape détruit les anciennes descriptions s'il y a lieu.

a) Description des variables de classification :

Une variable de classification est un nom donné à un élément de l'enregistrement.

<u>Position</u>	<u>Contenu</u>
1 - 3	nom de la variable de classification
4 - 5	position de début de la variable de classification à l'intérieur de l'enregistrement
6 - 7	position de fin de la variable de classification à l'intérieur de l'enregistrement
8 - 8	toujours le caractère '*'

REGLÉS

- décrire jusqu'à un maximum de 38 variables de classification.
- le nom de la variable de classification peut être formé de n'importe quel caractère.
- les positions de début et de fin de la variable de classification doivent être numériques
- les emplacements des variables peuvent se chevaucher
- la longueur d'une variable de classification ne devrait pas dépasser quatre (4) caractères sinon le système n'utilisera que les quatre derniers caractères de la variable lorsqu'il l'utilisera.
- la position 8 doit obligatoirement contenir un astérisque '*', ce qui indique au système qu'il s'agit d'une description de variable de classification.

b) Description des catégories :

Une catégorie est l'ensemble des valeurs, intervalles, d'une variable de classification nécessaires à la ventilation des données lors de la production d'un tableau statistique. Exemple : 0 à 5 ans, 6 à 14 ans, 15 à 49 ans et 50 ans et plus, pourraient être les quatre intervalles d'une catégorie pour l'âge.

<u>Position</u>	<u>Contenu</u>
1 - 3	nom d'une variable de classification
4 - 5	suffixe qui attaché au nom de la variable de classification crée un nom de catégorie
6 - 9	valeur la plus haute de l'intervalle
10 - 13	position d'impression ou numéro de compteur à l'intérieur de la catégorie
17 - 36	titre applicable à l'intervalle ou au numéro du compteur.

REGLES

- le nom de la variable de classification doit avoir été décrit précédemment
- le suffixe peut être composé de n'importe quel caractère
- la valeur inscrite doit avoir la même longueur que la variable et être cadrée à droite.
'01' n'est pas équivalent à '10'.
- la position d'impression ou numéro de compteur doit être numérique et différent de '0' et aussi cadré à droite, les '0' de début étant facultatifs.
- les valeurs de la catégorie doivent être en ordre croissant et doivent inclure toutes les valeurs possible de la variable de classification.
- le titre peut avoir 30 caractères mais seulement les douze premiers caractères seront imprimés si le titre est utilisé en colonne ou pour la catégorie ayant la plus basse hiérarchie pour les lignes.
Il est recommandé de cadrer les titres à droite sur les douze premiers caractères lorsque l'on désire les utiliser en colonnes.

XTALLY02: Cette étape permet de décrire les tableaux statistiques que l'on voudra produire. Cette étape peut être effectuée autant de fois qu'on le désire pour ajouter de nouvelles descriptions de tableaux statistiques.

NOTE : Lorsque l'on ajoute de nouvelles descriptions de tableaux statistiques, l'on fera attention de ne pas utiliser des noms de tableaux statistiques déjà utilisés, sinon le système lors de la production des tableaux statistiques utilisera la première description.

a) Descriptions des paramètres servant à la description des tableaux statistiques.

<u>Position</u>	<u>Contenu</u>
1 - 2	nom du tableau
3 - 7	nom de catégorie utilisée en colonne
8 - 8	inutilisé (ou ',')
9 - 13	nom de catégorie utilisée en colonne
14 - 14	inutilisé (ou ',')
15 - 19	nom de catégorie utilisée en colonne
20 - 20	inutilisé (ou ',')
21 - 25	nom de catégorie utilisée en ligne
26 - 26	inutilisé (ou ',')
27 - 31	nom de catégorie utilisée en ligne
32 - 32	inutilisé (ou ',')
33 - 37	nom de catégorie utilisée en ligne
38 - 38	inutilisé (ou ',')
39 - 43	nom de catégorie utilisée en ligne
44 - 49	inutilisé
50 - 52	nom de la variable 1 d'addition dont le contenu doit être cumulé.
53 - 54	position de début de la variable 1 d'addition à l'intérieur de l'enregistrement.
55 - 56	position de fin de la variable 1 d'addition à l'intérieur de l'enregistrement
57 - 57	inutilisé (ou ',')
58 - 60	nom de la variable 2 d'addition dont le contenu doit être cumulé
61 - 62	position de début de la variable 2 d'addition à l'intérieur de l'enregistrement
63 - 64	position de fin de la variable 2 d'addition à l'intérieur de l'enregistrement.

REGLES

- le nom du tableau peut contenir n'importe quel caractère
- décrire autant de tableaux statistiques que l'on veut
- les noms de catégorie doivent avoir été définis dans la première étape (XTALLY01). Le tableau est rejeté lorsqu'une catégorie est introuvée
- au moins une catégorie doit apparaître dans les emplacements prévus pour les catégories utilisées pour les lignes et pour les colonnes. Une pour les lignes et une pour les colonnes
- Il est suggéré de mettre des ',' dans les emplacements indiqués comme tel pour une meilleure apparence lors de l'impression des tableaux acceptés
- la hiérarchie des catégories se fait de gauche à droite, ce qui veut dire que la catégorie apparaissant en premier, soit pour les lignes, soit pour les colonnes, reçoit la plus haute hiérarchie
- le nom des variables d'addition peut être formé de n'importe quel caractère
- les positions de début et de fin des variables d'addition doivent être numériques, sinon la variable d'addition est rejetée sans avertissement et le tableau accepté
- la même variable d'addition peut apparaître et dans la variable 1 d'addition et dans la variable 2 d'addition simultanément
- la longueur des variables d'addition ne devrait pas dépasser huit (8) caractères, si cela se produit le système n'utilisera que les huit (8) derniers caractères lors de sa manipulation
- si un nom de variable d'addition est présent pour la variable 1 et que rien n'est inscrit pour la variable 2, le système cumulera le contenu de la variable seulement (un compteur contenant la somme).
- si un nom de variable d'addition est présent pour la variable 2 et que rien n'est inscrit pour la variable 1, le système cumulera la fréquence et le contenu. (deux compteurs sont créés, un pour la fréquence et un pour la somme).
- si un nom de variable est présent, et dans la variable 1, et dans la variable 2 d'addition, deux compteurs sont créés, (un compteur pour la somme de la variable 1 et un compteur pour la somme de la variable 2).
- si rien n'est inscrit dans les variables d'addition, le système ne fera que compter la fréquence (un seul compteur contenant la fréquence).

- le système imprime la liste des tableaux acceptés avec le nombre de compteurs nécessaires à chacun des tableaux. L'utilisateur devrait normalement y référer pour savoir si les variables d'addition ont été acceptées.
- le nombre de compteurs pour un tableau donné est le produit entre eux des nombres de compteurs pour chacune des catégories utilisées dans les tableaux, ce nombre est multiplié par deux lorsque la variable 2 d'addition est présente.

XTALLY03: Cette étape permet la production de tableaux statistiques. Les tableaux à produire sont sélectionnés parmi les tableaux statistiques acceptés lors de l'étape XTALLY02. A ce point-ci de la production des tableaux l'utilisateur peut choisir les enregistrements qui doivent être exclus du traitement.

- a) Description des paramètres pour la sélection des tableaux statistiques à produire.

<u>Position</u>	<u>Contenu</u>
1 - 2	nom du tableau à produire
3 - 15	condition 1 de rejet
16 - 28	condition 2 de rejet
29 - 41	condition 3 de rejet
42 - 54	condition 4 de rejet
55 - 67	condition 5 de rejet
68 - 80	condition 6 de rejet

NOTE Chacune des conditions de rejet a le format suivant :

- 1 - 2 début des données à comparer pour la sélection
- 3 - 4 fin des données à comparer pour la sélection
- 5 - 5 type de comparaison à effectuer sur les données
- 6 - 9 valeur 1 à utiliser pour la sélection
- 10 - 13 valeur 2 à utiliser pour la sélection.

REGLES

- le nom du tableau à produire doit être celui d'un tableau accepté par XTALLY02.
- deux articles (cartes) peuvent être utilisés par tableau à produire donnant ainsi la possibilité d'avoir un maximum de 12 conditions de rejet.
- introduire jusqu'à un maximum de 99 noms de tableaux à produire.
- les conditions de rejet ne sont acceptées que si les positions de début et de fin sont numériques.
- le tableau est produit même si les conditions de rejet sont invalides du fait de leur localisation.
- un "OU" est appliqué entre chacune des conditions de rejet. Ceci veut dire qu'aussitôt qu'une des conditions de rejet est satisfaite, l'enregistrement est rejeté sans examen des autres conditions.
- les conditions de rejet peuvent apparaître dans n'importe lequel des emplacements prévus à cette effet.
- la longueur des données à vérifier pour la sélection ne devrait pas dépasser quatre (4) caractères sinon le système n'utilisera que les quatre derniers caractères de la donnée.
- inscrire dans le type de comparaison un des deux caractères suivants :
 - "E" si l'on désire rejeter l'enregistrement lorsque les données de l'enregistrement sont égales à la valeur inscrite dans valeur 1 de la condition.
 - "B" si l'on désire rejeter l'enregistrement lorsque les données de l'enregistrement sont égales à la valeur inscrite dans valeur 1 ou qu'elles sont égales à la valeur 2 de la condition, ou finalement si les données de l'enregistrement sont comprises entre la valeur 1 et la valeur 2 de la condition de rejet.
- les valeurs 1 et 2 doivent être cadrées à droite.
- lors de l'impression des tableaux, le nombre d'enregistrements rejetés est mentionné.
- il est aussi possible d'avoir un listing des enregistrements rejetés (voir la procédure de contrôle).

XTALLY04 : Cette étape permet l'impression des tableaux produits par XTALLY03. Contrairement à la précédente étape elle ne permet l'impression que d'un seul tableau statistique à la fois. L'utilisateur devra donc effectuer cette étape autant de fois qu'il aura produit des tableaux statistiques lors de l'exécution de l'étape précédente (XTALLY03).

Cette étape peut recevoir trois (3) types de paramètres.

a) Paramètres type 1 : Ce premier paramètre est obligatoire et indique au système le tableau à imprimer. Il indique au système les espaces à insérer lors de l'impression.

<u>Position</u>	<u>Contenu</u>
1 - 2	nom du tableau à imprimer
3 - 5	inutilisé
6 - 8	nombre de lignes par page
9 - 9	option pour la génération des titres
10 - 10	caractère de saut applicable à la première catégorie utilisée pour les lignes.
11 - 11	caractère de saut applicable à la deuxième catégorie utilisée pour les lignes.
12 - 12	caractère de saut applicable à la troisième catégorie utilisée pour les lignes
13 - 13	caractère de saut applicable à la quatrième catégorie utilisée pour les lignes
14 - 25	titre utilisé pour le premier compteur si une variable 2 d'addition a été mentionnée lors de XTALLY02
26 - 37	titre utilisé pour le deuxième compteur si une variable 2 d'addition a été mentionnée lors de XTALLY02.

REGLES :

- le nom du tableau doit être un nom de tableau produit lors de l'exécution de l'étape XTALLY03
- ce paramètre est obligatoire
- le nombre de lignes par page doit être numérique, sinon la valeur par défaut est utilisée, soit 60 lignes par page
- les caractères de saut permettent d'indiquer le nombre de lignes à blanc que le système doit laisser lorsqu'il doit imprimer un titre de ligne
- un caractère de saut doit être numérique et différent de "0", sinon la valeur "1" est utilisée
- si le caractère "N" est inscrit dans l'option des titres, cela veut dire que le système générera les titres des colonnes et des lignes de façon automatique en se servant des titres mentionnés lors de la description des catégories à l'étape XTALLY01, sinon l'utilisateur doit prévoir les titres de colonnes lors de l'utilisation du paramètre suivant, type 2, et les titres des lignes avec l'utilisation du paramètre type 3
- les caractères de saut sont applicables aux deux options
- les titres utilisés lors d'une mention d'une variable 2 d'addition ne sont valables que pour l'option de génération automatique. Ces titres ne devraient contenir que onze (11) caractères et être cadrés à droite pour une meilleure apparence.
- à noter que l'utilisateur devrait toujours utiliser l'option de génération des titres automatique, lorsqu'il y a plus d'une page de largeur pour un tableau car le système ne prévoit pas la possibilité d'introduire des titres généraux ayant plus de 132 caractères de long.

b) Paramètre type 2: Ce paramètre permet l'introduction des titres généraux du tableau à imprimer, ou plus généralement appelés titres principaux, soit les en-têtes.

<u>Position</u>	<u>Contenu</u>
1 - 2	nom du tableau à imprimer
3 - 3	toujours le caractère "T"
4 - 4	numéro de la ligne du titre
5 -70	contenu du titre

REGLES :

- deux articles (cartes) peuvent être utilisés par titre donnant ainsi la possibilité d'écrire 132 caractères de titre par ligne
- le numéro de la ligne du titre doit être un nombre de "0" à "15" donnant ainsi la possibilité d'avoir six (6) lignes de titres par tableaux statistiques imprimés
- le premier caractère d'un titre peut servir de caractère de saut de ligne. Ce caractère servira de saut de ligne s'il est numérique et différent de "0", sinon un saut de une (1) ligne sera généré avant l'écriture du titre
- la règle précédente ne s'applique pas lorsqu'il s'agit de la ligne de numéro "0", celle-ci étant toujours imprimée du fait qu'elle contient la date d'impression et le numéro de page du tableau
- le titre résidant sur le numéro de ligne "0" ne peut contenir que 112 caractères
- si l'utilisateur n'utilise pas la génération automatique des titres, il doit utiliser ce paramètre pour introduire les titres de ses colonnes
- si l'utilisateur utilise la génération automatique des titres, le nombre de lignes disponibles est six (6) moins le nombre de catégories utilisées en colonne. Si une variable 2 d'addition est présentée, il doit encore soustraire 1 au nombre de lignes disponibles, le système utilisant ces lignes.

c) Paramètres type 3 : Ce paramètre permet l'insertion des titres des colonnes et des lignes si l'utilisateur a choisi l'option d'introduire les titres au moment de l'impression. Si l'utilisateur a choisi l'option de génération automatique, il peut utiliser ce paramètre pour remplacer le mot "TOTAL" généré par le système par un autre titre de ligne lorsqu'il s'agit d'imprimer la ligne total.

<u>Position</u>	<u>Contenu</u>
1 - 2	nom du tableau
3 - 4	toujours le caractère "R"
4 - 4	numéro hiérarchique de la catégorie à laquelle doivent s'appliquer les titres
5 - 34	titre de ligne
35 - 64	titre de ligne

REGLES :

- les numéros hiérarchiques doivent être numériques, supérieurs à "0" et inférieurs à "5"
- les numéros hiérarchiques doivent être introduits par ordre croissant
- le numéro hiérarchique indique à quelle catégorie la liste de titre doit s'appliquer
- il y a deux titres de ligne par carte. Le premier titre s'applique toujours à la ligne total de la catégorie, le deuxième titre s'applique au premier compteur de la catégorie et ainsi de suite dans l'ordre que les titres apparaissent sur les cartes
- si l'utilisateur choisit d'introduire les titres à cette étape de la production des tableaux statistiques, il doit s'assurer d'introduire dans le système autant de titres qu'il y a de compteurs dans chacune des catégories utilisées pour les lignes
- si l'option de génération automatique des titres est utilisée, cette carte est optionnelle et peut servir pour introduire un titre pour le total.

XTALLY05 : On se sert de cette étape une fois qu'un tableau statistique a été imprimé par l'étape XTALLY04 de façon à pouvoir manipuler les données du tableau. Elle permet donc de calculer des pourcentages, des moyennes, des inflations etc.

Les calculs peuvent être appliqués soit à des variables de classification (compteurs de fréquence), soit à des variables d'addition (compteurs de somme).

Les cartes paramètres sont identiques à celles utilisées par XTALLY04, exception faite de la carte type 1 qui se voit ajouter les paramètres nécessaires à l'exécution des calculs à effectuer.

a) carte paramètre type 1 :

<u>Position</u>	<u>Contenu</u>
1 - 37	identique à la carte type 1 de XTALLY04
38 - 38	inutilisé
39 - 46	variable "A" de calcul
47 - 47	inutilisé
48 - 55	variable "B" de calcul
56 - 56	inutilisé

<u>Position</u>	<u>Contenu</u>
57 - 59	opération demandée
60 - 60	inutilisé
61 - 67	spécification de pourcentage

REGLES :

- le tableau doit obligatoirement avoir franchi l'étape de l'impression
- les variables "A" ou "B" de calcul peuvent contenir chacune un nombre de huit (8) chiffres dont les trois derniers sont la partie décimale du nombre.
Ex: 1 s'écrit 1000, alors que , 001 s'écrit 1
- les nombres inscrits dans les variable "A" ou "B" de calcul doivent être cadrés à droite, les "0" de tête étant obligatoires
- un tableau comportant deux compteurs par classification est dit avoir un compteur "A" et un compteur "B"
- pour un tableau ne comportant qu'un seul compteur par classification, le nom du compteur importe peu
- les opérations suivantes peuvent être utilisées :
 - % calcul du pourcentage pour tous les compteurs du tableau
 - % calcul du pourcentage pour les compteurs "A" seulement
 - %B calcul du pourcentage pour les compteurs "B" seulement
 - A*B multiplication de A par B, A étant le résultat
 - A/B division de A par B, A étant le résultat
 - A+B addition de A et B, A étant le résultat
 - A-B soustraction de B de A, A étant le résultat
 - B/A division de B par A, B étant le résultat
 - B-A soustraction de A de B, B étant le résultat
- %, %A, %B ou espace, ont la même signification pour un tableau ne comportant qu'un seul compteur par classification.

- si aucune variable de calcul n'est inscrite sur la carte paramètre, les deux compteurs de chacune des classifications sont utilisés.
 - la règle du résultat ne s'applique que si l'opération utilise des compteurs du tableau, sinon le résultat est toujours un compteur du tableau
 - lorsqu'un nombre est inscrit sur la carte paramètre dans les variables "A" ou "B" de calcul, ce nombre a priorité sur le compteur de même nom du tableau
- pour les tableaux ne comportant qu'un seul compteur par classification, il faut faire attention lors de l'utilisation des opérations "/" et "-", et inscrire le nombre dans la bonne variable de calcul de la carte paramètre

Exemple : L'opération "A/B" devient: le compteur du tableau est divisé par le nombre de la carte paramètre si celui-ci apparaît dans la variable B

Si le nombre de la carte paramètre apparaît dans la variable A, alors le nombre de la carte paramètre est divisé par le compteur du tableau

- l'élément de spécification de pourcentage comprend sept (7) positions, ce qui permet d'indiquer les niveaux de pourcentage à calculer pour les catégories
- le nombre de positions utilisées correspond au nombre de catégories utilisées pour un tableau, si un tableau contient 3 catégories seulement, alors seulement les trois premières positions de l'élément spécification de pourcentage seront utilisées
- si l'élément de spécification de pourcentage est laissé en blanc, et qu'une opération de pourcentage est demandée, le pourcentage est alors calculé à partir du grand total
- un "Y" doit être inscrit dans la position correspondant à la catégorie dont on veut calculer le pourcentage
- plus d'une position de l'élément spécification peut contenir un "Y" à la fois
- le nombre de possibilités de calculs différents de pourcentage est 2^n , ou "n" est le nombre de catégories utilisées pour la production du tableau.

NOTE : Il est fortement suggéré à l'utilisateur de se référer aux exemples en annexe.

Utilitaires du logiciel COXTALLY :

Les programmes utilitaires suivants, XTALLYC6, XTALLY07, XTALLY98 et XTALLY99 servent à donner à l'utilisateur une plus grande flexibilité, ceux-ci n'ont aucunement besoin de cartes paramètres, exception faite de XTALLY98 qui utilise les mêmes cartes paramètres que le programme XTALLY04

a) XTALLY06 :

Cette étape permet de cumuler un, ou deux mêmes tableaux statistiques de façon à avoir les mêmes tableaux statistiques à un niveau plus élevé.

Un exemple est la production de tableaux statistiques pour chacun des districts d'un pays, si l'on exécute XTALLY06 après le passage de chacun des districts on aura finalement un tableau au niveau du pays.

b) XTALLY07 :

Cette étape contrairement à XTALLY06, permet de soustraire un tableau statistique d'un niveau plus élevé.

c) XTALLY98 :

Ce programme écrit les tableaux statistiques sur un fichier séquentiel de la même façon que XTALLY04 le fait **excepté** que les compteurs ne sont pas formatés (édités) donnant ainsi le loisir de pouvoir manipuler les compteurs de la manière qu'on le désire. Chaque enregistrement sur le fichier de sortie a le format suivant :

- 1 - 1 caractère de saut (0 signifie saut de page)
- 2 - 13 titre s'il y a lieu
- 14 - 103 10 compteurs de 9 positions chacun s'il y a lieu
- 104 - 132 inutilisé

d) XTALLY99 :

Ce programme permet de créer un fichier séquentiel à partir du dictionnaire de l'utilisateur. Ce fichier contient le dictionnaire sous la forme des paramètres nécessaires 1^{ère} étape XTALLY01.

Procédures d'exécution du logiciel COXTALLY :

Les procédures qui suivent sont utilisées pour l'exécution des différentes étapes nécessaires à la production des tableaux statistiques

Les procédures mentionnées sont écrites spécifiquement pour le système d'opération GCOS utilisé par les ordinateurs Honeywell-Bull Bull modèle 62, et DPS-4.

A chacune des étapes de la production de tableaux statistiques correspond une procédure d'exécution, celle-ci porte le même nom que l'étape qu'elle sert à exécuter.

Chaque procédure d'exécution comporte des paramètres de contrôle qui sont soit obligatoires, soit qui ont des valeurs par défaut que l'on peut modifier selon les besoins.

Les paramètres décrits ci-dessous sont communs à toutes les procédures d'exécution et ne sont pas documentés lors de la description de chacune des procédures.

Paramètres communs à toutes les procédures de contrôle :

Paramètre

Contenu

USE

Ce paramètre est obligatoire pour toutes les procédures de contrôle du logiciel COXTALLY.

Il permet au système d'identifier l'utilisation.

Le nom indiqué doit être le même tout au long de la production des tableaux statistiques.

Ce nom doit avoir un maximum de huit (8) caractères alpha-numériques, dont le premier est obligatoirement alphabétique.

IN

Indique au logiciel COXTALLY l'emplacement des cartes paramètres nécessaires à l'exécution de l'étape.

La valeur par défaut indique que les cartes se trouvent entre les \$input et \$endinput, soit avec le JCL.

WRK

Sert à indiquer au système le disque où réside le dictionnaire COXTALLY de l'utilisateur.

La valeur par défaut implique que le dictionnaire de l'utilisateur réside sur le disque du système d'opération de l'ordinateur.

Paramètres

Contenu

OUT

Permet à l'utilisateur de spécifier la disposition à prendre avec les listings produits par le logiciel COXTALLY.

La valeur par défaut indique que le logiciel COXTALLY écrira les listings produits sur le SPOOL, ce qui est la façon habituelle.

DVC

Ce paramètre définit le type d'imprimante sur laquelle sont imprimés les listings produits par COXTALLY.

La valeur par défaut de ce paramètre mentionne une imprimante "PRU002" (c'est le type d'imprimante utilisée à l'INSAE).

Procédure de contrôle XTALLY01 :

Cette procédure permet l'exécution de l'étape XTALLY01.

Format : \$XTALLY01 USE - utilisation

,IN $\frac{IN}{=}$
(fd)

,WRK $\frac{SD}{=}$
nom du disque

,SP $\frac{20}{=}$
espace

,OUT $\frac{OUT}{=}$
(urd)

,DVC $\frac{PRU002}{=}$
type d'imprimante

paramètres

contenu

SP

Sert à indiquer au système d'opération l'espace nécessaire du fichier qui devra contenir le dictionnaire COXTALLY de l'utilisateur

Cet espace est exprimé en terme de pistes allouées

La valeur par défaut, soit 20 pistes, est amplement suffisante pour permettre la production de très nombreux tableaux statistiques.

Procédure de contrôle XTALLY02 :

Cette procédure permet l'exécution de l'étape XTALLY02 qui introduit la description des tableaux statistiques de l'utilisateur.

Format : \$XTALLY02 USE =
 IN
 ,IN= (fd)
 OUT
 ,OUT= (urd)
 PRU002
 ,DVC= type d'imprimante
 SD
 ,WRK= nom du disque
 ;

NOTE : Tous les paramètres sont expliqués dans la présentation des procédures en page 79.

Procédure de contrôle XTALLY03:

Cette procédure permet la tabulation des données de l'utilisateur d'après les spécifications entrées via les étapes précédentes.

Format : \$XTALLY03 USE = nom de l'utilisation
 ,PGM = nom du programme
 ,IN = (fd)
 ,DATA = (fd)
 XTALDATA
 ,DDN =
 nom interne de fichier
 SD
 ,WRK =
 nom de disque
 SD
 ,TBL =
 nom de disque
 0
 ,VN =
 numéro de version
 10
 ,SP =
 espace
 NO
 ,PRINT =
 YES
 OUT
 ,OUT =
 (urd)
 PRU002
 ,DVC =
 type d'imprimante
 0000000
 ,REST =
 numéro de reprise

;

<u>Paramètres</u>	<u>Contenu</u>
PGM	nom du programme résultant de l'exécution de la procédure XTALLYNK.
DATA	paramètre servant à indiquer l'emplacement des données de l'utilisateur qui doivent servir à la production des tableaux statistiques.
DDN	indique le nom interne du fichier. C'est le même nom que celui apparaissant dans la clause "assign" de la routine de lecture de l'utilisateur.
TBL	indique le nom du disque où doivent résider les tableaux bruts une fois terminés.
VN	sert à donner un numéro de version aux productions de tableaux et ainsi avoir plusieurs passes sur ordinateur pour une même application.
SP	indique l'espace nécessaire pour le fichier contenant les tableaux bruts. Le nombre inscrit exprime des pistes allouées sur disque. Le nombre de pistes nécessaires est calculé de la façon suivante : il faut prendre la somme des compteurs des tableaux à produire diviser par 1454, le résultat étant arrondi à l'entier suivant.
PRINT	indique au programme de tabulation si l'on désire le listing des enregistrements rejetés ou non. YES : le listing des enregistrements rejetés est produit NO : il n'y a pas de listing des enregistrements rejetés.
REST	ce paramètre permet de relancer la tabulation à partir d'un enregistrement précis. ce paramètre est le complément à la méthode d'arrêt d'urgence du programme de tabulation. Cette procédure est expliquée en annexe.

Procédure de contrôle XTALLY04, XTALLY05 et XTALLY98 :

Toutes ces procédures utilisent les mêmes paramètres du fait qu'elles servent à imprimer les tableaux statistiques, exception faite de XTALLY98 qui doit recevoir une description de fichier pour le paramètre OUT.

Format : \$XTALLY04 USE = nom d'utilisation
 ou \$XTALLY05
 ou \$XTALLY98 SD

,WRK =
 nom de disque

,TBL =
 nom de disque

,VN =
 numéro de version

,OUT =
 (urd)

,DVC =
 type d'imprimante

,SP =
 espace

,T =
 T

Paramètres

Contenu

TBL

voir explication dans XTALLY03

VN

voir explication dans XTALLY03

SP

permet de spécifier l'espace disque du fichier de travail lorsque l'utilisateur désire introduire les titres de colonnes et de lignes. L'espace est exprimé en piste et il peut y avoir 228 titres par piste.

permet de choisir entre un fichier total ou détail.

N : fichier détail,

T : fichier total

Procédure de contrôle XTALLY06 et XTALLY07 :

Ces deux procédures permettent de soustraire ou d'additionner des des tableaux statistiques dans un fichier total.

Format : \$XTALLY06 USE = nom de l'utilisation
 ou \$XTALLY07

0

,VN =

numéro de version

0

,TN = numéro de version

,SP 10 espace

SD

,TBL =

nom de disque

NO

,NEW =

YES

Paramètres

Contenu

VN

voir explication dans XTALLY03

TBL

voir explication dans XTALLY03

TN

ce paramètre a la même signification que le paramètre VN, sauf qu'il s'applique à la version du fichier contenant le cumul des tableaux.

SP

indique l'espace nécessaire lors de la création du fichier total. Ce nombre doit être identique à celui utilisé lors de la production de XTALLY03.

NEW

ce paramètre est utilisé seulement par XTALLY06, et indique si le système doit créer le fichier total.

YES : création du fichier total.

NO : le fichier total existe déjà.

Procédure de contrôle XTALLY99 :

Cette procédure permet de vider le dictionnaire COXTALLY de l'utilisateur sur deux fichiers séquentiels le premier contenant la description des variables de classification et des catégories sous la forme des paramètres de XTALLY01, le deuxième contenant la description des tableaux statistiques sous la forme des paramètres de XTALLY02.

Format : \$XTALLY99 USE = nom de l'utilisation

,X01 = (fd)

,X02 = (fd)

,DELETE = YES

NO

Paramètres

Contenu

X01 Description du fichier qui doit recevoir les descriptions de variables de classification ainsi que les descriptions des catégories.

NOTE : Le RSZ doit être 80

X02 Description du fichier qui doit recevoir les descriptions des tableaux statistiques.

NOTE : Le RSZ doit être 80.

DELETE Sert à indiquer si le dictionnaire de l'utilisateur doit être détruit ou non.

YES : Destruction du dictionnaire

NO : Le dictionnaire est conservé

Utilisation de la procédure XTALLINK :

Cette procédure permet de créer le programme principal de tabulation de l'utilisateur.

Format : \$XTALLINK PGM = programme, RTN = routine;

PGM indique le nom du programme de tabulation de l'utilisateur qui est formé de la routine de lecture et du programme de tabulation du logiciel COXTALLY.

RTN nom de la routine de lecture des données à traiter. Cette routine est écrite par l'utilisateur.

Exemple de routine de lecture :

IDENTIFICATION DIVISION.

PROGRAM-ID nom.

ENVIRONMENT DIVISION.

CONFIGURATION SECTION.

SOURCE-COMPUTER. LEVEL-62.

OBJECT-COMPUTER. LEVEL-62.

INPUT-OUTPUT SECTION.

FILE-CONTROL.

SELECT fichier ASSIGN TO xtaldata.

DATA DIVISION.

FILE SECTION.

FD fichier LABEL RECORD STANDARD.

01 enregistrement PIC X (nnn).

WORKING-STORAGE SECTION.

01 SWT-OPEN PIC X (001) VALUE LOW-VALUE.

-

LINKAGE SECTION.

01 PASSED-RECORD PIC X (099).

PROCEDURE DIVISION USING PASSED-RECORD.

IF SWT-OPEN = LOW-VALUE

OPEN INPUT fichier

MOVE HIGH-VALUE TO SWT-OPEN.

READ fichier RECORD INTO PASSED-RECORD AT END

MOVE HIGH-VALUE TO PASSED-RECORD

MOVE LOW-VALUE TO SWT-OPEN

CLOSE fichier.

END-ROUTINE.

EXIT PROGRAM.

L'utilisateur n'a besoin que de modifier les endroits soulignés, il est bien évident que ce genre de routine peut devenir plus complexe du fait des fichiers à traiter.

Exemple de production de tableaux statistiques :

Dans cette annexe nous montrerons les différents paramètres nécessaires à la production de quatre tableaux statistiques différents, soit :

- a) Nombre de ménages et leur population selon le type de ménage (ordinaire ou collectif) et le nombre de personnes par ménage.
- b) Population totale selon les groupes d'âge, le sexe et l'état matrimonial.
- c) Population des ménages ordinaires selon le sexe, l'état matrimonial et le lien de parenté avec le chef de ménage.
- d) Population de six (6) ans et plus selon le sexe, les groupes d'âge et l'aptitude à lire et écrire.

De façon à bien illustrer nos exemples, nous assumons que le fichier avec lequel nous travaillons contient trois (3) types d'enregistrements différents qui contiennent un "M", un "P" ou un "C" dans la première position de l'enregistrement.

- a) M enregistrement contenant l'information sur les ménages, dont le type de ménage en position 25, et le nombre de personnes dans le ménage dans les positions 26, 27 et 28.
- b) P enregistrement contenant les informations sur les personnes dont le sexe en position 18, l'âge en position 19 et 20, le lien de parenté en position 17 et un code de dernière classe suivie en positions 32 et 33.
- c) C enregistrement contenant les informations sur les personnes vivant dans les ménages collectifs. Les informations sont les mêmes que pour les enregistrements de type P.

NOTE : Tout au long des exemples, les caractères écrits en MAJUSCULE font partie des cartes paramètres alors que les caractères en minuscule sont des commentaires. Les lignes de chiffres indiquent les positions sur la carte paramètre.

I) Description des paramètres nécessaires dans XTALLY01 :

000000000011111111112222222222333333333344444444445555555555
 12345678001234567890123456789012345678901234567890123456789

IDE0101*			type d'enregistrement
TYM2525*			type de ménage
NBR2628*			nombre de personnes par ménage
ETA2121*			état matrimonial
LIE-717*			lien de parenté
SEX1818*			sexe de la personne
AGE1920*			âge de la personne
DCL3233*			dernière classe suivie
TYM01	X	3	NON DECLARE
TYM01	0	2	COLLECTIF
TYM01	2	1	ORDINAIRE
TYM01	9	2	COLLECTIF
NBR01 001		1	UNE PERSONNE
NBR01 002		2	PERSONNES
NBR01 003		3	PERSONNES
NBR01 004		4	PERSONNES
NBR01 005		5	PERSONNES
NBR01 006		6	PERSONNES
NBR01 007		7	PERSONNES
NBR01 008		8	PERSONNES
NBR01 009		9	PERSONNES
NBR01 010		10	PERSONNES
NBR01 999		11	PLUS DE 10
SEX01	1	1	MASCULIN
SEX01	2	2	FEMININ
ETA01	X	5	NON DECLARE
ETA01	0	1	CELIBATAIRE
ETA01	5	2	MARIE
ETA01	6	3	VEUF
ETA01	7	4	DIVORCE
LIE01	X	5	NON DECLARE
LIE01	1	1	CHEF
LIE01	2	2	EPOUSE
LIE01	3	4	AUTRE
LIE01	4	3	ENFANT
LIE01	9	4	AUTRE
AGE01	XX	5	NON DECLARE
AGE01	05	1	0-- 6 ANS
AGE01	14	2	6-- 14 ANS
AGE01	49	3	15-- 49 ANS
AGE01	99	4	PLUS DE 50
AGE02	XX	4	NON DECLARE
AGE02	14	1	6-- 14 ANS
AGE02	49	2	15-- 49 ANS
AGE02	99	3	PLUS DE 50
DCL01	XX	4	NON DECLARE
DCL01	00	1	ANALPHABETE
DCL01	98	2	SCOLARISE
DCL01	99	3	LANG. NAT.

II) Description des paramètres nécessaires dans XTALLY02 :

000000000111111111222222222233333333334444444444555555555566666666
123456789012345678901234567890123456789012345678901234567890123456

TATYMO1, NBR01, PER2628
TBETA01, AGE01, SEX01,
TCETA01, LIE01, SEX01,
TDDCLO1, AGE02, SEX01,

III) Descriptions des paramètres nécessaires dans XTALLY03 :

000000000111111111222222222233333333334444444444
1234567890123456789012345678901234567890123456789

TA0101E P 0101E C
TB0101E M
TC0101B C M
TD1920B 00 050101E M

On notera que pour le tableau TD la première condition de rejet se fait sur l'âge, ceci vient du fait que nous assumons que notre fichier contient plus d'enregistrements sur des personnes de 0 à 5 ans que d'enregistrements sur les ménages, comme l'enregistrement est rejeté aussitôt qu'une condition est satisfaite, on sauvera du temps machine à l'exécution. Le même principe est appliqué pour le tableau TA, nous supposons qu'il y a plus d'enregistrements concernant les personnes des ménages ordinaires que d'enregistrements concernant les personnes des ménages collectifs.

IV) Description des paramètres nécessaires dans XTALLY04 pour l'impression du tableau TA

000000000111111111222222222233333333334444444444555555555566666666
123456789012345678901234567890123456789012345678901234567890123456

TA 060N2 NBR MENAGES POPULATION
TAT12 TABLEAU GENERAL 1 : NOMBRE DE MENAGES ET LEUR POPULATION SELON
TAT1 LE TYPE DE MENAGE ET LE NOMBRE DE PERSONNE PAR MENAGE
TAT22 TOUT LE PAYS
TAR1 T O T A L

V) Description des paramètres nécessaires dans XTALLY04 pour l'impression du tableau TB :

000000000111111111222222222233333333334444444444555555555566666666
123456789012345678901234567890123456789012345678901234567890123456

TB 060N21
TBT12 TABLEAU GENERAL 2 : POPULATION TOTALE SELON L'ETAT MATRIMONIAL
TBT1, LES GROUPES D'AGE ET LE SEXE
TBT22 POUR LE PAYS
TBR1 POPULATION TOTALE
TBR2 T O T A L

VI) Description des paramètres nécessaires dans XTALLY04 pour l'impression du tableau TC :

000000000111111111222222222233333333334444444444555555555566666666
123456789012345678901234567890123456789012345678901234567890123456

TC 060N21
TCT12 TABLEAU GENERAL 3: POPULATION DES MENAGES ORDINAIRES SELON
TCT1 L'ETAT MATRIMONIAL, LE LIEN DE PARENTE ET LE SEXE
TCT22 POUR LE PAYS
TCR1 POPULATION TOTALE
TCR2 T O T A L

VII) Description des paramètres nécessaires dans XTALLY04 pour l'impression du tableau TD :

000000000111111111222222222233333333334444444444555555555566666666
123456789012345678901234567890123456789012345678901234567890123456

TD 060N21
TDT12 TABLEAU GENERAL 4 : POPULATION DE SIX (6) ANS ET PLUS SELON
TDT1 "L'APTITUDE A LIRE ET A ECRIRE, LES GROUPES D'AGE ET LE SEXE
TDT22 POUR LE PAYS
TDR1 POPULATION DE 6 ANS ET PLUS
TDR2 T G T A L

VIII) Description des paramètres nécessaires dans XTALLY05 :

- a) Calcul du nombre moyen de personnes par ménage du tableau TA :

000000000111111111222222222233333333334444444444555555555566666666
123456789012345678901234567890123456789012345678901234567890123456

TA 060n2 NBR MENAGES POPULATION R/A
b) Calcul du pourcentage de la population du tableau TA:
TA 060n2 NBR MENAGES POPULATION %b
c) Calcul du pourcentage de la population selon la répartition d'après le total général du tableau TB :
TB 060n21 %
d) Calcul du pourcentage de la répartition des hommes et des femmes, les compteurs du tableau TB sont un pourcentage, soit du grand total masculin, soit du grand total féminin.
TB 060n21 % %b
e) Calcul du pourcentage de la répartition de la population dans chacun des groupes d'âge du tableau TB.
TB 060n21 % %b
f) Calcul du pourcentage de la répartition de la population dans chacun des sous groupes âge/état matrimonial.
TB 060n21 % %b

IX) Apparence des tableaux produits selon les cartes paramètres qui ont été introduites dans le logiciel :

a) Tableau ATA :

Tableau général 1: nombre de ménages et leur population pour le pays

	NBR ménages Total	Population Total	NBR ménages Ordinaire	Population Ordinaire
Total	n,nnn	nn,nnn	n,nnn	nn,nnn
Une personne	n,nnn	nn,nnn	n,nnn	nn,nnn
2 personnes	n,nnn	nn,nnn	n,nnn	nn,nnn
Plus de 10	n,nnn	nn,nnn	n,nnn	nn,nnn

b) Tableau TE :

Tableau général 2: population selon l'état matrimonial pour le pays

	Total	Célibataire	Marié
Population totale			
Total	n,nnn	n,nnn	n,nnn
Masculin	n,nnn	n,nnn	n,nnn
Féminin	n,nnn	n,nnn	n,nnn
0 - 5 ans	n,nnn	n,nnn	n,nnn
Total	n,nnn	n,nnn	n,nnn
Masculin	n,nnn	n,nnn	n,nnn
Non déclaré	n,nnn	n,nnn	n,nnn
Total	n,nnn	n,nnn	n,nnn
Masculin	n,nnn	n,nnn	n,nnn
Féminin	n,nnn	n,nnn	n,nnn

Records Read/Enregistrements lus : nn,nnn Rejected/Rejetés
nn,

c) Tableau TC :

Tableau général 3 : population des ménages ordinaires
pour le pays

	Total	Celibataire	Marié
Population totale			
Total	nn,nnn	nn,nnn	nn,nnn
Masculin	nn,nnn	nn,nnn	nn,nnn
Féminin	nn,nnn	nn,nnn	nn,nnn
Chef			
Total	nn,nnn	nn,nnn	nn,nnn
Masculin	nn,nnn	nn,nnn	nn,nnn
Féminin	nn,nnn	nn,nnn	nn,nnn
Non déclaré			
Total	nn,nnn	nn,nnn	nn,nnn
Masculin	nn,nnn	nn,nnn	nn,nnn
Féminin	nn,nnn	nn,nnn	nn,nnn

Records Read/Enregistrements Lus : nn,nnn Rejected /Rejetes
Nn

d) Tableau TD :

Tableau général 4 : population de six (6) ans et plus
pour le pays

		Analphabete	Scolarisé
Population de 6 ans et plus			
Total	nn,nnn	nn,nnn	nn,nnn
Masculin	nn,nnn	nn,nnn	nn,nnn
Féminin	nn,nnn	nn,nnn	nn,nnn
6 - 14 ans			
Total	nn,nnn	nn,nnn	nn,nnn
Masculin	nn,nnn	nn,nnn	nn,nnn
Féminin	nn,nnn	nn,nnn	nn,nnn
Non déclaré			
Total	nn,nnn	nn,nnn	nn,nnn
Masculin	nn,nnn	nn,nnn	nn,nnn
Féminin	nn,nnn	nn,nnn	nn,nnn

e) Tableau TA après l'opération B/A

Tableau général 1 : nombre de ménages et leur population pour le pays

	NBR ménages Total	Population Total	NBR ménages Ordinaire	Popu. Ordin.
Total	nn,nnn	moyenne	nn,nnn	moyenne
Une personne	nn,nnn	moyenne	nn,nnn	moyenne
2 personnes	nn,nnn	moyenne	nn,nnn	moyenne
Plus de 10	nn,nnn	moyenne	nn,nnn	moyenne
Records Read/Enregistrements Lus :	nn,nnn	Rejected/ Rejetes n		

f) Tableau TA après l'opération %B 1

Tableau général 1 : nombre de ménages et leur population pour le pays

	NBR ménages Total	Population Totale	NBR ménages Ordinaire	Popul. Ordin.
Total	nn,nnn	100.00	nn,nnn	X%
Une personne	nn,nnn	X%	nn,nnn	X%
2 personnes	nn,nnn	X%	nn,nnn	X%
Plus de 10	nn,nnn	X%	nn,nnn	X%
Records Read/Enregistrements Lus :	nn,nnn	Rejected/ Rejetes n		

g) Tableau TB après l'opération % :

Tableau général 2 : population selon l'état matrimonial pour le pays

	Total	Célibataire	Marié
Population totale				
Total	100.00	X%	X%	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	
0 - 5 ans				
Total	X%	X%	X%	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	

	Total	Célibataire	Marié
Non déclaré				
Total	X%	X%	X%	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	

Records Read/Enregistrements Lus : nn,nnn Rejected/
Rejetes n

- h) Tableau TB après l'opération % et spécification "YbB"
Tableau général 1 : population selon l'état matrimonial
pour le pays

	Total	Célibataire	Marié
Population totale	100.00	X%	X%	
Total	100.00	X%	X%	
Masculin	100.00	X%	X%	
Féminin	100.00	X%	X%	

0 - 5 ans

Total	X%	X%	X%
Masculin	X%	X%	X%
Féminin	X%	X%	X%

Non déclaré

Total	X%	X%	X%
Masculin	X%	X%	X%

Records Read/Enregistrements Lus : nn,nnn Rejected/
Rejetes n

- i) Tableau TB après l'opération % et spécification "YbB" :
Tableau général 1 : population selon l'état matrimonial
pour le pays

	Total	Célibataire	Marié
Population totale				
Total	100.00	X%	X%	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	
0 - 5 ans	100.00	X%	X%	
Total	X%	X%	X%	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	

	Total	Célibataire	Marié
Non déclaré				
Total	100.00	X%	X%	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	

Records Read/Enregistrements Lus : nn,nnn Rejected/
Rejetes n

j) Tableau EE après l'opération % et spécification "YBY "

Tableau général 1: population selon l'état matrimonial
pour le pays

	Total	Célibataire	Marié
Population totale				
Total	100.00	100.00	100.00	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	
Non déclaré				
Total	100.00	100.00	100.00	
Masculin	X%	X%	X%	
Féminin	X%	X%	X%	

Records Read/Enregistrements Lus : nn,nnn Rejected/
Rejetes n

LES "SWITCHES"

- I) Le logiciel COXTALLY permet de communiquer avec l'utilisateur au moyen de "switches" et c'est à partir de la console de l'ordinateur.

Nous allons ici décrire les trois "switches" utilisées dans le logiciel COXTALLY.

Ces trois "switches" sont utilisées par le programme de tabulation (XTALLY03), du fait que c'est le programme qui est le plus long à exécuter, ayant à lire les données de l'utilisateur.

- a) "SWITCH 1" : Cette "switch" permet au programme de tabulation de savoir s'il doit imprimer les enregistrements rejetés ou non. Si elle est "ON", il n'y a pas d'impression des enregistrements rejetés.
- b) "SWITCH 2" : Cette "switch" permet à l'utilisateur de savoir à quel enregistrement le programme de tabulation est arrivé. Lorsque cette "switch" est mise "ON", le programme de tabulation écrit à la console le numéro séquentiel de l'enregistrement qu'il vient tout juste de lire. Cette "switch" est remise sur "OFF" aussitôt que l'information demandée est donnée à l'utilisateur.
- c) "SWITCH 3" : Cette "switch" permet à l'utilisateur de stopper le programme de tabulation de façon normale, c'est à dire que tous les fichiers sont fermés et que le travail effectué n'est pas perdu. Lorsque cette "switch" est mise à "ON", le programme de tabulation indique le numéro séquentiel du dernier enregistrement qui a été traité au complet, ferme tous les fichiers et s'arrête normalement.

- II) Méthode de mettre les "switches" sur "ON" lorsque le logiciel fonctionne sur un ordinateur HP-62 ou DPS-4.

- a) Il s'agit d'effectuer la commande suivante sur la console : SETS Jnnn, SWx=ON

"nnn" est le numéro du JOB exécutant le programme de tabulation. "x" est le numéro de la "switch" que l'on veut mettre "ON".

Lorsque la "x" est égal à "3" le programme de tabulation est arrêté, et immédiatement après le job est annulé (avorté). Pour relancer le programme de tabulation voir la procédure de contrôle d'exécution XTALLY03 en page 82.

RESUME DES DISCUSSIONS

4. DISCUSSIONS

A. SUIVI DES PHASES DU TRAITEMENT

Généralement, afin de permettre la constitution du lot à traiter à l'étape suivante, les principales phases du traitement des données, à savoir le codage, la saisie, le traitement sur ordinateur, démarrent avec un décalage approprié. Par la suite, elles sont exécutées de façon concomitante, ce qui impose une organisation et un suivi aussi minutieux que possible. De ce fait, les débats ont porté sur les dispositifs à mettre en place en vue d'assurer une gestion efficace de ces tâches. En particulier, trois points ont attiré l'attention des participants : la circulation des dossiers, le contrôle de la production et le contrôle d'exhaustivité. Il en est ressorti les conclusions suivantes :

Circulation des dossiers :

Pour assurer une bonne circulation des documents, ceux-ci doivent nécessairement être regroupés par zone de dénombrement (ZD), élément de base du traitement. Une fiche suiveuse pourrait être collée ou agrafée sur chaque dossier de ZD contenant les informations analytiques sur l'identification géographique, la nature des tâches, les dates de début et de fin d'exécution des tâches, le nombre de questionnaires de ménages et d'individus, les visas des responsables.

Les numéros des dossiers transmis par lots d'un groupe à un autre seront consignés dans des registres exclusivement tenus par les responsables de groupes qui en assureraient le retour en salle d'archives.

Au niveau du traitement sur ordinateur, on veillera à produire pour chaque ZD, des statistiques permettant, après confrontation avec les informations de la fiche suiveuse, de décider que le dossier est complètement traité.

Contrôle de production

Les productions des agents codeurs et des opérateurs de saisie devront être relevées à la fin de chaque vacation. Il importe d'en faire une évaluation hebdomadaire par rapport aux prévisions et aux échéances. On pourrait même envisager la prise en charge d'un tel suivi par une chaîne informatique, lorsque le nombre d'agents est assez élevé.

Exhaustivité

Il a été conseillé de confectionner une sorte de tableau de bord où, pour chaque ZD, on marquerait la fin normale de chaque étape du traitement. On s'assurera alors qu'une phase est entièrement achevée par le noircissement complet de la partie correspondante dudit tableau.

B. DETERMINATION DU NOMBRE DE POSTES DE SAISIE

Ce calcul a fait l'objet de débats très intéressants. Une méthode de calcul a été étudiée qui prend en compte les éléments suivants :

C : volume du fichier (en caractères ou frappes)

M : nombre de mois de travail (en mois)

S : nombre de semaine de travail par mois (en semaines/mois)

J : nombre de jours de travail par semaine (en jours/semaine)

H : nombre d'heures de travail par jour (en heures/jours)

V : vitesse moyenne de saisie (en frappes/heures)

G : nombre de groupes de travail

P : nombre approximatif de postes de saisie

P est obtenu par la formule :

$$P = \frac{C(1 + T)}{V \times H \times J \times S \times M \times G}$$

où T représente le pourcentage de la partie du fichier qui est soumise à la vérification. Des exemples ont permis d'illustrer ce calcul.

C. EVALUATION D'UN PROGICIEL

Les participants se sont en particulier intéressés à la qualité et au mode d'acquisition des progiciels.

Plusieurs critères de choix de progiciels ont été exposés. Dans le cadre du traitement des recensements, les produits-programmes seront choisis en fonction :

- des facilités qu'ils offrent notamment à la production des tableaux et à l'analyse statistique
- des aspects techniques de mise en oeuvre (exigence en logiciel et hardware, méthode d'accès des fichiers etc);
- des possibilités de maintenance
- de la documentation d'accompagnement
- du coût.

Des indications ont été données à propos des organismes auprès desquelles les progiciels peuvent être obtenus.

D. METHODE DE REDRESSEMENT DES ERREURS - HOT DECK ET COLD DECK

Cette méthode de correction a attiré l'attention de la majorité des participants qui en entendaient parler pour la première fois. Ils ont pu suivre avec beaucoup d'intérêt les explications sur la démarche théorique de la méthode. De nombreux exemples ont été traités de façon détaillée sur leur demande.

Tout le long des débats sur ce point cependant, quelques participants ont fait part d'une certaine appréhension sur l'utilisation de cette méthode, car disent-ils, elle fournit des résultats différentes selon la manière dont les données sont classées et, qui plus est, une même valeur a une probabilité non négligeable d'être imputée sur une grande partie du fichier, ce qui aboutit à la concentration des résultats sur quelques modalités des variables.

Des explications supplémentaires ont été fournies sur la caractère aléatoire des imputations et sur le classement appropriés des données avant leur soumissions à la correction par hot deck.

Enfin, les participants ont manifesté le désir d'étudier de façon plus approfondie, cette méthode de correction.

En revanche, la méthode cold deck, par son approche conceptuelle leur a paru familière. Toutefois, sa mise en oeuvre requérant davantage de temps de traitement du fait de la double lecture des données, les participants ont émis des doutes sur son utilisation.

E. DEBAT SUR LES PROGICIELS

Les exposés sur les progiciels ont en général suscité de nombreuses questions sur leur fonctionnement, les possibilités offertes, les avantages des uns par rapport aux autres.

Les travaux pratiques initialement prévus auraient permis de répondre plus concrètement à certaines questions. Malheureusement, ils n'ont pas pu avoir lieu pour deux raisons :

- les progiciels n'ont pas été installés à temps
- les versions XTALLY et UNEDIT installés sur l'IBM/32 local étaient incomplètes.

Les participants ont néanmoins pu examiner les résultats de ces progiciels sur les listings produits à New York.

Il est revenu au cours des débats que le temps accordé à cette partie du séminaire est très insuffisant, les participants ayant marqué, de par leurs interventions, une importance toute particulière à chacun des progiciels.

Certains ont même émis le souhait de voir cette partie du séminaire constituer un atelier distinct tandis que d'autres, relevant l'interaction entre l'exécution des recensements et leur traitement informatique ont émis un voeu contraire. Mais ils se sont accordés à suggérer que la période réservée à l'étude des progiciels soit allongée lors des prochains séminaires qu'ils encouragent vivement la CEA à organiser.

Annexe I

1. OBSERVATION DES PARTICIPANTS SUR L'ORGANISATION ET L'EFFICACITE DE L'ATELIER

La dernière séance de l'atelier a été consacrée à l'audition des participants pour qu'ils expriment leurs observations et leurs suggestions sur les sujets traités ainsi que sur son organisation.

Des observations émises, il ressort que les participants sont satisfaits du déroulement de l'atelier dans son ensemble. L'initiative prise d'organiser l'atelier a été qualifiée de positive notamment le choix du moment qui correspondait approximativement au démarrage de la deuxième série de recensements de bon nombre de pays africains francophones et la pertinence des sujets.

Ils ont ainsi exprimé des remerciements à l'endroit du Gouvernement camerounais, de la CEA, du FNUAP, du Bureau de statistique des Nations Unies et du Bureau de recensement des Etats Unis et proposé qu'un tel atelier de formation soit organisé périodiquement tous les deux ans par exemple.

Les participants auraient également souhaité connaître et discuter plus amplement de l'expérience et des difficultés rencontrées par certains pays dans l'exécution des précédents recensements.

Par ailleurs, ils ont émis des critiques et fait des suggestions visant à améliorer l'efficacité des futurs ateliers éventuels, parmi elles, ont peu citer :

- l'absence de démonstration, à partir de données réelles, sur l'utilisation des progiciels étudiés en vue d'un choix plus conséquent.
- l'inadaptation de l'exposé sur l'analyse des données du recensement à l'auditoire de l'atelier composé en partie d'informaticiens et de statisticiens.
- la nécessité d'approfondir les thèmes abordés en séparant les informaticiens des statisticiens et des démographes.
- le choix d'un seul participant par pays, à la charge de la CEA est la cause de l'hétérogénéité des disciplines représentées.
- l'arrivée tardive des invitations de certains participants a également été déplorée.

Enfin, les participants ont tout de même reconnu la nécessité de ce dialogue entre informaticiens, statisticiens et démographes. Ils ont recommandé l'assistance mutuelle entre les pays en développement dans l'exécution des recensements et dans l'utilisation des progiciels statistiques.

Annexe I (suite)

L'équipe de coordination a pour sa part, fourni des éléments de réponses aux questions soulevées par les participants.

En ce qui concerne les modalités de l'organisation, plusieurs parties étaient impliquées : la CEA chargé de l'exécution du projet, le FNUAP qui finance, le Bureau de recensement des Etats Unis, le Bureau de statistique des Nations Unies qui devaient mettre des instructeurs à la dispositions de l'atelier pour la partie informatique, l'IFORD et enfin le pays hôte.

Les consultations entre ces différents parties ne sont pas toujours aisées, compte tenu des préoccupations de chacune d'elles.

S'agissant de la démonstration sur l'utilisation des progiciels, le manque de logiciel adéquat, sur place, a empêché la tenue de cette séance. Il a fallu recourir à la présentation des exemples de programmes sur des listings tirés à New York.

2. CONCLUSION

L'atelier de formation sur l'exécution des recensements de la population et sur l'utilisation des progiciels statistiques organisé sous l'égide de la Commission économique des Nations Unies pour l'Afrique (CEA) s'est déroulé, conformément au calendrier préétabli.

Les participants se sont beaucoup intéressés aux sujets traités, ce dont témoigne la pertinence des points de discussion qu'ils ont soulevés et leur assiduité aux séances.

A cause de la diversité des disciplines scientifiques d'appartenance des participants (démographes, statisticiens et informaticiens) il n'a pas été possible d'approfondir les aspects techniques de certains sujets notamment l'analyse des données et les mécanismes des progiciels CONCOR et CENTS-4. D'autant plus que les séances d'utilisation pratique de ces progiciels n'ont pas eu lieu.

Néanmoins, les explications fournies par les différents encadreurs, faisant surtout ressortir les aspects pratiques des problèmes abordés, ont permis de combler toutes ces lacunes.

Il faut rappeler que l'un des objectifs de cet atelier était de sensibiliser les responsables des futurs recensements généraux de la population, sur les techniques d'exécution de toutes les phases, pour qu'ils puissent aborder, en initiés, les problèmes que rencontrent leurs proches collaborateurs.

Annexe I (suite)

De ce point de vue, l'atelier a connu un succès total tel que l'ont exprimé les participants qui ont manifesté un sérieux et un intérêt particuliers tout au long de son déroulement.

Cependant, si l'atelier a atteint tous ses objectifs, il le doit également à la disponibilité des Autorités camerounaises à travers la Direction de l'Institut de statistique, de planification et d'économie appliquée (I.S.P.E.A.) et la Direction de la statistique et de la comptabilité nationale. Ces deux institutions ont mis à sa disposition tous leurs moyens matériels et humains.

En outre, l'apport obtenu d'autres organismes en production de documents et en techniques d'exploitation et de traitement des données des recensements, a été déterminant pour le succès de l'Atelier. Les organismes et institutions représentés étaient le Bureau de recensement des Etats-Unis, le Département de la coopération technique pour le développement (DTCD) le Bureau de statistique des Nations Unies, le Programme des Nations Unies pour le développement (PNUD) et l'Institut de formation et de recherche démographique (IFORD).

Annexe II

ATELIER DE FORMATION SUR L'EXECUTION D'UN RECENSEMENT ET L'UTILISATION DES PROGICIELS STATISTIQUES (YAOUNDE, 1-26 Octobre 1984)

Liste des participants

N° d'ordre	NOMS ET PRENOMS	PAYS	FONCTION ET ADRESSE
1.	Mr. SAHI FATAH ²	ALGERIE	Chef, Département informatique O.N.S. 8 et 10, rue des Moussebilines - ALGER
2.	Mr. BRAHIMI RABAH ¹	ALGERIE	Chef, Département analyse données démographiques O.N.S. 8 et 10, Rue des Moussebilines ALGER
3.	Mr. HOUSSOU RAYMOND D.	BENIN	Chef, Centre informatique INSAE, BP. 323 COTONOU
4.	Mr. OUEDRAGO T. MATHIEU	BURKINA FASO	Adjoint Chef, service statistique sociale BP. 374 - OUAGADOUGOU
5.	Mr. HAKOUA AMBROISE	CAMEROUN	Chef, service informatique à la Direction de la statistique - YAOUNDE
6.	Mr. AHANDA JEAN MARIE	"	Démographe, Direction de la statistique MINPAT
7.	Mr. BACKINY YETNA P.	"	Statisticien, service informatique YAOUNDE
8.	Mr. MOUSSIMA JEANNOT	"	Cadre B.C.R.E., Direction statistique YAOUNDE
9.	Mr. ALAM EMMANUEL	"	Démographe, Direction statistique YAOUNDE
10.	Mr. TCHUINOU DAVID	"	Cadre au B.C.R.E. MINPAT - YAOUNDE
11.	Mr. MIROIDI ALLAOUI	COMORES	Sociologue, Direction de la statistique BP. 131 MORONI
12.	Mr. BITEMO MICHEL ²	CONGO	Directeur de la démographie, Direction générale de la statistique
13.	Mr. KIMBASSA MOUKALA J. ²	"	Informaticien O.C.I. BRAZZAVILLE BP. 2031
14.	Mr. DERMAN YOUSOUF	DJIBOUTI	Statisticien économiste au B.C.R.
15.	Mr. KAIRE ALI ARREH ¹	"	Directeur B.C.R. BP. 1732 - DJIBOUTI
16.	Mr. MAGAYA JEAN PAUL	GABON	Directeur statistiques démographiques BP. 5402 - LIBREVILLE
17.	Mr. LEVY NTEM ALLOGO	"	Chef, Département ressources humaines, PDG, BP. 268 - LIBREVILLE
18.	Mr. KABA MAHMOUDOU	GUINEE	Chef, service exploitation et analyse données, BP. 221 - CONAKRY
19.	Mr. RABEMANANTSOA	MADAGASCAR	Chef, service technique recensement, BP. 485 INSRE - ANTANANARIVO
20	Mme KONARE KADIDIA	MALI	Superviseur régional, Division enquête BP. 12 - BAMAKO

1 : Ayant pris part à la première partie seulement

2 : Ayant pris part à la deuxième partie seulement

Annexe II (suite)

N° d'ordre	NOMS ET PRENOMS	PAYS	FONCTION ET ADRESSE
21.	Mr. NARAMBA TELESPHORE	RWANDA	Chef, Division statistiques courantes BP. 46 - KIGALI
22.	Mr. SADIO ABDOULAYE	SENEGAL	Chef, service statistiques socio- démographiques BP. 116 DAKAR
23.	Mr. AFODAH K. EGUIDA	TOGO	Chargé exploitation recensement, BP. 520 - LOME
24.	Mr. KASHALA MUKENDI	ZAIRE	Responsable informatique plan, BP. 7893 - KINSHASA 1
25.	Mr. PAUL THEODORE MBAGUIA	UDEAC	Expert démographe, UDEAC - CREP, BP. 1418 BANGUI

Annexe III

ATELIER DE FORMATION SUR L'EXECUTION D'UN RECENSEMENT ET L'UTILISATION DES PROGICIELS STATISTIQUES.

Liste des animateurs

a) Première partie

- | | |
|----------------------------|--|
| 1. Mr. Abdou Karim DIOP | - Coordonnateur, conseiller régional en Etat civil CEA |
| 2. Mr. Oumar BOCOUM | - Conseiller régional en cartographie CEA |
| 3. Mr. Antoine SIMONPIETRI | - Conseiller technique principal-projet UDEAC/CREP-CEA |
| 4. Mr. Pierre FOKOM | - Conseiller régional en démographie CEA |
| 5. Mr. Hamady SOW | - Conseiller régional en statistiques démographiques - CEA |
| 6. Mr. Julien AMEGANDJIN | - Directeur - IFORD |
| 7. Mr. HOUHOUGBE | - Professeur - IFORD |
| 8. Mr. MFOULOU | - Professeur - IFORD |

b) Deuxième partie

- | | |
|-----------------------|--|
| 1. Mr. Michel LACKNER | - Bureau statistique ONU |
| 2. Mr. Manuel CHUA | - Chargé de progiciels utilisés en Chine projet INT/83/P56 |
| 3. Mr. Jean Marc HIE | - Directeur, Bureau de recensement du Cameroun |
| 4. Mlle Nina BERLET | - Bureau de recensement des Etats-Unis |
| 5. Mr. D.E.BENZINE | - Conseiller régional en informatique CEA |