

DW 2466

TEST DU QUESTIONNAIRE DE
L'ENQUETE INTEGREE AU GHANA
LE PROGRAMME DE SAISIE

ACCRA, LE 28 MAI 1990

JEANNINE. A. COULIBALY

Dans le cadre du projet "Dimensions Sociales de l'Ajustement Structurel" (DSA), la Banque Mondiale est sur le point de mettre en place, un dispositif d'enquêtes auprès des ménages. Ce dispositif comprend une enquête prioritaire et une enquête intégrée. Le questionnaire de l'enquête prioritaire est léger tandis que celui de l'enquête intégrée est lourd. L'interview pour cette dernière se réalise en deux passages séparés d'un intervalle de quinze jours, tandis que celui de la première se fait en vingt minutes. Les données collectées fourniront des renseignements sur les différents aspects de la vie économique et sociale des ménages ainsi que les interactions entre des phénomènes tels que les revenus, les dépenses, la santé, l'emploi, l'éducation ...

Ces enquêtes doivent répondre aux besoins prioritaires de plusieurs utilisateurs et particulièrement à ceux des organismes de planification des pays participant à ce projet.

La préparation de l'enquête intégrée exige un travail préalable de mise au point des différents outils de collecte, de saisie, de traitement et d'analyse des données. A cet effet, l'équipe technique du projet DSA a organisé un test pour le questionnaire de l'enquête intégrée au Ghana du 15 Avril au 30 Mai 1990.

Le présent rapport porte essentiellement sur le test du programme de saisie. Tout d'abord, il présente une brève description des erreurs, et ensuite il explique l'architecture du programme avec ses grandes fonctionnalités. Enfin la dernière partie du rapport fournit quelques remarques et des recommandations.

1. LES DIFFERENTS TYPES D'ERREURS

Une évaluation correcte du programme de saisie, ne serait possible sans une présentation des différents types d'erreurs qui surviennent généralement lors d'une enquête. Celles-ci se produisent essentiellement au cours des étapes suivantes:

1.1 L'enquête

L'erreur survient de deux façons:

- L'enquêteur peut créer l'erreur soit dans la formulation de la question, soit dans l'enregistrement des réponses. IL est nécessaire que l'individu qui pose les questions soit en mesure de les comprendre d'abord, avant de pouvoir les transmettre correctement à son interlocuteur.

- La personne enquêtée peut fournir une réponse incorrecte soit parcequ'elle a mal compris la question, soit de façon délibérée. Lorsque l'interview devient trop long, l'enquêté commence à donner des réponses expéditives pour pouvoir se débarrasser de l'enquêteur au plus vite. Certains refusent catégoriquement de répondre aux questions.

1.2 La codification des réponses

On assiste de plus en plus à l'utilisation de questionnaires précodés dans les enquêtes. Des codes numériques ou alphanumériques sont affectés à l'avance aux réponses possibles. Puisque l'énumération de ces réponses n'est jamais exhaustive, on se retrouve souvent avec des cas de figure qui n'avaient pas été prévus. Généralement ces cas sont rangés dans la catégorie "Autres"; De telles réponses perdent leur valeur lors de l'analyse et de l'exploitation des données.

1.3 La saisie des données

Quelque soit le mode de saisie utilisé, il arrive toujours que des erreurs se glissent dans le fichier par une maladresse de l'agent de saisie ou même par une défaillance quelconque du système utilisé. De ce fait, le programme de saisie doit comporter le maximum de spécifications de contrôle.

1.4 Le redressement automatique des erreurs

Le programme de saisie doit être rigoureusement testé pour redresser les erreurs de logique qu'il peut contenir. Ces erreurs étant parfois difficilement décelables, il convient que le test prévoit les contrôles suivants:

- les contrôles de cohérence
- les contrôles de vraisemblance
- les messages d'erreurs...

Bien que le redressement automatique des données permette de gagner beaucoup de temps, il est bien loin de résoudre tous les problèmes pratiques de terrain. Dans tous les cas, la qualité du fichier final dépendra de la capacité du programme à prendre en compte les contrôles jugés nécessaires par les statisticiens.

II. LE TEST DU PROGRAMME SAISIE

2.1 Description du programme et de ses grandes fonctionnalités

a. Les fichiers

Le programme a été mis au point avec RODE-PC. C'est un progiciel spécialement indiqué pour la saisie des données sur micro-ordinateur. Le programme est conçu en deux fichiers (ROUND1 et ROUND2). Comme leurs noms l'indiquent, ces deux fichiers représentent le premier et le deuxième passage de l'enquête sur terrain.

Le ROUND1 contient les sections 1 à 7 tandis que le ROUND2 contient les sections 8 à 13.

Les sections sont réparties dans des sous-fichiers appelés formats à l'intérieur d'un ROUND. Le ROUND1 contient 25 FORMATS et le ROUND2 en contient 28. La raison pour laquelle le premier passage et le deuxième passage n'ont pas été regroupés dans le même fichier est qu'un fichier RODE-PC ne peut contenir plus de 32 FORMATS.

En dehors des fichiers ROUND1 et ROUND2, nous avons les fichiers annexes suivants:

- un fichier de procédures : "FUNCTION.AVL"
Lors de l'exécution du AVL, les routines sont chargées à partir de ce fichier externe.
- un fichier de variables globales: "TEMP.VAR"
Lors de l'exécution du AVL, les variables communes à tous les formats sont chargées à partir de ce fichier externe.

- un fichier de cohérence: "COHER.DAT"
Ce fichier contient les valeurs utilisées par le AVL pour les comparaisons afin d'assurer la cohérence entre les données.
- des fichiers d'impression: ".BOX"
Lors de l'impression des données, les routines d'impression sont chargées à partir de ce fichier pour l'encadrement des données imprimées sur listing pour les contrôles visuelles.

b. L'architecture du programme

La conception du programme de saisie étant modulaire, nous avons été en mesure de faire évoluer des FORMATS en y incorporant des modifications et des contrôles sans pour autant bouleverser les autres. Ceci est un élément positif dans la mesure où ce programme devra être adapté au cas spécifique du pays où l'enquête se déroule. Cependant, le problème qui reste posé est celui de la création d'un lien entre le ROUND1 et le ROUND2 afin que les contrôles de cohérence soient faits entre les données des deux passages.

2.2 Le test du programme et la formation des agents de saisie

a. La saisie

Le test du programme a démarré avec l'adaptation du questionnaire et la formation des enquêteurs. A partir du questionnaire modifié, le programme a été modifié de façon à prendre en compte les spécifications locales. En raison de l'utilisation aisée du RODE-PC et de l'architecture modulaire du programme, la modification de celui-ci et la formation des opérateurs de saisie s'est faite sans problème. Le jeu d'essai du test a été obtenu à partir de données collectées auprès de 180 ménages ghanéens.

L'utilisation d'un échantillon de données réelles nous a donné l'occasion de tester toutes les sections du questionnaire et d'y apporter un grand nombre de corrections. De nouveaux codes et des nouvelles variables ont été créés pour prendre en compte plusieurs cas de figures qui avaient été initialement omis. Nous pensons qu'avec les données des 180 ménages, le programme a été suffisamment testé. Les contrôles de base réalisés sont:

- le contrôle du caractère saisi (numérique ou alphanumérique).
- le contrôle de la donnée elle-même.
- le contrôle entre les données d'un même enregistrement.
- le contrôle entre les données appartenant à des enregistrements distincts dans le même fichier.
- le contrôle entre des enregistrements appartenant à des fichiers distincts.
- le contrôle de l'unicité de l'enregistrement.

b. L'impression des données saisies

Le programme de saisie fournit les sorties suivantes:

- un fichier de données sur disque (batch),
- un listing des données saisies avec une signalisation des erreurs.

L'impression des données se déclenche systématiquement au moment où l'agent de saisie finit de rentrer un enregistrement. De ce fait, l'imprimante doit être prête à tout moment durant la saisie. Si elle ne l'est pas pour une raison ou une autre, le système de saisie s'arrête.

c. Les corrections et les mises à jour

Les options "Modify" et "Update" du RODE-PC permettent à l'opérateur de saisie de corriger des données qu'il a déjà rentrées dans la machine. Dans sa version actuelle, le programme n'envoie pas les données à l'imprimante après que celles-ci aient été corrigées. L'impression n'est faite que lorsqu'on se trouve en mode "Enter". Il serait intéressant de faire l'impression des corrections afin de faciliter le travail entre le superviseur de l'équipe et l'agent de saisie. Par ailleurs, le dernier listing (celui des données corrigées) est toujours la preuve que le travail de saisie a été correctement fait.

d. La sauvegarde des données

Les fichiers "batch" contenant les données saisies doivent être régulièrement sauvegardés sur des disquettes. Selon le manuel de l'opérateur de saisie, les agents doivent faire ces sauvegardes en utilisant des commandes du DOS qui leur ont été spécifiées en annexe. Cette utilisation n'est pas toujours aisée pour eux. Il faudrait le faire avec précaution pour éviter les destructions accidentelles de fichiers.

e. L'extraction des données et la conception des fichiers d'analyse

La procédure d'extraction du RODE-PC permet d'extraire les données d'un fichier "batch" pour les stocker dans un fichier ASCII destiné à être utilisé par un logiciel d'analyse statistique. Dans la conception des fichiers d'analyse, le critère d'homogénéité des données est le plus important. C'est pourquoi un identifiant a été affecté à chaque enregistrement en sortie. L'identifiant est composé de:

- numéro de ménage
- numéro de la section
- numéro de l'individu

Pour l'instant, le manuel de l'opérateur de saisie ne fait pas allusion à l'extraction des données.

III. RECOMMANDATIONS ET REMARQUES

Dans ce chapitre, nous formulons quelques recommandations tout en apportant un point de vue sur les différents aspects du programme auxquels nous pensons que des améliorations pourraient être apportées.

3.1 L'architecture du programme

Bien que les contrôles inter-section à l'intérieur du ROUND1 et du ROUND2 soient réalisés, nous pensons qu'il est nécessaire que l'on crée un lien entre les deux passages. Ceci est d'autant plus important que certaines sections du second passage utilisent soit les données, soit les résultats du premier passage.

Si aucun lien n'est mis en place, il est évident que la cohérence des données entre les deux passages ne sera plus assurée. A première vue, la mise en relation des deux passages ne semble pas aisée dans le contexte du RODE-PC à cause des contraintes liées à l'utilisation de ce logiciel. Cependant, on pourrait envisager la création de fichiers externes pour tenter de résoudre ce problème.

3.2 L'impression des données saisies

Nous pensons qu'il serait intéressant de modifier légèrement la procédure d'impression de manière à la rendre optionnelle. En effet, il faut donner le choix à l'agent de saisie d'imprimer ou de ne pas imprimer les données pour les raisons suivantes:

- le programme dans sa forme actuelle, ne permet pas de continuer la saisie en cas de panne d'imprimante ou de manque de papier. Lorsqu'un tel incident se produit il faut stopper la saisie en attendant que la situation normale se rétablisse.

Le calendrier contraignant de l'enquête ne permettrait pas que la saisie s'arrête des problèmes aussi mineurs. Pour éviter que la saisie s'arrête pour ce genre d'incidents, il s'avère nécessaire de rendre l'impression optionnelle.

3.3 La sauvegarde des données et l'utilisation des commandes DOS par l'agent de saisie.

Il serait souhaitable de ne pas compter sur les connaissances d'un opérateur de saisie. Pour éviter les mauvaises manipulations, nous pensons qu'il serait judicieux de mettre à la disposition des agents de saisie un utilitaire qui l'aiderait à faire les opérations de formatage, de copie et d'effacement avec le minimum de risque. Des utilitaires tels que PC-TOOLS, NORTON etc... présentent un intérêt pour ce genre d'opérations.

3.4 Les ressources humaines

En plus du fait que les spécifications de contrôles doivent être mises au point par les statisticiens en collaboration avec les informaticiens, il serait souhaitable que les premiers définissent les caractéristiques des variables utilisées.

3.5 Le manuel de référence technique et la documentation du programme de saisie

Nous pensons que le programme de saisie devrait être accompagné d'une documentation exhaustive décrivant et expliquant les variables et les procédures utilisées. Ceci doit pouvoir permettre aux personnes qui n'ont pas participé à la conception du programme de pouvoir en assurer la maintenance sans problème.