

# LIMITES ET INTÉRÊT DES ASPECTS QUANTITATIFS D'UNE ÉTUDE SOCIO-ÉCONOMIQUE MÉTHODOLOGIE DE L'ENQUÊTE STATISTIQUE “ PLAINE DE TANANARIVE ”

PAR

Marc BIED-CHARRETON (\*)

## 1 — INTRODUCTION

Nous nous proposons ici d'exposer les méthodes statistiques qui ont permis de réaliser sur le plan quantitatif l'étude « Plaine de Tananarive » et de voir de quelle manière on pourrait concilier les impératifs méthodologiques du *quantitatif* avec ceux des *enquêtes qualitatives* plus traditionnelles en géographie ou en sociologie.

Disons tout de suite que l'enquête quantitative ne doit être perçue que comme un instrument d'analyse chiffrée et globale et comme un moment de l'étude ; elle doit être intégrée dans un ensemble qui comprend d'abord des pré-enquêtes précédant l'étude statistique proprement dite, puis des études ponctuelles plus qualitatives, simultanées ou postérieures, qui permettent de resituer dans leur contexte géographique les résultats chiffrés obtenus.

## 2 — DÉTERMINATION DES UNITÉS-ÉCHANTILLONS ET DU MODE DE SONDAGE - CRITIQUE

Cette « Plaine de Tananarive », a été découpée en deux parties : l'une dite « zone des tanety » (1), et l'autre dite « zone des vallées » (2). Seule cette dernière fut l'objet de l'enquête par sondage. Le cas de la zone des tanety sera envisagé au paragraphe 6.

---

\* Chargé de recherches ORSTOM. Section géographie. ORSTOM - TANANARIVE.

(1) *tanety* : collines.

(2) Pour plus de précisions sur les zones d'enquêtes, voir article de J. WURTZ, p. 83, *Réflexions à propos d'une enquête quantitative sur la région de Tananarive*.

La zone des vallées groupe environ 90 000 habitants et comprend la majeure partie des zones agricoles inondables des Plaines de Tananarive, ainsi que les franges de collines environnantes.

Il n'était donc pas question de procéder à un recensement intégral des individus dans une région aussi importante. Nous avons donc opéré par sondage, ce qui nous a permis par la suite d'extrapoler les résultats à la totalité de l'ensemble régional intéressé. Nous avons choisi le sondage aléatoire plutôt que des méthodes empiriques d'échantillonnage. En effet, le sondage aléatoire permet d'extrapoler avec rigueur et surtout d'exprimer avec précision l'erreur d'échantillonnage, c'est-à-dire l'erreur due au fait qu'une partie seulement de la population est enquêtée.

Par la suite, nous désignerons l'ensemble régional étudié par sondage par le terme d'« univers statistique d'enquête ». Cet univers statistique est composé de hameaux regroupés en villages administratifs.

Après pré-enquêtes, le hameau nous a paru l'unité échantillon la meilleure pour l'enquête démographique. La totalité des hameaux de la région enquêtée constitue l'ensemble des « individus », ou la « population » de l'univers statistique d'enquête. Les « individus » déterminés par le sondage sont appelés « unités-échantillons » (1).

La difficulté, en matière de sondage, consiste alors à dresser la liste complète de tous les hameaux de la région à étudier, c'est-à-dire la liste des « individus » de l'univers statistique. Dans cette liste qui doit servir de base de sondage on fera figurer un certain nombre de critères indispensables à la statification et au tirage au sort : par exemple, la population de chaque hameau. Dans la base de sondage doit figurer chaque « individu » de la population, c'est-à-dire chaque hameau de la zone étudiée, sans omission ni répétition. Tous les individus-échantillons (c'est-à-dire les hameaux tirés) doivent effectivement être observés : ceci est fondamental : chaque hameau désigné par le sondage devra obligatoirement être enquêté. Disons tout de suite que cela a posé un problème au cours de l'enquête : deux hameaux ont refusé l'enquête. Il a fallu user de toutes sortes d'arguments pour convaincre les habitants ; par la suite, l'habileté des enquêteurs a rétabli la situation.

D'autre part, l'établissement d'une telle liste aurait permis de tirer facilement les hameaux unités-échantillons à égales probabilités, donc d'avoir le même taux d'extrapolation pour chaque hameau. Le dépouillement et les calculs en auraient été grandement facilités et se seraient effectués de la même façon que pour une opération de recensement : les résultats de chaque hameau étant totalisés puis multipliés par un coefficient unique d'extrapolation.

Dans la pratique, il nous a été impossible de dresser la liste de tous les hameaux : ceux-ci sont trop nombreux, environ 2 500, et nous ne disposions pas de documents administratifs suffisamment précis. Il aurait fallu en dresser nous-mêmes la liste, enquêter de façon sommaire pour connaître la population de chaque hameau, et localiser ceux-ci sur photo-plan. C'eût été un travail énorme, fastidieux et impossible à réaliser dans les délais qui nous étaient impartis. D'autre part, nous devions aussi enquêter auprès des exploitants agricoles. Dresser la liste de toutes les exploitations de l'univers d'enquête pour l'utiliser comme base de sondage est évidemment utopique ; c'est pourquoi, étant donné le double objet de l'enquête : démographie et exploitations agricoles, et d'autre part, la difficulté quasi insurmontable de dresser la liste des hameaux ou celle des exploitants agricoles, nous avons établi plusieurs bases de sondage successives et procédé à un sondage à plusieurs degrés.

Dans un premier temps, nous avons dressé la liste des villages administratifs, cette liste servant de base de sondage du premier degré de tirage. On a donc d'abord tiré des villages que nous appellerons unités primaires échantillons. Ces villages tirés étaient rapidement enquêtés afin de déterminer le nombre

---

(1) Voir article de J. WURTZ.

et la taille des hameaux qui les composaient. La liste des hameaux par village tiré servit de base de sondage pour le deuxième degré de tirage ; les hameaux tirés, appelés unités secondaires échantillons, furent les unités échantillons de l'enquête démographique. Enfin, dans un troisième temps, on dressa la liste des exploitations agricoles dans chaque hameau tiré : cette liste servit de base de sondage pour le troisième degré de tirage ; les exploitations agricoles tirées, dites unités tertiaires échantillons, furent les unités échantillons de l'enquête « structures foncières et structures des exploitations ».

Cette méthode nous a ainsi permis de toucher rapidement les hameaux, unités-échantillons de l'enquête démographique, sans être obligés de connaître la totalité des hameaux de la région. Toutefois, on doit reconnaître que trois degrés de sondage introduisent des taux d'extrapolation variables et rendent les calculs plus compliqués.

### 3 — DÉCOUPAGE DE LA ZONE D'ENQUÊTE ET STRATIFICATION

La zone des vallées a été séparée en deux pour les besoins du maître d'œuvre :

- la strate « rive gauche de l'Ikopa », en amont de Tananarive, prioritaire du point de vue des aménagements ;
- la strate dite « rizicole dense ».

La stratification consiste à découper l'univers d'enquête en groupes homogènes : chaque individu appartient à une strate et à une seule ; on procède indépendamment à un tirage aléatoire dans chaque strate. Cette façon de regrouper les individus en groupes homogènes par rapport à un certain nombre de critères permet d'améliorer la précision des estimations. En outre, elle offre pour nous l'intérêt supplémentaire d'essayer de faire coïncider un groupe homogène avec une zone géographiquement homogène, elle aussi.

#### 1. Critères de stratification

Dans chaque strate, nous avons dressé la liste des villages administratifs, et noté, pour chacun, les critères suivants :

- population recensée par l'administration ;
- saison de riz ;
- pourcentage de contribuables payant un impôt sur le revenu.

Ces deux derniers critères ont servi de base à la sous-stratification que nous avons effectuée dans chaque strate.

Il y a deux saisons de riz dans la Plaine de Tananarive :

- le riz de première saison (*vary aloha*) est repiqué en août-septembre pour être récolté en décembre-janvier ;
- le riz de deuxième saison (*vary vakiambiaty*) est repiqué en novembre-décembre pour être récolté en avril-mai.

On ne pratique pas la double culture annuelle sur une même rizière. L'organisation hydraulique de la Plaine est telle que certaines parties sont réservées au vary aloha, d'autres au vary vakiambiaty ;

c'est pourquoi nous avons retenu ce critère qui permet de différencier géographiquement les sous-strates et qui joue un grand rôle dans le déroulement du calendrier agricole et les problèmes de main-d'œuvre.

Nous avons choisi le critère « pourcentage de contribuables payant un impôt sur le revenu » pour mesurer le degré relatif d'urbanisation.

A Madagascar, un individu n'est imposable sur le revenu qu'à partir de 90 000 FMG (1 FMG = 0,02 FF) de revenu annuel déclaré : il est assez exceptionnel que les activités agricoles procurent un revenu aussi élevé ; aussi, les individus payant cet impôt ont ils été déclarés « urbains ». Après dépouillement des rôles, nous avons classé les villages selon le pourcentage de contribuables payant l'impôt sur le revenu par rapport à la population totale de chaque village. Il est apparu une assez nette coupure autour du nombre de 2,5 %, qui a donc été choisi comme seuil de séparation : au-dessus de ce seuil, les villages furent déclarés « urbains », en dessous « ruraux ».

A l'usage, ce critère s'est révélé assez grossier au niveau des villages. En effet, les villages fiscaux ne correspondent pas toujours aux villages administratifs en ce sens qu'ils ne sont pas formés des mêmes hameaux. Entre nos listes de villages administratifs et les villages fiscaux nous eûmes souvent des distorsions importantes. Nous avons repris ce critère au deuxième degré de tirage en procédant à une sous-stratification de la base de sondage des unités secondaires par rapport à ce critère (voir 4.2).

Enfin, nous nous sommes efforcés d'obtenir des sous-strates aussi homogènes que possibles, spatialement parlant.

C'est à ce niveau, ainsi qu'à celui du choix des unités-échantillons que l'on peut tenter d'allier les exigences de l'enquête quantitative proprement dite et les préoccupations de l'enquête qualitative : nécessité, pour chaque unité-échantillon d'être aussi homogène que possible, sociologiquement et spatialement parlant, soucis de la régionalisation dans la détermination des strates et sous-strates ; à cet égard, comme nous le verrons au 3.2, les sous-strates C et D, représentant le centre de la Plaine, sont spatialement homogènes. De même, le hameau constitue une entité sociologique indéniable et sa projection dans l'espace, son terroir, est facilement repérable.

En dehors de ces deux derniers points, tout le reste de l'enquête est soumis aux règles impératives des lois statistiques si l'on veut obtenir des résultats globaux représentatifs de la zone entière. Les unités-échantillons sont déterminées par le mode de tirage choisi et l'on doit absolument s'en tenir aux unités ainsi fixées. Tous les hameaux sortis au tirage *doivent être étudiés* (cf. 2) et il n'est pas question d'en rajouter pour des raisons subjectives (accueil plus sympathique dans le hameau voisin, etc.) ou d'en retirer si l'enquête s'avère extrêmement difficile dans telle ou telle communauté. Le choix est mathématiquement fait, on ne pourra valablement extrapoler et connaître la précision des résultats en fonction du mode de sondage choisi que si l'on s'en tient rigoureusement aux résultats du tirage.

La plus grande rigueur au cours de l'enquête est évidemment nécessaire, notamment au niveau de l'uniformisation des réponses et du souci de prendre contact avec tous les individus ou de visiter toutes les parcelles. Il faut, de plus, réduire au maximum le coefficient personnel des enquêteurs, qui introduit un biais dans toute enquête de ce genre.

Chaque observation devra être réalisée avec le plus grand soin car les erreurs d'observation sont plus graves dans un sondage que lors d'un recensement ; mieux vaut en faire moins, mais de bonnes, que beaucoup de qualité douteuse, même si cela doit augmenter le coût de l'enquête.

Toute estimation établie à partir d'une enquête par sondage est affectée de deux erreurs : une erreur d'échantillonnage, provenant de ce qu'on n'a observé qu'un seul échantillon, et une erreur d'observation.

La résultante de ces deux erreurs est l'erreur totale. Le carré de l'erreur totale est proportionnel à la somme des carrés des erreurs d'échantillonnage et des erreurs d'observation.

Selon les objectifs de l'étude et les moyens dont on dispose (cf. 8), on s'efforcera soit de diminuer l'erreur d'échantillonnage en améliorant la précision du sondage, soit de diminuer l'erreur d'observation en améliorant les méthodes d'enquête.

## 2. Les sous-strates (1)

— LA STRATE RIVE-GAUCHE de l'Ikopa groupe, en 20 villages, 13 000 habitants vivant sur 2 000 ha dont 1 350 utilisables en riziculture, soit 60 % de la superficie de cette strate. Elle a été divisée en deux sous-strates :

— *sous-strate E* : 12 villages dont le pourcentage d'habitants payant l'impôt sur le revenu est supérieur à 2,5 % ;

— *sous-strate F* : 8 villages dont le pourcentage d'habitants payant cet impôt est inférieur à 2,5 %.

Ces deux sous-strates de la rive-gauche appartiennent à une zone où l'on cultive du vary aloha et du vakiambiaty.

— LA STRATE RIZICOLE DENSE regroupe 144 villages, 67 000 habitants vivant sur 24 000 ha, dont 54 % peuvent être utilisés pour la riziculture. Elle a été divisée en quatre sous-strates :

— *sous-strate A* : basse Ikopa, riz de deuxième saison ; les 18 villages sont tous ruraux ;

— *sous-strate B* : vallée de la Sisaony ; riz de deuxième saison ; 27 villages ruraux.

— *sous-strate C* : centre de la plaine, riz de première saison ou riz de première et de deuxième saison ; 87 villages ruraux ;

— *sous-strate D* : centre de la plaine ; 27 villages urbains. Le critère « saison de riz » n'a pas eu à intervenir pour la définition de cette sous-strate.

## 4 — MÉTHODES DE SONDAGE

Comme nous l'avons précisé dans le 2, le sondage comporte trois degrés :

— *Premier degré* : unité primaire (UP) : villages, tirés dans chaque sous-strate avec probabilité proportionnelle à la taille, de façon systématique ;

— *Deuxième degré* : unités secondaires (US) : hameaux, tirés à égales probabilités dans chaque UP échantillon ;

— *Troisième degré* : unités tertiaires (UT) : exploitations agricoles, tirées à égales probabilités dans chaque US échantillon.

La méthode la plus simple était exclue, la taille des UP étant très variable. Elle aurait consisté à tirer des unités primaires avec des chances égales, au taux  $f_1$ , puis des unités secondaires dans les UP échantillons avec un taux de sondage uniforme  $f_2$ . Le taux global, uniforme, aurait été  $f = f_1 \times f_2$ .

(1) Voir article de J. WURTZ : Réflexions à propos d'une enquête quantitative sur la région de Tananarive.

Si nous avons mieux connu le nombre de hameaux par village, nous aurions pu utiliser la méthode suivante :

- 1<sup>er</sup> degré : tirage de  $m$  UP en accordant à chacune d'elle une chance de sortie proportionnelle au nombre d'US qu'elle contient ;
- 2<sup>e</sup> degré : tirage dans chaque UP échantillon d'un nombre  $n_0$  d'US ;
- le nombre d'US échantillons aurait été :  $n = mn_0$  ;
- tous les hameaux auraient eu la même probabilité de sortie :

$$\frac{mn_0}{N}$$

$N$  étant le nombre total de hameaux, donc le même coefficient d'extrapolation.

Le tirage systématique à totaux cumulés permet de tenir compte de la taille des villages sans être obligé de connaître le nombre de hameaux par village. Le deuxième degré de sondage nous assurait de toucher des hameaux de taille inégale.

### 1. Tirage des UP échantillons

Avant le tirage des échantillons, les villages sont classés selon leur taille dans chaque sous-strate. La population de chaque village a été déterminée lors des pré-enquêtes ; il s'agit de la population recensée par l'administration en 1964 ou en 1965.

Dans le but de tenir compte de la taille des villages, on procède au tirage systématique par la méthode des totaux cumulés pour obtenir des UP dans chaque sous-strate (1).

Le nombre d'UP ainsi obtenu est le suivant :

Sous-strates	A	12 villages	2 UP tirées	
—	B	27	3 UP	—
—	C	81	8 UP	— (2)
—	D	27	3 UP	—
—	E	12	3 UP	—
—	F	8	4 UP	— } (3)

Il se révèle que le nombre d'US par UP tirée varie entre 1 et 9. La sous-stratification par rapport au critère « pourcentage d'imposables sur le revenu » réduit encore ce nombre.

Par suite de ce mode de tirage, le taux de sondage des villages n'est pas uniforme, il en sera de même après l'introduction du deuxième degré de sondage. Les coefficients d'extrapolation seront tous différents (voir 5).

(1) Voir 5.1.2, un exemple de tirage à totaux cumulés, dans la sous-strate F. Pour ce qui concerne le tirage systématique à totaux cumulés, on consultera DESABIE ; Théorie et pratique des sondages, Paris, Dunod, 1965, pp. 96 à 102 et 105-106.

(2) Le tirage systématique à totaux cumulés a désigné la même UP à deux reprises ; ce village étant composé d'un seul hameau, celui-ci verra ses résultats doublés.

(3) La rive gauche a été plus fortement sondée que l'autre strate, à la demande du maître d'œuvre.

## 2. Tirage des US échantillons

Dans chaque UP échantillon, on tirera deux hameaux, soit deux US, quand cela est possible. Quand il n'y a qu'un hameau par UP, celui-ci est automatiquement désigné comme US échantillon. Après sous-stratification par rapport au critère « pourcentage d'imposables sur le revenu », on tirera dans chacune des sous-strates ainsi déterminée une US, tirage à égale probabilité, sans remise.

On utilisera pour le deuxième degré de sondage l'estimation par le quotient (voir 5).

Le nombre d'US échantillon ainsi désigné est le suivant :

Sous-strate A :	4 US	Sous-strate D :	5 US
Sous-strate B :	5 US	Sous-strate E :	6 US
Sous-strate C :	10 US	Sous-strate F :	4 US

soit 34 US échantillons retenues pour l'enquête démographique. Le total de la population des 173 UP de l'univers se monte à 80 000 individus (population administrative des années 1964-1965). Le total de la population des US tirées se monte à 7 800 individus soit 9,7 % du total de la base de sondage. La population des US retenues est donnée par l'enquête dans les UP tirées au premier degré de sondage.

## 3. Tirage des UT échantillons

Une fois terminée l'enquête démographique dans les US échantillons, on dresse la liste des exploitants agricoles en les classant par taille croissante selon le nombre d'individus par exploitation ; cela constitue la base de sondage des unités tertiaires.

Nous nous sommes efforcés de tirer les exploitations à égales probabilités en choisissant un taux de sondage tel que le coefficient d'extrapolation soit le même pour toute la zone des vallées. Il y aura donc un taux de sondage des UT par US inversement proportionnel au coefficient d'extrapolation de chaque US dans sa sous-strate, et un coefficient d'extrapolation unique pour toutes les UT échantillons de l'univers.

Ce taux de sondage est égal à :  $\frac{C_i}{50}$  ;

$C_i$  représente le coefficient d'extrapolation global du village  $i$ , c'est-à-dire le produit du coefficient d'extrapolation des hameaux dans le village  $i$  par le coefficient d'extrapolation du village  $i$  dans sa sous-strate (voir 5).

Le coefficient d'extrapolation des UT est donc un nombre constant pour toute la zone, égale à 50. Ce nombre a été déterminé par récurrence en fonction du nombre d'exploitations que l'on se proposait d'étudier, soit environ 250.

## 5 — EXTRAPOLATION-VARIANCE

### 1. Extrapolation

#### 1.1. NOTATIONS (1)

Après avoir rappelé que le coefficient d'extrapolation est l'inverse de la probabilité de tirage, nous allons indiquer comment remonter, à partir des grandeurs observées dans les unités-échantillons, aux grandeurs totales de l'univers.

(1) Nous utiliserons les notations de DESABIE in « Théorie et pratique des sondages », p. 5.

Les 173 unités primaires constituent la population (1) de l'univers.

Cette population se compose de :

1, 2, ...,  $M$ , ... 173 villages

Un village  $\alpha$  se compose de  $N_\alpha$  hameaux :

1, 2, ...,  $\beta$ , ...,  $N_\alpha$

L'échantillon (2), soit les 24 villages tirés, est constitué de  $m$  villages :

1, 2, ...,  $i$ , ...,  $m = 24$ .

Dans un village  $i$ , on tire parmi les  $n_i$  hameaux le composant :

1, 2, ...,  $j$ , ...,  $n_i$ .

Population	Echantillon
$X_{\alpha\beta}$ = nombre de personnes dans le hameau	$x_{ij}$ = nombre de personnes recensées dans le hameau $ij$
$X_\alpha = \sum X_{\alpha\beta}$ = nombre de personnes du village $\alpha$	$x_i = \sum_j Sx_{ij}$ = nombre de personnes du village $i$ auquel appartient le hameau $ij$ .
$X = \sum_\alpha X_\alpha$ = nombre de personnes de la population	$x = \sum_i Sx_i$ = nombre de personnes de l'échantillon.

Les indices  $\alpha, \beta$  permettent d'identifier un individu de la population ; les indices  $i, j, \dots$  sont des numéros de tirage et identifient les individus de l'échantillon, c'est-à-dire ceux désignés par le tirage.

Le signe  $\Sigma$  indique une sommation relative à tout l'univers ; le signe  $S$  indique une sommation relative à l'échantillon.

Les populations initiales seront notées  $X'$ , soit les populations administratives constituant la base de sondage, et  $X''$ , soit les populations après enquête dans les UP échantillons tirés au 1<sup>er</sup> degré. Tous les villages de la zone étudiée sont donc inscrits dans la base de sondage du premier degré de tirage avec leur population initiale  $X'$ . Les villages tirés au premier degré ont été rapidement enquêtés : nombre de hameaux, population. Cette nouvelle population, qui diffère peu de la première, est appelée  $X''$  ; la somme des  $X''_{ij}$  de chaque hameau donne la population du village  $i$ ,  $X''_i$ . Ces populations sont obtenues auprès des autorités administratives villageoises, qui, de ce fait, sont prévenues de l'enquête qui suivra.

Les estimateurs des grandeurs  $X$  seront notés  $x'$ . Les estimateurs  $x'$  sont les résultats des extrapolations des grandeurs observées lors de l'enquête. L'enquête sera d'autant meilleure que  $x'$  se rapprochera de  $X$ . Ce degré de précision sera donné par le calcul de l'écart-type (voir 5). Dans ce qui va suivre, nous prendrons comme exemple de grandeur observée les nombres d'habitants par hameau, village, sous-strate et ensemble.

(1) La population est constituée par l'ensemble des unités (villages, hameaux, exploitations) de l'univers, c'est-à-dire de la zone d'enquête.

(2) L'échantillon se compose des unités-échantillons retenues par le tirage.



## 1.2. ESTIMATIONS

$A_\alpha$  est la probabilité de tirage du village  $\alpha$ .

$A_\alpha = \frac{X'_\alpha}{r}$  c'est-à-dire le quotient de la population initiale du village  $\alpha$  par la raison,  $r$ , du tirage systématique à totaux cumulés.

Cette raison  $r$ , par sous-strate, est égale au quotient de la population initiale totale de la sous-strate par le nombre d'UP tirées dans cette sous-strate.

*Exemple :* pris dans la sous-strate F :

L'objectif était d'atteindre environ 20 % de la population. La population moyenne par village se situant aux alentours de 530 habitants, il fallait tirer deux villages. La raison du tirage est  $\frac{4247}{2} = 2124$  (moitié de la population cumulée). Un nombre pris dans une table des nombres aléatoires donne le premier village, N° 2 210 (celui dont la population cumulée se rapproche le plus du nombre pris au hasard). L'UP échantillon suivante est obtenue en additionnant la raison au nombre pris au hasard, et ainsi de suite.

N° villages	Population initiale	Population cumulée
	$X'$	
2307	245	245
2315	288	533
2313	293	826
2210	303	1 129 (1 <sup>er</sup> village désigné)
2317	323	1 452
2302	571	1 823
2304	419	2 242 (2 <sup>e</sup> village désigné)
1905	2 005	4 247

Les deux UP tirées, N° 2 210 et 2 304, ont pour probabilités de tirage respectives  $\frac{303}{2124}$  et  $\frac{419}{2124}$ . Leur coefficient d'extrapolation  $\frac{1}{A_\alpha}$ , est l'inverse :

$$\frac{1}{A_{2210}} = \frac{2124}{303}.$$

a - L'estimation de la population du hameau  $ij$  est :

$$x'_{ij} = x_{ij} = X_{ij}.$$

C'est la population effectivement observée lors de l'enquête dans le hameau  $ij$ .

b - Estimation de la population du village  $i$ ,  $x'_i$  :

Pour remonter du hameau au village, on utilisera l'estimation par le quotient :

$$x'_i = \frac{X''_i}{\sum_j X''_{ij}} \times \sum_j x_{ij}$$

soit :

$\frac{\text{population initiale du village } i}{\text{population initiale des hameaux tirés dans le village } i} \times \text{population recensée dans les hameaux tirés du village } i$ .

Par une simple règle de trois, on rectifie donc la population du village dans les mêmes proportions que l'on a rectifié les populations initiales des hameaux après l'enquête proprement dite. Exemple :

	Populations initiales	Populations enquêtées
Village 0601	$X' = 350$ (base sondage 1 <sup>er</sup> degré) $X'' = 399$ (après 1 <sup>er</sup> passage dans le village)	
Hameaux tirés dans ce village N° 21	$X''_{21} = 15$ (après 1 <sup>er</sup> passage dans le village)	$x_{21} = 16$
N° 22	$X''_{22} = 85$ (après 1 <sup>er</sup> passage dans le village) $X''_{21} + X''_{22} = 100$	$x_{22} = 95$ $x_{21} + x_{22} = 111$

L'estimation par le quotient du nombre d'habitants du village 0601 donne :

$$x'_{0601} = 399 \times \frac{111}{100} = 442.$$

Le coefficient d'extrapolation des hameaux 21 et 22 dans leur village 0601 est  $\frac{399}{100}$  soit :  $\frac{X''_{0601}}{X''_{21} + X''_{22}}$

Pour chaque village  $i$ , on multipliera par  $\frac{X''_i}{\sum_i X''_{ij}}$  la somme des populations enquêtées dans les hameaux  $ij$  du village  $i$ .

c - Estimation de la population d'une sous-strate :

L'estimateur de la population d'une sous-strate est la somme des estimateurs des populations des villages de cette sous-strate multipliés par leur coefficient d'extrapolation.

$$x'_A = \sum_{i_A} \frac{x'_i}{A_i}$$

Exemple pris dans la sous-strate A :

Village	Hameaux	Population enquêtée $x$	Coefficient d'extrapolation des hameaux	Estimation de la population des villages $x'_i$	Coefficient d'extrapolation des villages	Estimation de la population de la s/strate $x'_A$
0601	21	16	3,99	(16 + 95)3,99	7,62	111 × 30,4
	22	95				
0802	23	105	1,95	(105 + 17)1,95	9,19	122 × 17,9
	24	17				
						5 448

Toutes les opérations, pour toutes les variables statistiques, ont été effectuées mécanographiquement sur IBM (voir 7).

d - Estimation de la population totale :

C'est la somme des estimations de la population de chaque sous-strate, soit :

$$x' = \sum_i \frac{x'_i}{A_i}$$

soit encore la somme des estimations des populations des villages  $i$  multipliées par leur coefficient d'extrapolation  $\frac{1}{A_i}$ ,

ou :

$$x' = \sum_i \left[ \frac{r}{X'_i} \times \frac{X''_i}{\sum_j X''_{ij}} \times \sum_j x_{ij} \right].$$

La méthode de sondage au deuxième degré permet la totalisation des résultats des hameaux (tirage à égales probabilités) dans leur village, mais on ne peut aller au-delà.

Ainsi, on additionnera la population des hameaux étudiés dans chaque village; ces nombres seront successivement multipliés par le coefficient d'extrapolation des hameaux dans leur village (estimation par le quotient) puis par le coefficient d'extrapolation de chaque village dans sa sous-strate. La somme totale des nombres ainsi obtenus donnera l'estimation du nombre d'individus de l'univers. On fera la même opération pour toutes les grandeurs observées.

## 2. Variance (1)

L'estimateur de la variance est donné par la formule :

$$v(x'_h) = \frac{1}{m_h} \frac{1}{m_h - 1} \sum_{i=1}^{m_h-1} \frac{1}{2} \left( \frac{x'_i}{A_i} - \frac{x'_{i+1}}{A_{i+1}} \right)^2 \quad \text{et :} \quad v(x') = \sum_1^k v(x'_h)$$

(1) cf. DESABIE, op. cit., chap. 16 et 17.

où :  $v(x'_h)$  est la variance de la sous-strate  $h$ ,  
 $m_h$  est le nombre d'UP tirées dans la sous-strate  $h$ ,  
 $x'_i$  est la population du village  $i$  avec estimation par le quotient,  
 $A_i$  est la probabilité de tirage du village  $i$ ,  
 $k$  est le nombre de sous-strates.

On utilisera cette formule selon que l'on voudra connaître la variance des grandeurs statistiques qui nous intéressent le plus : nombre d'habitants, de ménages, d'exploitations agricoles, superficies cultivées, etc.

La racine carrée de la variance donne l'écart-type  $\sigma$ , donc le nombre de chances de se trouver dans un intervalle de confiance, on dira : à 95 chances sur 100,  $X = x' \mp 2\sigma$  soit : il y a 95 chances sur 100 pour que la grandeur  $X$  de l'univers soit égale à  $x'$ , l'estimation, plus ou moins le double de l'écart-type. On calculera aussi le coefficient de variation  $\frac{\sigma}{x'}$ .

Ces calculs permettent de caractériser l'efficacité de la méthode de sondage choisie et de donner une estimation des erreurs dues à l'échantillonnage, mais ils ne donnent pas d'estimation sur les erreurs d'observation.

### 3. Cas du troisième degré de sondage

Dans chaque hameau, le taux de sondage des UT est égal à  $\frac{C_i}{50}$  ;  $C_i$  étant le coefficient d'extrapolation global du village  $i$ .

La détermination de ce mode de sondage nous a permis par la suite de faire les calculs d'extrapolation des grandeurs observées au 3<sup>e</sup> degré de façon simple, puisqu'il suffisait de multiplier tous les résultats par 50.

Le nombre d'exploitations dans un village  $i$  est :  $S_j \frac{X''_{ij}}{6}$  ; 6 étant le nombre moyen d'individus par exploitation, estimé a priori.

Dans les hameaux du village  $i$ , nous allons tirer des exploitations avec une probabilité égale à  $\frac{50}{C_i}$  ; le nombre d'exploitations enquêtées sera :

$$S_j \frac{X''_{ij}}{6} \times \frac{C_i}{50}.$$

L'estimateur  $y'$  d'une grandeur sera le produit de la grandeur observée dans l'exploitation  $e$  du hameau  $ij$  par le coefficient d'extrapolation du 3<sup>e</sup> degré, soit  $\frac{50}{C_i}$  et par le coefficient d'extrapolation global  $C_i$ . Le terme  $C_i$  saute donc.

$$y' = S_N [50y_{ije}] ;$$

$N$  = nombre total d'exploitations enquêtées.

L'estimateur  $y'$  est la somme des produits par 50 des valeurs observées dans toutes les exploitations enquêtées. On voit aisément l'intérêt d'avoir choisi ce taux de sondage au 3<sup>e</sup> degré ; les calculs pouvaient être effectués rapidement à n'importe quel niveau : hameau, village, sous-strate, ensemble de l'univers, étant donné les propriétés du produit : commutativité et additivité.

Ainsi, pour près de 10 000 personnes enquêtées en démographie, nous avons eu recours à l'ordinateur, tandis que pour 8 000 parcelles enquêtées, nous avons pu opérer à la main grâce à ce coefficient unique.

#### Variance du 3<sup>e</sup> degré de sondage

La variance du 3<sup>e</sup> degré de sondage est une expression simple (1) :

$$v(x') = \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^m \left( \frac{x'_i}{A_i} - x' \right)^2$$

ou

$$v(x') = \frac{1}{m} \frac{1}{m-1} \left[ S \left( \frac{x'_i}{A_i} \right)^2 - m x'^2 \right]$$

L'analyse des variances donne des résultats satisfaisants, confirmant les méthodes de sondage employées.

#### Exemple :

Sous-strate	Population avec intervalle de confiance à 95 %	Coefficient de variation $\frac{\sigma}{x'}$	Nombre d'exploitations agricoles avec intervalle de confiance à 95 %
A	5 500 ± 600	5,45	1 230 ± 200
B	8 900 ± 500	2,80	1 530 ± 90
C	48 400 ± 1200	1,10	5 200 ± 200
D	15 500 ± 1800	5,80	1 700 ± 130
Rizic. dense	78 400 ± 2000	1,27	9 650 ± 300
E	8 900 ± 400	2,24	950 ± 70
F	5 600 ± 200	1,78	960 ± 40
Rive gauche	14 500 ± 400	1,37	1 900 ± 80
Ensemble	93 000 ± 2200	1,2	11 500 ± 300 $\frac{\sigma}{x'} = 1,3 \%$

Superficies exploitées totales :  $\frac{\sigma}{x'} = 5 \%$ .

(1) DESABIE, *op. cit.*, p. 378.

Dans la strate rive gauche, l'enquête démographique a touché le 1/5 de la population ; dans la strate rizicole dense le 1/15. Pour l'ensemble de la zone des vallées, l'enquête démographique a porté sur le 1/10 de la population, l'enquête foncière et exploitation sur le 1/50 des exploitations de cette zone. Enfin, rappelons que la SCET effectuait une enquête « budget et rendement » sur des exploitations agricoles où nous avons déjà enquêté. Nous leur avons indiqué les taux de sondage nécessaires pour leur échantillonnage.

## 6 — CHOIX DES UNITÉS ÉCHANTILLONS DANS LA ZONE DES TANETY

L'enquête démographique devait toucher le 1/50 de la population. La faiblesse de l'échantillon ne permettant pas une extrapolation valable au niveau de cette zone, nous avons fait un choix raisonné de villages à l'intérieur desquels des hameaux ont été retenus, de telle façon que la somme totale de la population touchée par l'enquête soit voisine de 2 000 individus.

Dans ces hameaux où l'enquête démographique a été exhaustive, le 1/5 des exploitations agricoles fut tiré au sort.

Douze hameaux furent ainsi étudiés, totalisant 312 exploitations : 63 furent touchés par l'enquête foncière, soit le 1/5. Pour toute la zone des tanety, cela correspond environ au 1/300 des exploitations.

### Méthode de choix

Le choix raisonné des villages a été effectué sur les mêmes bases qu'un sondage par quotas.

Dans ce but, nous avons défini dans la zone des tanety quatre grandes régions :

#### 1 — Région nord :

Au nord de l'Ikopa, tous les villages de tanety de la Sous-Préfecture d'Ambohidratrimo à l'exception de ceux du canton d'Ampangabe.

#### 2 — Région sud-ouest :

Au sud de l'Ikopa, délimitée à l'est par la Sisaony, elle comprend tous les villages de tanety de la Sous-Préfecture d'Arivonimamo situés dans la zone d'enquête, ainsi que les villages de tanety des cantons d'Ampangabe, Ampitatafika, Androhibe, Fenoarivo, Ambohijoky.

#### 3 — Région sud-est :

Limitée à l'ouest par la Sisaony, au nord par l'Ikopa et par les villages de la zone des vallées situés dans le canton d'Ambohimangakely.

#### 4 — Région nord-est :

Cette région est comprise entre la Sous-Préfecture d'Ambohidratrimo, la ville de Tananarive et le nord de la région sud-est.

Après étude de la répartition des villages selon leur taille, nous avons retenu trois catégories : villages dont la population est inférieure à 200 habitants ; villages dont la population est comprise entre

200 et 400 habitants et villages de plus de 400 habitants. La population moyenne se situe aux alentours de 350 habitants. Les villages ont été classés selon leur taille et leur position géographique :

	N	S-O	S-E	N-E	Total
0-199	11	28 II	14	8	61
200-399	31 I	42 II	59 II	19 I	151
400 et +	32 I	14 I	14	7	67
Total .....	74	84	87	34	279

Les chiffres romains indiquent le nombre de villages qui ont été choisis à l'intérieur de chaque case ou groupe de cases. Dix villages ont été choisis, douze hameaux retenus parmi ceux-ci, en respectant la taille.

#### 7 — DÉPOUILLEMENT DE L'ENQUÊTE DÉMOGRAPHIQUE (1)

Le dépouillement a été effectué à la main au fur et à mesure de l'avancement des enquêtes sur le terrain.

Il était suffisamment rapide pour que puissent être signalées, le cas échéant, les données manquantes ou erronées fournies par les enquêteurs. Ceux-ci, poursuivant leur travail sur le même terrain pour l'enquête parcellaire, pouvaient alors contrôler ces informations et les rectifier rapidement. De ce fait, l'erreur d'observation sur la démographie a été réduite au minimum.

Nous avons pu faire concorder nos tableaux de dépouillement avec des fiches établies par l'INSRE (2) pour ses propres besoins. Cet ajustement a été possible grâce à une étroite collaboration avec cet organisme et notamment avec F. GENDREAU, démographe ORSTOM en poste à l'INSRE. Cet ajustement avait pour but l'utilisation d'une programmation réalisée précédemment par l'INSRE.

Nous avons dû plier nos propres exigences à celles de cette programmation mais cela nous épargnait de vérifier et d'extrapoler à la main les résultats de l'enquête démographique. Une programmation spéciale aurait largement dépassé nos possibilités financières. Nous n'avons eu qu'à payer la perforation, le triage et le passage sur machine. Nous avons pu nous accommoder des plans d'opération à effectuer, inclus dans la programmation INSRE et avons effectué les autres calculs par nos propres moyens.

(1) Pour le dépouillement de l'enquête structure foncière et structures des exploitations, voir J. WURTZ.

(2) I.N.S.R.E. : Institut National de la Statistique et de la Recherche Economique. Tananarive.

Nous avons neuf fiches de dépouillement :

- 1 — sexe — âge — origine — résidence ;
- 2 — sexe — âge — activité ;
- 3 — sexe — âge — activité — degré d'instruction ;
- 4 — sexe — âge — secteur d'activité ;
- 5 — répartition des ménages selon le nombre d'individus par ménage et l'activité du chef de ménage ;
- 6 — répartition des ménages selon le nombre d'individus par ménage, le sexe et l'âge du chef de ménage ;
- 7 — répartition des actifs selon le sexe, le secteur et le lieu d'activité ;
- 8 — répartition des chefs de ménage selon leur degré d'instruction, leur origine et leur secteur d'activité ;
- 9 — répartition des actifs selon le lieu, le secteur et la permanence de l'activité.

Ces fiches étaient accompagnées d'une fiche maîtresse (voir exemple) présentant le numéro de code de chaque hameau, son facteur d'extrapolation et la taille. Nous avons en effet groupé les hameaux en trois ensembles selon leur taille, indépendamment des sous-strates :

- Taille 1 : population inférieure à 150 habitants ;
- Taille 2 : population comprise entre 150 et 400 habitants ;
- Taille 3 : population égale ou supérieure à 400 habitants.

De même, dans le code hameau figurait, outre la sous-strate, l'appartenance au groupe « urbain » ou « rural », déterminée après une sorte de stratification faite a posteriori lors des totalisations après dépouillements, en nous basant sur les résultats de l'enquête elle-même : tous les hameaux dont plus de 65 % des ménages étaient exploitants agricoles furent déclarés « ruraux », les autres « urbains », ce seuil de 65 % étant significatif au niveau des 34 hameaux enquêtés. Les résultats extrapolés ont pu être totalisés séparément pour ces deux catégories, urbaine et rurale.

Nous avons obtenu, groupés en tableaux, une extrapolation et certaines totalisations pour l'ensemble de l'univers, par sous-strate, par taille de hameaux et par groupe de hameaux ; pour chaque sous-ensemble, un tableau par sexe et un tableau d'ensemble, soit en tout 250 tableaux.

Certaines opérations, notamment les pourcentages, n'ayant pu être intégrées au programme, ont été faites par la suite en même temps que l'interprétation de ces tableaux. Une machine à calculer DIELH nous fut d'un grand secours.

### **Exemples de fiche de dépouillement et de calculs effectués**

EXEMPLE N° 1 : FICHE 1 : pyramide des âges

*Sexe* : M ou F : regroupement en classes quinquennales 01 : 0 à 4 ans ; 02 : 5 à 9 ans ; 11 : 10 à 14 ans ; 12 : 15 à 19 ans, etc.

*Origine* : Hameau : H ; Ailleurs : A ;

*Caractère de la résidence* : P : permanente ; T : temporaire.

La fiche se présente comme suit :



FICHE DE DÉPOUILLEMENT n°  $\frac{1}{1}$ HAMEAU :  $\frac{1}{2} \frac{1}{3} \frac{1}{4}$  (numéro de code)

CLASSES D'ÂGE	SEXE MASCULIN				SEXE FÉMININ			
	Origine		Résidence		Origine		Résidence	
	H	A	P	T	H	A	P	T
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
01								
02								
11								
12 etc.								
	I	II	III	IV	V	VI	VII	VIII

— une fiche perforée par ligne ; la numérotation correspond aux colonnes de la fiche perforée.

Opérations effectuées par la machine :

— Vérifications :  $I+II = III+IV$   
 $V+VI = VII+VIII$

— CALCULS

$I+II = T = III+IV$  ;  $V+VI = T' = VII+VIII$

$I+II+V+VI = T'' = III+IV+VII+VIII$

$I+V = T'''$

$II+VI = T^4$

EXEMPLE N° 2 : FICHE 4 : Sexe — âge — secteur d'activité.

Hameau ; une fiche par sexe ; classes d'âges quinquennales = 15 : 15 à 19 ans ; 20 : 20 à 24 ans ; 25 : 25 à 29 ans ; etc. Total des actifs (individus de 15 ans et plus) ;

AU = activité agricole sur l'exploitation agricole familiale, sans autre activité ;

Activité agricole sur l'exploitation plus activité secondaire, par secteur d'activité :

1 = salariat agricole ; 2 = artisanat (à son compte) ;

3 = commerce (à son compte) ; 4 = employé du secteur privé ;

5 = employé du secteur public ; 6 = sans activité ; militaire, clergé.

Activité non agricole, par secteur d'activité (mêmes numérotations).

Salariat agricole, si celui-ci est l'unique activité.

FICHE DE DÉPOUILLEMENT N° 4  
1HAMEAU :           
2 3 4SEXE :     
5

(0) (I) (II) (III) (IV) (V) (VI) (VII) (VIII) (IX) (X) (XI) (XII) (XIII)

Classe d'âge	TOTAL	Au	ACTIFS AGRICOLES + ACTIVITÉ SECONDAIRE						ACTIFS NON AGRICOLES					SALARIÉS AGRICOLES
			1	2	3	4	5	6	2	3	4	5	6	
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
15														
20														
25														
30														
35														
45														
55														
65														
99														
*														

OPÉRATIONS : extrapolation et vérification :

$$0 = I + II + III + \dots + XIII.$$

La fiche maitresse se présente ainsi :

N° code hameau	Taille	Facteur d'extrapolation
500	1	006
511	2	006
502	2	004
503	3	004
514	3	003
515	3	003
.	.	.
.	.	.
etc.	.	.

*N° de code :*

le premier chiffre indique la sous-strate : 5 = sous-strate F,

le second chiffre indique le groupe de hameau : 0 = rural

1 = urbain,

le troisième chiffre est un numéro d'ordre à l'intérieur de la sous-strate.

*Taille :*

1 = inférieur à 149 habitants ;

2 = de 150 à 399 ;

3 = 400 et plus.

Un hameau était dépouillé manuellement et mis en fiche en quelques jours. L'ensemble a été perforé en une semaine ; le tri a duré plusieurs heures, le passage en machine environ deux heures.

La rapidité de telles opérations nous a permis de figurer à la main certains dépouillements : mouvements migratoires, immigration des chefs de familles et des individus, émigration depuis dix ans.

Avec des moyens financiers plus importants, nous aurions pu, après dépouillement à la main, de toute façon long et fastidieux, effectuer rapidement les calculs, les regroupements et certaines corrélations de l'enquête exploitation.

A l'avenir, il nous paraît fondamental d'utiliser pour les enquêtes lourdes et quantitatives tous les moyens mécanographiques et électroniques de dépouillement et de calcul.

## 8 — CONCLUSIONS

L'enquête statistique interdit toute finesse dans l'analyse des résultats et paraît donc incompatible avec les méthodes géographiques habituelles. Les extrapolations se font au niveau global ou à des niveaux intermédiaires de telle sorte que la notion d'espace sociologique ou géographique n'apparaît pas, ou peu. Cependant, seul ce type d'enquête permet l'approche globale et chiffrée d'une région, ainsi que la mise en évidence de certains phénomènes que l'analyse ponctuelle en profondeur interdit. Il est vain, par exemple, de généraliser les conclusions démographiques ou la taille des exploitations se rapportant aux résultats d'une enquête pratiquée sur des petites unités de l'ordre de 500 habitants. Par contre ces enquêtes monographiques mettent en évidence des mécanismes internes sociologiques ou géographiques qui ne sont pas du ressort des enquêtes par sondage.

L'enquête que nous avons effectuée sur toute une région a permis d'étudier notamment :

— les mouvements migratoires sur une grande échelle : mouvements saisonniers de main-d'œuvre temporaire, mouvements d'immigration définitive dans la Plaine, mouvements d'émigration, localisation de ces mouvements, secteurs d'activités ;

— l'activité de la population de l'ensemble de la Plaine, en relation notamment avec la proximité de Tananarive, d'où mise en évidence de la faiblesse relative de l'activité agricole ;

— la taille des ménages et toutes informations sur les chefs de famille et d'exploitation : degré d'instruction, origine, activité, etc.

De même, l'enquête exploitation a fait apparaître les phénomènes globaux caractéristiques de la Plaine : modes de faire-valoir, superficies exploitées et cultivées par exploitation ; distribution des exploitations selon leur taille, la superficie cultivée en faire-valoir direct ; emploi de main-d'œuvre salariée, équipement, systèmes d'appropriation et de transmission de la terre, et bien d'autres choses.

A ce niveau global de la Plaine, nous avons tenté d'ajouter des niveaux intermédiaires en utilisant au maximum les ressources que permettent stratification, mode de sondage et dépouillement.

Le niveau de la stratification et du choix des unités-échantillons est primordial : c'est là, en effet, que l'on peut essayer avec le maximum d'efficacité de concilier quantitatif et qualitatif : la stratification doit s'efforcer de tenir compte non seulement de critères communs à tout l'univers mais du facteur espace ; c'est ainsi que l'on peut isoler des strates ou sous-strates qui correspondent à des zones géographiquement homogènes. De même, le choix de l'unité-échantillon est déterminant, l'idéal étant d'obtenir des unités-échantillons correspondant à des unités spatialement ou sociologiquement significatives. Mais tout cela suppose deux préalables :

— des moyens financiers suffisants ;

— un temps de pré-enquêtes suffisamment long afin, d'une part, de déterminer avant le sondage lui-même quels sont les meilleurs critères de stratification et les unités-échantillons optimales et, d'autre part, de préparer correctement la base et le mode de sondage, les questionnaires et la stratégie d'intervention des enquêteurs dans les unités-échantillons (1). Il faut donc concilier les facteurs temps, argent et impératifs scientifiques, ce qui est bien difficile dans le cadre d'une convention ou d'une sous-traitance, où ces trois facteurs sont imposés.

Le choix du mode de sondage est également fonction du temps et des moyens dont on dispose (2), mais surtout des documents disponibles avant les pré-enquêtes et de ceux établis lors de celles-ci : pour le choix du hameau, nous avons dû passer par un premier degré de sondage pour obtenir des villages, puis effectuer un deuxième sondage à l'intérieur de ces villages. Des connaissances élémentaires dans le domaine des statistiques et de la théorie des sondages sont absolument nécessaires : l'idéal est évidemment de confier cette partie à un statisticien « disponible », c'est-à-dire capable d'intégrer dans sa propre discipline les objectifs d'autres disciplines moins « scientifiques » comme la sociologie ou la géographie, mais dont le rôle est déterminant quand aux buts à atteindre.

Enfin les possibilités de dépouillements et de calculs offertes par l'informatique moderne ne cessent de se perfectionner et doivent être utilisées au maximum. Par ailleurs, en aménageant le traitement des informations en fonction de notre bonne connaissance des conditions locales, nous avons pu mettre en évidence des niveaux intermédiaires entre la région et le hameau : par exemple, en privilégiant des sous-ensembles selon la taille des hameaux ou d'autres critères, nous avons fait ressortir des phénomènes intéressants sur le plan de la démographie, de l'activité de la population et des mouvements migratoires. De même, par l'utilisation de techniques cartographiques, on a essayé de rompre le cadre trop strict des sous-strates dans l'interprétation des résultats et la mise en évidence d'aires homogènes.

Nous concluons donc en insistant sur la nécessité de procéder à des pré-enquêtes importantes qui permettent par la suite d'intégrer unités statistiques et unités géographiques, et de donner ainsi une dimension supplémentaire aux études quantitatives. L'idéal serait d'obtenir un équilibre entre l'enquête globale et les enquêtes ponctuelles, l'une éclairant les autres et réciproquement. L'enquête ponctuelle de style

---

(1) Afin notamment de diminuer l'erreur d'observation.

(2) Un taux élevé de sondage suppose des moyens financiers importants, pouvant être disproportionnés avec l'amélioration attendue des résultats.

monographique permet, lors des pré-enquêtes, une meilleure détermination à la fois des unités-échantillons optimales et des critères de stratifications ; de plus, elle assure une meilleure adaptation des questionnaires au milieu rural, dans leur forme et dans leur fond. Plus la connaissance préalable de l'univers de l'enquête statistique est approfondie, meilleure sera cette enquête. Les monographies doivent être poursuivies lors du déroulement de l'enquête statistique elle-même car elles permettent de maintenir le contact avec la réalité villageoise et sa cohérence et d'offrir par la suite une meilleure interprétation des résultats d'ensemble. Quelques terroirs bien choisis peuvent ensuite servir de références aux unités soumises aux enquêtes statistiques si l'on a pu dégager des critères communs permettant, à partir d'un nombre limité d'informations, de généraliser sans trop de risque certaines notions essentiellement géographiques. On peut estimer ainsi, par exemple, la compacité des terroirs d'après l'éloignement des parcelles retenues pour enquête dans les exploitations tirées au sort.

## ANNEXE

### DÉFINITION DE TERMES STATISTIQUES

*Individu* : élément sur lequel on fait le dénombrement ou la mesure (hameau, exploitation agricole).

*Echantillon* : portion de population sur laquelle se fait effectivement le dénombrement : l'unité-échantillon de cette enquête démographique a été le hameau.

*Population* ou *univers* : c'est l'ensemble des individus sur lesquels on veut étudier un ou plusieurs caractères, ou variables statistiques (sexe, âge, mode de faire-valoir, etc.).

*Effectif d'un caractère* : nombre d'individus présentant ce caractère.

*Tirage à équiprobabilité* : chaque individu de la population a la même chance de sortie.

*Tirage avec probabilités inégales* : les chances de sortie des différents individus de la population sont inégales mais elles sont toutes connues et différentes de zéro.

*Variance* : c'est une caractéristique de dispersion de la distribution d'un caractère ; elle est égale à la moyenne des carrés des écarts à la moyenne.

*Ecart-type* : c'est la racine carrée de la variance.