PN-AAA-776
51585

Séminaire sur la Methodologie
de Recherche

Séminaire sur la Gestion des
Projets de Developpement

6 Juillet au 25 Juillet, 1987

T E X A S T E C H U N I V E R S I T Y

-=-=-=

S O M M A I R E

PREFACE

INTRODUCTION

CHAPITRE I — Introduction de concepts de base — 11

CHAPITRE II — 19
- Organisation et présentation de données
- Mesures dans l'analyse de données

CHAPITRE III — Introduction à l'inférence statistique et l'analyse — 43
- Estimation
- Test d'hypothèses - Methodes des grands nombres

CHAPITRE IV — Echantillonage — 74
- Echantillonage aléatoire
  - Simple
  - Systématique
  - Stratifié
  - En grappes

CHAPITRE V — Problèmes de contingence — 89
- Indépendance
- Mesures d'association
  Phi, Cramer, Gamma & Spearman

CHAPITRE VI — La recherche par l'enquête : un outil essentiel de génération d'informations nécessaires à la prise de décisions rationnelles — 101
- Objectifs et implications dans le cadre d'une évaluation de programme.
- Vers la quête constante de validité et de fiabilité
- Perspectives de la recherche par l'enquête

P R E F A C E

a) Remarque introductives :

L'Ecole Nationale d'Economie Appliquée est l'une des plus anciennes structures de formation mises en place au Sénégal après l'indépendance. Parallèlement à sa vocation première de former des cadres de hauts niveaux pour les différentes structures gouvernementales, l'Ecole Nationale d'Economie Appliquée, (E.N.E.A) est particulièrement engagée dans de activités complexes de recherches socio-économiques durant ces dernières années. Cependant, malgré l'inestimable temps accordé à ces activités et la présence d'hommes de bonne qualification, les résultats obtenus jusque là bien que très probants ont besoin d'intégrer davantage les potentialités qu'offrent les techniques quantitatives récentes pour une meilleure diffusion.

Aujourd'hui les autorités de l'ENEA, conscientes de la néccesité d'intégrer les outils de la statistique moderne, ont amorcé une phase de réforme globale dans l'enseignement de base de tous les collèges en plaçant les cours de méthodes statistiques parmi les matières fondamentales des programmes de formation.

Le document que nous sommes en train de développer va dans le sens d'une contribution à la modernisation des méthodes de formation de l'ENEA. Les sujets qui sont traités ici constituent une première approche dans la mise en forme de méthodes statistiques de base qui seront désormais intégrées dans tous les programmes de formation des différents collèges de l'ENEA.

Dans l'immédiat notre souci est de familiariser tout d'abord tous les collègues praticiens de l'ENEA dans le cadre d'un séminaire de courte durée sur certains concepts de base fondamentaux utilisés dans la littérature statistique et régulièrement usutés aujourd'hui dans toutes les littératures spécialisées de recherche.

b) <u>Besoin de formation statistique</u> :

Les résultats d'une enquête récente auprès d'enseignants et diplomés de l'ENEA.

Durant la période du 3 janvier au 30 mars 1985, nous avons eu à mener une petite enquête sur les besoins en formation sur les techniques quantitatives de recherche et particulièrement les méthodes statistiques à l'ENEA.

Les cibles concernés par cette enquête étaient de deux groupes. Le premier groupe était constitué par les enseigants de l'ENEA et le second groupe de travailleurs diplômés de l'ENEA. L'enquête était menée dans le souci de nous amener à mieux apprécier la nécessité de mettre en place de cours de méthodologie en général et de méthodes statistiques en

particulier dans les programmes des différents collèges compte tenu des réponses émises par chacun des enquêtés sur les différentes questions que nous leur avions posées. L'enquête aura aussi permis aux différents participants de trouver un cadre anonyme pour s'exprimer librement sur les problèmes de formation telsque conçus jusque-là à l'ENEA. L'échantillon, 38 personnes au total, était composé de 18 enseignants de l'ENEA et 20 travailleurs diplômés de l'ENEA et occupant des postes d'activités diverses à travers le pays.

Le questionnaire contenait deux catégories principales de questions. La première catégorie de questions était liée à la fréquence à laquelle les enquêtés estiment être confrontés à des tâches de collecte de données et le type de problèmes qu'ils déclarent considérer comme étant des obstacles serieux pour mener à terme leurs travaux statistiques. Pour le second aspect de cette première catégorie de questions, on peut citer la création d'échantillon(s) approprié(s), l'exécution d'enquête(s) et divers aspects de l'organisation, le traitement - présentation et l'analyse de données. La deuxième catégorie de questions concernait la necessité ressentie et le souhait emis de voir introduites les méthodes statistiques dans les programmes généraux de formation de l'ENEA. Le questionnaire s'était completé par une rubrique "Observations" dans laquelle chaque enquêté pouvait exprimer librement ses opinions et idées sur la nature, l'orientation et la mise en place de techniques quantitatives d'analyse et principalment des méthodes statistiques dans les orientations de l'école.

Un examen rapide des résultats obtenus montre ce qui suit :

A la question sur la fréquence d'utilisation de données statistiques dans leurs activités professionnelles régulières, tous les enquêtés à l'unanimité ont répondu que de telles situations étaient très fréquentes. 68% des répondants ont précisé qu'ils s'appuyaient beaucoup sur l'utilisation de données statistiques (entendre collection de données de base).

En terme de difficultés liées à l'utilisation de statistique, les procedures d'échantillonnage étaient mentionées dans la moitié de cas, suivaient la création et la mise à jour de base de données (30%) et l'exécution d'enquête (20%).

Lorsque les données statistiques existent, 39% du temps les gens éprouvent des difficultés pour les organiser et 63% du temps ils ne sont pas armés pour les analyser.

A la question sur la necessité de renforcer les méthodologies de recherche à l'ENEA par un accent plus marqué sur l'intégration de méthodes statistiques, 71% de gens ont répondu positivement. La nécessité d'introduire des cours de statistiques dans les programmes généraux de l'école a été soulignée plus de 70% du temps.

La rubrique "Observations" a fourni d'intéressantes réactions de la part de tous les enquêtés. Elle a donné l'opportunité à chacun d'exprimer ses idées sur l'intégration souhaitée de méthodes quantitatives d'analyse sociale à l'ENEA. Du point de vue des enseignants comme des diplômés de l'ENEA,

les méthodes statistiques devraient être partie intégrante de la formation en tant qu'elles constituent une base fondamentale et irremplaçable de toute(s) activité(s) de recherche(s). Cependant, le concensus sur la façon dont de tels cours devraient être intégrés a été difficile à réaliser. Du point de vue de la grande majorité, le cours de statistique devrait être donné à tous les étudiants compte non tenu de l'orientation de leur collèges d'appartenance. Un autre groupe, certes pas nombreux opte pour une spécification de l'enseignement de la statistique à chaque collège. Dans le cas extrême il a été suggéré de réunir tous les étudiants en première année pour un cours introductif sur les statistiques de base et d'attendre l'orientation aux collèges pour introduire une approche spécialisée des méthodes statistiques à chaque curriculum de collèges. Dans tous les cas la chose la plus importante demeure le voeux exprimé par tous de voir les méthodes statistiques enfin incorporées dans tous les niveaux et domaines d'enseignement de l'ENEA. Après tout la décision sur les modalités pratiques de mise en place de tel(s) cours reste du domaine des autorités supérieurs compétentes de l'Ecole.

Les 20 fonctionnaires diplômés de l'ENEA ont tous emis le voeux de pouvoir assister à des séminaires ou cours de recyclages à l'ENEA sur les méthodes de recherche pour mieux s'armer d'outils statistiques nécessaires à leur activités professionnelles.

c) MATERIEL DE FORMATION

Le matériel développé ici est spécialement conçu pour les formateurs de l'ENEA. Avec le support du Centre d'Etudes pour le Développement International (CAIDS) un séminaire d'une durée d'une à deux semaines est proposé à l'attention des cadres chercheurs de l'ENEA comme initiation aux techniques de recherche sociale. Bien que de très courte durée, nous espérons qu'une telle initiative sera particulièrement enrichissante et profitable à tous les participants.

Le matériel proposé ci-après comprend des discussions sur des concepts clés de la statistique appliquée aussi bien que des problèmes et cas d'études appropriés qui sont le reflet de ce qu'une situation de travail peut être lorsque des données statistiques de base nécessitent d'être correctement intégrées pour une meilleure prise de décision.

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

A ce jour les méthodes statistiques sont devenues d'applications courantes dans toutes les disciplines d'enseignement aussi bien que dans divers domaines des activités humaines. Cependant, il existe encore des méconnaissances appréciables sur la nature de la discipline statistique. La conception du profane est très différente de celle du professionnel statisticien. "Un statisticien est perçu comme une personne qui manipule des nombres pour prouver un point..., quelqu'un qui, avec l'aide d'un ordinateur magique peut rendre presque toute étude "scientifique"" (Blalock 1979, p 3). Cette méconnaissance des potentialités remarquables de cette discipline du reste en constante mouvance, est tellement installée dans l'esprit de l'individu "ordinaire" que toute approche de la science statistique ne peut se faire chez lui qu'avec beaucoup d'appréhension à priori.

Il est remarquable de noter cependant que pas une seule discipline ou domaine de travail ne peut se passer aujourd'hui d'intégrer les méthodes statistiques comme élément essentiel de compréhension et d'analyse des phénomènes socio-culturels et / ou économiques. Les méthodes statistiques ont été intégrées à des problèmes allant du business à la médecine, à l'agriculture et au développement. Un examen juste superficiel de la littérature récente des méthodes quantitatives de recherche, permet de situer

l'utilisation des méthodes statistiques en tout premier plan. "L'utilisation croissante de l'ordinateur a augmenté la richesse des données de base nécessaire aux preneurs de décisions; conséquemment, l'analyse intégrée de larges volumes de données s'est averée presque impossible sans l'utilisation judicieuse des méthodes statistiques" (Martre & Gilbreath 1980 p 5).

La place de la statistique dans le processus de recherche a pris une importance considérable étant entendu que les assomptions de base centrales à toute(s) question(s) de recherche requièrent d'être correctement évaluées et verifiées. La statistique, comme l'a du reste indiqué Blalock, est en réalité un outil à la fois flexible et très utile à l'analyse explicative qu'on ne l'imagine à priori (Blalock, 1979). La discussion proposée ici a été profondement simplifiée pour assurer une meilleure compréhension et un plus grand intérêt de la part du lecteur. Le niveau de mathématique requis à sa compréhension est presque nul ou du moins très minimal. De la part du lecteur et participant au séminaire, rien ne saurait remplacer une attitude positivement critique dans la lecture du matériel ainsi qu'une présence participative à toutes les séances qui seront proposées.

CHAPITRE I

REVUE DE CONCEPTS DE BASE

La discussion qui suit introduit certains concepts fondamentaux de base qui seront utilisés tout au long des exposés à suivre. Le chapitre donc est conçu principalement pour aider les participants au seminaire sur les méthodes statistiques à se familiariser dès le début au langage de base qui sous-tend la formulation et la conduite des méthodes statistiques.

A - DEFINITION

Le mot statistique a différentes connotations pour différents groupes de gens (Martre & Gibeath, 1980, p 2). Pour la grande majorité des gens, la statistique n'est rien d'autre qu'un outil de description, de comptage et d'assemblage. L'étude des statistiques en réalité va plus loin que cela. De façon générale, la statistique englobe quatre aspects fondamentaux que sont :

1. la collecte
2. l'organisation
3. l'analyse
4. l'interpretation de données quantitatives.

Alder et Roessler ont défini chacun de ces aspects de la façon suivante:

a) Collection de données : C'est le processus d'obtention d'unités de mesure c'est à dire des informations "individualisées" sur chaque variable d'intérêt. Dans une enquête en milieu rural par exemple, on pourrait compter et enregistrer les revenus annuels et une communauté bien déterminée.

b) Organisation de données : Elle consite à la présentation des données individuelles numériques ou observations sous une forme structurellement mieux organisée et donc plus facile à examiner dans le but de fournir des conclusions générales d'ensemble.

c) Analyse de données : C'est le processus d'extraire à partir de l'information préalablement organisée une information relevante capable d'informer de façon compacte et succincte sur l'entité d'étude.

d) Interprétation de données : Il s'agit de l'étape ultime de formulation d'un ensemble de conclusions à partir de l'analyse faite de résultats obtenus. Généralement, cette phase comprend la formulation de prédictions sur un univers statistique donné à partir d'informations recueillies sur une portion de cet univers d'intérêt (Alder & Roessler 1977).

La statistique en tant que science ne vise donc pas à analyser des informations individualisées mais plutôt un assemblage d'information sélectivement organisée sur un ensemble d'individus. Une enquête démographique par exemple ne devrait pas s'intéresser aux noms d'individus mais plutôt à des caractéristiques, jugées releventes à de futures analyses socioéconomiques telsque l'âge, le sexe, les catégories socio-professionnelles etc... Les groupes statistiques devront être clairement définis; l'on devrait être à mesure de déterminer si un élément donné rentre dans un goupe statistique donné ou non et sans aucune ambiguité (Sahler & Fourastié, 1981).

B - NOTE SUR L'INFERENCE STATITIQUE

Il s'agit d'utiliser une information "limitée" sur une portion de population, un échantillon pour développer des conclusions sur la population de base.

L'inférence statistique sera discutée dans un chapitre spécial à venir.

C - NOTION DE VARIABLE ET D'OBSERVATION

Une variable est une caractéristique mesurable d'une certaine population. Une observation est une mesure individuelle que prend une variable. On l'appelle aussi "Variate" (Aler & Roessler, 1977; Zar, 1974). Si par exemple nous décidions de collecter les revenus annuels paysans, nous définirons

X = (revenus paysans) comme étant la variable d'étude. Ainsi, des scores individuels ou mesures tels que 375.000F, 425.000F seront appelés "variates" ou observations.

NOTES SUR LES TYPES DE VARIABLES ET LES "NIVEAUX" DE MESURES

a) TYPES DE VARIABLES :

- Variable continue : Elle peut prendre n'importe quelle valeur entre certaines limites données. Deux exemples de variables continues sont la taille et le poids d'individus.

- Variable discrète ou "discontinue": Elle ne peut prendre que des valeurs spécifiques dans des limites spécifiées. Des exemples de variables discrètes sont le nombre d'actifs dans un carré, le nombre d'épouses par chef de carré ou de famille. Cependant, il est intéressant de noter que bien que les valeurs possibles d'une variable discrète soient généralement entières, cela n'est toujours pas le cas nécessairement (Zar, 1974, p 3). Un exemple classique au Sénégal serait le nombre de points assignés au fonctionnaire pour les retenus d'impots (1 - 1,5 - 2 - 2,5 - etc..).

b) NIVEAUX DE MESURES:

Nous distinguons quatre niveaux de mesures :

NOMINAL : Fréquemment dans les études sociales nous sommes confrontés à une tâche de classification des observations que nous obtenons.

Dans la classification nous essayons de regrouper les éléments que nous étudions par rapport à une caractéristique donnée d'intérêt mettant ainsi à jour les traits de similarité ou de différence entre les individus concernés. Notre but est donc de mettre en relief à travers ces catégories sélectivement définies les quels éléments sont homogènes et les quels autres présentent des différences (Blalock, 1979, p.15). Les catégories choisies doivent être exclusives et exhaustives. Des exemples de données nominales sont la distribution par sexe (masculin, feminin), les groupes réligieux(chrétiens, catholiques, musulmans et animistes).

- ORDINAL: Très souvent, nous sounaitons classer des catégories d'individus par ordre dit de priorité ou "d'importance". La décision de classer ces categories devrait cependant être faite de sorte que chaque élément puisse être rangé dans une et unique catégorie. Nous pourrions par exemple classer les fonctionnaires sénégalais dans trois catégories distinctes et les ranger comme suit :

Position	Rang
Cadres supérieurs	3
Cadres moyens	2
Cadres d'exécution	1

Les mesures de niveau ordinal ont tous les attributs assignés au niveau nominal auxquels s'ajoute la possibilité de les ranger en catégories définies.

- INTERVAL: Les mesures de niveau Interval ont les attributs des niveaux nominal et ordinal auxquels s'ajoutent les attributs d'addition et de soustraction entre observations.

- RATIO : Le niveau Ratio est le niveau ultime de classification de données. Il permet toutes comparaisons entre et parmi les catégories assignées à savoir l'addition ou la division d'observation. Nous pouvons par exemple comparer les scores d'âges d'individus en additionnant ou en divisant les scores obtenus sur des individus donnés.

c) CONCEPTS DE POPULATION ET D'ECHANTILLON

a) Population

La base de toute analyse statistique est le souci d'émettre des conclusions au sujet d'un groupe de mesures ou d'observations d'une variable d'intérêt. L'ensemble de mesures ou observations sur lesquelles nous tenons à tirer de telles conclusions constitue une population statistique ou univers (Zar 1974).

En d'autres termes une population est la totalité (un ensemble complet) de toutes les observations obtenues sur tous les individus présentant une certaine caractéristique commune, c'est à dire toutes observations possibles sur une ou des variables d'intérêt (Alder & Roessler, 1977; Fourastié, 1980). Comme exemple, l'ensemble des étudiants de 1, 2, 3ème années à l'ENEA constitue la population d'étudiants en formation au cours d'une année donnée, disons 1985.

b) Echantillon

Lorsque la population à étudier est très petite, il peut être plus pratique de collecter l'information sur chacun et tous les individus de cette population. Si nous voulons par exemple étudier et tirer des conclusions sur la distribution d'âges de tous les individus partis sur la lune, il ne serait pas irraisonnable de considérer tous les individus ayant participé à ce programme. Cependant, "en pratique les populations que nous proposons d'étudier sont souvent très larges de telle sorte qu'il peut être très difficile de vouloir considérer tous les individus dans une seule opération de collecte de données. Dans ce cas il devient plus logique de sélectionner un sous groupe d'une telle population, l'étudier et ensuite tirer des conclusions sur l'ensemble c'est à dire la population totale d'origine" (Zar 1974, p 15 - 16). Un tel sous groupe, représentatif de la population d'origine, est appelé échantillon.

L'échantillon est donc un ensemble d'observations sur une portion d'une population. Les caractéristiques ainsi obtenues sur cette portion ou sous groupe d'une population permettent de généraliser sur l'ensemble de la population. En fait il existe plusieurs façons de sélectionner un échantillon. Les procédures d'échantillonnage et l'information générée par un "échantillon" représentatif sont en réalité les conditions clés à toute inférence statistique. Dans une discussion ultérieure nous parlerons de la nature complexe des procédures d'échantillonnage et la nécessité plus qu'importante d'obtenir et garantir de correct(s), valide(s) et représentatif(s) échantillon(s).

CHAPITRE II : STATISTIQUE DESCRIPTIVE

L'ORGANISATION ET LA PRESENTATION DE DONNEES

MESURES DANS L'ANALYSE DE DONNEES

Abstraction

Très souvent en recherche socio-économique nous nous retrouvons avec une multitude de données, difficiles à apprécier convenablement si nous ne nous les présentons pas dans une forme resumée et succinctement informative.

En 1983, par exemple, 30 étudiants de l'ENEA devaient effectuer des enquêtes socioéconomiques dans la région de Yakar. Chaque étudiant avait à sa charge 3 villages équivalents à 200 carrés environ. A la fin chaque étudiant s'est retrouvé avec environ 200 questionnaires à traiter. Un groupe de 3 étudiants avait la tâche finale de regrouper toutes les informations receuillies par leurs collègues, de les organiser par département et de présenter l'information d'ensemble sur chaque département de façon raffinée, compacte et instructive. Il leur était nécessaire donc de déterminer par exemple quel était l'âge moyen des chefs de carrés enquêtés ou le nombre moyen d'enfants par carré visité, etc...

Il apparaît évident qu'avec une si riche base de données, il serait extrêmement difficile par une simple revue des données individuellement obtenues, de se faire une idée du contenu de l'ensemble de l'information ainsi receuillie. L'information, telle que présentée doit donc être traitée et ramenée sous une forme qui permet au chercheur de saisir convenablement et rapidement ce qu'elle contient. (Blalock, 1979)

La statistique descriptive nous permet d'organiser et de présenter de manière précise tout large groupe de données numériques sous une forme plus appropriée à la prise en connaissance rapide de l'information qu'il contient. La discussion qui suit portera d'abord sur certaines formes de tabulation et représentation graphique de données numériques et ensuite présentera certaines mesures sélectionnées permettant de resumer succinctement toutes les données quantitatives nécessaires à la description compréhensive de grandes masses de données.

II A. ORGANISATION ET PRESENTATION DE DONNEES

Les données issues d'enquêtes sont enregistrées en scores bruts. Une fois obtenues, de telles données doivent être classées et ensuite groupées pour une meilleure présentation et un meilleur resumé de l'information qu'elles contiennent.

II A.1. Tabulation : Forme tabulaire de présentation de données

La méthode la plus familière et de loin la plus efficace des méthodes de présentation de données est la distribution de fréquences. "Une distribution de fréquences est un réarrangement des scores bruts obtenus sous forme d'intervalles mutuellement exclusifs. Chaque intervalle considéré est affecté de tous les scores correspondant à cet intervalle précis" (Martre & Gilbreath, 1980, p.13). La forme tabulaire de présentation des données se veut d'éclairer rapidement sur la nature et le contenu de l'information proposée au lecteur ou analyste. Cependant, les règles ou décisions de regroupement de données sont quelque peu arbitraires. "Il n'existe pas de simples règles à la tabulation de données dans la mesure où les décisions d'organisation et de présentation d'une information donnée dépendent des objectifs visés, donc de l'analyste en dernier ressort" (Blalock, 1979 p.4).

La réalité cependant est qu' une distribution de fréquence doit être conçue de telle sorte que l'information centrale, c'est à dire les points ou caractères clés de la distribution puisse apparaître le plus clairement possible. L'idée est donc que n'importe quel lecteur qui jette un bref coup d'oeil à l'information ainsi proposée doit être à mesure de déceler immédiatement la ou les caractéristiques essentielles d'intérêt (Martre & Gilbreath, 1980). Les questions que le chercheur se voit confronter dès le début sont de savoir : Combien et quelle

étendue d'intervalles utiliser ? Dans la mesure où il n'existe aucune règle standard de décider sur une étendue et un nombre à priori déterminés d'intervalles, la décision finale dépendra en général du jugement personnel du chercheur.

Nous avons développé la stratégie qui suit pour permettre aux participants de disposer d'une méthodologie structurée pour la conception et la mise en forme d'une quelconque table de distribution de fréquences.

Etape 1 :

Classer les données par ordre de grandeur croissante ou décroissante. Supposons qu'on ait la distribution suivante 21, 35, 16, 17 et 26, nous pourrions procéder comme suit : 16, 17, 21, 26 et 35.

Etape 2 :

Déterminer les scores extremes; c'est à dire la plus petite et la plus grande valeurs enregistrées. Calculer l'écart entre ces deux extremes;

soit $R = Xh - Xs$

où

Xh = maximum de la distribution
Xs = minimum de la distribution

Dans notre exemple, nous aurions obtenu $R = Xh - Xs = 35 - 16 =$

Etape 3 :

Partager l'écart obtenu R en un nombre égal d'intervalles. Le nombre d'intervalles est ainsi donc une fonction de l'amplitude choisie. Il est à noter là que "l'expérience et la théorie suggerent que pour la plupart des types de données il est désirable d'utiliser entre 10 et 20 intervalles. Avec moins de 10 intervalles, beaucoup de détail sera perdu, tandisque avec plus de 20 intervalles la manipulation des données devient difficile et sans intérêt" (Hoel & Jessen, 1982, p.16).

A ce stade donc nous pouvons choisir n'importe quelle valeur entre 10 et 20 et diviser ce nombre choisi (n), tel que $10 < n < 20$, par l'inverse de l'écart précédemment déterminé, R. Plus simplement il convient de diviser l'écart R par la valeur n choisie entre 10 et 20. La valeur entière la plus proche dérivée de ce rapport constituera notre amplitude ou étendue d'intervalles.

Etape 4 :

Commencer par un intervalle qui contienne la plus petite valeur de la distribution ou minimum et qui soit un multiple simple de l'amplitude choisie.

Dans notre exemple la plus petite classe ou classe inférieure contiendra la valeur ou score 16.

Etape 5 :

Enregistrer et compter chaque score dans l'intervalle qui lui correspond (Blalock, 1979).

* Remarque : La procédure que nous venons de décrire n'est qu'indicative. Nous pourrions décider librement de choisir des amplitudes d'intervalles multiples de 5 ou 10 qui évidemment présentent l'avantage d'être plus maniables pour des calculs futurs. En réalité seul un jugement personnellement justifié suffit pour la détermination de l'étendue comme du contenu de tels tableaux.

RESUME

Les étapes ci-dessous listées sont applicables à la mise en forme de tout tableau de fréquences.

1 - Ordonner les scores obtenus
2 - Déterminer les scores extrêmes et l'écart.
3 - Déterminer une amplitude d'intervalle
4 - Définir et choisir des limites de classes utiles
5 - Enregistrer les scores dans les intervalles correspondants.

Une fois le tableau de distribution de fréquences est déterminé nous pourrions nous intéresser à déterminer la proportion de scores qui correspond à chaque intervalle défini référence à l'ensemble des scores enregistrés. De telles

proportions constituent ce qu'on appelle fréquences relatives. La plupart du temps ces proportions sont sous forme de pourcentages à côté de valeurs absolues correspondantes aux différents intervalles de la distribution.

II A.2. GRAPHIQUES DE DISTRIBUTION DE FREQUENCE

Bien que les graphes n'apportent pas plus d'information à la distribution condidérée que les tableaux eux-mêmes, ils ont cependant l'avantage du faire apparaître plus clairement les caractéristiques particulières de la distribution que les tableaux ne peuvent. Pour être effectivement informateur un tableau se doit d'être simple et faire ressortir les aspects significatifs de la distribution à étudier (Alder & Roessler, 1977).

La discussion qui suit est consacrée exclusivement à des données quantitatives. Le graphe de tel type de données est basé sur un système bidimentionnel de coordonnées. Les limites des classes choisies sont portées en horizontal et les fréquences ou nombre de scores (ou pourcentages) sont portées sur la verticale. Nous faisons une distinction entre histogramme des fréquences et polygone des fréquences.

Un histogramme est un graphe qui présente des données discrètes. L'histogramme comprend une série de barres verticales délimitant les limites de classes et dont la taille correspond au nombre de scores contenus dans chaque intervalle de la distribution considérée.

Un polygone de fréquences est constitué d'une ligne non discontinue pour la représentation de données continues.

* Remarque importante: Bien que l'histogramme est généralement conçu pour des données discrètes et le polygone de fréquence pour des données continues, cette distinction n'est toujours pas retenue dans la pratique. La subjectivité joue un rôle essentielle sur le choix décidé du chercheur de présenter ses données dans la forme qu'il estime la plus adaptée et la plus informée. (Martre & Gilbreath, 1980).

Une fois son graphe de fréquences défini, le chercheur peut vouloir l'ajuster d'une courbe ou, lorsqu'il s'agit de fréquences cumulées, développer une "ligne" qui renseigne rapidement sur des parties ou proportions de l'ensemble des scores. Une telle "ligne" est connue sous l'appellation d'"ogive" ou courbe des fréquences cumulées. Elle représente toutes les fréquences cumulées qu'elles soient relatives ou absolues. Elle peut être d'allure ascendante ou descendance soit que le cumul ait été fait vers le haut ou le bas de la distribution.

Voir l'étude de cas No 1 pour pratique.

Etude de cas No 1

Organisation et présentation des données

Revenus annuels paysans : Situation 1

Instructions aux participant

1 - Lire en entier le cas.

Il est demandé à chaque participant de choisir la ou les forme(s) de présentation convenables pour les données qui lui sont proposées. (Temps : 30 min)

2 - Les participants seront requis de discuter leurs travaux en groupe et de décider après concertation quelles formes de présentation leur semble plus appropriée(s) au cas proposé. (Temps : 45 min)

3 - Un élément sera choisi par groupe pour présenter le travail de son groupe.

Revenus annuels paysans

Le gouvernement du Sénégal est à la recherche de débouchés aux diplomés de l'université de Dakar. Chaque ministère devra étudier toutes les possibilités qui pourraient être exploitées pour trouver du travail au nombreux diplomés de l'université et à ce jour sans emploi. Le Ministère du Développement Rural est particulièrement intéressé de savoir si le milieu rural

sénégalais peut servir de milieu d'investissements profitables pour certains des diplomés d'université. Le Ministère du Développement Rural a choisi l'Ecole Nationale d'Economie Appliquée pour lui fournir dans les meilleurs délais des statistiques de base à partir desquels il pourra faire des propositions suggestives au gouvernement. Vous avez été chargé de collecter et présenter cette information. Vous avez conduit une enquête au niveau de la communauté rurale de "Keur Sidy". Entre autres données receuillies, vous avez la répartion suivante de 200 revenus annuels paysans.

200 REVENUS ANNUELS PAYSANS

(En Milliers de Francs CFA)

412	546	359	477	476	599	409	387
506	576	478	543	420	498	591	514
644	491	574	675	625	533	477	611
376	699	533	366	520	337	525	383
574	537	420	462	479	485	566	511
672	472	332	460	468	464	485	463
508	572	477	501	509	599	364	639
408	411	456	422	366	521	641	464
567	521	506	611	524	618	532	318
503	416	475	460	466	400	462	465
356	500	472	359	424	470	552	426
526	457	566	565	542	697	414	613
422	625	527	422	627	472	596	523
561	529	389	486	480	339	645	437
476	470	473	495	521	491	459	509
425	537	540	560	374	480	479	429
515	326	489	440	565	617	681	606
460	638	453	483	454	492	544	533
586	530	531	597	530	748	350	649
418	529	491	381	578	516	562	456
601	518	488	625	503	651	543	538
505	690	586	546	496	428	618	647
436	376	508	435	712	[illegible]	542	406
550	546	640	502	546	[illegible]	457	539
487	581	521	613	473	500	507	570

QUESTION

1 - Tracez un tableau approprié de la distribution des 200 revenus paysans receuillis.

Le tableau devra en outre inclure une distribution des fréquences absolues (nombre de scores par intervalle), une distribution des fréquences relatives et une distribution des fréquences relatives cumulées.

2 - Tracez l'histogramme des fréquences absolues et la courbe de fréquences cumulées ou "ogive".

3 - Montrez les différentes étapes de votre travail et soyez prêts à justifier vos choix et procédures.

II B. MESURES DE TENDANCE CENTRALE ET DE DISPERSION

Abstraction :

Lors d'une enquête en milieu rural, des étudiants de l'ENEA avaient collecté une large gamme de données allant de l'âge des personnes enquêtées au revenu annuel et la structure des budgets familiaux.

Supposons que vous êtes intéressés par la distribution des revenus et voulez en outre comparer les distributions de revenus sur 5 différents villages ou communautés rurales. Pour votre variable d'intérêt, le revenu annuel par carré, vous disposez de scores bruts.

Cependant vous êtes arrivés à organiser pour chacun de vos 5 villages ou communautés rurales les scores individuels précédemment obtenus sous forme de tableaux synthétiques c'est à dire en distribution de classes de fréquences absolues, relatives, etc... Maintenant, votre souci immédiat est de comparer vos 5 distributions disons en terme de performance à partir des seuls enseignements fournis par les revenus ainsi obtenus.

Vous avez donc besoin de déterminer et de calculer des mesures de fréquences pour chacune de vos 5 entités. Les mesures de référence les plus importantes auxquelles vous allez faire appel sont les mesures de tendance centrale et les mesures de dispersion.

II B.1. MESURES DE TENDANCE CENTRALE

Les mesures de tendance centrale les plus communément utilisées sont la moyenne, la médiane et le mode. Chacune de ces mesures-ci a ses propres propriétés, avantages et limitations.

II B.1.1. MOYENNE ARITHMETIQUE

Un étudiant du collège Animation a été intérrogé par ses parents à la fin de sa première année à l'ENEA "Comment vont les classes petit ?" Il répondit en disant "Assez bien; ma moyenne générale annuelle est 13,5, et je pense que j'aurais pû faire mieux si ma moyenne en comptabilité n'était pas si faible que 9/20". L'étudiant a simplement fourni une brève mais suffisamment indicative information à ses parents sur ses performances d'ensemble à l'issue de sa première année de collège à l'ENEA. La valeur 13,5 indique une mesure d'ensemble de tous les scores obtenus par l'étudiant alors que la valeur 9 indique la performance d'ensemble de l'étudiant en classe de compatabilité. Dans les deux cas, le terme "moyenne" sert de mesure typique sur les performances générales d'ensemble de l'étudiant. L'usage général du terme "moyenne" est une expression novice de la moyenne arithmétique, une mesure statistique de tendance centrale.

Considerons de nouveau notre collection de 200 scores obtenus lors de l'enquête de "Keur Sidy". Supposons que nous desirons maintenant obtenir une mesure typique sur ces 200 observations, c'est à dire le revenu typique d'un paysan du

village Keur Sidy. Nous pouvons considérer soit par une prise en compte de la distribution des données éparses telles que obtenues au début, ou considérer l'ensemble de ces données déjà organisées, regroupées en tableau de distribution de fréquences. Nous distinguons ainsi entre ce qu'on appelle des données non groupées et des données groupées.

* 1er cas : <u>Données non groupées</u>

Les 200 scores tels que obtenus dans la première situation sont appelés des données non groupées. La distribution est faite de <u>scores individuels</u> fournis dans un ordre quelconque. La moyenne arithmétique (U) pour de telles données consiste à prendre la somme de tous les scores (X1) ainsi collectés et de diviser cette somme par le nombre de scores considérés (N).

$$\text{Donc} \quad U = \frac{X1 + X2 + \ldots\ldots\ldots\ldots\ldots\ldots + XN}{N}$$

Une manière compacte de reécrire U est la suivante

$$U = \frac{\sum_{i=1}^{N} Xi}{N}$$ où $\sum_{=1}^{N} Xi$ Représente la somme de tous les scores considérés.

2ème cas : <u>Données groupées</u>

Considérons le cas où nous disposons déjà de la distribution d'intervalles de revenus. Pour obtenir la moyenne arithmétique d'une telle distribution nous avons besoin de modifier légèrement la formule donnée pour les données non regroupées.

Nous définirons d'abord le "milieu" des intervalles considérés c'est à dire $ci = \frac{Xi - Xi - 1}{2}$ où ci représente le "milieu" ou centre de classes.

Ensuite, nous multiplions chaque valeur ci par l'effectif ou fréquence (Ni ou Fi) correspondant à cette classe (i). Enfin, nous divisons la somme obtenue des Ci Fi par le total des scores obtenus, c'est à dire l'effectif total des observations constituant la distribution.

Donc la moyenne pour de telles données groupées s'obtient de la façon suivante :

$$U = \frac{f1c1 + f2c2 + \ldots\ldots\ldots\ldots + fici + \ldots\ldots\ldots\ldots + fmcm}{N = (f1 + f2 + \ldots\ldots\ldots\ldots + fi + \ldots\ldots\ldots\ldots + fm)}$$

ou

$$U = \frac{\sum_{i=1}^{N} fici}{N}$$

II B.1.2. LA MEDIANE : Cas de données non groupées

Supposez que vous avez été à même de ranger vos 200 revenus paysans par ordre croissant (ou décroissant) de valeurs. Une réaction évidente à ce stade est de vouloir localiser le score situé au milieu de cette distribution, c'est à dire le revenu situé à mi-chemin entre le plus petit score et le plus grand score.

La médiane est une mesure de position correspondant à la "valeur-milieu" de la distribution de scores ordonnés par valeurs croissantes (ou décroissantes).

Lorsque le nombre de scores ou observations est impair la médiane sera simplement le score qui partage la distribution en deux parts d'égal nombre de fréquences.

Lorsque le nombre de scores en observations est pair la médiane sera constituée par la moyenne arithmétique des deux scores situés au milieu de la distribution.

II B.1.3. LE MODE : Le cas de données non groupées

Lorsque les revenus dans notre exemple sont classés par ordre de grandeur croissante ou décroissante, il peut arriver qu'un ou plusieurs scores soient cités une ou plusieurs fois. Par exemple un revenu de 340.000F peut être cité 2 fois et un revenu de l'ordre de 450.000F cité 5 fois. Le mode est défini comme étant le revenu le plus cité c'est à dire le plus fréquent dans la distribution totale.

II B.2 MESURES DE DISPERSION

Supposez encore une fois que vous avez rangé vos 200 revenu paysans par ordre de grandeur. Vous souhaitez maintenant examine la variabilité de votre groupe de 200 scores. Vous pouvez par exemple évaluer la différence entre les scores extrêmes à savoir la plus faible et la plus élevé de vos 200 observations. vous pouvez aussi considérer la "distance moyenne" entre vos scores e le score moyen ou revenu moyen de la distribution. Les mesures d cette nature sont connues comme mesures de dispersion ou de variation. Les mesures de dispersion les plus communément utilisées sont l'écart total (ou range), la variance et l'écart-type.

II B.2.1. ECART TOTAL OU "RANGE"

Supposons qu'un étudiant de l'ENEA ait obtenu les notes suivantes en classe de sociologie : 8, 12, 14, 9, 17, 7, 16, 18 et 13. Une façon simple d'examiner ses performances en classe de sociologie serait de considérer ses notes extrêmes et d'en évaluer la différence. Une telle mesure de dispersion est connue sous le nom d'écart-total de distribution de fréquences ou "Range". C'est donc la différence entre le score le plus faible et le score le plus élevé d'une distribution donnée.

De la même manière, lorsque les scores ne sont disponibles qu'à l'intérieur d'un tableau de fréquences, l'écart-total sera obtenu en prenant la différence entre les centres de classes correspondant aux dernier et premier intervalles, de fréquence.

II B.2.2. VARIANCE ET ECART-TYPE

Référons nous encore une fois à la distribution des scores de notre étudiant à savoir 8, 12, ,15, 9, 17, 7, 16, 18 et 13.

Supposons maintenant qu'un autre étudiant de la même classe ait obtenu les notes suivantes : 9, 19, 15, 11, 17, 14, 15 et 12.

Si nous considérons la note moyenne de chaque étudiant nous obtiendrons 12,78 et 14,0 pour respectivement le premier et le second.

Si nous calculons l'écart-total ou "range" pour chaque étudiant nous obtiendrons respectivement 10,0 et 10,0 pour le premier et le second. Donc l'écart (10,0) ne montre pas les dissemblances entre les 2 distributions de notes.

Nous avons donc besoin de faire appel à une autre mesure de dispersion pour faire ressortir les particularités de nos 2 distributions. Nous pouvons donc constituer une mesure de variabilité de chaque distribution en référence au score moyen de cette distribution. Nous définissons, se faisant, une mesure de concentration autour des moyennes de chaque distribution de scores. Nous distinguerons là entre variance et écart-type.

La variance est la moyenne arithmétique de tous les carrés d'écarts à la moyenne. L'écart-type est simplement obtenu en prenant la racine carrée de la variance.

Pour calculer la variance d'un groupe de N scores dont la moyenne est U, nous commençons par déterminer les carrés de différences de chaque score par rapport à la moyenne U, ensuite nous les ajoutons les uns des autres et finalement nous divisons la somme obtenue par le nombre total N de scores dans la distribution. Ainsi la variance σ^2 s'exprime comme suit:

$$\sigma^2 = \frac{(X1-M)^2 + (X2-M)^2 + \ldots\ldots + (Xi-M)^2 + \ldots\ldots + (XN-M)^2}{N}$$

Nous constaterons que la variance prend en compte l'ensemble des scores de la distribution alors que l'écart-total ou range est obtenu seulement par la prise en compte des deux seuls scores extrêmes de la distribution. La variance donc édifie mieux sur la dispersion d'une distribution de fréquences. Cependant, du fait que la variance est exprimée en carré d'unités et l'écart-type dans les unités de la distribution originale, ce dernier est preféré pour exprimer la variation ou dispersion de distribution de fréquences.

<u>Note</u> : Lorsque les données disponibles sont déjà tabulées en distribution de classes de fréquences, la variance sera obtenue en substituant dans la formule précédente les scores Xi par les milieux de classes de scores à savoir fi et procéder ainsi de la même façon que dans le cas précédant. Dans une telle situation l'expression de la variance sera:

$$\sigma^2 = \frac{f1(c1-u)^2 + f2(c2-u)^2 + \ldots\ldots + fi(ci-u)^2 + \ldots\ldots + fM(c2-u)}{N = (f1 + f2 + \ldots\ldots + fi + \ldots\ldots fM)}$$

pour M intervalles de classes

ou simplement $\sigma^2 = \dfrac{\sum_{i=1}^{M} fi\ (ci - u)^2}{N}$

Pratique : voir étude de cas No 2

Etude de cas No 2

Mesures de tendance centrale et de dispersion

Revenus Annuels Paysans : Situation 2

Instructions aux participants

1. Lire en entier le cas decrit ci-dessous. Il est demandé à chaque participant de trouver des mesures indicatives pour resumer l'information contenue dans la distribution proposée (moyenne, mesure, médiane et variance par exemple).

(temps : 45 min.)

2. Les participants doivent discuter leur travail en groupes et effectuer les corrections appropriées à leurs résultats.

(temps : 15 min.)

3. Un membre par groupe sera désigné au hasard pour présenter son travail et résultats aux autres participants.

Enfin les représentants de groupes compareront leurs travaux et répondront aux questions en pleinière.

(temps : 30 min.)

Revenus Annuel Paysans

Le Ministre du Développement Rural avait demandé à l'ENEA de lui fournir une information suggestive sur les possibilités d'insertion de diplomés d'université en milieu rural. (voir le cas en situation No 1). Vous avez collecté 200 questionnaires et êtes à présent déterminés à calculer des mesures adéquates pour resumer succinctement l'information contenue dans vos 200 revenus ainsi obtenus :

200 REVENUS ANNUELS PAYSANS

(1,000 F)

318	414	460	479	507	531	565	613
326	416	460	480	508	532	566	617
332	418	462	480	508	533	566	618
337	420	462	483	509	533	567	618
339	420	463	485	509	533	570	625
350	422	464	485	511	537	572	625
356	422	464	486	512	537	574	625
359	422	465	487	514	538	574	627
359	424	466	488	515	539	576	638
364	425	468	489	516	540	578	639
366	426	470	491	518	542	581	640
366	428	470	491	520	542	586	641
374	429	472	491	521	543	586	644
376	435	472	491	521	543	591	645
376	436	472	495	521	544	596	647
381	437	473	496	521	546	597	649
383	440	473	498	523	546	597	651
387	453	475	500	524	546	599	672
389	453	476	501	525	546	599	675
400	456	476	502	526	550	600	681
406	456	477	503	527	552	601	690
408	457	477	503	529	560	606	697
409	457	477	505	529	561	611	699
411	459	478	506	530	562	611	712
412	460	479	506	530	565	613	748

QUESTIONS :

1 - Calculer toutes mesures de tendance centrale que vous jugez adéquat ment informative. Vous pourrez soit utiliser l'information telle qu présentée ci-dessus, ou vous reférer au tableau que vous avez const à l'issue du cas N° 1.

2 - Calculer toutes mesures de dispersion que vous jugez appropriées.

CHAPITRE III : INTRODUCTION A L'INFERENCE STATISTIQUE ET L'ANALYSE

Abstraction :

En 1975 le gouvernement du Sénégal était convaincu par l'USDA de la profitabilité d'introduire une nouvelle variété de mil dans la région du Sine Saloum. La nouvelle variété de mil avait été déjà testée avec succès dans plusieurs régions des Etats-Unis présentant des conditions climatiques et de culture similaires à celles du Sénégal. Au courant de l'année 1975 donc, 400.000 paysans avaient adopté la nouvelle culture mais en veillant du même coup à cultiver sur des terres adjacentes la variété traditionnelle de mil jusque là connue dans cette région. En 1976 un recensement agricole complet fut conduit à travers toute la région et des informations complètes recueillies sur les deux variétés de mil.

Dix années plus tard, en 1985, le Ministère du Développement Rural exprime son intérêt de vouloir évaluer les performances obtenues par la production de deux variétés de mil. L'Ecole Nationale d'Economie Appliquée a été saisie de la question pour mettre en oeuvre, dans un delai très limité, l'information nécessaire pour permettre aux autorités compétentes du gouvernement de prendre une décision sur la nouvelle variété expérimentée dans le cadre d'une stratégie de politique d'auto-suffisance alimentaire. Malheureusement cependant,

l'ENEA dispose de ressources humaines très limitées. De plus le temps imparti pour la réalisation de l'étude est très court et le support financier presque inexistant. Dans de telles conditions il apparait évidemment impensable de vouloir enquêter tous ceux qui sont impliquées dans la culture de ces deux variétés dans cette région, - de l'ordre de 400.000 paysans en 1975. La seule solution qui s'offre à l'ENEA est de sélectionner un groupe "raisonnablement" représentatif, de l'enquêter et ensuite de faire des "généralisations" sur tous ceux qui sont impliqués dans le programme dans cette région.

Dans le chapitre précedent nous avons discuté quelques techniques de base de la description de données. Lorsque de telles données portent sur une entité entière, une population, nous sommes dans une position de pouvoir tirer des conclusions globales avec le maximum de certitude possible. Dans la situation telle que décrite ci-haut, nous ne sommes plus dans une position de pouvoir disposer d'informations complètes sur chacun et tous les éléments de notre population (manque de temps, manque de support financier, manque de ressources humaines notamment). Nous <u>sélectionnons donc une proportion de la population que nous jugeons représenter</u> l'univers que nous voulons étudier et nous essayons de tirer des <u>conclusions correctes</u> sur l'ensemble des éléments de cet univers.

L'inférence statistique, est le processus d'utiliser une information limitée, un échantillon, en vue de développer des conclusions sur une entité plus large, une population statistiquement et correctement définie (Martre & Gilbreath, 1980 p.230). La question centrale à ce procédé est de savoir si l'échantillon en réalité représente correctement l'ensemble des éléments de notre population d'intérêt. Nous avons donc besoin de déterminer la chance que, par exemple, la moyenne d'échantillon peut être une représentation correcte de la moyenne de la population (Hickey, 1986 p.102). Dans la discussion qui suit nous considérerons deux aspects essentiels de l'inférence statistique que sont le processus d'estimation et, la formulation et conduite de tests statistiques.

III A. NOTES INTRODUCTIVES SUR L'ESTIMATION ET LE TEST D'HYPOTHESE(S)

III A.1 ESTIMATION : "Statistics et Paramètres"

III A.1.1 Mise en forme

Supposons que 100 paysans aient été sélectionnés sur l'ensemble de notre population d'int .ét telle que décrite plus haut. Le goupe de 100 paysans constitue un échantillon, supposé représenter correctement la population entière à étudier. Supposons que l'on ait pû obtenir la moyenne de production à partir de l'échantillon choisi. Nous pouvons donc utiliser une

telle information, notre moyenne d'échantillon, pour prédire la moyenne de production sur l'ensemble de la population de la région.

L'estimation consiste au processus d'utilisation d'informations d'échantillon(s) pour prédire ou estimer <u>les paramètres</u> d'une population donnée. Si les paramètres de la population sont u et σ^2(moyenne et variance), nous pouvons définir la moyenne et la variance de l'échantillon comme étant $\bar{x}$ et s^2. Il sont aussi connus sous le nom de <u>"statistics"</u> ou estimateurs.

<u>Note</u> : Les paramètres sont les <u>valeurs fixes</u>. Ils sont obtenus à partir de toutes les données d'une population. Les "statistics" ou estimateurs $\bar{x}$ et s^2, cependant, sont purement aléatoires. Ils sont obtenus à partir de sous groupe(s) de population aléatoirement choisi(s). $\bar{x}$ et s^2 donc varient d'un échantillon à un autre (Alder & Roessler, 1977, p.137).

III A.1.2.1 <u>STATISTICS</u> : Estimation et Point estimateur

Un estimateur réfère à la méthode d'estimation d'un paramètre d'une population alorsqu'un point estimatewest la valeur spécifique qu'un estimateur prend (Kutner Neter & Wasserman, 1983). Ainsi donc, un revenu moyen d'un échantillon de 100 paysans constitue un point estimateur dans la mesure où elle fournit juste une <u>approximation possible</u> du revenu moyen de l'entité globale.

III A.1.2.2. Moyenne d'échantillon

Le calcul de la moyenne d'un échantillon s'effectue de la même manière que pour la population (se reférer à la discussion sur la moyenne de population II B.1.1). Ainsi pour des données non groupées, la moyenne d'échantillon sera obtenue tout simplement comme suit:

$$\bar{X} = \frac{\sum_{i=1}^{n} xi}{n}$$

III A.1.2.3. Variance et Ecart type d'un échantillon

Le calcul de la variance et de l'écart type d'un échantillon ne se différencie pas beaucoup de celui des paramètres d'une population (voir la discussion en part II B.2.2). La principale et du reste unique modification des formules telles que définies avant, apparait dans le dénominateur. Ainsi la variance et l'écart type de données non groupées d'échantillon sont obtenus comme suit :

$$S^2 = \frac{\sum_{=1}^{n} (xi - \bar{x})^2}{n - 1} \quad \text{et} \quad S = \sqrt{\frac{\sum_{=1}^{n} (xi - \bar{x})^2}{n - 1}}$$

III A.2 TEST D'HYPOTHESE(S)

Supposons que lors d'une enquête exhaustive précédente, disons en 1980, le rendement moyen pour une culture donnée à été trouvée 1500 Kg par acre pour une région donnée du Sénégal. En 1985, à partir d'un échantillon de 100 exploitations nous essayons de réévaluer le rendement moyen pour la même culture dans la même région. Supposons qu'il a été clamé qu'en 1985 le rendement moyen pour la culture donnée n'était pas différent de ce qu'il a été trouvé en 1980 c'est à dire 1500 Kg par acre. Les nouvelles informations que nous obtenons de notre échantillon en 1985, c'est à dire à partir de l'échantillon de taille 100 exploitations, nous permettront de confirmer ou de contredire cette assertion.

Le processus de test donc est la mise à l'epreuve d'idées proconçues sur les paramètres d'une population. L'idée préconçue sur la valeur d'un paramètre d'une population est appelée hypothèse (Martre & Gilbreath, 1980 p.248). Les procedures de test visent donc à comparer un résultat obtenu à partir d'un échantillon ($\bar{x}$ par exemple) à la valeur hypothétiquement assignée à un paramètre de la population (u par exemple). Une telle procedure qui conduit à établir la vraissemblance ou la non vraissemblance d'une hypothèse est connue sous l'appelation test statistique (Alder & Roessler, 1972 ; Martre & Gilbreath, 1980 ; Hichey, 1986).

Dans l'exercice d'une évaluation objective de programme(s) donné(s), par exemple, les hyphothèses à tester se refèrent souvent aux niveaux quantifiés à réaliser ou toutes relations causales prévues d'être réalisées par le seul effet du (ou des) dit(s) programme(s). Sous certaines assomptions, disons, le programme d'immunisation infantile contre un lot de maladies données au Sénégal atteindra le seuil de satisfaction de 75 %. Les dites assomptions de même que le seuil envisagé sont guidés par l'expérience théorique et / ou pratique accumulée pour ce type de programme. Il s'agira, par conséquent, entre autres préoccupations de vérifier l'acceptabilité ou vraissemblance du seuil (de réalisation) envisagé à partir d'informations receuillies auprès d'un nombre réduit d'individus intéressés par le dit programme. Le seuil de réalisation envisagé sous certaines hypothèses est identifié à une hypothèse statistique.

Le processus de test d'hypothèse(s) comprend plusieurs aspects que sont :

1. Une claire formulation de l'hypothèse ou des hypothèses
2. Une formulation de l'hypothèse alternative.

<u>Note</u> : Le concept d'hypothèse nulle sera discuté plus loin).

Par exemple si notre hypothèse (hypothèse nulle) est que le rendement moyen d'une certaine culture est de 1500 Kg par acre, l'hypothèse alternative (ou hypothèse de recherche) pourra être que la <u>vraie</u> valeur du rendement est supérieure, inférieure ou simplement différente de ce qu'il a été déclaré être.

4. La définition d'un niveau ou seuil de signification,

5. Le calcul de l'estimateur d'intérêt ($\bar{x}$ par exemple); la mise en forme du correct relevant test statistique et l'analyse des résultats.

6. Conclusion et prise de décision sur l'acceptation ou le rejet de la valeur du paramètre retenue en hypothèse nulle.

N.B : Des assomptions sur le caractère aléatoire de la sélection de l'échantillon et sur la normalité de la population à étudier.

En conduisant un test statistique le chercheur est, à travers toute la procédure, mû par l'incertitude jusqu'à ce qu'il débouche sur la vraissemblante évidence d'une correcte conclusion. En d'autres termes, le chercheur est constament confronté avec la question suivante : Quelle confiance pourra supporter ma conclusion ?. Cette question est directement en relation au concept de probabilité qui justifie la mise en forme et décision d'un niveau de signification. Le niveau ou seuil de signification, donc, est fondamental à tout processus de test statistique tel que nous nous en apercevons tout au long de notre discussion. Conséquemment le test statistique, pour être possibiement conduit, a besoin de reposer sur des assomptions bien définies dont les plus importantes sont de loin la normalité de la population de reférence et le caractère aléatoire de la sélection d'échantillonage.

III B. LA DISTRIBUTION NORMALE

Abstractiont :

Vous avez conduit une enquête sur les carrés de production de riz dans la partie sud de la région de Casamance. L'information obtenue porte sur plusieurs centaines de carrés. Supposez que vous avez été à même d'organiser toute cette masse de données et d'isoler la production en plusieurs classes bien structurées. Vous avez ensuite construit un histogramme de fréquences (Production) que vous avez après ajusté par un polygone de fréquence.

Imaginez maintenant que vous décidez de reduire vos classes de production en de très petites classes telles que l'histogramme que vous obtenez maintenant n'est qu'une succession de très petits (étroits) rectangles. L'ajustement linéaire de ce nouvel histogramme donne une courbe presque en forme de cloche. Cette courbe approximative de votre histogramme des fréquences est dite une courbe normale.

De façon générale un histogramme d'une distribution de fréquences de moyenne M et écart type σ peut être raisonnablement ajusté lorsque la taille de distribution N est trés large par la courbe d'équation.

$$Y = \frac{NC}{\sigma\sqrt{2\pi}} e^{-(x-M)^2 / 2\sigma^2}$$

où C est la différence entre deux milieux de classes consécutives. Plus C est petit, mieux l'approximation sera en général (Alder & Roessler, 1977, p.123).

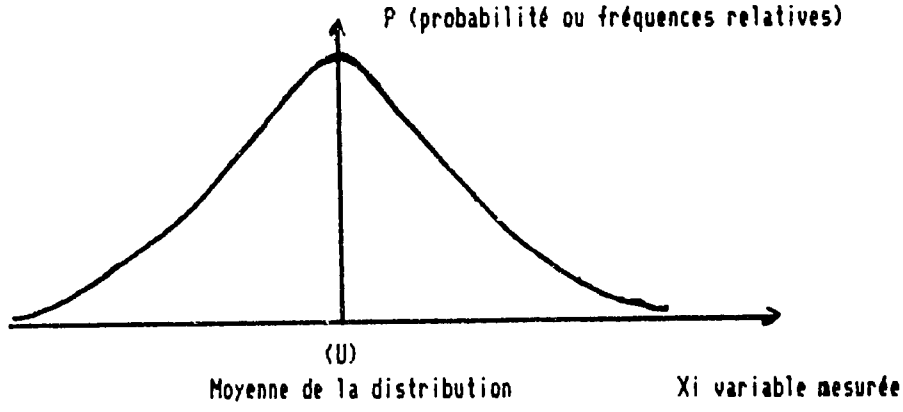
En réalité plusieurs types d'observations sont quasi normalement distribués. La proprieté de normalité est très cruciale à l'estimatation statistique et au test d'hypothese(s)

III. B.1 Caractéristiques de base de la Distribution Normale

a) La distribution normale est symétrique autour de la moyenne de la distribution, U

b) la distribution normale est en forme de cloche.

c) La surface totale sans la courbe normale est égale à 100%. Ainsi toutes les fréquences rélatives sont prises en compte.



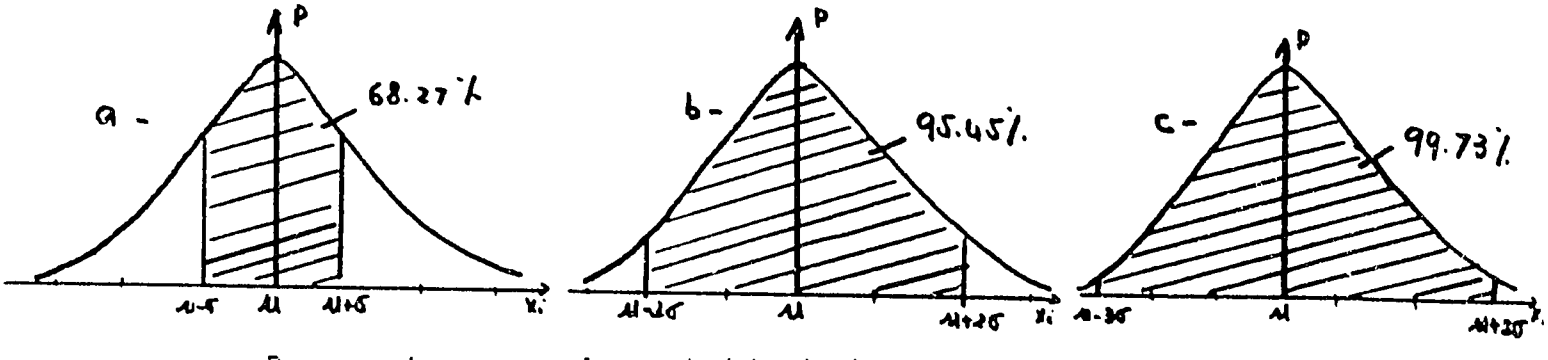
III.B.2 <u>Valeurs particulières (P ou F) sous la courbe normale</u>

Les études et travaux antérieurs ont prouvé que toutes distributions normales peuvent être caractérisées comme suit :

a) Environ 68% des obvservations sont contenues entre 2 écarts types autour de la moyenne c'est à dire $U - \sigma$ et $U + \sigma$.

b) Environ 95% des observations sont entre 4 écarts types autour de la moyenne c'est à dire $U - 2\sigma$ et $U + 2\sigma$.

C) Envrion 99% des observations sont entre 6 écarts types autour de la moyenne c'est à dire $U - 3\sigma$ et $U + 3\sigma$



Dans notre exemple précédent des revenus paysans, il peut être intéressant de déterminer l'intervalle (ou pourcentage) contenant disons 50 % de tous les revenus autour du revenu moyen observé. Heureusement, les statistiens ont développé une procédure simple pour résoudre un tel problème en introduisant ce qu'on appelle la courbe normale standardisée et la table normale.

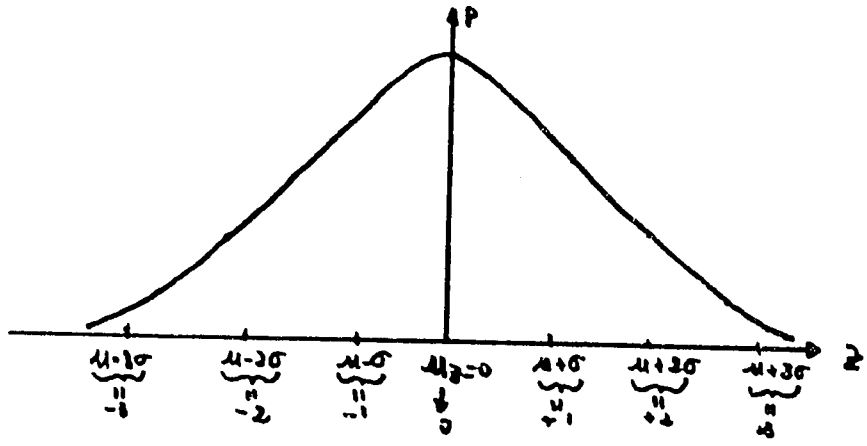
III. B.3 La courbe normale standardisée et la table normale

Supposez que vous avez collecté des informations sur deux populations distinctes de paysans. Imaginez que votre variable d'intérêt est le revenu paysan. Vous pouvez vouloir déterminer les limites dans lesquelles disons 50% ou 90% des revenus sont contenus dans les deux situations. Vous pouvez ainsi vouloir déterminer le pourcentage d'individus qui dans les 2 situations ont leurs revenus contenus entre disons 200.000F et 450.000F. Pour résoudre ces questions nous avons besoin de trouver une distribution de scores commune aux 2 distributions et ainsi opérer à de correctes comparaisons.

La distribution normale standardisée est commune à tous types de distribution normale de fréquences. La variable d'intérêt, connue comme "variable centrée réduite", est donnée comme suit :

$$Z = \frac{X - U}{\sigma}$$

. La courbe normale standardisée définie par cette variable Z, se caractérise par une moyenne $Uz = 0$ et un écart type $\sigma z = 1$

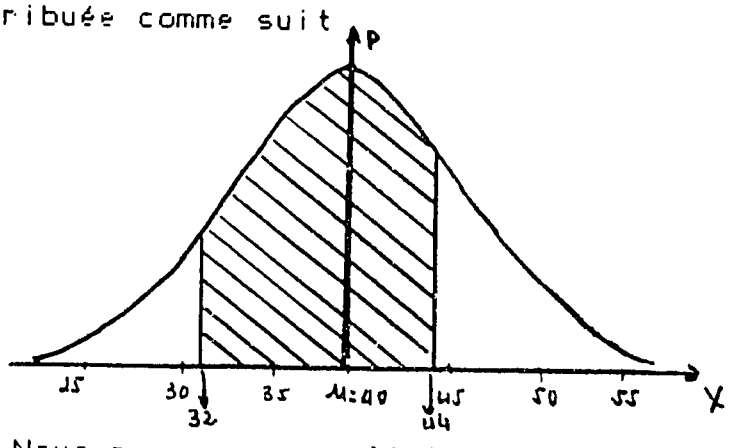


La distribution normale standardisée est tabulée. La table est appelée table normale stantardisée ou simplement table normale. La table normale fournit des probabilités ou pourcentages d'observations contenues entre des scores standardisés Z donnés.

<u>Exemples</u>

Ex 1 : Pour une population large de paysans dont l'âge moyen et l'écart type correspondant sont fournis comme suit :
$U = 40$ et $\sigma = 5$, trouvez le pourcentage d'individus dont les âges varient entre 32 ans et 44 ans.

* si X = âge d'un paysan, notre population sera distribuée comme suit



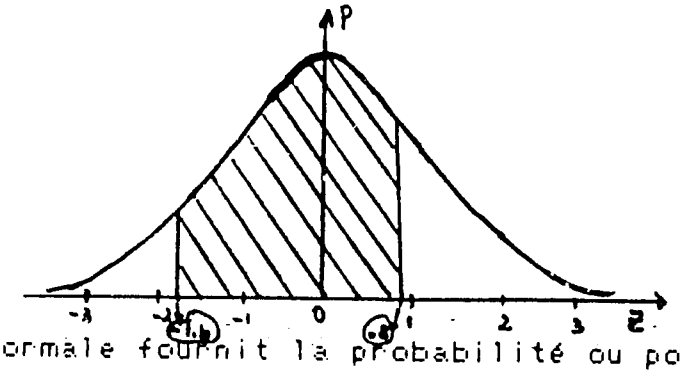
Nous ramenons nos limites d'âge (32 et 44) en unités standardisées, c'est à dire

$$Z1 = \frac{X1 - U}{\sigma} = \frac{32 - 40}{5} = -1.6$$

et

$$Z2 = \frac{X2 - U}{\sigma} = \frac{44 - 40}{5} = .8$$

Ainsi la distribution standardisée se présentera comme suit



La table normale fournit la probabilité ou pourcentage d'individus dont les âges sont contenus entre les limites 32 et 44. C'est à dire : $P(32 < X < 44) = P(-1.6 < Z < .8)$

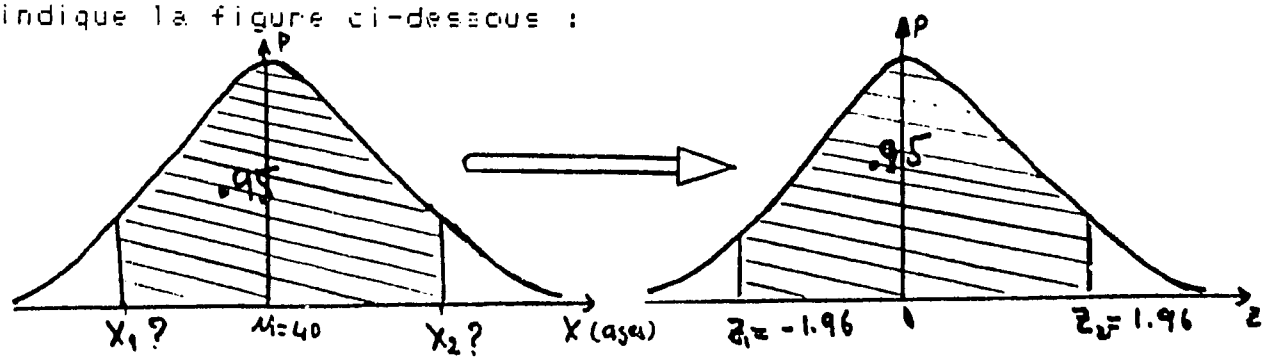
et

$$\begin{aligned} P(-1.6 < Z < .8) &= P(-1.6 < Z < 0) &&= .4452 \\ &+ P(0 < Z < .8) &&= \underline{.2881} \\ & &&= .7333 \end{aligned}$$

Ainsi donc 73.33% des individus enquêtés ont leur âge contenu entre 32 et 44 ans.

<u>Ex.2</u> : Pour une large population normalement distribuée de paysans avec âge moyen U = 40 (et σ = 5) trouvez les deux âges limites (X1 et X2) qui contiennent le milieu des 95% de scores.

* Etant donné que la table normale fournit les probabilités ou pourcentages de scores standardisés entre la moyenne UZ = 0 et les scores Z de part et d'autre de la moyenne, nous avons besoin de partager notre pourcentage (95%) en 2, de part et d'autre de la moyenne UZ. Nous obtenons la valeur .475. La valeur .475 corresponds àun score standardisé Z = + 1.96 comme l'indique la figure ci-dessous :



Prenant $Z = \dfrac{X - U}{\sigma}$ il s'en suit $(Z)(\sigma) = X - U$

d'où $X = U + (Z)(\sigma)$

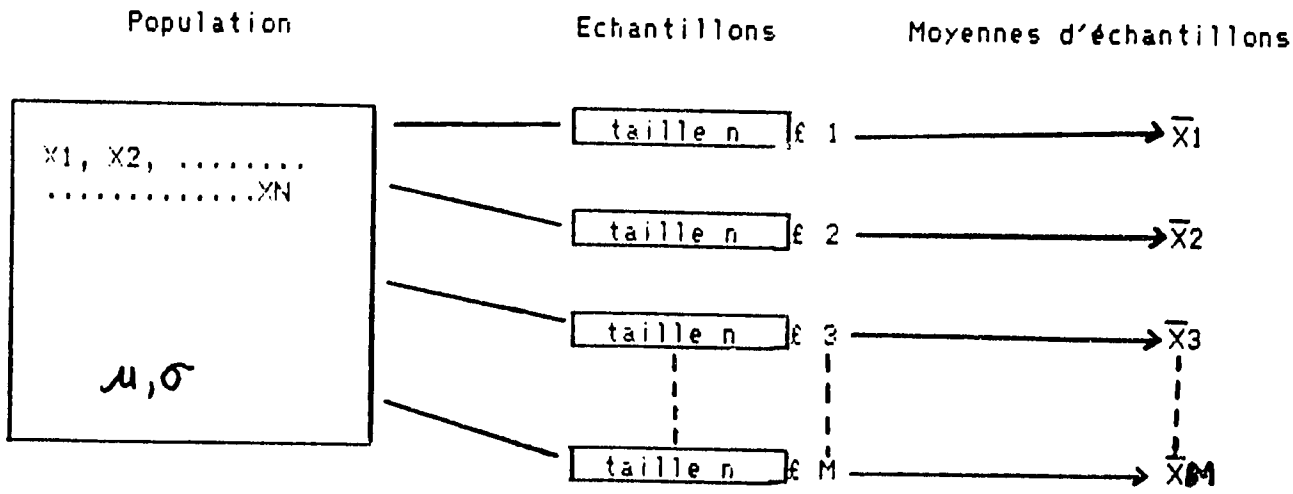
Pour $Z = -1.96$, $X1 = 40 + (-1.96)(5) = 30.2$

et pour $Z = +1.96$, $X2 = 40 + (1.96)(5) = 49.8$

Donc 95% de tous les individus enquêtés ont leur âge compris entre 30.2 et 49.8

III. C. <u>THEORIE DES LARGES ECHANTILLONS</u> : LA DISTRIBUTION DE LA MOYENNE

Considérons une large population de taille N (= 100.000) de moyenne U et écart type σ . Supposez que vous avez pu tirer tous les échantillons possibles de taille n (suffisamment large, disons n>30) comme suit.



Un nombre M d'échantillons a été tiré d'une population de taille N. Chaque échantillon est de taille n, la même pour tous les échantillons. En conséquence chacune des moyennes d'échantillons possibles est supposée être un estimateur de la moyenne de la population, u.

Dans la mesure où l'échantillonnage est aléatoire, les moyennes tirées de tels échantillons constituent des variables aléatoires.

Si nous calculons la moyenne de toutes les moyennes d'échantillons possibles (au nombre M dans notre exemple), nous obtiendrons une valeur égale à la moyenne de la population u, c'est à dire;

$$U_{\bar{X}} = \frac{\sum_{i=1}^{M} \bar{X}_i}{M} = U$$

Ce résultat ne parait pas surprenant dans la mesure où nous nous attendons à ce que la moyenne de toutes les moyennes d'échantillons soit la même que la moyenne de toutes les observations de la même population (Alder & Roessler, 1977).

La moyenne d'échantillon $\bar{X}$ est ainsi dite un <u>estimateur sans biais</u> de la moyenne de la population u.

De la même façon que pour la moyenne, nous pouvons aussi déterminer la variance et l'écart-type de la distribution de moyennes d'échantillons.

Moyennes d'échantillons	Variance de la Moyenne d'échantillon	Ecart-type de la moyenne d'échantillon
$\bar{X}_1$ $\bar{X}_2$ $\bar{X}_3$ - - - $\bar{X}_M$	$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

<u>Note</u> : a) Si la population d'origine des X's est normalement distribuée, les moyennes d'échantillons $\bar{X}$'s seront normalement distribuées.

b) Au fur et à mesure que la taille de l'échantillon accroit, la distribution de moyennes d'échantillon ($\bar{X}$'s) de tailles égales (n) tirées de la même population, tend progressivement vers une distribution normale. Ceci est connu comme étant le "théorème de la limite centrale.

La distribution de $\bar{X}$'s a pour graphe la courbe générale normale de moyenne $U\bar{x} = u$ et écart type $\sigma\bar{x} = \sigma / \sqrt{n}$ (Alder & Roessler, 1977, p. 140 - 142).

Problèmes

<u>Pm1</u> : A partir d'une large population d'éleveurs dont la distribution de têtes de bétail se caractérise par u = 150 et σ = 10. Vous avez sélectionné deux échantillons de tailles respectives :

(i) n1 = 16

(ii) n2 = 36

Trouver pour chaque cas les moyennes et écarts types $U\bar{X}$ et $\sigma\bar{X}$ et dessiner le diagramme approprié. Quels commentaires pouvez-vous faire sur ces distributions ? Justifiez vos commentaires en termes d'écart-types et tailles d'échantillons.

Pm2 : A partir d'une grande population d'éleveurs dont la distribution de têtes de bétail se caractérise par $u = 100$ et $\sigma = 8$, quelle est la chance de sélectionner un échantillon de 25 individus dont la moyenne de têtes est inférieure à 103,2.

III.D. TESTS D'HYPOTHESES - METHODE DES GRANDS ECHANTILLONS

Dans la partie III.A.2 nous avons discuté d'une situation où une investigation antérieure sur une culture donnée, le mil, avait revelé que le rendement moyen dans la région de reférence était de 1500 kg par acre.

En 1985 une enquête sur un échantillon de 100 exploitations a été conduite pour fournir une nouvelle information sur la culture de mil dans la région. Jusque là, il était assumé que le rendement moyen de la culture de mil dans la région n'était pas différent de son niveau de 1980 à savoir 1500 kg par acre.

L'enquête de 1985, bien que portant sur un nombre limité d'exploitations, doit permettre de confirmer ou remettre en cause l'assomption faite jusque là. Le problème qui nous est posé là est de vérifier la vraissemblance (vérité) ou non vraissemblance (fausseté) d'"opinion(s)" avancées et assumés "vraies" sur la valeur d'un paramètre donné (u par exemple).

III.D.1 HYPOTHESE STATISTIQUE

Une hypothèse statistique est un argument ou opinion emis sur un ou des paramètres de population. L'assomption faite sur le paramètre de la population peut être vraie ou fausse, c'est pourquoi on parle d'hypothèse.

Le processus de test d'hypothèse commence par la formulation d'hypothèses, l'une appelée hypothèse nulle (Ho) et l'autre appelée hypothèse alternative ou hypothèse de recherche (Hickery, 1986, p.151).

L'hypothèse nulle, Ho spécifie la valeur du paramètre de la population que nous souhaitons vérifier à partir d'information recueillie sur une partie (échantillon) de cette même population, à un point de temps donné.

Un exemple d'hypothèse statistique peut être l'une des formes suivantes :

(i) le rendement de la culture de coton au Sénégal est de 50 Kg par acre.

(ii) le revenu moyen paysan de la région du Sine Saloum est 250.000F par an

(iii) le revenu moyen paysan au Sine Saloum = le revenu moyen paysan en Casamance, soit 250.000F.

Conséquemment, l'hypothèse alternative ou hypothèse de recherche notée Ha, indique ce que le chercheur doit conclure sur le paramètre lorsque l'information fournie par l'échantillon ne supporte pas les arguments ou opinions précédemment émis sur le paramètre (Martre & Gilbreath, 1980)

L'hypothèse nulle Ho et l'hypothèse alternative Ha, couvrent toutes les valeurs possibles du paramètre à tester.

<u>Exemple de mise en forme de test d'hypothèse</u>

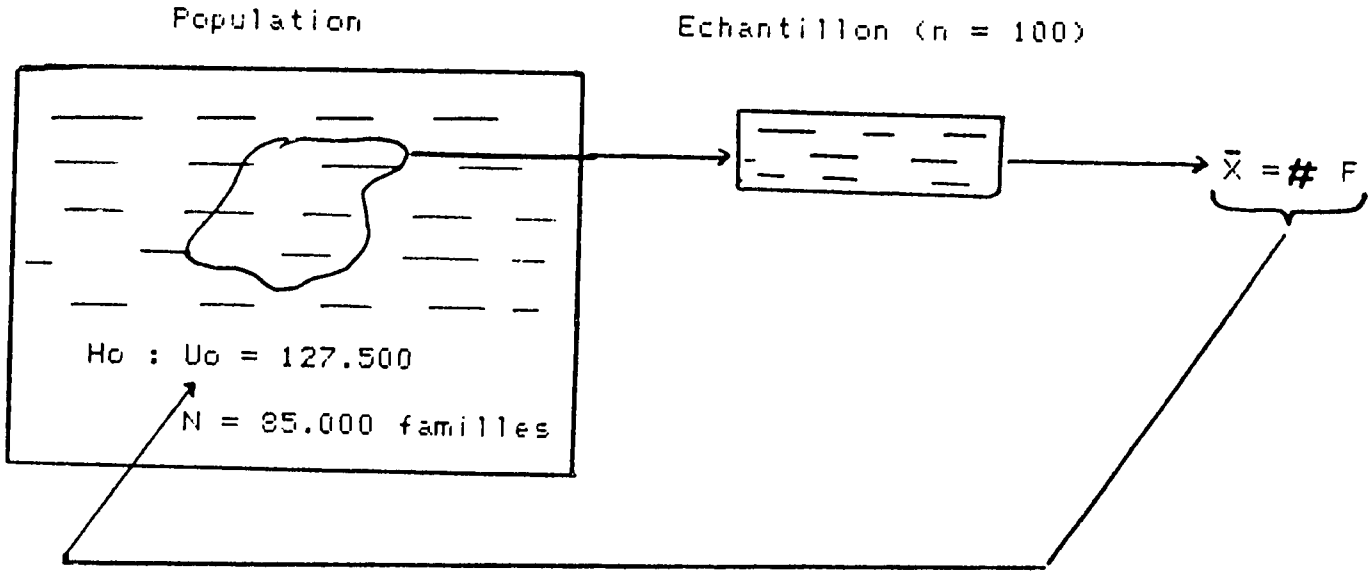
Ho : Le revenu moyen familiale à Dakar = Uo = 127.500F

Ha : U > 127,500

ou U < 127,500

Pour tester l'hypothèse sur U, nous avons besoin de selectionner un échantillon de n familles, et calculer ensuite $\bar{X}$ et le comparer avec la valeur d'hypothèse à savoir

Uo = 127.500F



Une fois la valeur de $\overline{X}$ déterminée, nous la comparons à la valeur d'hypothèse du paramètre. On parle en ce moment de test d'hypothèse.

III.D.2 Erreurs en conclusion et règle de décision

La règle de décision sur la valeur d'hypothèse du paramètre dépend du seuil de risque (ou niveau de signification)

Remarque : 1) S'il arrive que l'hypothèse à tester est en réalité vraie et qu'à partir des données d'échantillon nous concluons sa fausseté (rejet), nous disons qu'une erreur de type I a été commise.

2) S'il arrive que l'hypothèse à tester est en réalité fausse et qu'à partir des données d'échantillon nous concluons son acceptation, nous disons qu'une erreur de type II a été commise (Alder & Roessler, 1977).

En réalité le test d'hypothèse statistique est très conservateur dans la mesure où il est désigné pour minimiser la possibilité d'erreur de type I.

Le seuil de signification α , est désigné comme étant la probabilité de commettre l'erreur de type I. Par conséquent les procédures de test d'hypothèse sont toujours basées sur la sélection de seuils de signification convenables qui minimisent appréciablement le risque d'erreur de type I (Hickey, 1986).

Les assomptions qui sous-tendent le test d'hypothèse sont basées sur les critères fondamentaux de normalité de la population et de l'échantillonage aléatoire.

Le caractère normal de la population garantit la quasi normalité de la distribution des moyennes d'échantillons. Ainsi, compte tenu du niveau de signification α retenu, l'analyste peut déterminer la ou les régions sous la courbe normale qui constituent la ou les zones de rejet de l'hypothèse nulle.

Nous distinguons deux méthodes de détermination de région(s) de rejet :

a) lorsque une partie seulement de la courbe normale est utilisée, on parle de test <u>non symétrique</u>;

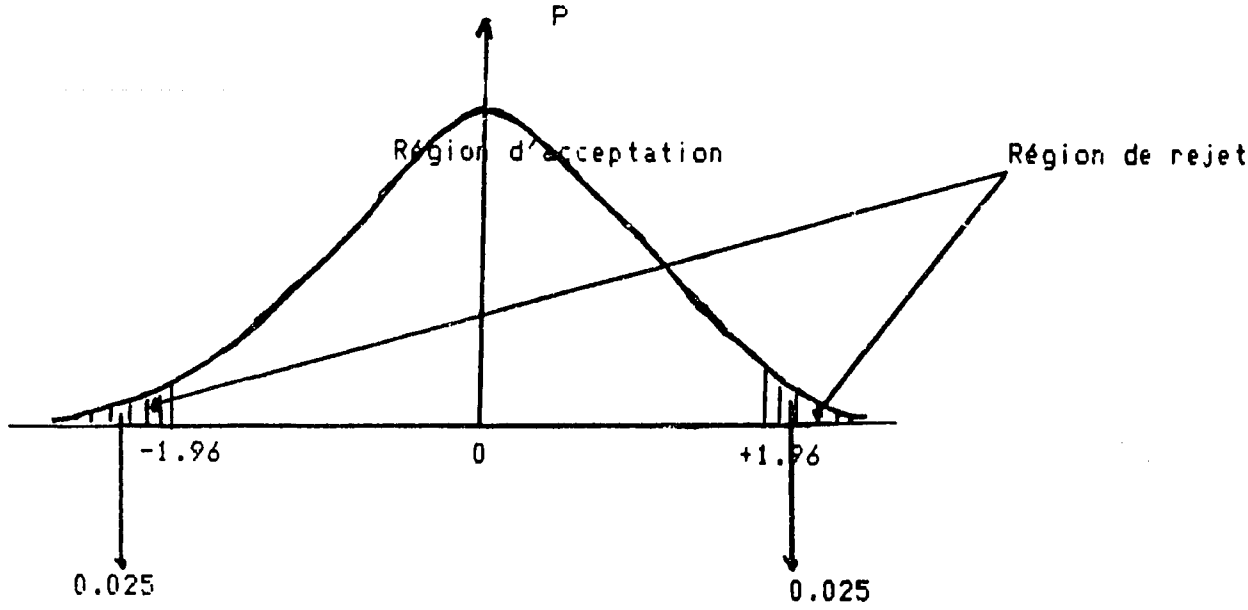
b) lorsque les deux régions de la courbe normale sont utilisées, on parle de test <u>symétrique</u>.

La décision d'utiliser un test symétrique ou non symètrique dépend généralement de la question de recherche posée; il s'agit de définir le sens de l'hypothèse alternative en fonction de l'hypothèse nulle avancée.

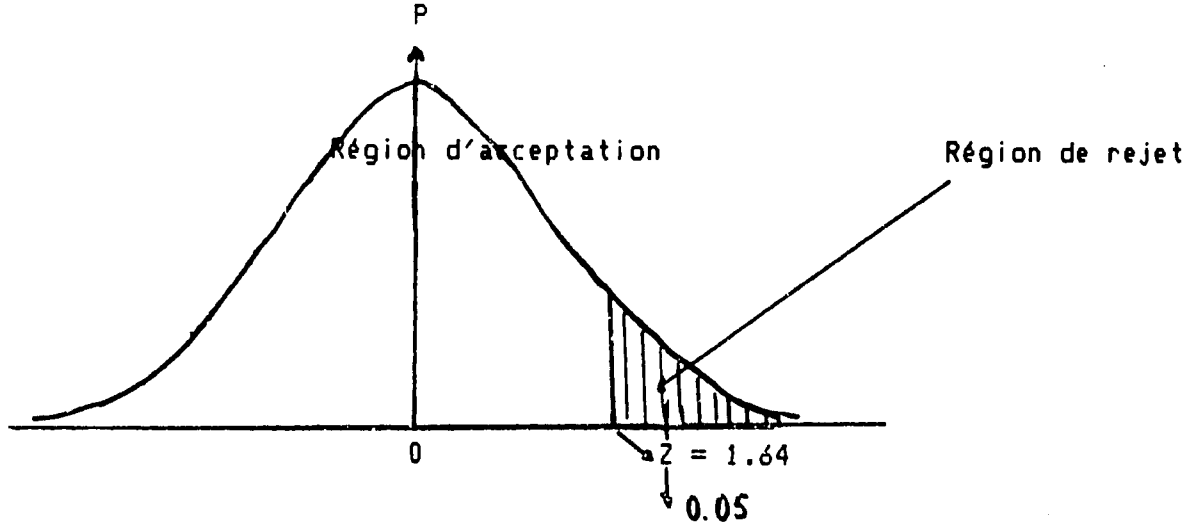
En cas de doute, un test symétrique est toujours recommandé (Alder & Roessler 1977, p.253 - 254).

Supposons qu'on ait retenu le seuil de signification $\alpha = .05$ nous pourrons définir nos régions de rejet comme suit :

* Cas d'un test symétrique



* Cas d'un test non symétrique



Pour déterminer les régions de rejet nous avons fait référence au seuil de risque α . α est en réalité la probabilité de rejeter l'hypothèse nulle. Par conséquent l'hypothèse sera déclarée vraie si la probabilité calculée correspondant à la sélection de la moyenne d'échantillon $\bar{X}$ est supérieure au seuil de signification .

III.D.3. Approche générale et logique du test d'hypothèse

Deux exemples numériques

Exemple No 1 : (Usage d'intervalle ou probabilités)

A l'issue d'une enquête précédente sur la culture de coton dans la région de Tambacounda, nous avions obtenu les résultats suivants :

Rendement moyen $U = 50$ Kg / acre

et $\sigma = 10$

Cependant au cours des cinq dernières années, une agence gouvernementale locale, la SODEFITEX a introduit l'utilisation d'un nouveau type d'engrais dans l'espoir d'accroitre substantiellement les rendements dans la région.

Le mois passé vous avez entrepris une enquête sur un échantillon de taille n = 49 exploitations et qui a généré le résultat $\bar{X}$ = 55.2 comme étant le rendement moyen sur ces exploitations visitées.

Au vu de ce résultat d'échantillon, nous pourrons affirmer que le nouvel "input" (intrant), l'engrais, n'a pas eu l'effet escompté à son introduction au seuil de signification α = .05 (*)

Ho : uo = 50

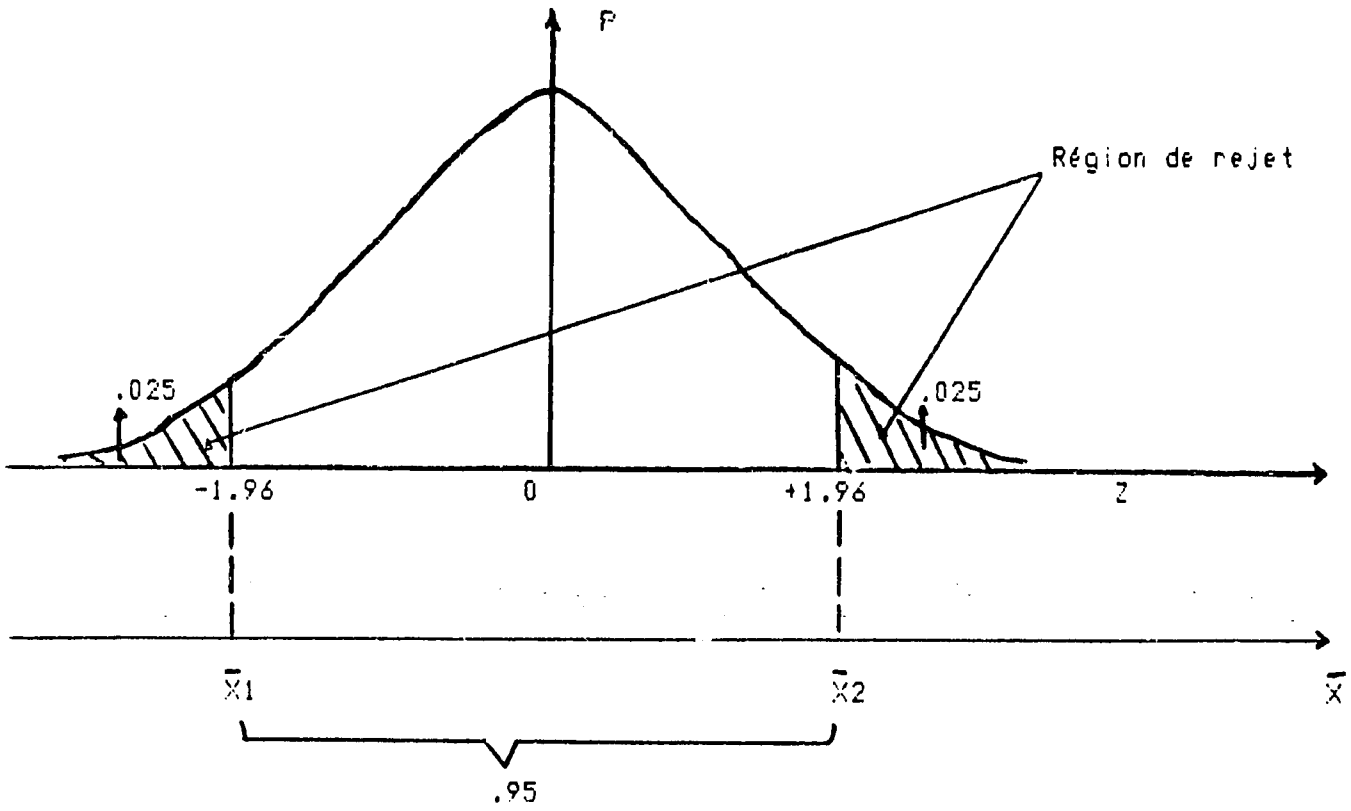
Ha : u > 50 } → test symétrique

ou u <

L'usage d'un test symétrique se justifie par le fait que si l'usage du nouveau type d'engrais n'a pas permis d'accroître les rendements, il peut même produire l'effet contraire c'est à dire précipiter la chute de rendement ne serait-ce que par l'effet d'incompatibilité avec d'autres facteurs présents.

(*) Les seuils ou niveaux de signification communément utilisés en recherche statistique sont .05, .01 et .001 bien qu'il n'y ait aucune restriction au choix d'un niveau quelconque de signification (Blalock, 1979, p.167). En réalité cependant, des niveaux de signification de l'ordre de .05, .01 ou .001 ont prouvé être satisfaisants pour la plupart de tests statistiques. Le plus petit est α, le plus difficile il sera de rejeter l'hypothèse nulle.

Pour α =.05 les régions de rejet seront définies comme suit:



Pour la taille de l'échantillon choisi n = 49, l'intervalle qui contiendrait la moyenne d'échantillon en adéquation avec les paramètres Uo = 50 et σ = 10 serait défini comme suit :

Etant donné $Z = \dfrac{\bar{X} - Uo}{\sigma/\sqrt{n}}$, il vient $\bar{X} = Uo + Z\ (\sigma/\sqrt{n})$

C'est à dire $\bar{X}1 = 50 + (-1.96)\ (10/7) = 47.2$

et $\bar{X}2 = 50 + (+1.96)\ (10/7) = 52.2$

Et dans la mesure où la valeur de la moyenne d'échantillon obtenue est de 55.2, elle serait située à droite de la valeur limite 52.8. Ce qui signifie que la valeur d'échantillon trouvée $\bar{X} = 55.2$ ne supporte pas notre hypothèse nulle. Nous en déduisons donc que

1) la vraie valeur du rendement moyen $\neq 50$
2) le rendement moyen est en réalité supérieur à la valeur 50 c'est à dire $u > 50$

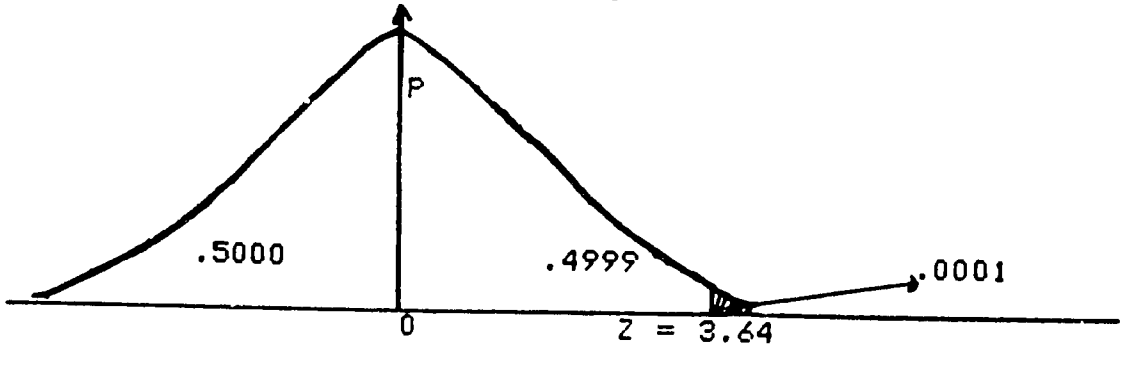
Note : Une autre manière d'examiner le problème est de comparer tout simplement le seuil de signification α à la probabilité correspondant au point estimateur $\bar{X}$ obtenu.

$$P(\bar{X} > 55.2) = P(Z > 3.64)$$

$$Z\ 55.2 \text{ étant } \frac{\bar{X} - U_0}{\sigma/\sqrt{n}} = \frac{55.2 - 50}{10/7} = 3.64$$

La table normale fournit la valeur de $P(Z > 3.64)$ qui est

$$.5000 - .4999 = .0001$$



La valeur .0001, c'est à dire $P = 35.2$ est de très loin inférieure du seuil de signification choisi ; ainsi l'hypothèse nulle est rejetée.

<u>Exemple No 2</u> : (Comparaison de Z - valeurs)

Le rendement moyen d'arachide pour la région du Sine Saloum a été établi en 1980, par suite d'un large recensement à la valeur U = 150 Kg par acre avec un écart type $\sigma = 8$.

A partir de l'année 1981 une nouvelle technique de culture d'arachide a été introduite dans la région avec pour but d'augmenter les rendements. Sur la base d'échantillonage en 1985 nous voulons vérifier si l'introduction de la nouvelle technique de culture a eu l'effet escompté au seuil $\alpha = .05$.

L'échantillon sélectionné est de taille 81 pour $\bar{X} = 147.1$ Kg par acre.

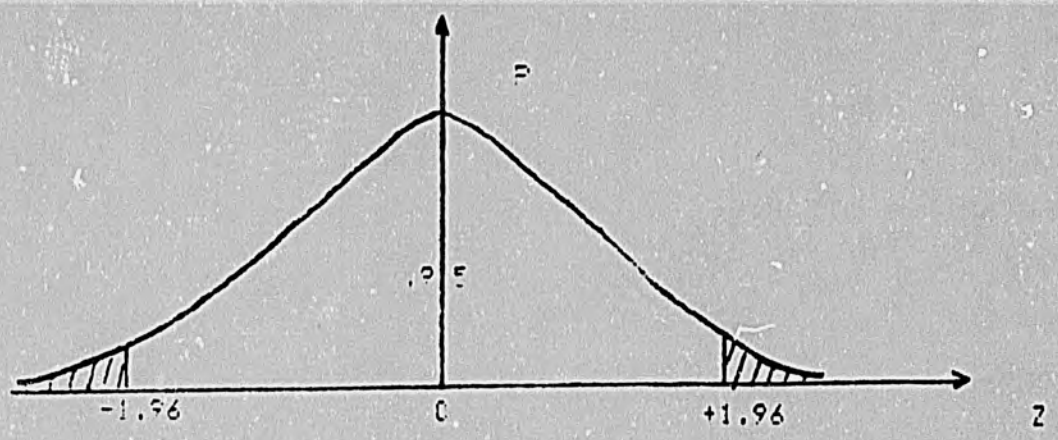
1. <u>Hypothèse nulle Ho</u> : $U_o = 150$
2. <u>Hypothèse alternative Ha</u> : $U > 150$! ! test symétrique $U < 150$!
3. <u>Assomptions</u> :

   a) La population d'exploitations est normalement distribuée;

   b) L'échantillon de 81 exploitations a été choisi de façon aléatoire.

4) Niveau de signification et régions de rejet

Pour α = .05, les valeurs de Z correspondantes dans le test symétrique seront -1.96 et +1.96.



5) Calcul et Analyse

La taille de l'échantillon est n = 81 exploitations. Le rendement moyen est trouvé être 147.1. Ainsi la valeur Z correspondant aux valeurs $\bar{X}$, σ et n est obtenue comme suit :

$$Z = \frac{\bar{X} - U_0}{\sigma/\sqrt{n}} = \frac{147.1 - 150}{8/9} = -3.2625 \quad (*)$$

6) Conclusions :

La valeur Z (-3.2625) "tombe" dans la région de rejet à gauche . Par conséquent l'hypothèse nulle est rejetée.

Donc : 1) Vrai $u \neq 150$

2) Vrai $u < 150$

Problème : La productivité moyenne par travailleur dans l'industrie est assumée être de l'ordre de 35 unités par 8 heures de travail. Pour $\sigma = 6$ et $\alpha = .05$, quelle(s) conclusion(s) peut-on tirer sur la valeur de 35 unités sur la base des résultats suivants obtenus sur un échantillon aléatoire de productivités de 20 travailleurs :

31; 32; 29; 34; 30; 33; 37; 29; 36; 33

34; 32; 27; 28; 31; 33; 34; 28; 38; 32

Note : - Utiliser la méthode des larges échantillons

- Utiliser le test symétrique

CHAPITRE II

ECHANTILLONAGE ET METHODES D'ECHANTILLONAGE ALEATOIRE

Abstrait :

Le ministère sénégalais du logement voudrait étudier la possibilité ou faisabilité de la mise en place d'un large programme d'habitat publique. Pour le faire, les spécialistes chargés du programme ont nécessairement besoin d'évaluer les besoins (tailles, nombre de pièces, etc...), les préférences (maisons, appartement, etc...) et les possibilités financières des futurs acquereurs. Ainsi deux solutions leur sont offertes :

- entreprendre un recensement complet de tous les ménages à travers le pays, ou
- prélever un nombre reduit de ménages par exemple 1 sur 2000.

Dans un pays de plusieurs milliers de ménages, il serait très difficile d'entreprendre une enquête exhaustive parceque exigeant d'énormes moyens techniques, financiers et humains. Cependant, en sélectionnant une portion représentative des ménages sénégalais, des économies très grandes de temps et de ressources pourront être faites et aussi déboucher à une information suffisamment riche pour permettre une généralisation statistique correcte sur l'ensemble des ménages sénégalais.

La discussion qui suit est exclusivement orientée sur les méthodes d'échantillonage aléatoire. Toute la discussion sur les inférences statistiques repose sur la garantie d'échantillonage aléatoire (ou probabiliste). En fait l'échantillonge aléatoire et la distribution d'échantillonage aléatoire représentent "la dernière marche d'échelle qui mène aux procédures inférentielles" (Mantre & Gilbreath, 1980; Blalock, 1979).

IV. A. <u>AVANTAGE DE L'ECHANTILLONAGE</u>

Les enquêtes d'échantillon sont très avantageuses aussi bien en terme de coûts, de rapidité que de précision. En réalité lorsqu'aucune difficulté particulière ne s'y oppose, les enquêtes d'échantillon peuvent être menées en très peu de temps et générer des résultats concrèts en l'espace de quelques jours. Les enquêtes de marché, tout comme le sondage d'opinions, sont par exemple exécutées en très peu de temps et produisent généralement de très bons résultats. compte tenu de la petite taille d'entités à étudier, nous sommes en position de pouvoir mesurer simultanement différents caractères d'une population avec beaucoup plus de précision que lors d'un recensement sur les mêmes objectifs. De plus, "l'expérience en matière de recherche a prouvé que des études basées sur la select on d'échantillons représentatifs généralement produisent des informations plus exactes que s'il s'agissait de recensements dû en particulier à l'apparition d'erreurs presque inévitables qui apparaissent lors d'enquêtes exhaustives" (Hoel & Jessen, 1982 p.166).

Les erreurs liées aux procédures exhaustives peuvent être nombreuses et souvent difficiles à évaluer : duplication, erreurs d'appréciation ou d'enregistrement, etc...

IV. B. METHODES D'ECHANTILLONAGE ALEATOIRE

L'échantillonage aléatoire correspond à toutes les méthodes de sélection d'échantillons pour lesquelles chaque élément de la population à une probablité connue (différente de zero) d'être retenu. Les variables observées par suite de ce processus, sont dites variables aléatoires. Le critère de sélection aléatoire refère à l'assignation d'individus (ou unités d'analyse telles que groupes, villes, pays) d'une population à des groupes expérimentaux ou groupes de comparaison de façon telle que pour chaque assignation à un tel ou tel autre groupe, chaque élément ou individu de la population mère a la même probabilité non nulle d'être sélectionné "(Nachmias, 1979 p. 27 - 28). La sélection aléatoire ou l'échantillonage aléatoire, est en fait la propriété - clé qui nous autorise à procéder à des généralisations sur la population de reférence.

Il existe quatre types "primaires" d'échantillonage aléatoires que sont l'échantillonage aléatoire simple, l'échantillonage sytématique, l'échantillonge stratifié et l'échantillonage par grappes.

IV. B.1 L'échantillonage aléatoire simple

Lors d'une discussion précédente nous avions mis l'accent sur la nécessité de garantir à chaque élément de la population une chance égale d'être sélectionné dans l'échantillon. Ainsi, "lorsque à chaque tirage les éléments restants ont la même chance d'être sélectionnés au tirage suivant compte non tenu des éléments tirés précédemment, nous parlerons d'échantillonage aléatoire simple" (Blalock, 1979, p.554). Les mécanismes du tirage aléatoire simple peuvent être de diverses formes. Cependant, nous nous limiterons ici à l'usage de tables de nombres au hasard. Avant d'aller plus loin sur l'usage de telles tables nous mentionnerons un certain nombre de précautions de base à prendre pour garantir les bases d'une bonne sélection. Les précautions peuvent être résumées en trois points :

a) être sûr que tous les éléments de la population sont listés une fois et une seule fois,

b) associer un nombre à chaque individu ou unité de la population,

c) s'assurer que la liste est exhaustive et correcte. La liste doit être nécessairement mise à jour, purifiée et présentée de nature à garantir une chance égale à chacun et tous les éléments qui la constituent.

Une fois ces précautions prises, nous tournons vers la procédure de tirage.

Les tables de nombres au hasard ont été établies sous la base de génération de nombres aléatoires. Ceux des tables les mieux connues sont :

1. Table of 105.000 Random Decimal Digits, Interstate Commerce Commission, Washington, 1949

2. A Million Random Digits, The Rand Corporation, Free Press, 1955. (Martre & Gilbreath, 1980, P.175 - 177)

Pour utiliser une table de nombres au hasard nous pouvons aussi bien opérer en lignes qu'en colonnes et commencer par nimporte quel point de la table.

Ainsi pour obtenir un échantillon aléatoire simple de taille n, nous pourrons procéder comme suit :

1. Assigner à chaque élément de la population un nombre de 1 à N.

2. Sélectionner au hasard un point de départ sur la table.

3. Enregistrer les nombres retenus au fur et à mesure qu'ils sont lus de la table en éliminant toutes valeurs lues supérieures à N

4. Lorsque l'échantillonage est effectué sans remise , éliminer les nombres répétés. (Martre & Gilbreath,1980)

Supposons par exemple que nous voulions tirer un échantillon de taille $n = 400$ à partir d'une population de taille $N = 10,000$ électeurs.

Nous affectons à chaque électeur un nombre compris entre 0001 et 10,000. Ensuite pour obtenir un échantillon de 400 électeurs il et simplement nécessaire de sélectionner 400 nombres de 4 chiffres à partir de la table des nombres au hasard (Hoel & Jessen, 1982 p.168).

<u>Remarque</u> : Lorsque l'échantillonge est sans remise l'assomption de non indépendance est quelque peu violée. Cependant, il est intéressant de noter qu'une telle situation n'est pas aussi sérieuse qu'on peut à priori penser dans la mesure où l'échantillon sélectionné est juste une petite fraction de la population à étudier et que donc la chance qu'un élément quelconque de cette population d'être sélectionné deux ou plusieurs fois est plutôt très faible (Blalock, 1979, p.557)

Lorsque le problème devient sérieux cependant nous aurons besoin d'introduire un facteur ou coefficient de correction aux estimateurs obtenus.

Ainsi l'espérance mathématique ($E(\bar{X})$) et la variance ($\sigma^2_{\bar{x}}$) de la moyenne de l'échantillon ($\bar{X}$) lors de tirage sans remise , seront obtenues comme suit :

$$E(\bar{X}) = U_{\bar{X}} = E(X) = U$$

et $$\sigma^2_{\bar{x}} = \frac{\sigma^2_x}{n} \quad \frac{N-n}{N-1}$$

où $\frac{N-n}{N-1}$ est appelé facteur de correction, N étant la taille de la population et n la taille de l'échantillon choisi.

IV. B. 2 L'échantillonage systématique :

L'échantillonage ou tirage systématique est de loin plus simple que l'échantillonage aléatoire simple lors de situations où nous disposons de liste très longue ou lorsque nous souhaitons sélectionner un échantillon de taille très large.

La procédure de tirage est très simple. Il n'est pas nécessaire d'utiliser une table de nombres au hasard. Il s'agit plutôt de parcourir la liste en sélectionnant chaque Jème élément une fois un point de départ choisi sur la liste.

Si par exemple nous voulions constituer un échantillon de taille n = 100 à partir d'une liste de 10.000 éléments, nous pourrions sélectionner chaque dixième élément à partir d'un point de départ déterminé au hasard sur la liste. Si notre point de départ est le 5ème élément sur la liste, notre échantillon sera constitué des éléments suivants : 5, 15, 25, 35, 45, 55 ...

Il s'agit donc dans cet exemple de simplement partir de la position 5 et de sélectionner tous les dixièmes éléments suivants.

Remarque : Lorsque les éléments de la liste de reférence sont classés suivants des critères déterminés telsque ethnicité, affiliation réligieuse, le tirage systématique peut introduire de sérieux biais et doit par conséquent être évité.

IV. B.3 L'échantillonage stratifié

L'échantillonage stratifié est très utilisé en recherche sociale particulièrement lorsqu'il s'agit d'étudier des comportements ou performances de groupes sociaux. Dans des situations telles que celles-là, la stratification apporte plus d'exactitude par une meilleure représentation de catégories qui autrement seraient simplement sous représentées.

Dans un dessin (ou plan) d'échantillonage stratifié, les individus ou éléments sont assignés en catégories ou groupes en fonction de caractéristiques spécifiques telles que réligion ou ethnicité par exemple. Ensuite des échantillons indépendants sont tirés de chaque catégorie ou groupe et l'ensemble de ces échantillons-groupes constitue l'échantillon définitif à retenir pour la population à étudier.

Dans chaque groupe ou catégorie, le tirage peut être effectué par une sélection simple ou par sondage systématique. Les fractions d'échantillonage de chaque groupe peuvent être de

tailles égales ou de tailles inégales. Nous parlons respectivement d'échantillonage stratifié proportionnel ou d'échantillonage stratifié disproportionnel (Blalock, 1979; Martre & Gilbreath, 1980).

IV.B.3.1 Echantillonage stratifié proportionnel

Supposons que par exemple, nous voulions étudier les attitudes des populations réligieuses casamançaises vis à vis des fêtes de mariage. L'affiliation réligieuse des personnes enquêtées était constituée comme suit: 50.000 musulmans, 20.000 catholiques, 15.000 protestants et 5.000 animistes. Si nous tirions un échantillon de taille n = 1.000 individus nous ne pourriont assurer une correcte représentation de tous les groupes réligieux. La proportion d'animistes par exemple pourrait être relativement exagérée ou très faible. Donc pour prévenir un tel biais probable, nous pourrons établir notre échantillon de 1000 individus en sélectionnant tout simplement 1/100 de chaque groupe réligieux. Ainsi, nous obtiendrons 600 musulmans, 200 catholiques, 150 protestants et 50 animistes.

Remarque : Les catégories ou groupes doivent être connus et établis à priori, bien avant le tirage. De plus une stratification de cette nature-là ne peut être possible que pour des variables pour lesquelles l'information catégorisée est disponible sur la liste à partir de laquelle le tirage sera effectué (Blalock, 1979).

IV.B.3.2 Echantillonage stratifié disproportionnel

Pour obtenir un échantillon stratifié disproportionnel nous aurons simplement besoin des fractions différentes d'échantillonage pour chaque groupe ou strate de la population de référence. Supposons que nous veuillons comparer les degrés de religiosité (pratiques religieuse) des personnes appartenant aux quatres groupes réligieux précedemment définis. Pour garantir des comparaisons sensées nous pourrons sélectionner un nombre égale d'individus pour chacune des catégories réligieuses soit 1000 par exemple. Ainsi, nous obtiendrons les fractions suivantes : 1/60, 1/20, 1/15 et 1/5. Le choix de fractions différentes permet d'améliorer la représentativité de l'échantillon pour le problème de recherche en question.

Remarque : Comme nous le savons, l'objectif ultime d'un échantillonage aléatoire est de générer des points estimateurs "corrects" des paramètres de la population tels que la moyenne u et l'écart-type σ . Lors d'échantillonage stratifié la formule d'estimation de la moyenne de la population, par exemple, doit être établie en prennant en compte chacune des moyennes générées par les différentes strates retenues. Il s'agit de :

1. Calculer chaque moyenne de strate,
2. De pondérer chacune de ces moyennes par le coéfficient de tirage retenu pour la strate correspondante et,

3. De calculer la somme des moyennes de strates pondérées. Ainsi donc, la moyenne générale $\bar{X}$ de l'échantillon sera définie comme suit :

$$\bar{X} = \sum_{i=1}^{K} wi \ (\bar{X}i)$$

et conséquemment, la variance de la moyenne d'échantillon $\bar{X}$ sera obtenue comme étant

$$S^2_{\bar{X}} = \sum_{i=1}^{K} wi^2 \ S^2_{\bar{X}i}$$

où wi = pondération pour la i ème strate

et $\bar{X}i$ = moyenne de la i ème strate

IV.B.4 L'ECHANTILLONAGE PAR GRAPPES

Dans un échantillonage par grappes nous ne nous intéressons pas directement aux individus ou éléments composant la population de référence.

L'échantillonage par grappes tend à résoudre les disparités régionales attachées aux unités d'intérêt (Borey & Thiombiane, 1981). "Dans un tirage de premier degré, chaque grappe est complètement recensée. On parle en générale dans ce cas d'échantillonage à un seul degré dans la mesure où la sélection est faite une seule fois dans le processus de tirage" (Blalock, 1979; Martre & Gilbreath, 1980).

Dans un tirage à plusieurs degrés, la sélection se passe à plusieurs niveaux. Supposons par exemple que nous voulions étudier la fertilité de la femme au Sénégal.

Nous pourrions dans un premier lieu sélectionner un certain nombre de villes à travers le pays. Une seconde sélection consisterait à sélectionner des quartiers au niveau de chaque ville retenue. Et en fin, une dernière sélection consisterait à prendre par exemple chaque i ème habitation de chaque quartier et ensuite recenser toutes les femmes vivant dans ces habitations.

L'échantillonage par grappes est très recommandé dans des situations telles que :

a) il n'existe aucune liste complète des éléments à étudier et lorsque l'établissement d'une telle liste peut s'averer très cher et long à faire.

b) il existe une répartition géographique très dispersée des éléments à étudier.

Malgré sa simplicité et tous les avantages qui peuvent lui être inhérents, l'échantillonage par grappes peut, au bout du compte, entrainer l'analyste à des difficultés sérieuses lorsqu'il s'agit d'obtenir les points estimateurs d'intérêts de certains paramètres de sa population. De telles difficultés sont à la fois inhérentes au degrée d'homogeneité à l'intérieur des grappes que de l'échantillon dans son ensemble. Ainsi, l'on ne saurait être suffisamment prudent lorsqu'il s'agit d'analyser ou d'interpreter les résultats obtenusd'échantillonage par grappes.

Etude de cas No 3

Dessin d'échantillonage aléatoire

Ministère du Développement Social

Instructions aux participants

1 - Lire l'étude de cas en entier.

Les participants sont demandés de travailler individuellement sur chacune des questions posées

(temps : 15 min)

2 - Les participants discuteront leur travail au sein de petits groupes de travail. La discussion doit aboutir à une solution-type de groupe.

(temps : 30 min)

3 - Un membre sera désigné par groupe pour présenter aux participants la démarche / résultats de son groupe.

Ministère du Développement Social

Recemment un nouveau ministre a été nommé à la tête du Département du Développement social. Avant sa nomination, le futur ministre a été saisi du laxisme et de la totale inéfficacité de plusieurs structures à l'intérieur de ce département.

Supposez que vous avez été disigné pour conduire une étude au niveau de ce département sur les attitudes des agents face à leur travail ou responsabilités.

Le tableau qui suit donne la distribution des travailleurs, agents du Ministère du Developpement Social en 1980.

Sexes / Catégories	Hommes Nombre	Hommes %	Femmes Nombre	Femmes %
Cadres Supérieurs	401	5	41	1
Cadres Moyens	829	11	230	7
Employés	840	11	1.377	44
Ouvriers - Manoeuvres	5.402	71	1.260	40
Dactylo et chauffeurs	140	2	254	8
Autres	8		2	
	7.620	100	3.104	100

QUESTIONS

1. Choisir le dessin d'échantillonage que vous jugez le plus approprié.

2. Aurez-vous besoin d'informations supplémentaires pour déterminer votre échantillon ? justifier.

3. Expliquer toutes les étapes de votre dessin d'échantillonage et justifier le choix de la méthode

CHAPITRE V PROBLEMES DE CONTINGENCE

Lors de discussions antérieures (II.A.1) nous avions souligné le nécessité d'organiser l'information statistique sous forme de tableaux qui puissent resortir de manière compacte mais claire les caractéristiques importantes de la population à étudier.

Le tableau de contigence fournit une information articulée sur deux variables de classification, l'une placée en ligne et l'autre en colonne. Généralement, c'est la variable dite "indépendante" qui est placée en haut du tableau et la variable dite "dépendante" placée sur le côté gauche du tableau. Une fois que deux variables nominales (ou ordinales) sont classées et croisées, nous pouvons déterminer si oui ou non il existe aucune relation entre elles. Ainsi, le test du Chi-Deux pourra être utilisé pour déterminer s'il existe ou non une relation significative entre deux critères de classification.

V.A. LE TEST DU CHI-DEUX D'INDEPENDANCE

Par usage du test du X^2 nous cherchons à évaluer si des fréquences observées diffèrent réellement de celles qui auraient dû être obtenues sous certaines conditions. On parle généralement dans le second cas de fréquences attendues en reférence aux composantes ou totaux marginaux.

Considerons un échantillon de 95 paysans classés en fonction de leurs niveaux de revenu et intentions de vôte pour l'adoption de la proposition du budget comminautaire.

Niveau de revenu	Réponse (vôte) Pour	Contre	Tl = Total ligne
Inférieur	32	12	44
Moyen	14	22	36
Supérieur	6	9	15
Tc = Total colonne	52	43	TG = 95

Ce tableau est défini comme étant un tableau de contigence 3 X 2 (3 lignes et 2 colonnes). Le problème est de savoir maintenant s'il existe ou non une relation entre Niveau de Revenu et Preférence au Vôte. Ainsi, l'hypothèse nulle sera définie comme suit :

Ho : Il n'existe pas de relation entre Niveau de Revenu et Preférence au Vôte.

Les nombres contenus dans les différentes cases du tableau sont les fréquences observées (Oi). Ces fréquences sont obtenues par échantillonage. Sous l'assomption maintenant que l'hypothèse nulle est correcte et qu'aussi l'échantillon est aléatoirement choisi, nous procédons à une redistribution des totaux marginaux aux différentes cases du tableaux.

La distribution des totaux marginaux génère ce que nous appelerons des "fréquences attendues" notées ei. Ces fréquences attendues sont obtenues en multipliant pour chaque case les totaux de ligne et de colonne correspondants et divisant le produit ainsi obtenu par le total général ou somme totale de observation ($ei = \frac{Tc \times Tl}{G}$). Ainsi, dans notre exemple, les fréquences attendues pour toutes les cases du tableau seront obtenues comme suit :

Niveau de revenu	Réponse (vote) Pour	Contre	Tl = Total ligne
Inférieur	$\frac{52X44}{95} = 24.1$	$\frac{43X44}{95} = 19.9$	44
Moyen	$\frac{12X36}{95} = 19.7$	$\frac{43X36}{95} = 16.3$	36
Supérieur	$\frac{52X15}{95} = 8.2$	$\frac{43X15}{95} = 6.8$	15
Tc = Total colonne	52	43	95

Le tableau suivant donne la distribution des fréquences observées (Oi) et des fréquences attendues (ei) à chaque case.

Niveau de revenu	Pour	Contre	TL = Total ligne
Inférieur	O=32; e=24.1	O=12; e=19.9	44
Moyen	O=14; e=19.7	O=22; e=16.3	36
Supérieur	O=6 ; e= 8.2	O= 9; e= 6.8	15
Tc = Total colonne	52	43	G= 95

Finalement, le X^2 calculé sera obtenu comme suit :

$$X^2 = \sum \frac{(Oi - ei)^2}{ei}$$

Le X^2 calculé sera comparé à la valeur X^2 de la table au seuil de signification α et au nombre de degré de liberté ν (ou df) tiré du tableau.

Pour tous tableaux de contingence le nombre de degrés de liberté d f s'obtient comme suit:

$$df = (l - 1)(c - 1) \quad \text{où } l = \text{nombre de lignes}$$

$$\text{et} \quad c = \text{nombre de colonnes}$$

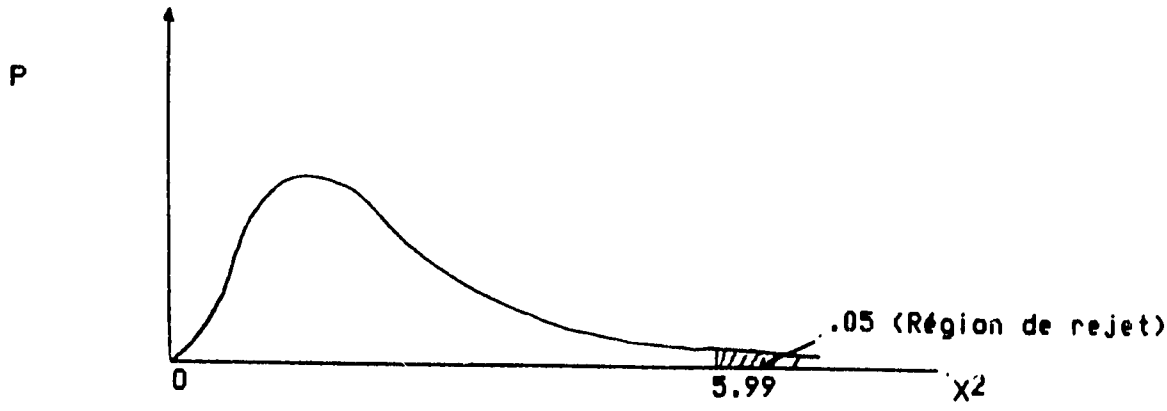
Dans notre exemple, le tableau de calcul du X^2 se présentera comme suit :

Cases	Oi	ei	Oi-ei	$(oi - ei)^2$	$(oi - ei)^2 / ei$
1	32	24.1	7.9	62.41	2.5896
2	14	19.7	-5.7	32.49	1.6492
3	6	8.2	-2.2	4.84	.5902
4	12	19.9	-7.9	62.41	3.1362
5	22	16.3	5.7	32.49	1.9932
6	9	6.8	2.2	4.84	.7118
					$X^2 = 10.6702$

Le nombre de degré de liberté associé sera

$$(l - 1)(C - 1) = (3 - 1)(2 - 1) = 2$$

Ainsi $X^2_{.05}$ avec ν (= df = 2) sera 5.99 (voir table X^2) en annexe



Etant donné que X^2 (calculé) = 10.67> X^2 (table) = 5.99, nous rejetons l'hypothèse nulle.

Nous concluons en disant qu'il existe une <u>relation significative</u> entre niveau de revenu et la preférence au vote.

Note : L'usage du test du X^2 repose sur un échantillonage aléatoire et assume des observations indépendamment distribuées. En outre toutes les fréquences attendues doivent être au moins égales à 5 faute de quoi il est nécessaire de "combiner certaines cases" pour satisfaire cette exigence.

Une fois l'hypothèse nulle rejetée, nous assumons qu'il existe une relation entre les deux variables en considération. Par consequent il peut être nécessaire d'évaluer le degré de cette relation.

On parle alors de mesure d'association entre variables de classification

V. B. MESURES D'ASSOCIATION POUR DES VARIABLES NOMINALES PHI ϕ ET CRAMER'S V

Le coéfficient ϕ tout comme Cramer's V sont des mesures du degré de relation entre des variables classées et croisées.

L'usage ϕ ou V dépend en réalité de la taille du tableau de contigence.

ϕ est utilisé pour des tableaux de taille 2 x 2 alors Cramer's V est conçu pour des tableaux de taille supérieur à 2 x 2.

V.B.1. Le coefficient ϕ

ϕ mesure le degré de relation entre deux variables d'un tableau de contigence et pour lesquelles une au moins est de niveau nominal.

ϕ est obtenu comme suit

$$\phi = \sqrt{\frac{X^2}{n}} \quad \text{où} \quad X^2 = \sum \frac{(oi - ei)}{ei}$$

ϕ varie de 0 à 1. Son interprétation reste un problème de jugement personnel et / ou de l'expérience que l'on a du type de relation que l'on veut analyser.

Note : $\phi = r^2$ pour tout tableau 2 x 2

V.B.2. Cramer's V

Comme nous l'avons mentionné plus haut, Cramer's est conçu pour des tableaux de taille superieur à 2 x 2. En fait Cramer's n'est rien d'autre qu'une modification du coefficient Phi.

C'est à dire $$V = \sqrt{\frac{X^2}{nt}} \quad \text{où} \quad X^2 = \sqrt{\frac{(oi - ei)^2}{ei}}$$

n = taille de l'échantillon

t = minimum de (l - 1) et (c - 1)

Tout comme le coefficient ϕ, Cramer's varie entre 0 et 1. Son interprétation dépend de l'analyste. Cependant une valeur de V proche de 1 suggère une forte relation et une valeur proche de 0 suggère une faible relation.

<u>Problème</u> : A partir des résultats des élections législatives passées vous avez isolé un échantillon de 518 électeurs que vous avez classés suivant leur affiliation religieuse et leur preférence de vote, comme suit :

Parti	Musulmans	Catholiques	Protestants & Animistes	Total
Parti socialiste	126	61	38	225
Parti démocratique	71	93	69	223
Parti maxiste	19	14	27	60
Total	216	168	134	518

<u>QUESTIONS</u> : 1. Définir l'hypothèse nulle et faire usage du test du X^2

2. Si la conclusion de votre test suggère une certaine relation, donner une indication du degré de cette relation.

V.C. MESURES D'ASSOCIATION POUR DES VARIABLES ORDINALES : Gamma (γ) et Spearman's (φ)

Gamma (γ) tout comme Spearman's (φ) sont tout aussi bien des mesures du degré que de la direction de la relation entre 2 variables ordinales classées et / ou croisées.

V.C.1. GAMMA (γ)

Gamma mesure le degré d'association entre 2 variables ordinales par une prédiction de l'ordre suivi par une variable dépendante à partir de la connaissance de l'ordre affectant la variable indépendante correspondante (Hickey 1986 p. 252).

Dans la discussion qui suit, nous nous intéressons exclusivement à un cas spécial de Gamma nommé "Yule's Q", conçu principalement pour des tableaux de taille 2 x 2.

Le calcul de ce coéfficient simplifié de Gamma s'effectue comme suit :

$$Q = \frac{AD - BC}{AD + BC}$$

1 Où A = Valeur contenue dans la case supérieure gauche

B = " " " supérieure droite

C = " " " inférieure gauche

D = " " " inférieure droite

Supposons, en guise d'exemple, qu'à la suite d'une enquête nous avons croisé dans le tableau cidessous le statut des paysans enquêtés et le niveau de leur revenu, comme suit :

Niveau Revenu	Occupation (Statut)		Total
	Employé	Propriétaire exploitant	
Bas	20 (A)	5 (B)	25
Elevé	15 (C)	25 (D)	40
Total	35	30	65

"L'Occupation" ou Statut représente la variable indépendante alors que le "Niveau Revenu" correspond à la variable dépendante. Toutes les 2 variables ci-dessus décrites, sont catégorisées à savoir de la catégorie supposée "faible" à la catégorie "forte" et ce dans le même sens.

Notre mesure Gamma, ou plus exactement Q, sera obtenue comme suit :

$$\gamma = Q = \frac{AD - BC}{AD + BC} = \frac{20(25) - 5(15)}{20(25) + 5(15)} = \frac{425}{525}$$

D'où $Q = .739$

Cette valeur de Q (= .739) indique qu'il y'a 74 % de reduction d'erreur dans l'usage de connaissance sur l'occupation ou statut de l'individu pour prédire son niveau ou catégorie de revenu. En outre, puisque Q est positif, l'ordre donné à la variable

dépendante est le même que celui donné à la variable indépendante. En d'autres termes si un individu doté d'un statut plus élevé qu'une autre, il lui correspondra probablement une classe ou catégorie de revenu supérieure.

<u>VC2</u> : Spearman's (φ)

Dans plusieurs situations de recherche nous sommes amenés très souvent à ordonner les données que nous obtenons en considérant par exemple 2 variables différentes simultanément dans un même tableau. La question est donc de savoir si l'ordre suivi par une variable est le même que celui de la variable qui lui est placée en correspondance.

Spearman's (φ) est une mesure d'association entre 2 variables ordinales qui nous permet de voir si l'ordre de rangement donné à la variable dépendante suit l'ordre de rangement préalablement assigné à la variable indépendante.

φ est défini comme suit :

$$\varphi = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$ où d mesure la différence entre 2 observations du même palier ou niveau après rangement des 2 variables.

Tout comme (γ), Spearman's Ro (φ) peut prendre toutes valeurs entre - 1 et + 1. Il informe à la fois sur le sens (direction) et le degré d'association entre 2 variables ordinales.

Problème : A l'issue d'enquêtes antérieures dans la région du Sine Saloum, l'information suivante portant sur la taille de population et le revenu annuel total de 15 communautés rurales a été présentée ci-après :

Population et revenu annuel de 15 communautés rurales de la région du Sine Saloum, en 1985

Communauté	Population en 1.000	Revenu annuel en 10^6 FCFA
1	15	37
2	18	42
3	21	29
4	14	28
5	18	36
6	32	45
7	17	33
8	23	49
9	9	38
10	9	17
11	26	56
12	13	29
13	15	31
14	19	38
15	26	61

Question : Calculer rho (φ) pour mesurer le degré d'association entre les 2 variables ci-dessus décrites.

Note - guide :

1. Arranger les scores buts fournis par ordre de magnitude croissante ou décroissante. Lorsque 2 scores tombent à la même place (rang), déterminer leur moyenne. Par exemple lorsque 2 scores tombent au rang 6, au lieu de leur donner les rangs 6 et 7, faire la moyenne de 6 et 7 à savoir 6.5 et affecter

aux scores la position 6.5. Ainsi le prochain score prendra le rang 8 et ainsi de suite.

2. Calculer d, la différence de rangs affectés aux différents scores des 2 variables.

3. Calculer φ à partir de la formule fournie en haut et déterminer ainsi le sens et le degré de la relation entre les 2 variables ainsi classées.

Le tableau de calcul pourra être de la forme suivante :

Communauté	Population	Rang 1	Revenu	Rang 2	d = Rang1 + Rang2	d^2
						-
1	16	-	37	-	-	-
2	18	-	42	-	-	-
3	21	-	29	-	-	-
4	14	-	28	-	-	-
5	18	-	36	-	-	-
6	32	-	45	-	-	-
7	17	-	33	-	-	-
8	23	-	49	-	-	-
9	9	-	38	-	-	-
10	9	-	17	-	-	-
11	26	-	56	-	-	-
12	13	-	29	-	-	-
13	15	-	31	-	-	-
14	19	-	38	-	-	-
15	26	-	61	-	-	-

Ainsi $\varphi = 1 - \dfrac{6 \sum d^2}{n(n^2 - 1)}$ avec n = 15

POP	REV	R1	R2
16	37	6	8
18	42	8.5	11
21	29	11	3.5
14	28	4	2
18	36	8.5	7
32	45	15	12
17	33	7	6
23	49	12	13
9	38	1.5	9.5
9	17	1.5	1
26	56	13.5	14
13	29	3	3.5
15	31	5	5
19	38	10	9.5
26	61	13.5	15

LUMN 3: R1
UMN 4: R2

EGREES OF FREEDOM: 13

SPEARMAN R: .729148275 CE

T: 3.8415422

LUMN 3: R1
LUMN 4: R2

EGREES OF FREEDOM: 13

Va Pop = 49

SPEARMAN R: 1

T: 118146.704

LUMN 3: R1
LUMN 4: R2

Va Pop = 19

GREES OF FREEDOM: 13

SPEARMAN R: 1

T: 118146.704

100a

CHAPITRE VI : LA RECHERCHE PAR L'ENQUETE : UN OUTIL ESSENTIEL DE GENERATION D'INFORMATIONS NECESSAIRES A LA PRISE DE DECISIONS RATIONNELLES

INTRODUCTION

De nos jours l'enquête est probablement la méthode de recherche la plus usutée dans le domaine des sciences sociales (Babbie, 1973). Le développement de l'enquête et particulièrement de l'enquête par sondage, en tant que procédé pour générer des connaissances édifiantes concernant différents aspects de la vie humaine peut être considéré comme l'une des grandes réussites des sciences sociales au cours des deux dernières décennies. Mais en réalité, la technique de recueil de données par le biais d'enquête a une longue histoire. "Les anciens empires d'Egypte et de Rome organisaient des recensements périodiques pour fixer les taux de taxe d'impôts, le niveau d'enrolement militaire et aussi pour générer des informations nécessaires à la prise de décisions adminitratives. Mais ce n'est qu'au XVIII ème siècle que l'enquête à grande échelle fut employée de manière organisée et mieux pensée pour l'étude de problèmes sociaux" (Warwick & Lininger, 1975, p.2).

On pourrrait considérer les premières années qui suivirent la seconde guerre mondiale comme le tournant qui a progressivement orienté la recherche par l'enquête vers des perspectives plus larges et plus dynamiques. "Des études modèles sur des groupes d'intérêts spéciaux, les comportements sociaux et autres aspects de la vie humaine, ont marqué de façon à la fois évolutive et révolutionnaire l'usage de l'enquête comme outil central et essentiel de la recherche. Le travail est devenu comparatif et interdisciplinaire. Et, de façon particulière, l'accent est désormais mis sur des études empiriques et vers un effort de neutralité sur les jugements à priori (Zisk, 1981, p.318). En fait la recherche par l'enquête s'est avérée révolutionnaire grâce à la combinaison de plusieurs facteurs dont probablement le plus important est l'intégration et l'utilisation des procédures aléatoires. Depuis, les procédures d'échantillonage, et de sondages aléatoires en particulier, sont devenues les principales sources de génération de données nécessaires à l'étude des comportements sociaux, politiques et économiques (Alwin, 1978). La recherche par l'enquête a été désormais intégrée et employée dans des disciplines aussi variées que les sciences politiques, la sociologie, l'agriculture, la médécine, la géographie, pour ne citer que quelques unes. Et, dans son évolution et expansion rapide, la recherche par l'enquête est venue prendre une place de choix sinon centrale dans bien de domaines de la vie autres que ceux supposément reservés aux universitaires. Dans une certaine mesure, aujourd'hui tout habitant des régions développées a été une fois

au moins dans sa vie, touché par une enquête, qu'elle soit d'opinions ou de de toutes autres natures (Babbie, 1973). Progressivement, dans les pays en voie de développement, les agences spécialisées de gouvernement, en particulier, ont mis l'accent sur l'usage de la recherche par l'enquête pour la conduite d'études de (pré)planification et de travaux d'évaluation de programmes ou projets de développement de tous genres.

VI.A - LES OBJECTIFS DE LA RECHERCHE PAR L'ENQUETE ET SES IMPLICATIONS DANS LE CADRE D'UNE EVALUATION DE PROGRAMMES

Malgré son usage de plus en plus répandue aujourd'hui, la recherche par l'enquête n'est pas aussi simple qu'elle n'apparait à priori. La recherche par l'enquête ou l'enquête tout court, ne consiste pas tout simplement à poser des questions et à compter des réponses. En fait la recherche par l'enquête est d'autant plus complexe que les problèmes d'intérêts sont divers et souvent complexes et obscures à cerner. Néanmoins, la recherche par l'enquête s'est averée être appropriée à l'étude de problèmes socio-économiques divers. Et, un usage très important de la recherche par l'enquête a été la mise en forme et le développement d'informations exploratoires fondamentales à l'étude préliminaire ou compréhensive de programmes et projets de développement. Par conséquent, "autant le champ d'application s'est élargi, autant la controverse quant à son emploi et la confusion sur ses objectifs et implications se sont accrues" (Warwick & Lininger, 1975, p.5).

Cependant, la litérature récente en matière de recherche par l'enquête suggère des tentatives réussissantes vers une connaissance meilleure de nos politiques en matière de développement.

Les travaux et efforts quotidiens immenses investis sur divers aspects du processus de la recherche à savoir la formulation de la question de recherche pour une situation donnée, la définition des objectifs et des indicateurs appropriés associés, la mise en forme des questions pertinentes de même que l'interprétation scientifiquement menée des données générées par l'enquête, ont permi aujourd'hui, à n'en pas douter, d'améliorer sensiblement nos connaissances de notre environnement de même qu'ils ont fait de la technique un outil presque irremplaçable de meilleure perception de bien nombre de phénomènes socio-économiques.

Du fait la recherche par l'enquête appliquée à l'évaluation de programmes, pour répondre correctement à l'objectif fondamental consistant à la prise de décisions rationnelles, a besoin de reposer sur un certain nombre de préalables dont:

- la mise en forme correcte et sans ambiguité des buts/objectifs du programme. Tout aussi important est, à ce stade, d'assurer le conscensus de la part de toutes les parties impliquées autour de ces buts/objectifs
- la définition, le recensement et la consignation de tous indicateurs nécessaires à l'appréciation des objectifs définis
- La pré(planification) objective et dûment documentée de tous aspects déterminants du programme. Une fois ces préalables remplies, et seulement alors, le chercheur sera en position de définir sa question de recherche de même que toutes les hypothèses pertinentes nécessaires à la conduite de la recherche par l'enquête et pour l'évaluation envisagée.

VI.B - VERS L'ATTEINTE DES BUTS ASSIGNES A LA RECHERCHE PAR L'ENQUETE : LA QUETE CONSTANTE DE VALIDITE ET DE FIABILITE

Dans leur quête constante de connaissances <u>valides et fiables</u> des politiques contemporaines de développement, particulièrement dans les régions en voie de développement, les spécialistes en sciences sociales ont investi des efforts appréciables en vue d'améliorer les aspects critiques et souvent controversés du processus de la recherche. Ces aspects ont souvent pour noms:

- la mise en forme claire de la question de recherche d'intérêt,

- la définition et l'identification effective des objectifs du programme et donc de la recherche à entreprendre,

- le récensement de tous indicateurs appropriés et nécessaires à la la réponse à fournir plus tard,

- le choix pensé de l'échantillon représentatif des sujets intéressés par la question,

- la mise en forme d'instrument(s) nécessaires à la collecte d'informations,

- la réduction autant que possible de tous risques de subjectivisme ou de contamination des réponses par l'enquêteur ou l'enquêté.

Chacun de ces aspects et problèmes ci-dessus énumérés constitue en lui seul un point vital et déterminant sur l'issue de la recherche à entreprendre.

Des chercheurs comme Fisk (1981) et Warwick & Lininger (1975) ont richement discuté la nécessité de rendre nos questions de recherche et toutes hypothèses associées aussi claires et spécifiques que possible afin que le développement d'indicateurs appropriés et adéquats de même que l'instrument qui permet de les mesurer puisse être envisagé avec le maximum de pertinence possible. Kerlinger (1973) de même que Babbie (1973) se sont longuement étendus sur la complexité et le rôle fondamental de la mise en forme de questions en tant qu'elle constitue un facteur déterminant sur aussi bien le volume que la qualité des réponses qui seront fournies.

La sélection ou le choix d'échantillon(s) approprié(s) à un sujet de recherche donné a toujours constitué un souci permanent de perfection de la part des chercheurs. Les limitations souvent dénoncées des procédures non aléatoires, leur manque de rigueur ou de base scientifique à l'inférence statistique, par exemple, ont progressivement amené la plupart des chercheurs à se tourner vers les méthodes de sélection aléatoire, communement appelées échantillonages probabilistes.

Il est certain que les procédures aléatoires ont permi aujourd'hui un examen raffiné et à la fois sophistiqué de bien de problèmes sociaux ou économiques, de générer des données plus fiables et de permettre ainsi une meilleure garantie d'une base scientifique de généralisation de résultats et conclusions tirées sur de petit(s) nombre(s) de sujets.

Un autre domaine particulièrement vital à la recherche par l'enquête a été (et reste) de savoir comment réduire le taux de réticence chez les enquêtés intérressés par un programme donné et quelles stratégies adopter afin de minimiser le plus possible les biais souvent introduits au moment précis du remplissage du questionnaire.

En d'autres termes il s'agit de trouver la méthode "d'interview" la moins sujette à la contamination de l'instrument d'enquête et à l'interprétation falacieuse des questions et donc à la production de réponses tendencieusement erronées. Les travaux d'érudits tels que Mizes (1984), Tucker (1983) et Osenberger (1986) indiquent sans nul doute que le taux de réticence aux enquêtes de même que les erreurs d'interprétation de question, et donc de production de réponses erronnées, tendent à baisser très remarquablement avec l'usage d'interview directe, c'est à dire de la méthode dite "face-à-face" entre enquêteur et répondant.

Experts et utilisateurs de la recherche par l'enquête, concients des effets potentiels des caractéristiques de l'enquêteur sur les résultats d'enquêtes, ont particulièrement

mis l'accent sur la formation. Singer et al (1983) et Warwck & Lininger (1975), par exemple, ont montré que les caractéristiques intrinsèques de l'enquêteur lui même peuvent de façon déterminante influencer la nature et qualité de réponses. Par conséquent, suggèrent-ils, un effort de formation suivie d'une phase de "pretest" pourront contribuer remarquablement à améliorer la qualité des réponses en situation réelle d'enquête.

Techniquement parlant, spécialistes de la recherche par l'enquête ont reconnu la nécessité absolue d'avoir un instrument d'enquête, le questionnaire, proprement et structurellement bien conçu pour pouvoir servir à l'objectif ultime de générer l'information ou les données nécessaires à l'examen de la question de recherche et donc satisfaire aux besoins de l'évaluation envisagée. Cet instrument doit permettre, tout aussi bien dans sa structuration que par la pertinence des questions qui le composent, de satisfaire valablement l'objectif ultime de la recherche entreprise à savoir la génération d'informations valides et fiables quant au problème d'intérêt posé. Du fait, la mise en forme d'un instrument de génération de données doit tenir compte à la fois de la forme, du contenu tout comme de l'ordre de questions qui le composent. La litérature récente en matière de recherche est très riche d'expériences quant à l'énonciation et à la forme des questions, le contenu et la localisation de questions délicates ou souvent perçues comme "tabou".

VI.C - PERSPECTIVES DE LA RECHERCHE PAR L'ENQUETE

La discussion ci-entreprise est suggestive de la place prépondérante que progressivement les méthodes de recherche par l'enquête sont venues occuper dans l'examen de nos problèmes quotidiens et particulièrement de l'évaluation de nos programmes de développement. Elles ont été appliquées à des domaines aussi variés que la sociologie, l'agriculture et la médecine, pour ne citer que ceux-là. Dans son développement et expansion rapides, la recherche par l'enquête s'est averée être non seulement un outil puissant de description mais il est apparu aussi comme étant l'un des outils scientifiques les plus aptes aujourd'hui à répondre aux questions telles que "comment ceci..... étant donné que .....?", "Qu'est-ce qui est ..... si ceci .....?". La recherche par l'enquête est ainsi devenue un instrument, un outil puissant d'études scientifiquement élaborées de politiques (Johnson & Joslyn, 1986). Cependant, certaines personnes (humanistes et étudiants la plupart du temps) continuent de soutenir le contraire à savoir que "la quantification de relations humaines est deshumanisante", donc par consequent rejettent-ils toute idée de quantification du comportement social humain. Cette opinion n'est heureusement pas partagée par la grande majorité de penseurs en matière de recherche. Babbie, par exemple, de suggérer que "si chacun considérait toujours les relations humaines seulement de manière quantitative, alors ces relations seraient probablement déshumanisées dans le processus et, ce faisant, tous instincts humanitaires innés chez

l'homme seraient perdus. Mais même le savant ou erudit le plus déterministe et grand adepte de la quantification, ne mène pas sa vie en de tels termes; et plus important est que les problèmes humains requèrant une profonde fondation morale et appelant à la compassion dans leur solution, sont ceux-là même qui nécessitent avant tout une quantification, c'est à dire des données précises (Babbie, 1973 p.370).

Aujourd'hui la recherche par l'enquête a pris un tournant particulièrement encourageant parce que progressivement en quête de parfaire nos connaissances des politiques et autres changements qui entourent et guident notre existance. D'outil descriptif, la recherche par l'enquête s'est constamment tournée vers une demarche intégrée d'exploration, d'explication et de prédiction de bien des phénomènes ou activités socio-économiques. Grâce à la recherche par l'enquête il est aujourd'hui possible de bien cerner nos programmes de développement de la phase de planification à l'évaluation de leurs impacts en passant par le suivi régulier de leur exécution.

CONCLUSION

De façon générale la litérature récente sur la recherche par l'enquête est riche d'expériences révolutionnaires qui ont fait d'elle, désormais, un outil presque irremplaçable par lequel nous pouvons mieux initier et gérer nos politiques de développement. Par la génération d'informations fiables et valides, cet outil nous met progressivement dans une position d'appréhension meilleure et scientifiquement pensée de bien des facteurs déterminants inhérents à nos conditions d'existence.

Il est certain qu'à travers cette discussion-ci nous ne pouvons pas recenser les problèmes divers suscités ou centraux à son usage. Notre ambition a été de souligner que certes des problèmes existent mais qu'aussi, dans son évolution, la recherche par l'enquête s'est constamment améliorée pour constituer aujourd'hui le meilleur outil d'inquisition mais aussi de base de décisions nécessaire à nos politiques. Et, parce que toujours en évolution, la recherche par l'enquête reste un processus continuel par lequel nous tendons à achever le but ultime d'une connaissance toujours plus fiable et plus valide de nos politiques de développement. Un domaine d'application particulièrement enrichissant est l'évaluation de nos programmes de développement. Et sur ce domaine d'application bien précis, la recherche par l'enquête a besoin d'être mieux assimilée pour mieux répondre à mieux élucider nos choix et politiques de développement.

ETUDE DE CAS : LA RECHERCHE PAR L'ENQUETE POUR UNE EVALUATION DE PROGRAMMES DE DEVELOPPEMENT

Les programmes de riz au pays du Sahel

Instructions aux participants

1. Lire en entier le cas décrit ci-après.
   Se rappeler que le cas d'étude proposé à été conçu exclusivement pour des besoins académiques; ce faisant aucune discussion quant à la consistence des points ou nature des programmes reférés ne saurait être pertinente.

2. Les participants sont demandés de travailler et mener leur reflexion en groupes (nombre de groupes = 3).

3. Chacune des questions posées à la fin de la discussion du cas fera l'objet de débats et synthèse après rapport des delegués de groupes.

4. Les temps approximatifs de réflexions-travaux et de présentation, discussion et synthèse sont donnés plus loin dans la partie travail demandé.

LES PROGRAMMES DE RIZ AU SENEGAL

I - ENONCE DE LA SITUATION ET LA QUESTION DE RECHERCHE

Depuis l'indépendance en 1960 le gouvernement du pays du Sahel essaie de reduire progressivement les dépenses sur les importations de riz en encourageant la production locale de cette céréale qui constitue de loin la principale composante alimentaire de la population (Lele, 1978; Eicher 1981; Platon, 1981). L'expansion de la production locale de riz a toujours occupé une place de choix dans toutes les stratégies de développement en général et de la politique d'auto-suffisance alimentaire initiées par le gouvernement depuis l'indépendance. Les projets de production et d'expansion du riz ont nécessité des investissements énormes au cours des années 1960 et 1970 par le biais principalement d'Agences Nationales Locales de Développement Agricole (Tuluy 1981; Kirtley, 1985; FAO, 1982).

La stratégie de développement du secteur du riz au pays du Sahel tourne autour de deux points essentiels à savoir

- l'expansion de la culture irriguée au nord et,
- la modernisation des systèmes traditionnels de cultures au sud (Craven et Tuly 1981; Nelson 1976; Eicher et Baker, 1982)

Cependant, malgré les efforts immenses investis pour développer la production, les importations de riz sont restées très importantes représentant en moyenne près de 75 % de la consommation totale du pays (FAO 1982 p.67). DO aux dépenses énormes occasionnées par la seule importation de riz, le gouvernement s'est vu confronté au dilemne suivant :

1. Continuer d'investir d'importantes ressources (tant matérielles, humaines que financières) sur la production et l'expansion de riz dans l'espoir que les objectifs assignés seront certainement atteints, ou

2. arrêter ces programmes et/ou réorienter les efforts vers d'autres types de production ou autres programmes de développement.

LA QUESTION DE RECHERCHE

Comme déjà indiqué plus haut, les objectifs des programmes de riz au pays du Sahel étaient fondamentalement d'augmenter la production nationale en vue de réduire donc sensiblement les importations; ensuite d'améliorer et de développer cette production et, en même temps, améliorer l'accessibilité au riz produit localement par une production de grandes quantités qui concoureront à achever plutard la politique visée d'auto-suffisance alimentaire.

Le plan de recherche par l'enquête proposé se veut de fournir toutes informations nécessaires de base aux responsables et planificateurs qui leur permettront de prendre des décisions judicieusement pensées quant au secteur du riz. Il s'agira pour ces derniers donc de disposer de données de base en vue d'une "formative évaluation", c'est à dire "une évaluation qui permettra de choisir entre: reviser, améliorer ou reorienter les programmes mis en oeuvre" (Warwick Lininger 1975, p.12)

Le plan de recherche par l'enquête proposée vise fondamentalement à :

1. fournir des données comparatives de base sur les productions et importations de riz au pays du Sahel.

2. Fournir des renseignements précieux sur tous les éléments concernant ou déterminant la production de riz.

3. Fournir l'information nécessaire sur la manière dont la politique nationale du riz affecte les budgets des foyers du pays du Sahel, à savoir dans quelle mesure en particulier la production et l'expansion de riz domestique a été effective quant à son abondance, son accessibilité à la famille moyenne locale.

II - LE PLAN "OU MONTAGE" PROPOSE DE LA RECHERCHE PAR L'ENQUETE

A. Population-cible et échantillonage

La recherche par l'enquête proposée a pour mission première de fournir l'information nécessaire pour un examen judicieux de la culture et l'expansion du riz de même qu'elle servira au diagnostic de l'impact socio-économique des programmes du gouvernement pour le secteur du riz sur les budgets et comportements des familles du pays de Sahel. Les données nécessaires sur les importations et productions annuelles, la taille de la population, les dépenses de recherche et d'encadrement, de même que les prix à la consommation sont disponibles auprès des ministères du Commerce, de l'Economie, de la Recherche Scientifique et des diverses agences gouvernementales de développement en rapport ou engagés dans la culture du riz. Par conséquent, toutes données nécessaires à l'établissement de series temporelles depuis 1961 à aujourd'hui pourront être et seront fournies par ces ministères ou agences mentionnés ci-dessus.

La recherche par l'enquête proposée ici a deux composantes: une enquête sur les exploitations de riz et une enquête nationale des comportements alimentaires de la famille du pays du Sahel. Les procédures de sélection des échantillons nécessaires à ces deux enquêtes sont rapidement discutées en appendix I.

L'unité d'analyse pour l'enquête sur les exploitations de riz est le foyer rural exploitant tandisque l'unité d'analyse pour l'enquête nationale est le ménage typique du pays du Sahel.

Dans les deux situations, des échantillons stratifiés seront choisis. La raison d'une telle sélection est de permettre dans le premier cas une meilleure représentation des petits comme des grands foyers exploitants de riz et, dans le second cas une correcte représentation des ménages ruraux et urbains du pays du Sahel.

La technique d'enquête retenue dans les deux situations sera l'interview directe supportée par un questionnaire standard.

B. Aspects critiques du plan proposé de recherche par l'enquête

B.a. Applicabilité au contexte du pays du Sahel

Comme pour tous autres pays en voie de développement, l'entreprise d'une recherche par l'enquête au pays du Sahel reste un défi majeur. Cependant, au pays du Sahel des bases solides permettant une conduite proprement acceptable d'enquêtes existent. Pour une chose, la population rurale du pays participe regulièrement à des recensements périodiques aux fins de la collecte de la taxe rurale.

Deuxièmement, depuis l'indépendance, plusieurs opérations d'enquêtes ont été menées à travers le pays telles que le recensement démographique, les enquêtes sur la mortalité infantile et sur la fertilité féminine initiées par l'office nationale de statistique.

Troisièmement, les agences locales de développement agricole ont méné un suivi régulier de l'évolution de la culture du riz basé souvent sur des recensements réguliers (Tuluy, 1981). Nous assumons, par conséquent, que la recherche par l'enquête dans un contexte tel que celui-ci ne présente aucune difficulté ou contrainte majeures pour générer l'information nécessaire à l'examen correct de la question posée.

B.b. Administration de recherche par l'enquête proposée

Le plan de recherche par l'enquête ci proposée sera administré par le Centre de Recherche de l'Ecole Supérieure d'Agriculture qui fera en outre appel à certaines compétences ponctuelles de l'office national de la statistique.

Il n'y aura pas d'enquête(s) pilote(s) ou "prétest". L'option non-prétest est mûe par le souci de ne pas contaminer l'instrument utilisé en éveillant très tôt la sensibilité de certains sujets qui pourraient être retenus plus tard pour l'enquête définitive.

B.c. Option d'enquête

La recherche par l'enquête est basée sur une interview directe. Il s'agit d'une interaction directe entre enquêteurs et répondeurs par le biais d'un questionnaire standard écrit en français.

Dans le contexte actuelle du pays du Sahel il n'existe pas de bases d'enquêtes postales ou téléphoniques. Les équipements téléphoniques du pays sont réduits aux seuls centres urbains pour l'essentiel. En outre, le bas niveau d'alphabétisation française particulièrement en milieu rural (World Bank, 1980), doublé d'un service postal pas encore bien au point, ne permettent pas d'envisager d'enquêtes postales.

L'enquête directe, dite de face-à-face, reste la seule et unique voie de génération de l'information nécessaire à la réponse adéquate à la question de recherche.

Néanmoins cependant, des grands efforts devront être mis sur la formation des enquêteurs, et particulièrement la traduction du questionnaire d'enquête en langue nationale aux sujets ne parlant pas ou ne comprenant pas bien le français. Cet aspect ne constitue cependant pas un blocage sérieux dans la mesure où l'administration de l'enquête dispose déjà d'équipes d'enquêteurs particulièrement bien formés dans la traduction et l'usage de la langue nationale pour la conduite d'interviews directes.

C - Assomptions sous-tendant la recherche par l'enquête proposée

La recherche par l'enquête ci-dessus proposée est fondamentalement conçue pour l'identification, la détermination et la quantification des incidences des programmes de riz. L'étude proposée repose sur des assomptions suivantes que:

- une planification objective et dûment documentée des programmes fut entreprise au démarrage des dits programmes

- un suivi régulier et une collecte systématique de données ont été entrepris par les agences de développement intéressées sur des aspects essentiels des programmes

- le personnel qualifié, l'équipement et le financement nécessaires sont réunis pour la conduite correcte de la présente étude.

APPENDIX I

APPENDIX I - Note brève sur la nature de l'échantillonage retenu

- Enquête sur les foyers d'exploitants rizicoles

L'unité d'analyse est le foyer exploitant rizicole. Les exploitions de riz sont toutes concentrées au Nord et au Sud du pays. Des listes complètes d'exploitants ou foyers d'exploitants rizicoles sont disponibles auprès des agences de développement intéressées. De même des informations sur les tailles et localisation d'exploitations rizicoles sont disponibles. La procédure d'échantillonage est basée sur une stratification proportionnelle aléatoire. Les deux régions sont assumées jouer un rôle central "similaire" dans le cadre des programmes de riz. Pour chacune des deux régions un nombre déterminé d'exploitations de tailles données sera choisi en reférence à la répartition de celles-ci pour l'ensemble de toutes les exploitations que compte chacune des régions.

Une fois les strates correctement définies et identifiées avec les unités et tailles correspondantes, un sélection aléatoire simple sera utilisée pour déterminer l'échantillon final à enquêter.

Enquête nationale, budget et comportements alimentaires

L'unité d'analyse est le ménage du pays du Sahel. Des listes de ménages et carrés de même que leurs localisations sont disponibles auprès de l'Office National de la Statistique et du Ministère de l'Intérieur. La sélection ou échantillonage sera basée sur la stratification. Les strates seront de deux natures à savoir les ménages urbains d'une part et les ménages ruraux d'autre part. Le tirage se fera comme dans le cas précédent.

TRAVAIL DEMANDE

A) Refléchir en groupe et fournir tous éléments de réponses au points 1, 2 listés ci-dessous :

(Temps : 2h30 mn)

1. Discuter la question de recherche (les objectifs) en reférence aux programmes considérés. Quelles reserves ou questions suscitent les modes de sélection et d'enquête proposés ?

2. Considérant les deux volets d'enquête séparement, définir tous les indicateurs pertinents pour une correcte évaluation des dits programmes.

Discussion plénière (1h30mn)

B) En vous reférant à la discussion précédente sur la complexité de la formulation des questions, la structuration du questionnaire, les questions dites sensibles ou tabous (l'âge, la composition de la famille, les revenus et dépenses ou attitudes politiques, etc....),

(Temps : 2h30mn)

1. lister pour chaqu'un des deux volets de l'enquête toutes les sections essentielles du questionnaire.

2. Concevoir un questionnaire complet pour la recherche à entreprendre (les questions devront être organisées et présentées en tenant compte d'abord de leur pertinence par rapport au problème ci-posé, ensuite de manière à éliminer ou diminuer sensiblement tous risques de refus ou de réserve de la part des répondants).

Présentation, discussion et synthèse

(temps : 1h30 mn)

REFERENCES I Ch. I à V

Alder, Henry L., & Roessler, Edward B. Introduction to Probability & Statistics. Sixth edition, 1977, W. H. Freeman and Co., New York, 426p.

Blalock, Hubert M., Jr. Social Statistics. Revised Second Edition, 1979, McGraw-Hill Book Co., 625p.

Borey & Thiombiano. Initiation al la Statistique Descriptive. 1981, Cooperation Rep. française, 733p.

Hickey, Anthony A. An Introduction to Statistical Techniques for Siocial Research. First edition, 1986, Random House, Inc., New York, 363p.

Hoel G. Paul, & Jessen, J. Raymon. Basic Statistics for Business and Economics. 3rd Edition, 1982, John Wiley & Sons, New York, 629p.

Kutner, Nater & Wasserman. Applied Linear Regression Models. Richard D. Irwin, Inc., 1983.

Martre, Joseph G. Van, & Gilbreath Glenn H. Statistics for Business and Economics. Bisiness Publications, Inc., 1980, 589p.

Nachmias, David. Public Policy Evaluation. St. Martin Press, New York, 1979, 195p.

Sahler, Benjamin, & Fourastie, Jacqueline. Probabilities et Statistiques. Serie J. Guinet, 2nd edition. Dunod, 1981, 260p.

Zar, H. Jerrold. "Biostatistical Analysis." Prentice-Hall. Biological Sciences Series, Englewood Cliffs, Inc., New Jersey, 1974, 620p.

REFERENCES II Ch.VI

1. Alwine Duane F. (1978) Survey Design and Analysis : Current issues. London : Sage Publications

2. Alder, Henry L. & Roessler, Edward B. (1977) Introduction to Probability & Statistics. Sixth Edition: W.H. Freeman and Co., New York.

3. Babbie, Earl R. (1973) Survey Research Methods Belmont, California : Wadsworth Publishing company, Inc.

4. Blalock Hubert M. Jr. (1979) Social statistics Revised Second Edition : McGraw - Hill Book, Co.

5. Craven, K. and H. A.Tuluy (1981) "Rice Policy in Senegal" pp 201-228 in Pearson et al (eds), Rice in West Africa. Stanford, Ca : Stanford University Press

6. Eicher, C.K. & D.C. Baker (1982) Research on Agricultural Developement in Subsaharan Africa. Lansing, MI : Michigan State University Press.

7. Eicher CK (1981). Facing up to Africa's Food Cisis Foreign Affairs Vol 61, No 1, pp 154-176

8. FAO (1982) Rice Development and Rainfed Rice Production. Food and Agr. Org. of the UN. Plant Prod. and Prod. Paper, No 41, Rome.

9. Hickey, Anthony A. (1986) An Introduction to Statistical Techniques for Social Research. First Edition. Random House, Inc, New York.

10. Johnson, Janet B. & Joslyn, Richard A. (1986) Political Sciences Methods Library of congress. Congressional Quarterly, Inc.

11. Kerlinger, Fred N. (1973) Foundations of Behavioral Research New York : Holt, Rinehart and Winston, Inc.

12. Kirtley & Michael Aubine (1985) "Senegambia, A Now and Future Nation" National Geography. August 1985

13. Lele, Uma (1975) The Design of Rural Development Johns Hopkins

14. Martre, Joseph G.Van & Gilbreath Glenn H. (1980) Statistics for Business and Economics: Business Pulication, Inc.

15. Mizes et al (1984) "Incentives for Increasing Return Rates" Public Opinion Quarterly (POQ) vol 45. 1984

16. Nelson H. (1976) Area Handbook for Senegal Washington D.C. US. Government Printing Office

17. Osenberger et al (1983) "Interviewers' Various Refusal Rates" POQ, No 50 1983

18. Platon, P. (1981) "Les objectifs de l'OMVS". Marchés Tropicaux et Méditérraneens. No 1849.

19. Singer et al (1983) "The Effects of Interviewer Characteristics" POQ. 47, 1983.

20. Tucker (1983) "Interviewer Effects in Telephone Survey" POQ 49, 1982.

21. Tourangeau & Smith (1982) "Finding Subgroups for Survey" Po Q 49, 1982.

22. Tuluy H.A. (1981) "Costs and Incentives in Rice Production in Sénegal" pp.229-262, in Pearson et al, (eds), Rice in West Africa. Stanford, Ca : Stanford University Press.

23. Warwick, C. & Lininger, C.A. (1975) The Sample Survey: Theory & Practice New York, Mc Graw - Hill Book Company.

24. World Bank (1980) Poverty and Human Development : A world Bank Publication. Oxford University Press.

25. Zisk Betty H. (1981) Political Research D.C. Heath and Company.

A : LES ELEMENTS CARACTERISTIQUES D'UNE EVALUATION

L'évaluation de programme(s) fait particulièrement appel aux outils de recherche car la pratique de l'évaluation consiste essentiellement en une collecte systématique de données sur les activités, caractéristiques ou résultats de programme(s), ou de produits aux fins de réduire les incertitudes, améliorer l'efficacité et prendre des décisions par rapport à ceux que ces programmes affectent directement. On le voit donc, il s'agit essentiellement dans le cadre d'une évaluation d'arriver à comparer des objectifs assignés à un programme et les résultats réalisés par le même programme. Il y a 4 éléments essentiels caractéristiques de toute évaluation.

1 - La collecte systématique de données

2 - La variété des domaines d'évaluation

3 - Les différentes utilisations possibles d'une évaluation

4 - Les différentes raisons pouvant justifier une évaluation

Suivant leur but et les activités sur lesquelles elles portent, on peut identifier différentes catégories d'évaluation.

CATEGORIES D'EVALUATION	BUTS	ACTIVITES ADRESSEES
Evaluation pour la Planification	Guider la Planification et l'implantation du du programme	Analyse : - du contexte - de préinstallation - de faisabilité
Evaluation "formative"	Amélioration, modification, et management du programme	Développement et processus du programme
Evaluation d'impact	Déterminer les résultats et effets pour des décision sur la reduction, l'expansion ou la continuation du programme	Efficacité, résultat
Evaluation de contrôle	Suivre fonctionnement programme	- controles périodique - nature des services rendus

Cette catégorisation, illustre encore une fois qu'une évaluation peut souvent s'intéresser à un aspect spécifique d'un programme. En effet, tout programme mis en place intéresse différentes personnes :

- ceux qui ont financé le programme
- les bénéficiaires directs du programmes
- les gestionnaires du programme
- la communauté en général
- etc...

Ces différentes personnes ou institutions peuvent être, dans le cadre d'une évaluation, interessées chacune par un ou des aspects bien particuliers du programme. Exemple, dans le cadre d'un programme de soins de santé en milieu rural, les financiers du programme peuvent effectuer une évaluation pour voir si le

plan de financement des activités est bien mené. Une évaluation auprès des bénéficiaires essaierait plutôt, elle, de voir dans quelle mesure les conditions de santé de la population se sont améliorées.

B - LES ETATPES D'UNE EVALUATION

Une évaluation peut être subdivisée en trois grandes phases, chacune d'elles étant constituée d'un ensemble d'actions spécifiques.

PHASE 1 : SPECIFICATION DU THEME D'EVALUATION

Tout programme mis en place s'est fixé des objectifs et envisage la mise en oeuvre de moyens pour atteindre ces objectifs. La première étape d'une évaluation consiste donc à fournir des éléments d'information sur le programme que l'on se propose d'évaluer. L'accent sera mis sur les aspects suivants:

1. Présentation du programme et de ses objectifs

Les objectifs du programme constituent en effet les références devant guider le travail de l'évaluateur. Ces objectifs doivent être traduits en termes très opérationnels. Exemple, notre projet de santé peut se fixer comme objectifs opérationnels :

- vacciner 45% des enfants de 0 à 5 ans entre 1987 et 1989
- construire 10 cases de santé dans la communauté rurale avant 1989.
- etc...

131

2. Revue de la littérature sur le domaine d'évaluation

Ce travail présente plusieurs avantages :

- Il permet de prendre connaissance de travaux antérieurs effectués en la matière;
- Il fournit des références pouvant servir ultérieusement de base de comparaison;
- Il facilite les recherches ultérieures à entreprendre dans le cadre de l'évaluation;
- Il permet de documenter les hypothèses à tester.

3. Spécification du type d'évaluation

Le type d'évaluation à adopter influence ou oriente la méthode d'évaluation. Il existe différents types d'évaluation suivant que le programme n'a pas encore demarré, est en cours, ou est terminé. Les types d'évaluation les plus courants sont mentionnés dans les pages qui précèdent.

4. Détermination du but de l'évaluation

Chaque programme d'évaluation est sous-tendu par une ou plusieurs questions auxquelles on tente d'apporter des réponses aussi précises et exhaustives que possible.

132

Exemple : Un programme de soins de santé implanté en milieu rural s'était fixé comme objectif d'immuniser en deux ans, tous les enfants de 1 à 5 ans contre la rougeole dans la zone (P_o = 100%). Différents objectifs peuvent être assignés à un programme d'évaluation de ce projet.

Si le type d'évaluation est celui de <u>coût d'éfficacité</u>. On s'interessera essentiellement à voir combien il a coûté pour immuniser un enfant. Plutôt que de voir si les objectifs ont été atteints, la préoccupation ici est de voir ce que chaque franc investi a permis de réaliser.

Une évaluation <u>orientée vers le résultat</u>, par contre s'interessera à déterminer si oui ou non les objectifs ont été atteints plutôt que de voir combien il a coûté pour atteindre le résultat.

Une évaluation d'impact s'interesserait à voir tous les effets directs ou indirects du programme sur la communauté (ou des effets spécifiques, tels que impact sur l'envirionnement).

Il est essentiel de bien spécifier au départ les buts qu'on assigne au progamme d'évaluation.

5 - <u>Développer un modèle d'intervention</u>

Le modèle d'intervention ou d'impact prend la forme d'une affirmation sur les relations présumées entre un programme et ses objectifs. Elle est donc une assomption ou hypothèse élaborée à

partir de la question de recherche qui détermine en fait le but de l'évaluation.

Exemple : L'introduction de nouvelles méthodes pédagogiques dans une école visait à améliorer les capacités de lecture des enfants. La question de recherche à laquelle tenterait de répondre une programme d'évaluation pourrait être : Les nouvelles méthodes pédagogiques ont-elles permis d'améliorer les capacités en lecture des enfants ?

Le modèle d'intervention pourrait, lui être formulé sous forme d'hypothèse de recherche : <u>les nouvelles méthodes pédagogiques ont permis d'améliorer les capacités en lecture des enfants</u>.

La formulation d'une hypothèse met en exergue une relation de causalité entre deux ou plusieurs variables :

a) la ou les variable(s) indépendante(s)
(ex: les méthodes pédagogiques)

b) la ou les variable(s) dépendante(s)
(la capacité en lecture des enfants)

Cependant, puisque toute hypothèse est tentative, ce n'est que par la recherche que l'on pourra vérifier la teneur de l'hypothèse (voir chap.III pour plus de détail sur la formulation des hypothèses).

134

PHASE 2 : ELABORATION DES PROCEDURES D'EVALUATION

1. Choix d'une méthode de recherche

Différentes approches existent pour mener une recherche dans le cadre d'une évaluation. Quelques unes seront passés en revue :

a) La méthode expérimentale

groupe expérimental	X_1	X_2	$d = x_2 - x_1$
groupe de contrôle	X'_1	X'_2	$d' = x'_2 - x'_1$

Cette approche consiste à :

- choisir un groupe expérimental bénéficiant du programme implanté;
- choisir un groupe de contrôle présentant les mêmes caractéristiques que le groupe expérimental à la différence que le groupe de contrôle ne bénéficie pas du programme;
- évaluer les changements intervenus au niveau des deux groupes entre le moment où le programme est implanté et le moment où l'évaluation est effectuée;
- estimer l'effet exact du programme sur le groupe expérimental en comparant d à d' (d - d')

Si cette approche peut garantir une certaine validite interne, il reste cependant qu'elle soit couteuse et demande du temps. Par ailleurs il y a des difficultés pratiques liées au choix et au suivi des groupes.

b) La méthode quasi-expérimentale

Dans cette approche on effectue deux observations à deux moments différents sur un groupe expérimental et un groupe de contrôle. Il s'agira de comparer les changements intervenus auprès du groupe bénéficiaire à ceux du groupe similaire non bénéficiaire du programme. Dans notre programme d'introduction de nouvelles méthodes pédagogiques on peut choisir deux groupes d'enfants : ceux fréquentant les classes où les méthodes a été introduites et ceux dans les classes ne bénéficiant pas de l'innovation. On compare les performances des deux groupes pour tirer une conclusion sur l'efficacité de la méthode. Si l'approche est moins couteuse, elle est également moins rigoureuse et ne garantie nullement une validité interne car elle ne permet pas d'isoler les effets de facteurs dus aux changements endogènes, aux tendances à long terme, aux éléments interférants qui peuvent influencer les résultats.

c) La méthode avant et après (before and after)

Ici, l'observation porte exclusivement sur le groupe bénéficiant du programme. On compare la situation avant le programme à celle après

X_1	X_2

$d = X_2 - X_1$

Ici, également deux problèmes essentiels se posent : l'absence d'éléments de comparaison et le problème de la validité (dans qu'elle mesure peut on attribuer le changement intervenu au programme ?). En outre cette approche exige que des informations exhaustives et fiables sur le groupe avant l'implantation du programme, existent.

D'autres méthodes telles que les ajustements statistiques, le prétest, post test, etc... existent mais on fera l'économie de telles approches.

2. Sélection des critères ou d'indicateurs d'évaluation

En reférence avec les hypothèses élaborées, il s'agit en rapport avec les différentes variables identifiées de repertorier tous les indicateurs pouvant permettre de bien mesurer ces variables. Ces indicateurs doivent être objectivement vérifiables.

Si l'introduction d'une nouvelle variété de mil dans le cadre d'un projet agricole vise à améliorer les rendements à l'hectare, plusieurs indicateurs peuvent être identifiés. Parmi ceux-ci on peut noter :

- les quantités de semences / ha
- la densité du semis
- les quantités d'engrais utilisées
- le nombre de labour
- etc...

L'identification des indicateurs est une étape importante car elle permet réellement de guider le travail de collecte de données. Les questions qui figurent dans le questionnaire d'enquête sont d'ailleurs élaborées en fonction des différents indicateurs identifiés.

3. Les mesures d'évaluation

En faisant de la recherche on est toujours confronté à des problèmes de mesures. Contrairement aux sciences dites naturelles, mesure des indicateurs dans le cadre d'une recherche en science sociale, ne porte pas toujours sur des aspects physiques facilement mesurables. Il existe différents niveaux de mesures, abordés en détail dans d'autres chapitres. Ces niveaux de mesures seront simplement rappelés ici :

- niveau nominal : Il s'agit d'une simple classification (ex: typ de menage)
- niveau ordinal : En plus de la classification, il y a un certa souci d'ordre (ex: favorable / défavorable)
- niveau intervalle: Il y a une classification, un ordre et des intervalles mesurables entre différents niveaux de mesure (âge, taux de rentabilité, etc)
- niveau ratio : Presque identique au niveau intervalle à la différence qu'il existe un zero absolu dans c type de mesure.

4. Sélection d'une méthode d'échantillonnage (cf.chap.IV)

5. Décider de la méthode de collecte de données (cf.chap.VI)

PHASE 3 : CONDUITE DE L'EVALUATION

1. Tester les outils de collecte de données

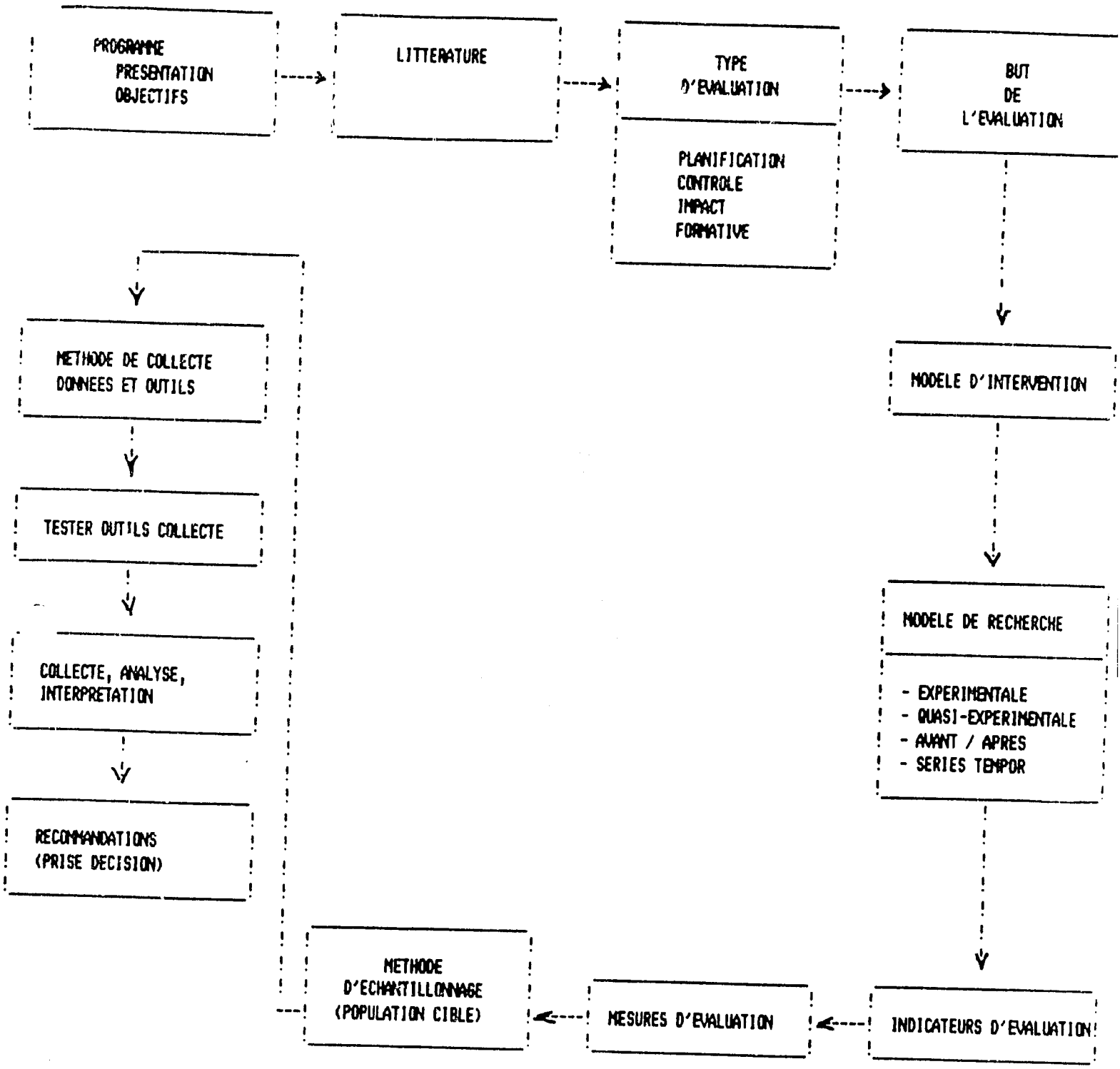
2. Collecter, analyser et interpreter les données

3. Faire des recommandations : L'objectif ultime de toute recherche est de permettre de prendre une décision rationnelle. Dans le cadre de l'évaluation, l'analyse des données recueillies devra permettre à l'évaluateur de faire des recommandations dans le sens :

- d'une amélioration du programme
- de l'expansion du programme
- de la reduction ou de la réorientation du programme
- ou même de l'arrêt pur et simple du programme
- etc...

Toutefois ces décisions seront toujours prises en rapport avec le type de programmes implantés. Il est par exemple très rare qu'une évaluation recommande l'arrêt d'un programme social, quelque soient les performances d'un tel programme. Plutôt, les moyens d'améliorer le fonctionnement de ce programme attireront l'attention de l'évaluateur.

SCHEMA D'UN PROCESSUS D'EVALUATION



140

Table C Areas under the normal curve

Fractional parts of the total area (10,000) under the normal curve, corresponding to distances between the mean and ordinates which are Z standard-deviation units from the mean.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0159	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2518	2549
0.7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3718	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4083	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1.6	4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4762	4767
2.0	4773	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4980	4980	4981
2.9	4981	4982	4983	4984	4984	4984	4985	4985	4986	4986
3.0	4986.5	4987	4987	4988	4988	4988	4989	4989	4989	4990
3.1	4990.0	4991	4991	4991	4992	4992	4992	4992	4993	4993
3.2	4993.129									
3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

SOURCE: Harold O. Rugg, *Statistical Methods Applied to Education*, Houghton Mifflin Company, Boston, 1917, appendix table III, pp. 389–390, with the kind permission of the publisher.

Table I Distribution of χ^2

df	Probability .99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	0.0157	0.0628	0.0393	0.0158	0.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

For larger values of df, the expression $\sqrt{2\chi^2} - \sqrt{2df - 1}$ may be used as a normal deviate with unit variance, remembering that the probability for χ^2 corresponds with that of a single tail of the normal curve. A slightly more complex transformation, which gives far better approximations for relatively small values of ν, is

$$Z = (\sqrt[3]{\chi^2} - \sqrt[3]{\nu}(1 - 2/9\nu))/\sqrt[3]{\nu}\sqrt{2/9\nu}$$

(See A. C. Acock and G. R. Stavig: "Normal Deviate Approximations of χ^2," *Perceptual and Motor Skills*, vol. 42, p. 220, 1976.)

SOURCE: Table I is reprinted from Table IV of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (1948 ed.), published by Oliver & Boyd Ltd., Edinburgh and London, by permission of the authors and publishers.

Best Available Document

Table B Random numbers

10 06 73 25 33	76 52 01 35 86	34 67 35 48 76	80 95 90 91 17	39 29 27 49 45
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02	00 82 29 16 65
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 47 64	35 08 03 36 06
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97	04 43 62 76 59
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77	12 17 17 68 33
66 06 57 47 17	34 07 27 68 50	36 69 73 61 70	65 81 33 98 85	11 19 92 91 70
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39	23 40 30 97 32
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47	18 62 38 85 79
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09	83 49 12 56 24
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44	35 27 38 84 35
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33	50 50 07 39 98
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01	52 77 56 78 51
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 54 06 10	68 71 17 78 17
88 68 54 02 00	86 50 75 84 01	36 76 66 79 51	90 36 47 64 93	29 60 91 10 62
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68	23 47 83 41 13
65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86	40 21 81 65 44
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53	14 38 55 37 63
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37	96 28 60 26 55
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90	94 40 05 64 18
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22	54 38 21 45 98
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23	37 08 92 00 48
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	53 14 03 33 40	42 05 08 23 41
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81	22 22 20 64 13
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39	28 70 72 58 15
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82	07 20 73 17 90
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93	42 58 26 05 27
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18	33 21 15 94 66
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92	92 92 74 59 73
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59	25 70 14 66 70
23 52 37 83 17	73 20 88 98 37	68 93 59 14 16	26 25 22 96 63	05 52 28 25 62
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35	65 33 71 24 72
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91	23 28 72 95 29
35 96 31 53 07	26 89 80 93 54	33 35 13 54 62	77 97 45 00 24	90 10 33 93 33
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92	78 56 52 01 06
46 05 88 52 36	01 39 09 22 86	77 28 14 40 77	93 91 08 36 47	70 61 74 29 41
32 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57	85 39 41 18 38
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23	97 11 89 63 38
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06	84 96 28 52 07
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 01 79 00 33	20 82 66 95 41
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56	05 01 45 11 76

SOURCE: The RAND Corporation, *A Million Random Digits*, Free Press, Glencoe, Ill., 1955, pp. 1–3, with the kind permission of the publisher.

Best Available Document

Table B Random numbers (*Continued*)

98 08 62 48 26	45 24 02 84 04	44 99 90 88 96	39 09 47 34 07	35 44 13 18 80
33 18 51 62 32	41 94 15 09 49	89 43 54 85 81	88 69 54 19 94	37 54 87 30 43
80 95 10 04 06	96 38 27 07 74	20 15 12 33 87	25 01 62 52 98	94 62 46 11 71
79 75 24 91 40	71 96 12 82 96	69 86 10 25 91	74 85 22 05 39	00 38 75 95 79
18 63 33 25 37	98 14 50 65 71	31 01 02 46 74	05 45 56 14 27	77 93 89 19 36
74 02 94 39 02	77 55 73 22 70	97 79 01 71 19	52 52 75 80 21	80 81 45 17 48
54 17 84 56 11	80 99 33 71 43	05 33 51 29 69	56 12 71 92 55	36 04 09 03 24
11 66 44 98 83	52 07 98 48 27	59 38 17 15 39	09 97 33 34 40	88 46 12 33 56
48 32 47 79 28	31 24 96 47 10	02 29 53 68 70	32 30 75 75 46	15 02 00 99 94
69 07 49 41 38	87 63 79 19 76	35 58 40 44 01	10 51 82 16 15	01 84 87 69 38
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74	64 27 85 80 44
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50	68 47 66 46 59
73 18 95 02 07	47 67 72 52 69	62 29 06 44 64	27 12 46 70 18	41 36 18 27 60
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71	93 82 34 31 78
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95	07 70 61 78 13
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53	31 22 30 84 20
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 04 75	94 11 90 18 40
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95	77 76 22 07 91
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14	83 48 34 70 55
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06	94 54 13 74 08
77 51 30 38 20	86 83 42 99 01	68 41 48 27 74	51 90 81 39 80	72 89 35 55 07
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51	65 34 46 74 15
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46	31 85 33 84 52
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91	08 00 74 54 49
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42	43 86 07 28 34
33 71 34 80 07	93 58 47 28 69	51 92 66 47 21	58 30 32 98 22	93 17 49 39 72
85 27 48 68 93	11 30 32 92 70	28 83 43 41 37	73 51 59 04 00	71 14 84 36 43
84 13 38 96 40	44 03 55 21 66	73 85 27 00 91	61 22 26 05 61	62 32 71 84 23
56 73 21 62 34	17 39 59 61 31	10 12 39 16 22	85 49 65 75 60	81 60 41 88 80
65 13 85 68 06	87 64 88 52 61	34 31 36 58 61	45 87 52 10 69	85 64 44 72 77
38 00 10 21 76	81 71 91 17 11	71 60 29 29 37	74 21 96 40 49	65 58 44 96 98
37 40 29 63 97	01 30 47 75 86	56 27 11 00 86	47 32 46 26 05	40 03 03 74 38
97 12 54 03 48	87 08 33 14 17	21 81 53 92 50	75 23 76 20 47	15 50 12 95 78
21 82 64 11 34	47 14 33 40 72	64 63 88 59 02	49 13 90 64 41	03 85 65 45 52
73 13 54 27 42	95 71 90 90 35	85 79 47 42 96	08 78 98 81 56	64 69 11 92 02
07 63 87 79 29	03 06 11 80 72	96 20 74 41 56	23 82 19 95 38	04 71 36 69 94
60 52 88 34 41	07 95 41 98 14	59 17 52 06 95	05 53 35 21 39	61 21 20 64 55
83 59 63 56 55	06 95 89 29 83	05 12 80 97 19	77 43 35 37 83	92 30 15 04 98
10 85 06 27 46	99 59 91 05 07	13 49 90 63 19	53 07 37 18 39	06 41 01 93 02
39 82 09 89 52	43 62 26 31 47	64 42 18 08 14	43 80 00 93 51	31 02 47 31 67

INSTRUCTIONS AUX APPLICATIONS INFORMATIQUES

Mercredi 22: La présentation des données (se référer à la page 42 chap. 11 pour application).

1) Insérer la disquette "**Program**" et presser "**Return**"

2) Taper "**Data**" et presser "**Return**".

3) Appuyer sur l'une des flèches verticales.

4) Entrer les données une à une en appuyant sur le clavier après chaque donnée.

5) Appuyer sur "**Esc**" et donner un nom à la variable. Ensuite appuyer sur "**Return**".

6) Attendre l'animateur pour d'autres instructions.

Jeudi 23: Problèmes de Contingence se référer au cas "Be sube".

1) Suivre instructions du mercredi 22 de 1 à 4.

2) Passer à une nouvelle colonne en appuyant sur l'une des touches horizontales en bas du clavier et entrer les données correspondantes. Répéter la procédure autant de fois qu'il y a de variables.

3) Presser "**Esc**" et donner un nom à chaque variable en appuyant sur "**Return**" après chaque nom assigné.

4) Attendre l'animateur pour d'autres instructions.

Vendredi 24: Mesure d'association se référer au cas de la page 99 chap. V.

1) Suivre les mêmes instructions que jeudi.

2) Attendre l'animateur pour de nouvelles instructions.

INSTRUCTIONS AUX APPLICATIONS INFORMATIQUES

Mercredi 22: La présentation des données (se référer à la page 42 chap. 11 pour application).

1) Insérer la disquette "**Program**" et presser "**Return**".

2) Taper "**Data**" et presser "**Return**".

3) Appuyer sur l'une des flèches verticales.

4) Entrer les données une à une en appuyant sur le clavier après chaque donnée.

5) Appuyer sur "**Esc**" et donner un nom à la variable. Ensuite appuyer sur "**Return**".

6) Attendre l'animateur pour d'autres instructions.

Jeudi 23: Problèmes de Contingence se référer au cas "Be sube".

1) Suivre instructions du mercredi 22 de 1 à 4.

2) Passer à une nouvelle colonne en appuyant sur l'une des touches horizontales en bas du clavier et entrer les données correspondantes. Répèter la procédure autant de fois qu'il y a de variables.

3) Presser "**Esc**" et donner un nom à chaque variable en appuyant sur "**Return**" après chaque nom assigné.

4) Attendre l'animateur pour d'autres instructions.

Vendredi 24: Mesure d'association se référer au cas de la page 99 chap. V.

1) Suivre les mêmes instructions que jeudi.

2) Attendre l'animateur pour de nouvelles instructions.

SEMINAIRE SUR

L' EVALUATION DES PROJETS

DE DEVELOPPEMENT

NOTIONS GENERALES SUR LES TECHNIQUES D' ENQUETE

MODULE 1

CONCEPTION ET SELECTION DES ECHANTILLONS

ANALYSE D'OUTILS D'ANALYSE SOCIALE SELECTIONNES

1. ENQUETES

Objectif

L'enquête échantillon est un condensé des informations qu'on retire d'une portion représentative de la population afin d'identifier et de mesurer la structure et les attributs de la population entiére.

Analyse

Les enquêtes sont une méthode expéditive pour rassembler un maximum d'informations à un cout minimal. L'ampleur de l'échantillon peut etre modelée aux fins de se conformer au temps et au budget disponibles. Il existe deux instruments communs de mesure que l'on peut utiliser: les questionnaires et les interviews.

L'analyste doit employer la plus grande partie de ses efforts á la conception de l'enguete. Cette derniére requiert une capacité à établir des échantillons qui soient représentatifs de la population et ce, au niveau de la rédaction des questionnaires et de l'analyses statistique des donées.

Il y a plusieurs considérations à faire lorsque l'on prend des décisions concernant un projet. Ces décisions sont:

a) le type de problème
b) le coût de l'enquête
c) le temps nécessaire pour obtenir les données, et
d) l'ampleur de l'échantillon nécessaire

Ces éléments doivent etre rassemblés lors des phases de planification du projet qui sont une période pendant laquelle on estime qu'une enquete est la méthode adéquate pour fournir les informations nécessaires aux fins d'élaboration et d'exécution satisfaisantes du projet

Définitions

1) Une population-échantillon est une série d'événements ou de gens qui possédent certaines caractéristiques, par exemple, toutes les femmes mariées d'un district d'une région.

2) Un échantillon est un sous-ensemble d'une population globale d'une zone délimitée et qui est reliée à une activité un à groupe spécifique.

Les attributs de ce sous-ensemble sont censés être les mêmes que ceux de la population entière en question.

3) Les attributs d'un projet sont les caractéristiques sociales, économiques et politiques du projet.

4) Un instrument de mesure ou de quantification est une méthode spécifique pour obtenir l'information qu'on désire à partir de quelqu'un.. Par exemple, ce sera un questionnaire qui pourra entrainer une réponse affirmative ou négative; cela peut etre également des formes variées d'accords ou d'avertissements (avec des nuances), une gradation d'un groupe de réponses possibles, etc...

5) Un recensement complet est une enquete de tous les événements ou de tous les membres d'une population donnée.

6) L'observation des participants.
C'est la collecte de l'information désirée relative aux modéles de comportement, aux rapports sociaux, aux pratiques culturelles ainsi qu'aux croyances et au statut économique; c'est également l'inter-action directe qu'on a avec un groupe (par le biais de cette collecte d'information) sur une certaine période de temps.

7) Cadre d''chantillonnage: La liste d'éléments à partir de laquelle un
---------- échantillon de probabilité est selectionné.

CAS D'ECHANTILLONNAGE No 1

LA COMPAGNIE DE TOMATES DE SONATAM

* * * * . * * * * * * * * * * *

Directives destinées à l'instructeur

1 - Ce cas doit être employé en tant qu'exercice préliminaire aux fins de permettre au seminariste de commencer à identifier quelques-uns des problémes majeurs attachés à l'échantillonnage, à savoir la représentativité (en termes de types, de caractéristiques géographiques etc...) et de taille ou ampleur de l'échantillon, de sélection aléatoire et autres caractéristiques liées à un "bon échantillon.

2 - Cet exercice doit être suivi d'une discussion et de l'utlisation du second cas d'échantillonnage qui est censé apporter une variété de dimensions sociales dans la conception de l'échantillon.

CAS D'ECHANTILLONNAGE No 1

LA COMPAGNIE DE TOMATES DE SONATAM

****

Instructions destinées au séminariste

---

1 - Lire le cas décrit plus bas. Essayez de décider quelles procédures vous utiliserez afin d'assurer la sécurité des produits que vous êtes chargé d'inspecter (limite de temps: 15 minutes).

2 - En travaillant en petits groupes, essayez de mettre au point une stratégie pour échantillonner les produits de cette compagnie (limite de temps: 20 minutes).

3 - Un membre de votre groupe sera chargé de présenter vos résultats à la classe.

LA COMPAGNIE DE CONSERVERIE DE LA SONATAM

Le Ministre de la Santé a loué vos services en tant qu'inspecteur de produits alimentaires tranformés. Votre première tâche est d'inspecter la qualité des tomates mises en conserve et vendues au Sénégal par la SONATAM, l'une des plus grandes productrices de tomates en conserve. La compagnie a des usines et des entrepots à St. Louis, Rufisque et Tambacounda. La saison de mise en conserve vient juste de se terminer et les conserves sont stockées dans les entrepots, dans l'attente d'approbation pour vente et distribution. En arrivant au Siége National de la Compagnie à Rufisque, on vous emmène à l'entrepôt principal de l'usine de Rufissque. Dans l'entrepot, il y a des milliers de caisses contenant les deux principaux produits de la compagnie, à savoir Tomate en pate et la sauce de tomate avec des oignons. Le Directeur de l'usine semble être trés coopératif. Il demande au magasinier de vous donner à vous l'inspecteur un carton de tomates en conserves. Il vous dit ensuite ce dont vous aurez besoin pour conduire l'inspection de leur production, c'est porter le carton à votre laboratoir de test de Dakar, ou voir une ou plusieurs des conserves contenues dans la caisse et entreprendre n'importe quel test que vous voudrez sur son contenu. De cette manière, vous pouvez etre rassuré de la qualité et de la sécurité du produit vendu par al SSONATAM à travers le pays.

Questions

Dans la sélection de l'échantillon de conserves à examiner, est-ce que vous, en tant qu'inspecteur, suivriez vòus la procédure suggérée par le Directuer ou voudriez-vous faire quelque chose d'autre?

Si vous optez pour quelque chose d'autre, qu'aimeriez-vous faire afin de vous rassurer, vous et votre service, de la qualité et de la sécurité de produits de tomate en conserve de la SONATAM?

154

Les types d'échantillon

1) Un échantillon au hasard simple est un échantillon qu'on tire d'une population et qui laisse à chaque membre de cette population des chances égales d'etre choisi.

2-a) Un échantillon au hasard stratifié sélectionne un nombre proportionnel à partir de chacun des groupes à l'intérieur d'une population totale particuliére; l'échantillon pourrait etre par exemple les paysans s'une paroisse. On sait que 10% de ces paysans ont des revenus bruts de 450.000 FCFA et plus par an, que 30% d'entre eux ont des revenus qui s'étalent de 250.000 FCFA à 449.990 par an et que 60% ont des revenus inférieurs à 250.000 FCFA par an. Si l'on prend la décision d'effectuer un échantillon au hasard stratifié de 100 paysans et qui soit basé sur le revenu annuel brut, on aura 10% qu'on choisira au hasard à partir du groupe des 450.000 FCFA et plus an, 30% de celui des 250.000 à 449.990 par an et 60% à partir du groupe à revenu bas.

2-b) Un échantillonnage aléatoire stratifié peut également être utile dans les cas où il existe de petits mais importants sous-groupes d'une population et qui ne pourraient etre adéquatement échantillonnés si l'on utilisait des procédures d'échantillonnage simple. Par exemple, si nous essayons d'analyser les attitudes des gens d'une région vis-à-vis d'un projet de détail donné et supposer que la population soit composée 50% de Volof 45% de Séréres et 5% de Peuls. Si l'on dresse un échantillon purement au hasard de 100 personnes, nous aurions à inclure 4 ou 5 interviews avec les peuls dans notre échantillonnage. Mais nous désirons etre à même de généraliser nos résultats pour chaque groupe éthnique. Nous serions peut-etre capables de faire cela pour les Wolofs et Séréres, mais 4 à 5 interviews seraient trop peu pour que l'on puisse généraliser sur les attitudes des Peuls. Nous aurons alors à stratifier notre échantillon (sur-échantillonner les Peuls), à savoir que novs pourrions dresser un échantillon aléatore séparé environ 30 membres de la population Peul. Nous serions alors capables de comparer les attitudes des trois différents groupes.

"La Fonction ultime de la stratification 'est d'organiser la population en sous-groupes homogénes (avec une hétérogénéité entre les sous-groupes) et de sélectionner le nombre approprié d'éléments consititutifs de chaque sous-groupe".

3-a) Un échantillonnage par grappes est une procédure de s'lection dans laquelle les éléments de l'échantillon sont choisis à partir de la population par groupes ou grappes plutot qu'individuellement. Les grappes utilisées sont souvent des regroupements pré-existants naturels ou administratifs d'une population tels que des blocs d'immeubles, des quartiers, des subdivisions politiques, des zones géographiques etc... L'idee fondamentale d'échantillonnage par grappes peut être illustré grâce à une étude hypothétique sur la santé d'enfants d'une communauté rurale de deux à six ans. La population est compossée de tous les enfants vivant dans les limites de la communauté rurale . Les chercheurs (enquêteurs) pourraient utiliser les archives de recensement, les livres de l'Etat Civil et les chefs de village, afin de préparer une liste de "tous" les enfants vivant dans la Communauté rurale. Il existera toujours des problémes évidents à mettre au point une liste complète et exacte. Un échantillon aléatoire simple pourrait alors etre etabli à partir de la liste de population. le résultat en serait cependant un échantillon étalé sur l'entiére communauté rurale, ce qui, dans certains cas, est une région très vaste. Dans une étude de santé, un tel échantillon pourrait s'avérer etre à la fois inconvenant et couteux, car il serait nécessaire de procéder à des examens médicaux, des tests de laboratioire et dans des villages trés dispersés.

L'alternative consisterait à prendre un échantillon de tous les villages et évaluer la santé de tous les enfants dans ces villages. Les villages dans cet exemple seraient des grappes, c'est-à-dire des regroupements de petits éléments sélectionnés ensemble. Cette procédure d'échantillonnage aurait l'avantage de réduire les coûts et de simplifier le travail sur le terrain. Simultanément, elle résultera probablement en des erreurs d'échantillonnage plus grandes que pour un échantillon aléatoire simple comprenant le même nombre d'éléments (ici des enfants).

3-b) Echantillonnage à progressions multiples

Une autre possibilité serait un compromis entre un échantillonnage aléatoire simple et de grandes grappes tels que des village entiers. La premiére progression ou étape serait de sélectionner au hasard des villages, avec un choix d'un plus grand nombre que lors de précédents échantillonnages par grappes. La deuxiéme progression ou étape consisterait à réaliser un échantillon de probabilité concernant les enfants au sein de villages sélectionnés afin d'arriver au nombre voulu d'éléments.

LIMITES

CONDITIONS

Ces conditions doivent être réunies lorsqu'une enquête (ou un sondage est effectué :

1 - On doit s'assurer qu'un échantillon au hasard peut être sélectionné et qu'il est représentatif de la population totale.

2 - On doit s'assurer que les critéres demandés ne sont pas partiaux.

3 - On doit également être sur que les conclusions que l'on tirera des données seront, d'un point de vue statistique, valables par rapport à la population concernée.

SONDAGES

Le sondage peut être utilisé pour:

1 - définir les caractéristiques, problémes, opinions et les attitudes d'un certain groupe sélectionné, par exemple les familles ayant des enfants en age d'aller à l'école, ou alors des enfants, des paysans, des méres de famille, des ouvriers industriels, ou des fonctionnaires ;

2 - quantifier les données :

a) donnant un un poids aux différentes réponses à une question et en réduisant les réponses à des pourcentages ; ou

b) en modelant les questions de façon à permettre à chaque répondant une gamme de choix. Voici un exemple de ce genre de questions :

"Quelle faveur connait la méthode de contraception DUI

A - Trés acceptable

B - Acceptable

C - Innaceptable

Afin de quantifier cela, on peut, soit réduire les réponses à un pourcentage des réponses totales, soit leur attribuer un poids numérique déternimé tel que 3 pour a, 2 pour b, et 1 pour c ; pour ensuite faire la somme de chacune des réponses pondérées.

L'enquête peut également être utilisée pour :

3 - Servir de base à une évaluation des facteurs sociaux qui affectent le projet et permettre ainsi aux planificateurs de projet de venir à bout de ces problémes. Cela peut entrainer la prise de décision de ne pas entreprendre le projet en question.

4 - Avoir une idée précise de l'efficacité du projet une fois achevé en mesurant son impact sur les groupes visés et les groupes non visés qui se trouvent dans la zone du projet (par exemple lorsqu'il s'agit de nouveaux emplois ou d'imposition supplémentaire etc...).

5 - Quantifier les données statistiques utilisées dans le diagnostic financier et économique des projets.

MODULE 2

ELABORATION D'UN QUESTIONNAIRE

ELABORATION D'UN QUESTIONNAIRE

****

ELABORATION D'UN QUESTIONNAIRE

****

Introduction

Les recherches d'enquête, lorsqu'elles sont systématiquement exécutées, ressemblent à la méthode expérimentale classique sur de nombreux plans. L'effort de base est conçu de façon à collecter des informations relatives aux attitudes, opinions, connaissances et perceptions. Dans ce sens, la nature des questions posées et le questionnaire sont des aspects cruciaux de l' effort de recherche et de collecte des données. La recherche d'enquête est basée sur l'une des notions les plus fondamentales de l'étude du comportement, à savoir le stimulus et la réponse.

S ----------> R

Question ----------> Réponse

Dans des situations expérimentales, nous appliquons un stimulus ou une série de stimulus à un sujet et nous essayons ensuite de mesurer la réaction au stimulus de la part du sujet expérimental. Par exemple, lorsque vous subissez un examen médical, le médecin peut tester vos reflexes en tapant sur votre genou à l'aide d'un maillet en caoutchouc. Le coup sur votre genou est le stimulus, le mouvement involontaire de votre jambe est la réponse. Au cours de la recherche d'enquête, il existe deux types fondamentaux de stimulus:

1) les questions et le questionnaire
2) l'interviewer

Dans ce bref aperçu, nous étudierons le premier type.

Chaque question doitetre traitée comme un stimulus pour lequel nous cherchons une réponse, celle de l'interviewé. Afin d'être à même de pouvoir faire quelque chose de sensé à partir des réponses des interviewés, il est important que nous nous assurions que le stimulus fourni par chaque question est exactement le même pour chaque personne interviewée. Si le stimulus varie d'une personne à une autre. les réactions (réponses) ne seront pas comparables parce que le stimulus aura été différent. Autrement dit, afin d'évaluer les réponses des gens, nous devons connaitre ce à quoi ils répondent.

Si l'on pose une question à l'interviewé "A" et une autre à l'interviewé "B", nous ne pouvons pas comparer leurs attitudes ou n' importe quelle autre information que nous aurions souhaité obtenir à partir de cette même question. Les points et exemples suivants sont censés aider à éviter ces erreurs.

1 - La question doit etre claire et non ambigue

Les questions doivent être trés claires et spécifiques. Souvent, ce qui peut sembler parfaitement clair au chercheur peut ne pas l'être pour l'interviewé. Par exemple si nous demandons: Que pensez-vous des nouvelles stratégies du gouvernement sénégalais? Il se peut que l'interviewé demande à connaitre de quelles nouvelles stratégies vous parlez. la question n'est pas vraiment claire. En outre, si le répondant "Y" dit qu'il pense que les stratégies sont trés bonnes, il se peut qu'il parle des nouvelles politiques ou stratégies agricoles. Le répondant "Y" peut également dire qu'il pense que les politiques sont trés bonnes, mais il se peut qu'il ait à l'esprit la politique étrangère. Ainsi, parce que la quesiton est à la fois floue et ambigue, nous ne pouvons pas vraiment interpréter les réponses. "X" et "Y" répondaient en fait à différents stimuli. Par conséquent, ces réponses ne sont pas comparables.

En plus, des mots trés imprécis tels que "souvent", "beaucoup", "quelques, doivent être évités parce qu'ils ne sont pas clairs et qu'ils peuvent donc pouvoir signifier des choses fifférentes à différents interviewés. Si nous nous intéressons de savcoir si les femmes en grossesse recherchent les soins médicaux cur une base reeguliéren nous pouvons leur demander si elles voient un docteur, souvent, parfois ou jamais. Un probléme peut surgir lorsque nous essayons d'interpréter les résultats. Alors qu'une personne donnée peut penser que le fait de rendre visite à un docteur une fois tous les deux mois est "souvent", un autre répondant pourra, lui, penser que rendre visite à un docteur toutes les deux semaines est "parfois". En posant une telle question, il vaudrait peut-etre mieux que l'interviewé choisisse à partir d'une série de réponses plus claires telles que "une fois par semaine", "une fois tous les quinze jours" etc...

2 - N'utilisez pas des questions à double sens

Quelquefois, lorsque nous essayons de poser une question, nous en posons, sans nous en rendre compte, deux ou plus en même temps. Une fois encore, nous rencontrons des problémes dans l'interprétation des résultats parce que nous ne savons pas à quel stimulus l'interviewé répond. Par exemple, nous pouvons demander aux répondants d'être en accord ou en désaccord avec l'affirmation auivante : le gouvernement du Sénégal doit stopper les investissements au sein des sociétés de développement régional et le gouvernement du sénégal doit fournir plus de fonds pour les CER. Si quelqu'un dit qu'il est en désaccord, nous ne savons pas exactement avec quoi il est en desaccord. Il esiste trois possibilités :

1) il est à la fois en désaccord avec le fait de cesser les investissements au niveau des sociétés de développement régional et avec le fait d'augmenter les fonds pour les CER;

2) il est en désaccord avec la cessation des investissements dans les sociétés de developpement régional mais il est d'accord avec le fait d'accroitre les ressources pour les CER; ou

3) il veut maintenir les fonds pour les "SRD" mais il n'est pas pour le fait d'accroitre les fonds pour les "CER".

3 - Les questions doivent être formulées d'une maniére impartiale

Dans la formulation des questions, on doit éviter de faire référence à des personnalités importantes ou à des autorités trés respectées, références qui suggéreraient la position prise par ces individus. Autrement, le répondant serait amené à répondre en accord avec la position prise par l'autorité plutôt que par rapport à sa propre opinion sur le sujet. Par exemple, la question posée concernant les opinions sur les nouvelles politiques du gouvernement au Sénégal comporte une telle partialité intrinséquemment. Les gens peuvent dire qu'ils soutiennent la politique simplement parce qu'ils savent que c'est la politique du gouvernement, même d'ils devaient être en désaccord avec cette politique si la question avait été posée autrment, grâce à une série d'options à partir desquelles ils pourraient chosir librement.

En plus, il existe des partialités associées à certains mots-clé et à ceraines phrases et qui peuvent influencer les réponses et les diriger vers une réponse particulière ou loin de cette derniére, plutôt que de nous donner une image exacte des attitudes des gens. Par exemple, si nous voulons connaitre la réaction des paysans par rapport aux projets d'un pays en voie de développement donné, nous ne poserons probablement pas la question de la facon suivante: "Est-ce vous pensez que les gens locaux ont des "inputs" appropiés au sein des projets de développement parachutés dans votre région par le PVD (Pays en Voie de Develpment) "X" ?" Les réponses seront automatiquement influencées par l'inclusion du terme "parachuté" dans la question. Ainsi notre stimulus est un terne partial plutot qsue la réaction à un type de projet.

4 - Les questions doivent être courtes

Même si les questions doivent être précises et claires, les interviewés perdent souvent de l'intérêt, ou alors oublient ce que l'on attend d'eux lorsque la question est trop longue. De courtes questions qui peuvent être posées repidement et auxquelles on peut répondre rapidement sont d'habitude préférables.

5 - Les questions négatives ne doivent pas être employées

Les questions négatives sont souvent mal interprétées par les répondants. Le résultat en est que beaucoup d'interviewés ratent ou interprétent mal le stimulus qui est communiqué par le chercheur. Les réponses sont donc interprétables parce que nous ne savons pas à <u>quoi</u> l'interviewé répond. Par exemple, nous pourrions demander aus interviewés s'ils sont en accord ou en désaccord avec l'affirmation suivante : "Le développment à la base" <u>n'est pas</u> une stratégie de développement viable. Nombre de personnes ne remarqueront pas la nature négative de la question et diront alors qu'ils sont en accord (plutôt qu'en désaccord), signifiant par là qu'ils pensent que le développement à la base est en fait une politique viable. Une fois de plus, les résultats seraient ininterprétables.

6 - Les traductions doivent être uniformes et doivent faire l'objet d'un consensus préalable

Lorsque les questions sont d'abord rédigés dans une langur, en français par exemple, et que les questions doivent être posées dans une ou plusieure langues nécessaires à l'interview des conseillers locaux ou paysans du Sénégal, une "meilleure" traduction qui aura fait l'objet d'un consensus préalable doit être produite avant de réaliser n'importe quelle interview. On ne doit jamais laisser le soin aux intervieweurs de faire les traductions sur le terrain. Deux traductions indépendantes de questions ou plus doivent être faites par des personnes dont la langue en question est la langue maternelle. Ces traductions doivent ensuite être comparées et parfaites et l'on doit s'assurer que les nuances incluses dans le questionaire original sont retenues dans les nouvelles versions. Les traductions des langues qui ne sont pas écrites par une grande majorité sont mieux réalisées grâce à des magnétophones. Les intervieweurs peuvent alors utiliser des cassettes pour apprendre le questionnaire. Lorsque les traductions ne sont pas faites à l'avance et lorsqu'elles n'ont pas fait l'objet de consensus préabable, les intervieweurs peuvent poser des questions trés différentes. Une fois de plus, des stimulus différents provoqueront une variété de réponses qui seront alors largement non interprétables.

7 - Questions "ouvertes" par rapport aux questions "fermées"

Des questions "ouvertes", c'est-à-dire des questions auxquelles le répondant fournit sa propre réponse, ont l'avantage de procurer une large variété d'informations, particuliérement pour ce qui est d'études exploratoires. Mais ils ont le désavantage amer d'être trés difficile à codifier et à analyser, une fois que l'enquête est terminée.

Les questions "fermées" -à savoir les questions pour lesquelles le chercheur fournit les réponses à partir desquelles le répondant doit choisir-ont l'avantage d'être plus faciles à poser et de fournir des données qui sont plus faciles à codifier et à analyser. La fragilité principale de ces questions surgit lorsque le chercheur n'a pas inclu toutes les réponses appropriées dans une question. Ce problème peut être résolu si l'on ajoute une réponse telle que "autre", "priére de spécifier", à beaucoup de questions.

ELABORATION D'UN QUESTIONNAIRE

Cas No1

ANALYSE SOCIALE SU PROJET D'AMENAGEMENT RURAL INTEGRE DANS LE DEPARTEMENT DE "LIGGEE"

****

Directives à l'instructeur

1. Le cas doit être utilisé à la suite de la lecture relative à l'élaboration d'un questionnaire et à la constitution d'échelles.

2. L'objectif fondamental de l'exercice est de fournir au participant une expérience pratique sur les problémes relatifs à la rédaction de questionnaires et à la facture des questions.

3. Chaque seminariste devra travailler individuellement pendant la première phase de l'exercice (limite de temps: 30 mn).

4. La deuxième phase de l'exercice consistera à un travail, en petits groupes de 3 à 5 participants. A ce moment, les questions individuelles seront regroupées et une série commune de questions sera établie. Chaque groupe sera alors appelé à présenter les résultats de son travail aux autres différents groupes pour critique. Les rôles de présentateur et de critique seront alors renversés (limite de temps: travail en groupe - 30 mn; présentation - 15 mn par groupe).

5. Une discussion globale devra alors être entreprise et les problémes majeurs concernant la construction de questionnaire devront être réaccentués à ce niveau.

Instructions au participant

1. Lire la briève description du cas.

2. En travaillant individuellement, établissez une série de questions (8-10), qui répondront aux besoins en matiére de données exprimés dans le cas.

3. Durant la seconde phase de l'exercice, vous travaillerez en petits groupes et essaierez de mettre au point un questionnaire qui aura l'approbation de tous. Les résultats de votre travail en groupes seront alors présentés à un autre groupe pour une critique constructive. Votre groupe pourra, à son tour, critiquer le travail d'un autre groupe. Aprés cette phase de critiques, on vous donnera une courte période pour améliorer vos questions.

ELABORATION D'UN QUESTIONAIRE

CAS No1
PROJET D'AMENAGEMENT RURAL INTEGRE DANS LE DEPARTEMENT DE LIGGEE.

****

Le Département de "Liggee" est situé dans une région qui a été extremement secouée par les problèmes de sécheresse et d'exode rural qui en a résulté.

Tradionnellement, cette zone trés hétérogéne sur le plan ethnique a toujours beaucoup dépendu de la culture des arachides, mais la production de bétail est également de quelque importance. La production de mil et autres céréales a décliné avec les années, à la fois à cause de la sécheresse et de la poussée des agents locaux à convaincre les paysans de produire des cultures commerciales. Un important project d'aménagement rural intégré pour le département de "Liggee" est actuellement en phase d'élaboration. Avant d'accepter de financer le projet, l'Institut pour le Développement Rural Intégré en Afrique (ITRDA) a demandé qu'une étude soit réalisée parmi les participants potentiels du projet afin de déterminer:

1. Les types de projets pour lesquels ils sont le plus intéressés.

2. Du point de vue de la participation au projet, les genres de groupes de travail qui existent et qui ont la préférence (individuel, collectif, etc...).

3. Comment les populations locales voient leurs problèmes de développement, particulièrement la nature de leur perception et des relations qu'ils entretiennent avec les agents locaux de gouvernement qui seront chargés de las aider à exécuter le projet.

Pour l'enquête, l'échantillonnage par grappes a été sélectionné en tant que méthode plus appropriée. Vingt-cinq villages ont été sélectionnés et

un échantillon aléatoire de la population adulte dans chacun des villages a été établi. Votre tâche est de concevoir l' instrument d'enquête nécessaire pour faire la collecte de ces données.

BAREMES // ECHELLES GRADUEE

Trés souvent, les résultats de l'analyse peuvent ne pas se prêter à une quantification à proprement parler, dans ce sens qu'il existe déjà une valeur absolue qui est intrinséquement attachée à certains aspects d'un projet. Cela n'empêche cependant pas l'analyste de faire une détermination quantivative qui sera basée sur des jugements absolus ou relatifs de la part des répondants à une série de questions préfabriquées -par exemple, ils se peut qu'on envisage d'implanter un projet au sujet duquel on se pose des questions quant aux réactions sociales et culturelles probables ainsi qu'aux réactions basées sur des considérations économiques, dans ce cas, l'on doit essayer de mesurer les réactions absolues et relatives quant aux aspects sociaux du projet. On peut utiliser les critéres à cet effet.

Objectif des barémes (critéres)

Les barèmes sont utilisés pour mesurer la reeaction absolue ou relative à un projet donné ou à un projet, soit sous la forme d'un rejet absolu, ou d'une acceptation totale, soit sous la forme d'un niveau relatif au jegement, c'est-à-dire une classification numérique d'une réaction par rapport à plusieurs options ou possibilités données.

Utilisation des barémes

Les barémes sont employés pour quantifier des jugements subjectifs, soit en forme absolue, soit sur une base relative. Un exemple de forme absolue de jugement est une question qui offre deux choix, c'est-à-dire "oui" ou "non", "acceptable" ou "inacceptable". Une fois que ces jugements

sont faits, alors ils peuvent être quantifiés, à savoir par exemple 70% de personnes ont répondu "Oui" et 30% "Non". Les choix sont alors quantifiés et peuvent être utilisés dans les analyses.

Il y a trois genres de barémes qui peuvent être utilisés :

* Les échelles nominals indiquent la portion relative d'un groupe de possibilités basées sur la réponse des membres d'une population à plusieurs choix qu'ils peuvent faire sur une échelle qui comporte un jugement relatif ou absolu. Cela permet à l'anayste de quantifier les facteurs qui peuvent être hautement subjectifs ou qualitatifs et de les incorporer dans l'analyse, ce qui sera utile au premeur de décision.

* Les échelles ordinales sont utilisées pour classer par ordre d'importance une série d'objet similaires selon une échelle de critéres qui prend en compte les traits ou caractéristiques comparatifs des objets tout en n'indiquant pas le degré de différence. Ces échelles montrant seulement que quiques-uns de ces objets excédent ou dépassent d'autres lorsque l'on considère tous ces critéres et cela, en termes de "bon", "mieux" et "excellent". Par exemple, trois possibilités de projet peuvent être examinés et on demandera aux répondants de les classer selon un ensemble de critéres. Ils doivent les classer en ternes de "bon", "mieux" et "excellent". S'il y a 40 répondants qui classent les trois options de projet de façon qu'une de ces options ait une majorité de réponses qui indiquent "excellent", on classera cette option en tant que "noméro 1". l'option qui a le plus de réponses de"excellent ou "mieux", mais en revanche, plus de "mieux" que d' "excellent" est classée seconde etc...

* Les échelles à intervalles reflétent le rang d'un facteur par rapport à un autre et également le degré dépassement d'un certain facteur par rapport à un autre. La différence entre les facteurs correspond à la distance (qui existe) sur l'échelle qui se trouve entre eux, ce qui les différencie. Par exemple les coûts financiers, pour couvrir les besoins de trois projets, s'étalent de 0 à 800.000, mais le coût du projet A est de 610.000, celui de projet B de 500.000, et celui du projet C de 550.000. Sur une échelle à intervalle, le projet B serait le plus bas de

50.000, par rapport à C, et en deuxième lieu par ordre d'infériorité, viendrait C par rapport à A et ce, de 60.000 F. L'échelle à intervalles n'indique pas seulement la portion relative sur une échelle, ce qui représente le coût de chaque option de projet, mais également l'lintervalle qui existe entre les projets que l'on quantifie.

Les échelles de gradation permettent une comparaison quantitative en rédusant les réponses à un ordre de grandeur, un taux ou une affirmation numériquement définissable, ce qui accroit la possibilité de prise de décision. Ces différentes échelles de grandeur sont indiquées à l'illustration 1 et donnent un exemple des résultats obtenus dans une enquête qui essayait de déterminer l'acceptablilité de diverses méthodes de contraception.

Dan cet exemple des trois genres d'échelles de classification indiquées à l'illustration 1, l'échelle nominale a été utilisée pour les réponses à la question:

Quell est votre réaction envers ces sept méthodes de contraception?
(indiquez si favorable ou d'éfavorable)

1.	Abstinence	Favorable	----------
		Défavorable	----------
2.	Méthode rythmique	Favorable	----------
		Défavorable	----------
3.	Le matin suivant la prise des pillulues	Favorable	----------
		Défavorable	----------
4.	Méthode intra-utérin	Favorable	----------
		Défavorable	----------
5.	Préservatif	Favorable	----------
		Défavorable	----------
6.	Pillule	Favorable	----------
		Défavorable	----------

170

L'échelle ordinaire fut employée lorsqu'on donnera aux interviewés la même liste de méthode de contraception possible et qu'on leur demanda de les classifier de "1" à "7" étant la note la plus élevée ou préférée, et "7" la plus basse ou la moins préférée.

Par exemple : Prière de classifier les méthodes suivants de contraception selon vos préférences d'emploi. Classez-les de 1 à 7, (1 étant le maximum, et 7 le minimum).

( ) Abstinence

( ) Méthode rythmique

( ) le matin suivant la prise de la pillule

( ) Méthode intra-utérin

( ) Préservatif

( ) Pillule

( ) Stérilisation.

Le troisième genre d'échelle de classification qui peut etre utilisé est l'échelle à intervalles. Lorsque cette dernière échelle est utilisée, on donne au répondant l'éventail de choix où il peut puiser (dans ce cas précis, les 7 méthodes de contraception) et on lui demande de les placer sur une échelle de valeur relative, ainsi qu'indiqué à l'illustration 1. Dans ce cas, une échelle de 10 à 1 est employée, 10 étant le maximum et 1 le minimum. L'échelle indique le rang de même que les distances qui existent entre deux méthodes données (degré qui existe entre le fait qu'une méthode est préférée à une autre).

Procédure d'emploi des échelles de classification :

1. Déterminer les facteurs à classifier.

2. Elaborer le rythme de collecte des données afin que les répondants fournissent les données sous une fois qui soit utilisable.

3. Déterminer la meilleure échelle pour classer les différents facteurs. Cela dépendra du point de vue de l'analyste quant au degré de jugement subjectif dont les répondants sont capables, cela étant basé surtout sur leur connaissance des facteurs.

4. Mettre les données disponibles sous forme de tableaux et les rassembler .

5. Construire une échelle de classification qui pourra être utilisée afin de résumer graphiquement et sous une forme quantifiée, les données.

6. Mettre les données une fois rassemblées dans les échelles de classification.

ILLUSTRATION NO 1
****

Exemples d'échelles des valeurs (graduées) utilisées pour servir à distinguer entre les facteurs et à quantifier ces dernières par rapport à des méthodes de contraception

Echelle nominale	Echelle ordinale	Echelle à intervalle
Favorable Méthodes Défavorable 6, 1, 5, 4	Premier choix : 4 Deuxième " " : 3 Cinquième " " : 7 Sixième " " : 2 Septième " " : 5	10 Stériliser 9 6 5 D U I 4 3 Matin suivant la prise de la pillule 2 Posologie 1 Abstinence

Les nombres identifient les différentes méthodes de contraception

173

ELABORATION D'UN QUESTIONNAIRE

CAS No 2

Directives à l'instructeur

Cet exercice est destiné à donner au participant une meilleure idée de l'importance de la traduction des questions et de la production et traduction d'un instrument d'enquete qui aura fait l'accord de tous avant que l'on commence le travail d'enquete à proprement parler. En plus de l'exemple inclu dans l'exercice, de nombreux autres exemples auxquels vous êtes familier peuvent être ajoutés pour compléter le cas. Les éléments du "Type Likert" avec des gradations dans l'accord et dans le désaccord servent de bons outils de test pour la traduction.

ELABORATION D'UN QUESTIONNAIRE

CAS No 2

Instructions aux participants

1. Lire le bref cas décrit plus loin
2. Identifier le problème et répondre aux questions qui sont à la suite du cas.

CAS No 2

LES OPINIONS DES PAYSANS DE LA COMMUNAUTE RURALE DE "DIAKHASSE"

Le Ministère de l'Intérieur essaye de déterminer comment les populations locales voient le rôle des conseillers ruraux. Il a requis l'une des Ecoles Nationales de mettre au point une étude qui lui fournira quelques réponses. Après avoir conçu un questionnaire et avoir formé des enqueteurs, Mr. Amadou Sy, le Directeur de l'étude décida d'entreprendre un test du questionnaire avant de réaliser l'enquete proprement dite. La Communauté Rurale de "DIAKHASSE", qui, en de nombreux points, est représentative des Communautés Rurales qui existent dans le pays, fut sélectionnée pour servir d'essai à l'instrument d'enquete. Trois enqueteurs sont envoyés pour interviewer chacun des dix paysans qui ont fait l'objet de sélection aléatoire. Lorsque les résultats du test furent examinés, l'on découvrit qu'il y avait une grande différence entre les enqueteurs, dans les réponses aux mêmes questions. Par exemple, à l'une des questions concernant le rôle du conseiller en tant que représentant de l'opinion locale (voir la question plus loin), l'intervieweur "A" a trouvé que 90% étaient en faveur du premier rôle et 10% en faveur du deuxième rôle. Les enqueteurs "B" et "C" ont, quant à eux, trouvé que 20% étaient en faveur du premier rôle et 80% en faveur du deuxième rôle. Chacun des enqueteurs avait exactement la même question en français et tous avaient comme langue maternelle le Wolof. Les interviewés étaient tous des locuteurs wolofs. Le directeur de l'étude devait décider que cela n'est pas seulement la résultante de différences liées à l'échantillonnage aléatoire, mais que cela devait être dû à différents stimulus appliqués aux interviewés de la part des intervieweurs. Comment pouvez-vous expliquer cette différence, dans les résultats de son test de questionnaire ?

La question suivante a été utilisée afin de déterminer les opinions de la population, quant au rôle du conseiller :

Supposons qu'il y ait deux conseillers dans une communauté rurale comme la votre. Le premier essaie de déterminer exactement ce que veulent les gens qui l'ont élu et plaide pour un budget de la communauté rurale qui reflète les désirs des électeurs et ceci même si personnellement il n'est pas d'accord avec leurs priorités.

Le second conseiller estime qu'il doit utiliser son propre jugement et ne pas se fier à l'opinion publique. Il milite pour un budget de la communauté rurale qui, selon lui, répond le mieux aux besoins et aux intérets de la population, même si ses priorités à lui ne concordent pas avec celles de la population ;

Selon vous, Quel est le meilleur conseiller ? (encerclez)

1. le premier
2. le second
3. il ne sait pas.

Questions

1- En travaillant individuellement, traduisez la question de la page précédente en votre langue maternelle.

2- Lorsque l'instructeur vous en donne l'ordre, posez votre question à l'un de vos camarades participants. Ensuite l'un des autres participants vous pose sa question.

3- Comparez les traductions de ces questions.
Qu'est-ce qui aurait pu provoquer les résultats trouvés dans le test original du questionnaire?

4- Essayez de traduire la question suivante et les réponses dans votre langue natale.
Le gouvernement n'est pas concerné par les gens comme nous (encerclez)

1. entièrement d'accord 2. d'accord 3. pas d'accord
4. désaccord total 5. il ne sait pas

5. Comparez les traductions de cette question, surtout les nuances dans les réponses.

CAS D'ECHANTILLONNAGE No 2

CONSEILLERS DE COMMUNAUTE RURALE

****

Directives detinées à l'instructeur

---

1 - Ce cas doit être utilisé en qualité de suivi apporté au cas de la Compagnie de Tomates de la SONATAM. Entre l'examen des deux cas, on devra analyser en détail la fonction fondamentale de l'echantillonnage ainsi que les critéres qui y sont associés. Une brochure on un autre matériel de lecture adéquat concernant l'échantillonnage devra précéder l'analyse de ce cas.

2 - L'exercice devra initialement être réalisé individuellement pour ensuite être conduit en petits groupes. Le travail en petits groupes devra être suivi d'une discussion en classe.

3 - Cet exercice conduit aux exercices traitant la conception d'un instrument d'enquête.

CAS D'ECHANTILLONNAGE No2

CONSEILLERS DE COMMUNAUTE RURALE
****

Instructions aux seminaristes
---

1 - Lire le cas décrit ci-dessus. En vous basant sur votre connaisance des techniques d' échantillonnage et du milieu rural, élaborez une série de procédures pour dresser un échantillon des conseillers ruraux à interviewer.

2 - Après avoir travaillé individuellement sur ce problème, l'on vous divisera en petits groupes afin que vous mettiez au point une stratégie finale de recherche. L' on vous demandera de présenter les résultats des délibérations de votre groupe sur un papier -graphe.

Questions

1 - En utilisant les données fournies ci- dessus, votre connaissance du milieu rural du Sénégal et votre compréhension des problèmes d'échantillonnage , mettez au point une conception d'échantillonnage afin de répondre aux problèmes et questions posés ci-dessus. Assurez vous de garder à l'esprit que les ressources disponibles pour cette enquête sont limitées et que cette dernière doit donc être réalisée de la manière la plus efficiente possible.

2 - Dans l'accomplissement de la présentation de votre groupe, soyez prêt à défendre et à ajuster les choix que vous faites dans votre conception d'échantillonnage.

REVISION

PROCEDURE A SUIVRE POUR PREPARER ET FAIRE UNE ENQUETE

Il y'a un certain nombre de procédures qui sont nécessaires:

Echelon 1 : Il faut décider pourquoi l'enquête est entreprise et les résultats qui sont censés être engendrés par elle.

Echelon 2: Décider de la population qui fera l'objet de l'enquête en accordant l'attention nécessaire sur ce que l'on attende de l'enquête et à la coopération probable ainsi qu'à l'accessibilité de la population sélectionnée pour l'enquête.

Echelon 3: Choisir la méthode d'échantillonnage à utiliser, par exemple: échantillonnage au hasard, échantillonnage stratifié etc...

Echelon 4: Déterminer la meilleure ampleur à donner à l'échantillon, aprés avoir pris en considération:

a) la vraissembiance et la représentativité souhaitées,

b) le coût de la collecte des donnés. Plus l'enquête est importante plus chère elle sera;

c) le coût de traitement des données; et

d) les variations dans la population ainsi que les données obtenues l'indiquent.(b)

LE PROJET DE FORMATION DE CONSEILLER RURAL

On vous a chargé d'étudier la faisabilité d' initier un projet destiné à améliorer la qualité des éfforts de developpement local par le biais de cours de formation à court terme pour les conseillers de communauté rurale. Il a été décidé qu' une enquête - faisant partie de l'étude de faisabilité du projet - et portant sur les conseillers ruraux - doit être ensuite évaluée la proposition de projet. Les données dont on a besoin traitant de leur "background" , expérience, attitudes, compréhension des opérations de la communauté rurale et leurs perceptions de leurs besoins de formation.

Il y'a quelque 317 communautés rurales au Sénégal et environ 5500 conseillers élus aux deux tiers et choisis par les coopératives au tiers. Les conseils ont des tailles qui varient selon la population de la communauté rurale (<5000=12 conseillers, 5001 à 10 000= 15 conseillers; 10 001 à 15000 = 18 conseillers; plus de 15 000 = 21 conseillers). La réforme créant les conseils a été instituée à différentes périodes dans les différentes régions (Thies, 1972; Kaolack et Fatick, 1974; Diourbel et Louga, 1976; Ziguinchor et Kolda, 1978; St louis, 1980; Tambacounda, 1982; Dakar, 1984). Le nombre de communautés rurales varie selon les régions:

Thies.....31
Kaolack.....41
Fatick.....35
Diourbel.....32
Louga.....48
Ziguinchor.....25
Kolda 43
St Louis.....28
Tambacounda.....32
Dakar.....2

Echelon 5: Décider de la procédure qui sera utilisée pour faire la collecte des données par exemple questionnaire et interview personnels, questionnaire envooyé par la poste et interview téléphonique avec questionnaire etc...

Echelon 6: Familiariser les enquêteurs avec la technique utilisée.

Echelon 7: Elaborer le questionnaire et faites en un test préliminaire.

Echelon 8: Mettre au point la conduite de l'enquête y compris les arrangements pour le transport.

Echelon 9: Vérifier le travail des enquêteurs en contactant plusieurs des interviewés pour savoir si vraiment le contact a eu lieu. La "vérification sur le tas" et la surveillance sont nécessaires.

Echelon 10: Commencer à analyser les résultats dès que possible et remodeler le questionnaire et les méthodes d'interviews si nécessaire.

Echelon 12: Mettre autant de matériaux que possible sous une forme illustrée, c'est à dire sous une forme graphique et tabulaire, sous forme d'histogrammes etc... Calculer les pourcentages et faire l'analyse tabulaire nécessaire.

Echelon 13: Faire les tests statistiques nécessaires afin de permettre aux utilisateurs de tirer des conclusions sur la vraissemblance des résultats.

Echelon 14: Résumer les données et developper les conclusions et implications des données qui doivent être utilisées par les planificateurs et les exécutants du projet . Inclure le questionnaire et la procédure d'échantillonnage dans les annexes ainsi que dans un résumé des réultats de l'enquête.

SEMINAIRE SUR LES METHODES de RECHERCHE

A. PRESENTATION de DONNEES

tableau de distribution de fréquences

1- Ordonner les scores
2. Determiner les scores extremes et l'ecart entre eux.
3. Determiner une amplitude d'intervalle
4. Definir et choisir les limites de classes utiles
5. Enregistrer les scores dans les intervalles correspon

B. MESURES de TENDANCE CENTRALE et de DISPERSION

BI. Population

BI1. tendance centrale (la moyenne)

* données non groupées :

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

* données groupées :

$$\mu = \frac{\sum_{i=1}^{m} f_i c_i}{N}$$ où

et f
du

BI2. Dispersion (la variance)

* données non groupées : $\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$

* données groupées $\sigma^2 = \frac{\sum_{i=1}^{m} f_i (x_i - \mu)^2}{N}$

B II. Echantillon

B II 1. tendance centrale (moyenne)

* données non groupées :

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

* données groupées

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{n}$$

B II 2. Dispersion (variance)

* données non groupées

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

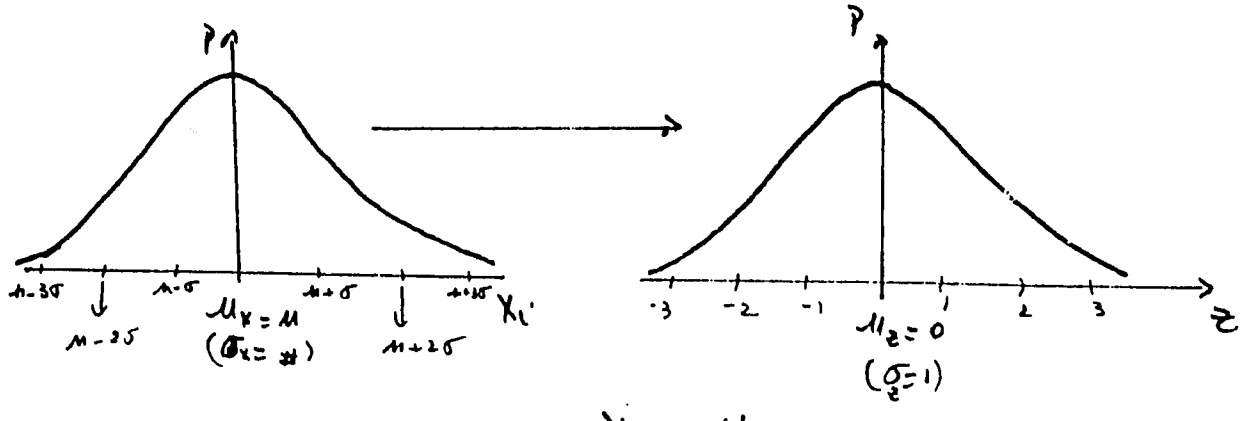
* données groupées

$$s^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{n-1}$$

C. ESTIMATION et TEST d'HYPOTHÈSES

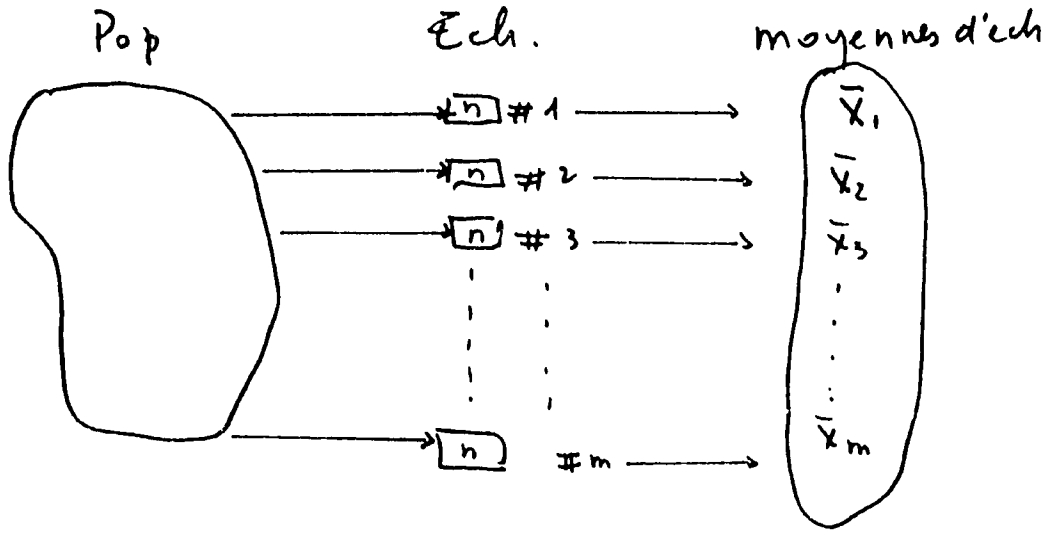
théorie des Grands Echantillons

C I. distribution d'observations



$$Z_X = \frac{X_i - \mu}{\sigma}$$

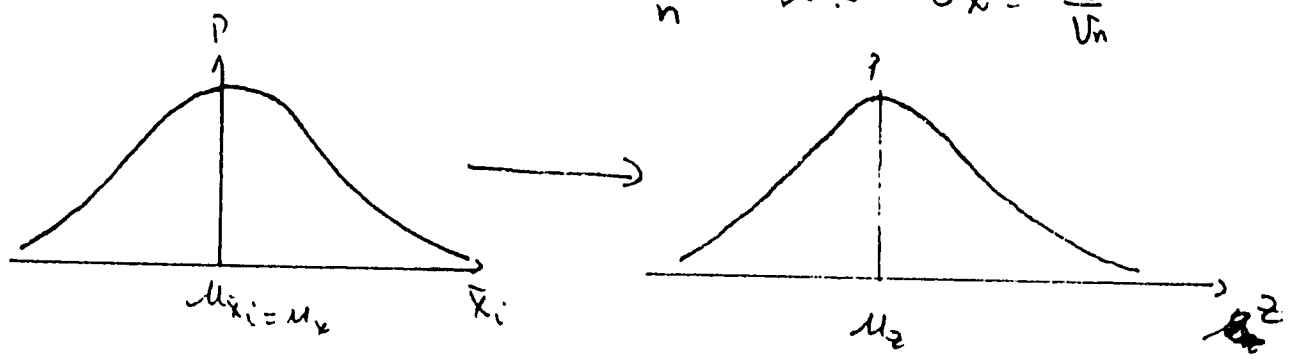
CII distribution de moyennes d'échantillon



$\bar{X}_i$ = Variable aleatoire

$\mu_{\bar{X}} = \mu_{(x)}$ ($\bar{X}$ est dit estimateur sans biais de $\mu_{(x)}$)

et $\sigma^2_{\bar{X}} = \dfrac{\sigma^2_{\bar{x}}}{n}$ i.e. $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$



$$Z_{\bar{x}} = \frac{\bar{X}_i - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X}_i - \mu}{\sigma/\sqrt{n}}$$

<u>Remarque 1</u> . X_i normalement distribués $\rightarrow$ $\bar{X}_i$ norm. distribuées

<u>Remarque 2</u> : Au fin et à mesure que $n \nearrow$, la distribution des $\bar{X}_i \sim N$ (théorème central lim

D- PROBLEMES de CONTINGENCE

DI. TEST d'INDEPENDANCE

H_0 : pas de relation

$$X^2 = \sum \frac{(O_i - e_i)^2}{e_i}$$

où O_i = f. observées
e_i = fréquences attendues

DII Mesures d'Association

* niveau nominal

$$\phi = \sqrt{\frac{X^2}{n}}$$

n = nombre d'observations

$0 \leq \phi, V \leq 1$

tab sup (2x2) $V = \sqrt{\frac{X^2}{nt}}$ où t = minimum de $(L-1)$ et $(C-1)$

* niveau ordinal

Gamma $\gamma = \frac{N_r - N_s}{N_r + N_s}$ formule générale

$$Q = \frac{AD - BC}{AD + BC}$$

	V1	
V2	A	B
	C	D

Spearman's

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$-1 \leq \rho \leq 1$
γ, Q