INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ECONOMIQUE

LA STATISTIQUE ET SES APPLICATIONS DANS UN PAYS EN VOIE DE DEVELOPPEMENT

par M. G. THEODORE

Administrateur à l'I.N.S.E.E.

TOME I

Année Scolaire 1962-1963

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES

ECOLE NATIONALE DE LA STATISTIQUE

ET DE L' ADMINISTRATION ECONOMIQUE

LA STATISTIQUE ET SES APPLICATIONS DANS UN PAYS EN VOIE DE DEVELOPPEMENT

par

M. G. THEODORE

Administrateur à l'I.N.S.E.E.

T O M E I

Année Scolaire
1962-1963

LA STATISTIQUE ET SES APPLICATIONS DANS UN PAYS EN VOIE DE DÉVELOPPEMENT

T O M E I

S O M M A I R E

			Pages
Chapitre I	-	GÉNÉRALITÉS	1
Chapitre II	-	UNITÉ STATISTIQUE ET MÉTHODES STATISTIQUES	23
Chapitre III	-	LE QUESTIONNAIRE	41
Chapitre IV	-	LA MISE EN ORDRE DES OBSERVATIONS. LE TABLEAU STATISTIQUE	63
Chapitre V	-	LES MÉTHODES GRAPHIQUES	83
Chapitre VI	-	LES CARACTÉRISTIQUES DES DISTRIBUTIONS	99
Chapitre VII	-	LES SÉRIES CHRONOLOGIQUES	119
Chapitre VIII	-	LES INDICES	141
Chapitre IX	-	HASARD ET PROBABILITÉ	167

TABLE DES GRAPHIQUES

DIAGRAMMES ..... 85 et 87

VARIATIONS SAISONNIERES 121

COVARIATION ET CORRÉLATION 127

o

o o

LA STATISTIQUE ET SES APPLICATIONS DANS UN PAYS EN VOIE DE DÉVELOPPEMENT

TABLE DES MATIERES

Pages

Chapitre I - GÉNÉRALITÉS I

Chapitre II - UNITÉ STATISTIQUE ET MÉTHODES STATISTIQUES 23

Chapitre III - LE QUESTIONNAIRE 41

Chapitre IV - LA MISE EN ORDRE DES OBSERVATIONS. LE TABLEAU STATISTIQUE 63

Chapitre V - LES MÉTHODES GRAPHIQUES 83

Chapitre VI - LES CARACTÉRISTIQUES DES DISTRIBUTIONS 99

Chapitre VII - LES SÉRIES CHRONOLOGIQUES 119

Chapitre VIII - LES INDICES 141

Chapitre IX - HASARD ET PROBABILITÉ 167

Chapitre X - LES SONDAGES 187

Chapitre XI - STATISTIQUES DÉMOGRAPHIQUES - GÉNÉRALITES - STATISTIQUES D'ETAT.. 219

Chapitre XII - STATISTIQUES DÉMOGRAPHIQUES - MOUVEMENTS DE POPULATION 249

Chapitre XIII - STATISTIQUES DÉMOGRAPHIQUES MOUVEMENTS DE POPULATION (fin)..... 283

Chapitre XIV - LES STATISTIQUES DÉMOGRAPHIQUES DANS LES RÉPUBLIQUES - LA POPULATION IMMIGRÉE - LA POPULATION ORIGINAIRE.. 309

Chapitre XV - LES STATISTIQUES DÉMOGRAPHIQUES - L'ACTION MENÉE ACTUELLEMENT POUR UNE MEILLEURE CONNAISSANCE DES POPULATIONS DES PAYS D'OUTRE-MER 335

		Pages
Chapitre XVI -	LES STATISTIQUES DU COMMERCE EXTERIEUR	363
Chapitre XVII-	PRIX ET INDICES DE PRIX	383
CONCLUSION		417

TABLE DES GRAPHIQUES.

DIAGRAMMES	85 et 87
VARIATIONS SAISONNIERES	121
COVARIATION ET CORRELATION	127
PYRAMIDE DES AGES	229
TAUX DÉMOGRAPHIQUES	291

=:=:=:=:=

Chapitre I

GÉNÉRALITÉS

1.1 Préambule.

1.2 L'ignorance ou le scepticisme du grand public.

1.3 La statistique. Approche générale.

1.4 Historique rapide.

1.5 La coopération.

1.6 L'enregistrement continu des observations.

1.7 Complexité de la connaissance et conclusion.

"Quand on est mal renseigné
on ne peut que mal raisonner".

Louis XIV. Mémoires.

1.1 Préambule.-

Depuis quelques amées, les Instituts, les cours, les conférences foisonnent : ils visent soit à l'explication des problèmes concernant les pays sous-développés, soit à bâtir ou à repenser les concepts économiques nécessaires à ces pays. Beaucoup d'idées circulent, beaucoup de théories sont émises puis réfutées, mais on constate malheureusement que les matériaux de base font souvent défaut et que de brillantes controverses ne sont souvent que des jeux de l'esprit.

Cet état de choses a pour raison profonde le manque de renseignements chiffrés valables, cette absence de documentation a dans certains cas des excuses, mais en général elle est le fait d'un manque d'adaptation des méthodes de recherches mises en oeuvre dans les pays économiquement développés. Le but réel du présent document est de préciser les applications de la méthode statistique pour des pays en voie de

développement. On s'apercevra rapidement que si sur le plan technique cette méthode reste unique quels que soient ses domaines d'application, les concepts parfois, les approches et les relevés diffèrent souvent selon que l'on se trouve en Europe ou en Afrique.

1.2 - L'ignorance ou le scepticisme du grand public.-

Il ne s'agit donc pas d'exposés théoriques ou de recommandations formelles : c'est un contact, une rencontre au cours desquels les difficultés journalières seront éclairées sous une optique nouvelle.

Mais tout de suite, il faut dire combien l'embarras est grand lorsque l'on veut convaincre un auditoire de l'intérêt et de l'importance de la statistique : l'opinion publique est dans notre pays rebelle au chiffre.

En effet, le Français, ne disons pas le Français Moyen, car on pourrait penser que l'on commence déjà à faire de la statistique, le Français si facilement conquis et donc pénétré de respect pour toutes les techniques et toutes les sciences fussent-elles les plus élaborées : telle la physique nucléaire ou la microscopie électronique, ce français n'a qu'une très vague idée de ce que l'on appelle la statistique. La presse ne ménage pas ses sarcasmes, les articles ironiques des journaux sont légion et il n'est pas rare d'en trouver de la veine suivante :

" Je n'ai rien contre la statistique, c'est le gagne pain des statisticiens lesquels font partie de notre standing national. J'ignore seulement s'il faut y voir les obscurs pionniers de la cité future ou les peigneurs de girafe du secteur public." (1).

Si le lecteur, animé de la meilleure volonté, s'arme de courage, malgré ces fâcheuses prémisses, et tente de pénétrer les arcanes de la documentation statistique ou économique, il découvre des documents compacts, bourrés de tableaux de chiffres, arides par leur aspect général et par leur contenu et vite découragé il les referme à jamais ayant senti combien la communication est malaisée entre le tableau de chiffres et le lecteur.

(1) Les imbéciles dans la statistique. Jacques PERRET. Arts - 13 Avril 1955.

Ou bien encore, une question précise étant d'actualité, s'il tente par ses lectures personnelles de se forger une opinion, il peut lire des articles signés par des informateurs également sérieux, truffés de chiffres identiques, qui défendent des théories parfaitement contradictoires. Dans ce cas encore, le doute saisit le chercheur et il s'aperçoit vite, pour reprendre une phrase prononcée durant une conférence sur le problème du vieillissement de la population qu' "un chiffre est une matière délicate : si on le torture, il avoue tout ce que lui demande son bourreau. Mais cet aveu arraché est sans valeur car il se rétracte aussitôt" (1).

Le chiffre peut être sollicité de mille manières : les commentaires qui accompagnent les résultats d'une élection en sont un exemple : pour les uns, l'accent sera mis sur des gains en pourcentage, pour d'autres ce seront les chiffres absolus qui seront évoqués.

C'est sur ce cas extrêmement classique que l'on peut comprendre les deux étapes de l'analyse numérique : au cours de la première, il s'agit simplement d'une élaboration anonyme, tandis que la seconde est toujours une prise de position personnelle et souvent passionnelle. On doit alors citer une expression significative d'un inspecteur général de l'I.N.S.E.E. (G. BOURNIER) faisant allusion à ces utilisations abusives de la documentation chiffrée :

" La statistique ne ment qu'aux mains du profane".
On reviendra souvent sur ces deux points fondamentaux :

- le responsable de la naissance du chiffre doit être d'une honnêteté intellectuelle absolue, ceci explique pourquoi le statisticien passe souvent pour être susceptible et avoir mauvais caractère ;

- quant à l'utilisateur des matériaux numériques, il doit être informé de la source, de la précision et de la signification des statistiques qu'il met en oeuvre et plus encore peut-être que le statisticien, il doit être d'une impartialité rigoureuse.

Ces deux conditions ne sont pas toujours réalisées et il ne faut pas s'étonner si la faveur va à tous les

(1) 22-24 Avril 1948. Comptes rendus Fascicule 1, page 47.

sarcasmes qui visent la statistique. Ce discrédit et cette méfiance tiennent d'abord de ce que le soi disant bon sens refuse au chiffre toute expression et il lui préfère un raisonnement subjectif, elles tiennent, bien sûr, à l'imperfection de certaines données numériques et aux erreurs d'interprétation, et puis il y a aussi la marque d'un individualisme qui croît préserver son libre arbitre : faisant allusion à des travaux de revenu national, un de nos Académiciens écrivait : "cette fausse science, ce charlatanisme d'Etat prêt d'ailleurs à se mettre aux ordres d'une tyrannie quelconque de droite ou de gauche, nous prépare en toute hypothèse la suppression de la sagesse du sens de la vie et de la mesure, de la liberté".

Nous voici bien au coeur du problème, on peut y répondre aisément : "la liberté économique demeure pour chaque Français en période de pression inflationniste d'épargner ou de consommer ; néanmoins le comportement global de nos compatriotes peut être mesuré et prévu avec une certaine exactitude. L'inquiétude manifestée par une certaine école vise davantage une crainte d'envahissement de la société par les techniques, filles de la connaissance" (1).

Il est bien vrai que trop souvent l'intérêt personnel se sent menacé par les investigations statistiques et que le réflexe de sauvegarde de ces intérêts se manifeste pleinement.

Mais ce n'est pas tout ; à toutes ces raisons s'ajoute ce que l'on a appelé le quiproquo du statisticien (G. BOURNIER) Prenons un exemple, il existe des magistrats, des inspecteurs du travail, des sociologues, qui ont chacun un champ d'action particulier, il ne leur viendrait pas à l'idée qu'un statisticien décide d'entreprendre une étude sur un régime foncier, ou sur l'application de tel article du code du travail, ou encore sur les rapports entre le langage et la parenté. Inversement, on trouve tout naturel qu'un non-staticien puisse entreprendre une enquête démographique sur telle ou telle région d'Afrique Centrale ou qu'un sociologue décide de se lancer dans une enquête extensive sur des budgets ruraux. L'opinion est en effet fort répandue en France que la statistique est une voie publique ouverte à tous, alors qu'en réalité un grand écriteau "Voie privée" devrait être placé à l'entrée, à l'intention

(1) Au service des tyrannies ? F-L. CLOSON.
Le monde - 27 Novembre 1953.

des éventuels visiteurs. N'importe qui se croît autorisé à faire de la statistique et la fait mal. C'est le quiproquo du statisticien. Ce n'est pas parce qu'une discipline traîte de matières qui intéressent la vie quotidienne que l'on peut, sans préparation, en faire usage et en discuter convenablement.

Le décompte des études inexactes, mal conduites dans les domaines démographique et économique (études sur le revenu national et la balance des payements notamment) n'est plus à faire. Le statisticien est un technicien, même si cela ne se voit pas toujours sur le terrain. Il y a une technique des relevés que cache le caractère très matériel du rassemblement des informations.

Toutes les incompréhensions qui viennent d'être énumérées ont pour conséquence le scepticisme indéracinable qui accueille les travaux statistiques les plus sérieux : tandis que le paradoxe de DISRAELI a encore son petit succès dans les salons plus ou moins mondains : il satisfait à la fois esprit critique et individualisme, la phrase suivante de Mr. BEVIN livrée ici directement traduite, a beaucoup moins de retentissement : "Le monde moderne sera civilisé lorsqu'il sera sentimentalement ému par la statistique". "The modern world will be civilized when it vill be emotionally moved by statistics". (cf. Mac ARTHUR "Il n'y a pas de gouvernement sans statistique").

Sous son apparente emphase, l'homme d'Etat anglais voulait faire comprendre tout simplement que les hommes politiques feraient moins d'errours s'ils avaient en leur possession les éléments nécessaires à leur gestion.

1.3 - La Statistique - Approche générale.-

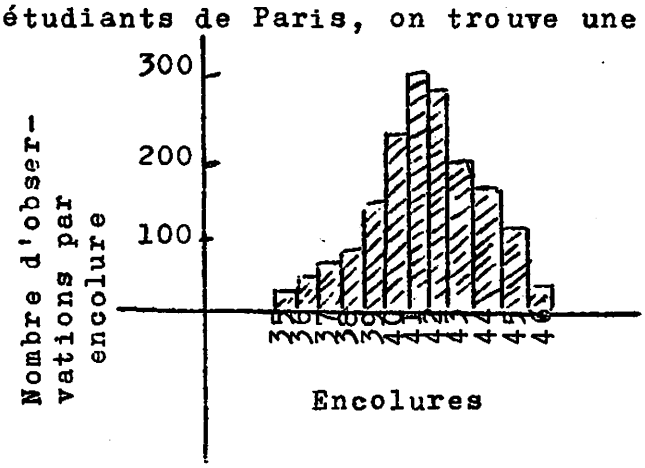
Le terme de statistique a été évoqué à plusieurs reprises, il convient maintenant d'en préciser son contenu. Bien entendu sans faire intervenir pour l'instant une quelconque notion de calcul de probabilités, autrement dit sans le secours d'un appareil mathématique qui pourrait rebuter.

Prenons un exemple simple :

Un besoin vital est celui du vêtement. Si l'on a besoin d'acquérir une chemise, on peut se rendre chez un

chemisier à façon, il est nécessaire dans ce cas de prendre différentes mesures. Cette condition remplie le travail une fois exécuté, on entre en possession d'une chemise bien adaptée.

Mais en général, il suffit de pénétrer dans une chemiserie pour être immédiatement satisfait. Analysons ce résultat : si on considère plus précisément une mesure déterminée, celle de l'encolure, si le chemisier se trouvait avoir en stock un assortiment tel qu'il possédât des quantités de chemises égales pour chaque pointure, il s'apercevrait à l'expérience que certaines dimensions sont demandées plus fréquemment que d'autres et qu'une partie de son assortiment est épuisée tandis qu'une autre, au contraire, lui reste en stock et risque à la longue de s'abîmer. On peut concevoir qu'il ait le désir d'avoir en magasin les proportions optima ; si le statisticien entre alors en jeu, il peut, pour l'aider, demander aux différents élèves qui l'écoutent de mesurer leur encolure : 37, 39, 38, etc... ; il pourra calculer une valeur moyenne de l'encolure pour ce groupe l'année prochaine, ou aujourd'hui, dans une autre classe il peut répéter cette mesure et étendre ses observations ; finalement si on prend tous les étudiants de Paris, on trouve une répartition des encolures telle que celle qui est figurée sur le graphique ci-contre. La valeur moyenne de l'encolure pour ce nouvel ensemble différerait un peu de la valeur moyenne obtenue dans cette classe, mais, au fur et à mesure que le nombre des observations va s'accroître, on s'aperçoit que la valeur moyenne est pratiquement stabilisée et qu'elle varie dans des limites très étroites, de même la forme générale de la répartition demeure constante à partir d'un nombre assez important de mesures.



Le chemisier une fois informé (si tant est que sa propre expérience ne l'a déjà instruit), établira son

stock de telle manière qu'à partir d'une quantité maximum pour une valeur moyenne de l'encolure, le nombre de ses chemises soit, pour chaque taille, égal à une certaine proportion du total, et qu'aux valeurs extrêmes il ne possède que quelques unités en stock.

Cet exemple très matériel et très simple est un exemple d'une part de la permanence statistique de la valeur d'un caractère (ici il s'agit de la valeur moyenne de l'encolure), d'autre part d'une distribution ou d'une répartition de fréquences.

On perçoit ici une des différences fondamentales qui peuvent exister entre la statistique et la comptabilité. La comptabilité livre un renseignement avec une précision absolue ; au contraire, la statistique fournit un ordre de grandeur, la valeur vraie se trouvant dans un certain intervalle entourant la valeur observée, puisqu'en général on ne peut dénombrer toutes les unités qui composent le domaine étudié.

Un comptable s'intéressera aux entrées individuelles dans une formation hospitalière, tandis que le statisticien s'intéressera globalement aux entrées selon le sexe, l'âge, la maladie constatée, etc., Pour entreprendre une statistique, un nombre suffisant d'observations est nécessaire et c'est le moment de conter cette anecdote chinoise d'un statisticien jeté en prison, et dont le supplice permanent était de ne pouvoir compter dans sa geole qu'un exemplaire des objets qui s'y trouvaient.

Pour être plus sérieux, une définition plus précise de la statistique doit être donnée :

" La Statistique est la science qui a pour objet de recueillir et de coordonner des faits nombreux dans chaque espèce, de manière à obtenir des rapports numériques sensiblement indépendants des anomalies du hasard".

Cette définition de COURNOT, qui date de 1843, attire donc l'attention sur les trois tâches principales du statisticien.

Il recueille des chiffres.

Il coordonne, c'est-à-dire il cherche les liaisons entre les différents phénomènes.

Il établit les rapports numériques ou les valeurs caractéristiques qui peuvent varier avec les séries d'observation qu'il se procure mais qui, en définitive, pour un nombre important d'observations, demeurent infiniment voisins des chiffres fondamentaux réels.

Cette définition n'est pas unique, car le nombre des définitions de la statistique dépasse la centaine ; plus simplement, on peut dire que la Statistique est l'ensemble des méthodes ayant pour but l'étude numérique des collectivités.

Pour en terminer avec les définitions, il faut remarquer que le mot statistique recouvre donc deux notions distinctes:

la Statistique en tant que technique d'observation, de mesure ou de raisonnement ;

mais aussi, le résultat final des recherches précédentes, et c'est là un des traits originaux du statisticien que de rassembler la documentation en vue de l'action ultérieure : c'est un géographe des temps modernes.

1.4 - Historique rapide.-

Le dénombrement était connu bien avant le début de notre ère : il y a plus de quatre mille ans, les Chinois établissaient des tableaux chiffrés relatifs à la production agricole, et la Bible fait déjà allusion à des recensements de population. On ignore si l'hostilité anti-statistique date de ce temps, mais l'Ancien Testament conte qu'inspirés par l'Esprit du Mal, David et Saül avaient opéré un dénombrement. Le peuple en fut puni par une épidémie qui fit périr 70.000 personnes en trois jours ; le recensement de la population fut alors considéré par la chrétienté comme une pratique impie et dangereuse (1).

(1) Ce sentiment demeure parmi certaines populations, puisque le statisticien britannique MARTIN note en 1953 dans son compte rendu à l'I.I.S. : "Une superstition existe parmi certaines tribus que le recensement de population aurait une influence diabolique". En d'autres parties de l'Afrique (mise à part la crainte de la fiscalité) on note de semblables craintes (nombre exact de têtes de bétail, enfants).

Les Romains connurent également les recensements. Se reporter à Saint Luc : "En ce temps-là, un édit de César Auguste ordonna de recenser toute la terre". Mais si l'on consulte certaines archives, (les recensements étaient fort réguliers puisqu'une inscription à Ankara mentionne que trois recensements ordonnés par Auguste furent exécutés en - 28, - 8 et + 14) il faut être prudent dans l'interprétation des chiffres successifs, car les Censeurs, comme on les appelait, recensaient seulement les citoyens romains, et il n'y a pas correspondance entre l'accroissement de la population de Rome, ou de l'Empire et celui du nombre des citoyens Romains. D'autres chiffres qui pourraient faciliter certaines estimations de population sont difficiles à utiliser : l'unité de recensement est très variable : on connaît par exemple tantôt le nombre des personnes touchant tous les mois une allocation de blé gratuite, tantôt la quantité de blé emmagasinée pour l'année à cet effet, mais ce blé n'était distribué qu'aux indigents et la proportion de ceux-ci par rapport à la population totale variait certainement. Déjà, on est dans l'obligation de critiquer les rares chiffres disponibles.

Si l'on franchit plusieurs siècles, on apprend que Sully en 1602 avait constitué un "Cabinet complet de Politique et de Finances" ; cet organisme était chargé de dépouiller les rapports sur toutes les branches de l'Administration. Colbert procède ensuite en 1664 à une vaste enquête sur les Provinces. A la fin du XVIIème siècle, il est prescrit à chacun des Intendants du royaume de dresser un mémoire sur la Généralité placée sous son administration.

A l'étranger, au XVIIème siècle, il y a encore simple compilation des données connues et il n'y a aucune critique des données fournies, en revanche on note une nouvelle tendance qui vise à fournir une description raisonnée des Etats d'Europe. Conring (considéré depuis comme le chef de l'école allemande, professeur à Helmstadt) est le premier qui introduisit cet enseignement original qui demeure encore cependant très descriptif.

Une école concurrente se développe parallèlement, elle s'attache à la recherche des caractères purement numériques des collectivités et tend à dégager certaines lois dites naturelles. Cette école est l'école anglaise, ou l'école des Arithméticiens Politiques. On doit citer HALLEY (1662-1742), l'astronome, qui construisit la première table de mortalité pour la ville de Breslau.

C'est à cette époque qu'apparaît le mot "Statistique". L'origine en remonterait à 1749, et il aurait été employé par un professeur allemand ACHENWALL qui employa le terme "Statistik", dérivé de l'italien "Statista" qui signifie "Politique". (1)

Si l'on revient en France, les principales étapes sont les suivantes :

a) Sur le plan administratif : en 1788, (2) NECKER constitue un bureau spécial "la Balance du Commerce". Ce bureau doit "recueillir les renseignements sur la production naturelle, le travail et tous les faits économiques pouvant éclairer le Gouvernement". Ce bureau disparaît en 1793. A la même époque, LAVOISIER entreprend certains travaux - certains sont bien connus : Le nombre total des socs de charrues étant déterminé pour l'ensemble de la France, il en déduisait, par l'utilisation d'un certain coefficient, l'étendue des terres arables.

Il est créé, vers l'année 1790, au Ministère de l'Intérieur, un bureau chargé de la Statistique Générale. Il faut rappeller à cette occasion quelques citations de François de NEUFCHATEAU relatives à l'établissement des tableaux de population "Transmettre des résultats d'une rigoureuse exactitude .... l'Arithmétique Politique se fonde sur ces éléments ... C'est une belle Science que celle dont les opérations ont pour but des recherches utiles à l'Art de gouverner".

Le premier annuaire statistique régional est celui de l'annuaire du Bas-Rhin par BOTTIN. 1801 est une date qu'il faut garder en mémoire car cette année est marquée par le premier dénombrement de la population française.

Avec des fortunes diverses (La Restauration en particulier ne fut guère favorable à la statistique) le bureau de statistique générale créé en 1790, demeure jusqu'en 1907, date à laquelle le décret du 14 Août fixe l'organisation de la Statistique Générale de la France. Dix ans plus tard, est créé en 1917 le Service d'Observation des Prix : ce service est ensuite fondu en 1919 dans la Statistique Générale de la France. A cette époque, les effectifs sont encore

---

(1) Un verbe grec statizens se traduit par constater.

(2) Voir "De l'Administration des Finances de la France" 1784, NECKER.

peu nombreux : on compte 48 statisticiens, contrôleurs, dames employées et soixante auxiliaires.

En 1936, les effectifs sont un peu plus importants : 175 personnes, mais, pour donner une idée de la pauvreté des moyens mis à la disposition de la S.G.F., on peut les comparer aux effectifs des services similaires à l'étranger, à la même époque :

Allemagne	2.358	personnes
Tchécoslovaquie	676	"
Pologne	315	"
Espagne	300	"

Cette comparaison peut apparaître quelque peu superficielle, elle est toutefois édifiante.

En 1941, l'Institut de la Conjoncture (créé en 1938) et la S.G.F. sont fusionnés dans le Service National des Statistiques. Cette création est inséparable du nom du contrôleur général CARMILLE, à qui revient le mérite d'avoir pu dans les circonstances difficiles développer la statistique administrative française.

Le contrôleur général CARMILLE voulait recréer clandestinement les services de recrutement de telle manière qu'à la Libération on puisse, sans perdre de temps, reconstituer les grandes unités militaires nécessaires. Ce but nécessitait la création de nombreux ateliers mécanographiques, afin de transcrire et de conserver sur cartes perforées les caractéristiques des diverses personnes susceptibles d'être mobilisées. Monsieur CARMILLE fut emprisonné par les Allemands, transféré au camp de Dachau où il est mort. Mais l'oeuvre amorçée fut maintenue et en 1946 l'I.N.S.E.E. (1), pour la Métropole et la France d'Outre-Mer était créé. C'est le régime actuel. L'organisation est la suivante : l'I.N.S.E.E. comprend une direction générale à Paris et dix-huit directions régionales en province. A cet Institut, est rattachée une Ecole d'Application chargée de former les cadres administrateurs et attachés, et d'instruire les fonctionnaires et les auditeurs étrangers qui désirent suivre les cours et de former aussi des statisticiens-économistes pour les entreprises.

---

(1) Institut National de la Statistique et des Etudes Economiques.

Cinq ministères possèdent des services particuliers : Ministères du Travail, de la Production Industrielle, de la Reconstruction, de l'Agriculture et enfin de la France d'Outre-Mer. Le Ministère des Finances possède un service d'études économiques et financières (S.E.E.F.).

Sur le plan universitaire, l'I.N.S.E.E. collabore avec l'Université pour l'organisation de l'enseignement statistique (Institut de Statistique de l'Université de Paris, Facultés de Droit et des Sciences Economiques, Ecole Normale Supérieure de l'Enseignement technique,...).

b) Sur le plan technique, les progrès ont été également très importants. Après Lavoisier, c'est Laplace (1749-1827) qui doit retenir l'attention: Laplace utilisant pleinement les travaux de Bernouilli (1654-1705) mit sur pied le calcul des probabilités. Plus tard, en contact avec de nouvelles disciplines, la statistique se morcelle, se spécialise : avec Quetelet (1796-1874) par exemple, elle s'oriente nettement vers l'étude des ensembles très nombreux qui deviendra l'essence même de la statistique. Le champ de son action est maintenant très divers et on ne peut citer que quelques-unes de ses applications :

Statistique biométrique, en particulier l'étude de l'hérédité et l'étude des aptitudes physiques et mentales (psychotechnique et psychologie expérimentale) - Statistique psycho-sociologique des comportements de groupes.

Statistique stellaire, statistique appliquée à l'étude des phénomènes radioactifs.

Statistique appliquée à l'Industrie : contrôle des fabrications, étude des conditions des cahiers des charges, à l'aéronautique, à l'électronique, etc...

Statistique appliquée aux recherches agronomiques : analyse des champs d'expériences, comparaison des variétés, des rendements, etc...

On peut rappeler encore l'importance de la statistique médicale (efficacité de médicaments, effets comparés de vaccins, etc...)

Sont citées pour mémoire la statistique économique et la statistique démographique qui seront étudiées plus en détail dans les prochains chapitres.

Tous ces motifs d'étude sont relativement austères, mais d'autres domaines sont plus anecdotiques : Dans la "Méthode statistique" de NICEFORO, on voit des relevés sur la hauteur des statues romaines du Musée du Louvre, les temps de coureurs cyclistes, la longueur des mots dans différentes langues et l'importance de la période dans deux oeuvres de Balzac.

Dans un autre article, on a observé l'analogie des mouvements d'une mouche avec les circuits d'un chauffeur de taxi, à l'intérieur de la capitale. La liste serait sans fin de toutes les applications de la méthode statistique, toutefois, malgré ses lettres de noblesse, le grand public même le plus averti, n'a pris en France, conscience de l'intérêt de la statistique que fort tardivement et il ignore qu'une formation très sérieuse est nécessaire pour qu'un statisticien puisse exercer convenablement son métier.

1.5 - La Coopération.-

En effet pour connaître et être familiarisé avec l'ensemble de ces méthodes, il est nécessaire de passer deux ans à l'Ecole d'Application de l'I.N.S.E.E., après avoir réussi un concours d'un niveau équivalent à la licence (L'I.N.S.E.E. recrute également des élèves à la sortie de l'Ecole Polytechnique). On peut donc concevoir que l'approche des méthodes statistiques ne soit pas évidente.

1.5.1- La coopération entre services.

Est-il alors logique ou même possible, et est-il efficace d'évoquer les problèmes que posent les méthodes statistiques lorsqu'elles s'appliquent à des pays en voie de développement devant un auditoire qui ne soit pas composé uniquement de scientifiques ? On doit répondre par l'affirmative pour plusieurs raisons :

En réalité, le statisticien ne travaille pas dans sa tour d'ivoire ou en vase clos, son domaine est étendu et divers, il faut donc qu'une liaison s'établisse entre son action et celle des autres disciplines.

Si cette liaison est absente, ce manque de coopération se traduit alors par deux phénomènes : le statisticien se livre à des recherches dont les priorités successives lui

sont ignorées, il entreprend des études accessoires qu'il estime essentielles.

L'utilisateur, de son côté travaille à grande peine sur des problèmes que le statisticien a déjà résolus, et il arrive parfois que ces recherches soient entachées de certaines faiblesses.

Aussi, en évoquant ces problèmes, trois objectifs peuvent être visés :

a) Dégager des notions élémentaires mais précises, qui, au moment où s'exercent les recherches personnelles, empêcheront les erreurs grossières.

b) En matière de politique sociale ou économique on est finalement amené à prendre des décisions. Ces décisions doivent être guidées par la connaissance de la population et de l'économie. Or cette connaissance ne peut être anarchique, mais elle doit au contraire être instruite et étayée par des enquêtes et des observations quantitatives. Il faut alors savoir, ce qu'il est possible de demander au statisticien, dans l'état actuel de la technique et de la documentation numérique. La méconnaissance de cette règle élémentaire a conduit dans le passé à de graves mécomptes :

- Circulaire en date du 3 janvier 1955 demandant pour le 15 février 1955, pour chaque cercle le budget familial moyen de deux familles africaines.

- Demande d'un député d'Algérie demandant par âge et catégorie les salariés non inscrits à une caisse de retraites.

- Circulaire prescrivant un recensement exhaustif des aveugles.

- Questionnaire envoyé aux entreprises en 1951 demandant d'année en année jusqu'en 1955 et par catégorie les besoins de personnel pour renouvellement et les besoins supplémentaires.

Toutes ces demandes qualifiées abusivement de statistiques indisposent évidemment les destinataires.

c) Un troisième objectif est celui de démontrer la nécessité de la coopération entre le statisticien et les autres fonctionnaires. Une première raison est la variété des

travaux statistiques. L'ingénieur, le médecin peuvent dans une certaine mesure, travailler sans avoir des contacts très étroits avec d'autres disciplines, au contraire le statisticien ne peut être un solitaire, son action exige un travail d'équipe, et bien qu'il soit un spécialiste, le rassemblement des données ne peut échoir à lui seul : cette observation s'applique tant aux missions d'enquête sur le terrain qu'aux études administratives courantes.

L'investigation du statisticien doit donc être éclairée et il doit y avoir imbrication extrême entre les travaux du statisticien et ceux des autres chercheurs. Cette observation est générale, et mériterait d'être dite pour sa seule importance, mais elle est encore plus essentielle dans les pays en voie de développement que pour les pays déjà développés.

En France, en effet, le statisticien métropolitain peut, pendant de longues années, se spécialiser selon ses goûts personnels ou au hasard d'une affectation : il acquiert à la longue une connaissance du milieu, compétence qui deviendra égale à celle des fonctionnaires avec lesquels il a à travailler. Au contraire, le statisticien d'Outre-Mer touche par nécessité et par vocation à tous les domaines, Il doit se familiariser avec les différentes ethnies, connaître le rythme des saisons, le cycle des végétations, les conditions du marché du travail, les prix des matières premières, la politique économique, la fiscalité, etc...

Il en résulte que si les services administratifs et techniques ont besoin du statisticien pour préciser leurs informations, le statisticien de son côté pourra difficilement se passer du concours de ces services. La tâche demeure immense et la collecte de la documentation chiffrée consiste finalement en une vaste "chasse à courre".

1.5.2- La coopération entre la capitale et la brousse.

Jusqu'à une époque récente, les statistiques diffusées par les services n'ont pu être considérées que comme les sous-produits de la gestion administrative. On doit donc si l'on désire progresser s'orienter vers une politique d'étude d'organisation systématique (fichiers établis ts C.D.) et de recherches en profondeur. Or, les pays africains sont généralement très étendus, la densité démogra-

phique est faible, enfin l'armature administrative est à mailles assez lâches. Il en résulte qu'une collaboration doit régner entre services des chefs-lieux mais aussi entre le statisticien du chef-lieu et l'administrateur ou le technicien de brousse, car, si à distance le premier peut coordonner les recherches et leur tracer un cadre, il ne peut rien sans la coopération du second qui, à l'autre bout de la filière, est à pied-d'oeuvre et se trouve sur place à la source même des informations.

C'est grâce à ces contacts que les échanges seront les plus fructueux : l'utilisateur sera à même de savoir ce que le statisticien possède et pourra apprécier la précision ou la valeur des informations fournies, il ne s'étonnera plus de certaines imprécisions. A son tour, il pourra lui rendre service : grâce à sa connaissance du pays, il pourra critiquer à bon escient certaines séries numériques défectueuses et de concert, s'épaulant l'un l'autre, ils pourront établir le grand inventaire critique du pays et cela sans fard, car le statisticien aura besoin d'alliés dans ce devoir de vérité.

Cet inventaire sera toujours marqué par son caractère laborieux, le statisticien doit non seulement être impartial, mais il doit être patient, ce côté ingrat de la recherche est une raison supplémentaire pour éviter que l'action sur le terrain soit une action en franc-tireur. Bien sûr, et on ne peut l'éviter, on s'aperçoit que pour un chercheur patient, obstiné, compétemt, il existe des dizaines d'utilisateurs habiles mais verbeux qui pillent sans vergogne des données encore partielles pour étayer des théories émises plus ou moins à priori (utilisation erronée des termes de l'échange, emprunt à des enquêtes démographiques non extensives, etc...).

Cette collaboration mettra en lumière toutes les difficultés de la recherche et démontrera que la statistique n'est pas une connaissance facile d'accès, simpliste dans ses méthodes, ouverte à n'importe qui, n'importe quand, n'importe comment. S'il est bien certain qu'il ne faut pas avoir une dévotion aveugle pour le chiffre, qu'il faut nuancer la brutalité des chiffres, il n'empêche que la désinvolture n'est pas de mise et qu'il faut se méfier de tout travail qui ne soit pas entouré de toutes les garanties .

Enfin, lorsqu'il y a effort de création, le chercheur doit prendre tous les contacts utiles afin que l'effort ne soit pas stérile parce qu'inutilisable.

1.6 - L'enregistrement continu des observations.-

Cette création, ces progrès et ce développement de la connaissance quantitative seront facilités par ce que l'on pourrait appeller l'enregistrement continu des observations.

Pendant très longtemps, les fonctionnaires se sont contentés d'expédier leurs propres affaires sans trop se soucier de savoir si après leur passage, une trace restait de leur activité. L'inventaire de la documentation chiffrée est souvent décevant, malgré l'abondance et la richesse des services. Périodiquement les termites ou des rangements périodiques (se terminant par des mises au pilon ou des feux purificateurs) ont fait disparaître de précieuses archives. Dans le même temps, faute d'organisation personnelle, des renseignements utiles glanés au cours des tournées n'étaient pas connus et finalement que ce soit du point de vue de la connaissance démographique ou du point de vue de la documentation économique, au fil des années rien de systématique n'a été rassemblé. La condition essentielle d'un progrès dans cette voie est l'adoption d'une attitude active et la mise sur pied d'une organisation préservant les archives utiles. Il faut que l'on se rende compte de la gravité de cette ignorance lorsqu'elle s'étend à l'échelon d'un territoire ou d'un groupe de territoires :

- méconnaissance de la vocation agricole de tel groupe de villages (enquête 1957-1958 dans le delta vif du Niger) ;

- ignorance de la gravité de la crue du Niger en 1954-1955 et de ses effets désastreux sur la riziculture locale, entraînant par la suite des importations massives de ce produit.

- absence de prévision de la récolte de certaines cultures industrielles, telle la campagne record 1956-1957 d'arachides conduisant à des doubles emplois avec les importations d'origine étrangère.

Ces exemples sont récents mais on pourrait rappeller qu'en 1936 la semaine de quarante heures a été édictée sans que l'on connaisse exactement le niveau de l'emploi et la part moyenne du salaire dans le prix de revient ; qu'en 1945 la création du franc C.F.A. a été réglée sans investigation antérieure valable, il est anecdotique de souligner que la valeur du franc C.F.A. à l'époque (1,7) est égale à la moyenne arithmétique du franc métropolitain et du franc Pacifique (2,4).

Qu'en 1953 l'augmentation des salaires anormalement bas a été décidée en France alors que l'ignorance de la pyramide des salaires était totale.

Et que dans les années récentes, l'aide à l'exportation fut décidée sans statistiques préalables ... elles furent établies après.

1.7 - Complexité de la connaissance et conclusion.-

Le problème de la connaissance devient chaque jour plus complexe : Si l'on se cantonne dans le domaine de l'outre-mer, il y a cinquante ans, on n'assistait pas alors aux mouvements de migrations observées maintenant. Le problème de l'attraction des centres ne se posait pas ; enfin dans l'ensemble des échanges, le facteur monétaire intervenait beaucoup moins, car en brousse, on vivait encore au système du troc et la progression actuelle du secteur monétaire rend le pays plus sensible aux influences extérieures : la mise en oeuvre des plans de développement a augmenté la circulation monétaire, bouleversé les habitudes alimentaires de certaines couches de la population, tout ceci mérite observation et il faut se convaincre qu'il n'y a plus place pour l'empirisme et que pour agir un matériel statistique minimum s'impose.

Ce matériel numérique est encore rudimentaire sinon inexistant, et il convient de la créer. Malgré la charge du travail courant, il faut comprendre que la création de la documentation chiffrée est une oeuvre commune. Il serait absurde d'espérer sa génération spontanée comme certains de ces industriels qui, en France, refusent de donner leur chiffre de production et qui s'étonnent ensuite que le statisticien ne leur livre pas les chiffres de leur concurrent. (Cette suspicion était telle, qu'avant la guerre, chaque industriel d'un certain syndicat déposait anonymement son chiffre dans une urne.)

La création de ce matériel est indispensable si l'on ne veut pas subir passivement les fluctuations de la conjoncture, et si l'on souhaite réellement administrer, il faut pressentir l'avenir par les inventaires démographique et agricole, car le premier problème est de connaître la production disponible, le nombre actuel des rationnaires et les perspectives de leur accroissement.

Il est certes difficile de tout planifier, de tout enserrer dans un cadre rigide de chiffres : dans toute action administrative il se glisse toujours un peu d'empirisme, mais isolé ou pratiqué comme seule méthode de gouvernement, cet empirisme sera toujours trahi : car faute d'une organisation numérique valable et précise, finalement même dans une discussion objective, c'est l'élément impalpable, l'opinion subjective qui l'emportera.

Et cela est malheureusement pratiqué quelque soit le domaine d'étude : tel l'influence de la polygamie sur la natalité en matière démographique, tel en économie la création d'une caisse de compensation et de son influence sur le niveau des prix des marchandises importées dans un territoire donné. Souvent le facteur politique l'emporte sur l'économique, mais il faut en être persuadé, en définitive c'est l'économique qui se venge du politique !

Que l'on s'en réjouisse ou qu'on le déplore, la seconde moitié du XXème siècle sera pour les pays africains et asiatiques une période durant laquelle, quelles que soient les voies ou les méthodes de planification en vue d'une croissance rapide, l'action pour être efficace et réfléchie devra être éclairée par la connaissance statistique de ces pays. Ce thème peut d'ailleurs être illustré en conclusion par un propos du représentant français à l'organisation mondiale pour l'Agriculture et l'Alimentation, qui montre combien la statistique est auxiliaire indispensable non de la seule pensée scientifique mais plus encore de l'action économique et de l'action politique :

" L'un des aspects les plus frappants du XXème siècle
" est l'apparition de ces grands plans techniques, écono-
" miques, financiers qui tendent à influencer profondé-
" ment sur tous les caractères des sociétés : plans dres-
" sés à l'échelle d'une nation, ou d'un groupe de nations
" qui tendent à embrasser peu à peu l'activité de toutes
" les nations.

" Le premier stade prévu est toujours un immense inven-
" taire de plus en plus fouillé, de plus en plus précis. La
" technique qui permet de le faire va chaque année en s'amé-
" liorant. Il est bien remarquable en particulier que l'in-
" ventaire des besoins inéluctables soit désormais basé
" sur des connaissances scientifiques de plus en plus exac-
" tes. Inventaire des besoins que fait apparaître le déve-

" loppement général de la population. De plus en plus pré-
" cis, l'inventaire des ressources naturelles est poussé
" partout et de même l'inventaire des hommes : non seule-
" ment de leur nombre, mais de leurs capacités de vivre,
" d'agir et dans leurs diverses formes d'activité.

Pr A. MAYER

"Population" 1er Trimestre 1951.

BIBLIOGRAPHIE SOMMAIRE

La statistique Collection que sais-je ? Vessereau

Association amicale des ingénieurs de la statistique et des Etudes Economiques : bulletin n° 25, janvier-mars 1955. De la statistique au camouflage. A. Caffot
bulletin 33 - janvier juin 1957.
Aperçu rétrospectif sur la statistique française.

Journal de la Société de Statistique de Paris (Avril mai juin 1953). Le mouvement brownoïde de l'homme et des animaux.

La méthode statistique et ses applications aux sciences naturelles, aux sciences sociales et à l'art.
Alfredo Niceforo 1925.

Grandeurs et misères de la statistique dans les T.O.M.
Francis Closon.
Le Monde - 9, 10 et 12 janvier 1958.

Etablissement d'un service central statistique dans les pays en cours de développement G. Duon.
Bulletin d'information de l'I.N.S.E.E.
n° 5 mai 1952.

Discours de S.S. le Pape Pie XII. 28ème session de l'Institut international de statistique.
Bulletin hebdomadaire n° 292 du 5 décembre 1953.

Ouverture de l'année scolaire de l'Ecole d'Application de l'I.N.S.E.E. année 1947 G. Bournier.

Communication à l'Académie des Sciences Coloniales
1943 G. Bournier.

Misère et progrès de la Statistique Coloniale
Bulletin mensuel d'information du Ministère de la France d'Outre-Mer.
1946 G. Bournier

L'approche statistique M.G. Kendall
Economica Mai 1950.

=:=:=:=:=:=

Chapitre II

UNITE STATISTIQUE ET MÉTHODES STATISTIQUES

2.1 L'unité statistique.

2.2 Les méthodes statistiques.

2.3 L'Importance de la définition de l'unité statistique.

2.4 Les méthodes d'approche ou d'observation des phénomènes.

2.4.1 - Les méthodes directes.

2.4.2 - Les méthodes indirectes.

" Le pont aux ânes du statisticien :
" définir l'unité statistique et s'y tenir.
" Bannir les doubles comptes et faire la
" chasse aux unités vagabondes ".

Inspecteur Général DUMAS

On distingue dans le domaine de la recherche statistique plusieurs stades :

Le premier est le dénombrement à une époque voulue des données ou des faits étudiés, il s'y rattache les méthodes d'élaboration des statistiques.

On peut ensuite étudier les variations de ces données numériques, variations soit dans le temps, soit dans l'espace. On a ainsi l'occasion de mettre en évidence certains déterminismes ou certaines relations entre séries numériques différentes.

Le troisième stade consiste en des tentatives d'explication de ces déterminismes.

A cette dernière recherche se rattachent alors :

- les jugements sur épreuves comparatives ;

- les appréciations de différences ;

- l'établissement d'échelles de valeur et de normes.

Mais à la base, la qualité du dénombrement conditionne la validité des calculs ultérieurs.

Rappelons qu'une définition de la Statistique (parmi beaucoup d'autres), est qu'elle est constituée par les méthodes applicables à l'étude numérique des ensembles de faits nombreux, des phénomènes de masse.

Ces faits, ces phénomènes, peuvent être d'une nature quelconque ; cependant, pour que les méthodes statistiques puissent leur être applicables, il est nécessaire que ces évènements ou ces observations soient définies sans ambiguïté. C'est tout le problème de la définition de l'unité statistique.

o

o o

2.1 - L'unité statistique.-

Tout être vivant, tout objet inanimé, tout concept, tout phénomène peut être pris comme unité statistique. Deux unités statistiques seront de même espèce, si on peut leur attacher un caractère commun : ce caractère commun devant répondre à une définition précise. Dans le chapitre précédent l'unité statistique était la mesure de l'encolure d'une chemise. En démographie l'unité statistique pourra être un mariage, en économie, un budget familial sera une unité statistique.

2.1.1- Le caractère attaché à l'unité statistique peut être mesurable ou quantitatif : tonnage d'une campagne agricole d'un produit déterminé, mais il peut aussi ne pas être mesurable (qualitatif) : couleur des yeux, état matrimonial par exemple.

En dépit de leur caractère commun, les unités statistiques de même espèce font l'objet de mesures ou d'observations variées concernant d'autres caractères qui leur sont attachés. On conçoit ainsi qu'une unité statistique peut être plus ou moins complexe selon sa définition et les différents caractères destinés à être mesurés. Une naissance vivante est une unité statistique relativement simple, tandis qu'une entreprise, un budget de famille présentent des aspects très divers. La définition d'une unité statistique peut donc être parfois délicate : dans une enquête agricole, on a à préciser la notion d'exploitation agricole: l'exploitant n'est pas forcément le propriétaire des terres. La superficie de l'exploitation doit-elle comprendre les terres laissées en jachère ? Un jeune frère qui aide à la culture des champs de son frère aîné, mais qui néanmoins cultive deux ou trois jours par semaine ses propres champs, doit-il être compris dans le personnel de la première exploitation et dans l'affirmative comment ?

Un autre exemple emprunté celui là au domaine démographique est la définition de l'unité familiale élémentaire : en Guinée on peut rencontrer les trois types "classiques" : ménage au sens moderne compris dans le "carré" - segment de lignage réuni dans une "cour" - famille étendue habitant la "concession ". Au Togo, en pays Kabré, une terminologie empruntée à des pays voisins fait appel à la notion de soukala, alors qu'en réalité sur le terrain, on a intérêt à se raccrocher à l'enclos ou au vestibule qui se subdivise lui-même en courette ("ouao").

L'enquête nutritionnelle pose elle aussi ses problèmes : dans certains pays la mobilité des cuisines, les échanges de plats, le manque de fixité des rationnaires oblige à retenir une unité statistique relativement complexe: le "rectangle" et ses campements.

2.I.2- On définit les ensembles statistiques, appelés aussi : champ, population, univers. Un ensemble statistique est constitué d'unités statistiques de même nature, l'étude visant les autres caractères différents.

Un producteur de cacao est une unité statistique bien précise, l'ensemble des producteurs de cacao représente un ensemble statistique bien déterminé.

Cet ensemble peut être subdivisé en considérant la superficie cultivée par chaque producteur et l'on est amené à définir trois sous-ensembles :

- producteurs ayant moins de 2 Has en plantation
  " " de 2 à 4 Has en plantation
  " " 5 Has et plus en plantation.

Ces trois sous-ensembles sont eux aussi bien définis.

Il ne faut cependant pas croire que l'addition de sous-ensembles issus d'un ensemble commun le reconstitue toujours :

Il arrive que les champs portent en association plusieurs cultures :

ignames - arachides
riz - coton.

Si l'on additionne sans précaution, les superficies indiquées sous chaque rubrique individuelle on risque des doubles emplois tel que la superficie totale dépasse la superficie physique cultivée.

Lorsque l'effectif d'un ensemble statistique est susceptible de variation, on dit qu'il est fluctuant : dans l'exemple ci-dessus, les nouveaux travailleurs entrent dans l'univers considéré, tandis que les décédés ou les paysans trop vieux sortent de l'ensemble statistique.

Dans ce cas, cet ensemble est appelé aussi ensemble renouvelé. Si l'on considère, une population sans émigration, dont les naissances compenseraient exactement les

décès , on aurait affaire à une population fermée stationnaire.

Cette notion d'unité statistique, ayant été étudiée, on va passer en revue les méthodes statistiques applicables à l'étude numérique des ensembles statistiques.

2.2. - Les Méthodes Statistiques.-

On distingue les méthodes d'élaboration d'une part et, les méthodes d'analyse et d'interprétation d'autre part.

2.2.1 - Les méthodes d'élaboration sont extrêmement importantes, elles concernent :

a) les méthodes de relevés et d'observation des faits
Avant de les mettre en oeuvre, un premier stade de travail doit être celui de la préparation des relevés. Par exemple l'établissement d'une statistique d'état civil requiert :

- la préparation des modèles de bulletins
- la mise sur pieds de l'organisation administrative requise
- l'enregistrement de chaque événement.

b) La descriptive numérique ; en considérant l'exemple précédent, le dépouillement portera sur les naissances du sexe masculin d'une part, et les naissances du sexe féminin d'autre part . On procédera ensuite au classement de ces observations, classement simple si on considère seulement l'appartenance à l'un ou l'autre sexe, plus compliqué si l'on envisage d'entreprendre à la fois la répartition par sexe et selon l'âge de la mère par exemple. On entreprend enfin le dénombrement des unités de chaque groupe ainsi créé.

c) Au troisième stade enfin, intervient la présentation des résultats statistiques sous forme de tableaux numériques et de graphiques .

2.2.2.- Après les méthodes d'élaboration, on trouve les méthodes d'utilisation ou d'interprétation des résultats ; ces méthodes font intervenir l'analyse statistique. Elles ont pour but de déterminer les paramètres qui caractérisent les permanences statistiques constatées, d'en mesurer la précision, de caractériser numériquement les liaisons constatées et finalement déduite dans la mesure du

possible, des explications aux phénomènes et de faciliter la prévision : on observe par exemple une certaine permanence dans la proportion des sexes à la naissance, et l'on peut envisager certaines hypothèses pour expliquer cette permanence.

Il faut insister ici surtout sur les méthodes d'élaboration car, on ne le répètera jamais assez, une statistique dépend de la valeur et du degré de confiance que l'on peut accorder à la documentation de base. Il faut connaître dans quelles conditions sont nés les chiffres, car cette connaissance subordonne toutes les études postérieures. L'étude économique exige des chiffres synthétiques, le crédit que l'on peut accorder à ces chiffres est lié directement à la valeur des chiffres de base dont ils dérivent . Le calcul d'un indice de prix ou d'un indice de volume aboutit en définitive à un seul nombre ... et ce nombre dépend d'une masse d'observations dont il est nécessaire d'être assuré de la qualité.

Il faut réfléchir à cette naissance du chiffre et savoir faire le partage entre le chiffre légitime et le chiffre "enfant trouvé" . Il faut avant d'utiliser un chiffre, s'enquérir de ses origines et connaître son état-civil. Plus on est loin de la source des chiffres, plus on se trouve désarmé et plus on a tendance à leur faire confiance surtout s'ils sont officiels et imprimés . Il y a quelques années , l'édition de province d'un journal parisien publiait un indice des prix à la consommation familiale avec une coquille typographique: 147,5 au lieu de 145,7 . Les journaux de province s'étant emparé de l'information, on eut par la suite les plus grands peines à faire admettre par le public qu'il y avait erreur.

Rassemblement des données, dépouillement, analyse, tout ce processus commun à tout renseignement statistique exige une organisation énorme. Rares sont les chiffres synthétiques qui peuvent provenir du seul effort personnel . L'effort collectif est seul payant et l'organisation, qui a pour charge la création du chiffre, demande tant de soins que, sur le plan intellectuel, ce travail est digne de mérite et se place au même niveau que celui de l'étude pure.

2.3- L'importance de la définition de l'unité statistique.-

Avant de passer aux méthodes d'élaboration, il faut revenir sur la définition de l'unité statistique. "Si l'on prend la construction d'automobiles, on peut hésiter entre les sorties de chaîne, les entrées au parc, les sorties de parc, les livraisons et les facturations" (1). Une définition précise de ce que l'on observe est primordiale. Soit la statistique des permis de bâtir : établie à partir des renseignements fournis par les services municipaux, on s'aperçoit que certaines constructions administratives peuvent échapper. Soit encore la longueur des voies d'un réseau ferré, fait-on ou non entrer en ligne de compte les voies de manoeuvre ? Dans certaines publications les voies de $0^{m},60$ Decauville sont comprises. Dans le domaine des Travaux Publics, on peut trouver selon la source, des longueurs très différentes pour des tronçons qui ont fait l'objet de rectifications de tracés (Conakry Mamou : 310 ou 285 Kms.) et les ingénieurs se plaçant dans l'optique de l'obtention des crédits préfèrent conserver les valeurs anciennes !

La statistique du chômage est un exemple classique de l'intérêt limité de certaines séries : dans ce cas, la statistique ne touche qu'une partie des chômeurs : l'intéressé doit avoir présenté sa demande auprès d'un fond de chômage (il n'en existe pas dans toutes les communes) et il doit avoir au moins trois ou six mois de résidence dans ladite commune.

La définition de l'unité statistique aura donc une influence importante sur la valeur représentative du relevé et sur l'utilisation même des nombres présentés. Ainsi, pour reprendre le cas des chômeurs secourus, faute de renseignements complémentaires, on ne peut se rendre compte de la question dans son ensemble, et, si l'on suit l'évolution du chômage d'une année à l'autre, la comparaison risque d'être faussée par suite de l'ouverture de nouveaux fonds de chômage, ou d'une nouvelle règlementation de l'attribution des allocations.

De plus deux séries de définitions différentes sont actuellement suivies : nombre de demandes d'emploi non satisfaites, nombre de chômeurs secourus et les variations de ces deux séries ne sont pas forcément concordantes. Il

---

(1) La statistique est-elle utile aux entreprises ? R. DUMAS - Bulletin d'information de l'I.N.S.E.E. - janvier 1957.

s'y ajoute que l'on ne peut saisir le sous-emploi caché (tel celui de l'industrie textile travaillant à certaines époques de vingt à vingt-cinq heures par semaine) (1).

Les modifications ou les distorsions du domaine d'étude.

Les changements de définition de l'unité statistique ou les modifications du domaine d'étude sont la cause de nombreux mécomptes. On va en citer des exemples empruntés aux horizons les plus divers :

Les mères de quatre enfants ont été dans certains territoires exemptées depuis longtemps de l'impôt personnel : au Cameroun, cette mesure fut étendue en 1948 aux mères de trois enfants et plus. Si l'on entreprend sans précaution une statistique sur les effectifs par grande tranche d'âge, on risque de trouver une diminution qui ne s'explique pas par une fuite devant l'impôt, mais simplement par une modification de la législation : les mères exemptées étant classées avec les jeunes filles. Dans le tableau ci-dessous, cette erreur se combine certainement avec des erreurs de calculs :

Statistique de la population au Cameroun
(milliers d'habitants)

	Hommes	Femmes	Garçons	Filles	Total
1943	1.117	899	409	300	2.725
1944	825	977	516	475	2.793

Source : Rapport annuel à l'Organisation des Nations Unies.

En juillet 1946, la Direction de la Comptabilité Publique a introduit progressivement les opérations des Caisses d'Alsace-Lorraine. Elle le fit sans le spécifier dans sa publication ; or de juillet à octobre, on assista à une progression des dépôts sans qu'il y eût de raisons plausibles. C'est d'ailleurs cette expansion, qui permit de découvrir cette introduction.

En matière de commerce extérieur, les statistiques métropolitaines depuis avril 1948 comprennent les échanges

(1) En 1949, le Bureau du Census estimait à quatre millions le nombre de chômeurs aux Etats-Unis, Malenkov à la même époque annonçait quatorze millions. Le premier chiffre se rapportait aux chômeurs complets, le second aux travailleurs n'accomplissant pas un travail à plein temps.

de la Sarre. Dans le même ordre d'idées, de 1940 à 1946 l'A.E.F. et le Cameroun ont été liées par une Union douanière. Cependant les deux pays continuaient à établir des statistiques séparées . Il en résulte que les entrées de marchandises d'importation par le poste de Garoua, même lorsqu'elles étaient destinées à l'A.E.F. étaient reprises pour le compte du Cameroun. Inversement, le cacao du Gabon exporté par Kribi était porté au crédit du Cameroun. Enfin, l'or du Cameroun était envoyé à Brazzaville et porté au compte de l'A.E.F.

Un cas comparable est celui des Comores :

Depuis le 1er janvier 1952, l'archipel ne fait plus partie du territoire douanier de Madagascar. Si l'on compare les échanges en 1949 et durant une année récente, il faut pour cette dernière éliminer les échanges internes à ces deux pays, qui en 1949 n'entraient pas en statistique.

Un autre exemple d'extension artificielle du champ statistique est le suivant :

Avant 1933, la statistique des automobiles circulant en France était liée à la perception d'un impôt trimestriel. Depuis 1933, un volet détaché de la carte grise est envoyé à un fichier central. Bien que la déclaration d'un véhicule détruit soit obligatoire, on comprend facilement qu'avant 1933,le propriétaire s'empressait de déclarer la mise à la ferraille tandis que maintenant on est beaucoup moins certain de cette déclaration (1). Un phénomène semblable de boule de neige se produit outre-mer. Dans les fichiers de la police, les arrivants se déclarent en général mais, au au départ, ils omettent de remettre leur carte d'identité. Ainsi, au 1er janvier 1952, le Togo accusait 1399 non-originaires présents, trois mois plus tard, le recensement ne comptait que 1086 personnes.

En A.O.F. les services d'Emission de la Banque d'Afrique Occidentale notaient les numéros des séries émises dans les différents territoires. Une fois par an, la B.A.O. examinait les vieux billets retournés à ses guichets et pouvait se rendre compte des transferts internes à l'A.O.F., elle rectifiait alors brutalement sa statistique.

(1) La nouvelle imposition fiscale de la vignette a peut-être "amélioré" les choses.

Billets en circulation A.O.F.

PAYS	Août	Sept.	Oct.	Nov.	Déc.
	Millions de Francs C.F.A.				
1951					
Sénégal, Soudan, Mauritanie	5391	4879	4647	4959	6516
Guinée	361	324	214	749	893
Côte d'Ivoire, Haute-Volta	12985	12695	12701	12452	13155
Dahomey, Niger	2723	2675	2683	2954	3279
Togo	1177	1182	1235	1250	1287
	22637	21755	21480	22364	25130
1952					
Sénégal, Soudan, Mauritanie	5976	5476	4923	8551	9780
Guinée	1055	1074	1067	1217	1373
Côte d'Ivoire, Haute-Volta	13956	13601	13731	10141	11499
Dahomey, Niger	3619	3493	3384	3527	3913
Togo	1445	1486	1529	1579	1720
	26051	25130	24634	25015	28285

Mais ceci n'avait pas grand sens .....

En A.E.F., les cartes d'abonnement de la proche banlieue de Pointe-Noire n'ont pas été comptées dans le trafic voyageurs du chemin-de-fer Congo Océan jusqu'en février 1952. En mars 1952, le nombre total des voyageurs transportés se gonfle subitement, sans que le nombre de voyageurs kilométriques progresse dans les mêmes proportions. La ventilation des voyageurs montre alors le changement de l'enregistrement.

1952	Voyageurs (1000)	Voy. Kilom. (millions)	Voyageurs (1.000) montée	descente	banlieue
Janvier	37	2,9	15	14	8
Février	40	3,9	16	15	9
Mars	130	4,2	18	17	95
Avril	124	4,0	17	16	91
Mai	129	3,5	15	14	100
Juin	126	3,9	16	16	94
Juillet	129	4,2	18	18	93
Août	125	4,2	18	17	90
Septembre	121	4,1	19	17	85
Octobre	106	3,5	16	16	74
Novembre	108	3,3	15	14	79
Décembre	109	3,7	17	16	76

Au Cameroun, une statistique du personnel des cadres généraux est dressée, mais en 1952, une rupture de définition intervient : les contractuels ne sont plus compris et on obtient la série ci-dessous :

T.P.	1951 :	270	1952 :	52
Enseig.t	1951 :	138	1952 :	39

Un autre exemple de l'extension du champ d'un ensemble statistique peut être donné par les opérations de la Chambre de Compensation des Banquiers de Paris :

On a tenté en comparant les volumes des règlements effectués en 1938 et en 1954, de déduire une évaluation de la perte du pouvoir d'achat du franc, or :

le nombre des participants à la Chambre est passé de 46 à 52, s'y ajoutent 124 sous-participants ;

les effets de commerce sont obligatoirement domiciliés depuis la loi du 4 septembre 1947 (il n'y a plus de tirage à domicile) ;

les prêts de banque à banque (open-market) se règlent depuis 1951 par la voie de la Chambre.

La modification du domaine d'étude est donc un phénomène fréquent, qui peut ne pas apparaître toujours très clairement :

Soit parce que la définition même du phénomène est très floue : les maladies mentales semblent plus développées chez les hommes que chez les femmes. Une difficulté apparaît dans l'interprétation des chiffres : il est vraisemblable que les hommes sont plus fréquemment décelés en raison de la responsabilité qui leur incombe et du handicap créé par ce désordre, tandis que les femmes peuvent plus facilement demeurer dans le milieu familial.

Soit parce que la réglementation administrative doit être parfaitement connue pour éviter les erreurs d'interprétation : dans les années 1956-1957, la répartition des recettes budgétaires des pays d'outre-mer appartenant au même groupe a été remaniée à plusieurs reprises (droits de sortie au crédit du pays, droits d'entrée au bénéfice du groupe, alors qu'antérieurement, tous les droits étaient inscrits au budget du groupe) il en résulte que l'observation de la structure d'un budget à cette époque exige des précautions particulières.

Le comble est atteint lorsque deux sources d'information sont disponibles et que la statistique semble porter sur le même domaine, alors qu'en réalité les deux séries concernent deux matières qui ne peuvent être confondues :

Devises procurées par le tourisme selon l'Office des Changes ou la Direction Générale du Tourisme (1).

Statistiques du bétail selon qu'elles proviennent des régies financières ou des services vétérinaires.

Statistiques des investissements privés d'après le dépouillement des journaux d'annonces obligatoires ou des calculs indirects des spécialistes du revenu national.

Effectifs des fonctionnaires selon qu'ils résultent des sources officielles .... ou de journalistes malveillants (2).

---

(1) Statistiques. Revue du Touring Club de France, page 81, mars 1954.

(2) L'I.N.S.E.E. avait ainsi présenté le recensement du 1er avril 1950 :

1.205.000 fonctionnaires à temps complet
59.000 fonctionnaires à temps incomplet
500.000 agents des collectivités locales.

Ces chiffres comprenant les offices et les services à caractère industriel et commercial et les militaires de carrière.

.../..

On voit sur des exemples précis comment les fausses interprétations sont dues

- à un désaccord sur les définitions

- ou à l'ignorance ou non de la véritable définition adoptée.

En résumé, il faut toujours agir en sorte que les nombres que l'on a à comparer concernent des faits identiques, que cette comparaison ait lieu dans les mêmes conditions et avec le même instrument, c'est à ce prix seulement que des observations valables peuvent être formulées.

Cette recommandation est fondamentale car la valeur de tout travail statistique est liée à la netteté des définitions de base.

2.4- Les méthodes d'approche ou d'observation des phénomènes.-

Une fois bien définie l'unité statistique à mesurer ou à dénombrer, il se pose le problème du choix de la méthode de rassemblement des observations.

Ces méthodes peuvent se ramener à deux catégories :

- les méthodes directes
- les méthodes indirectes.

2.4.1.- Les méthodes directes -

Les méthodes directes consistent à observer séparément chacune des unités à dénombrer. Cette observation peut revêtir divers aspects.

a) Elle peut être instantanée et générale, et avoir le caractère d'une photographie de l'ensemble observé, c'est un inventaire ou plus précisément, un recensement exhaustif.

(suite renvoi (2) de la page précédente)

Un article paru dans un journal du matin sous le titre "Le fabuleux train de maison de l'Etat" indique : 1.516.000 fonctionnaires non compris les militaires de carrière, les offices et les services.

Ce procédé est fréquent dans le domaine agricole, démographique, de l'industrie, des établissements commerciaux, etc...

b) Elle peut être instantanée, mais non exhaustive, c'est ce que l'on appelle une enquête statistique qui peut être :

b_1)- une simple investigation partielle dont les résultats ne sont pas valables pour l'ensemble statistique étudié. Ainsi parmi l'ensemble des marchandises importées, on peut étudier les marges pour les matériaux de construction, dans ce cas la marge moyenne observée ne sera pas extrapolable à l'ensemble des importations.

b_2)- mais on peut faire appel à un autre genre d'enquête partielle, c'est-à-dire procéder par sondage : Le schéma général est le suivant : soit une urne remplie de nombreuses billes en nombre N de différentes couleurs. Ces billes ne peuvent être toutes examinées en raison de leur grand nombre, mais on suppose qu'elles sont bien mélangées de telle manière que les couleurs soient bien réparties dans l'urne. On tire une poignée, c'est-à-dire un sous-ensemble de billes et on examine leur couleur. Le problème est le suivant : que peut-on déduire d'après la composition de l'échantillon qui vient d'être tiré de la composition totale de l'urne ? Cet aspect de la statistique inductive, on le sait, a de nombreuses applications : contrôle des fabrications, sondage d'opinion publique, prospection de marchés, enquête sur l'emploi et les salaires, etc...

Cette méthode comporte de nombreuses variantes, mais toutes consistent à choisir pour une enquête un certain nombre d'unités d'après un principe mécanique quelconque indépendant du but de l'enquête. Préciser que le tirage est mécanique, signifie le tirage est indépendant de l'observateur.

b_3)- une autre méthode dite méthode représentative fait appel au contraire à l'initiative de l'observateur. Elle consiste à choisir un certain nombre d'unités statistiques de telle manière que l'échantillon obtenu présente pour ce qui regarde des éléments caractéristiques déjà connus par des enquêtes antérieures les mêmes moyennes et les mêmes proportions que l'ensemble.

Si l'on imagine d'entreprendre une enquête démographique partielle portant sur le quart de la population d'un district, on peut réaliser une enquête partielle sur les villages situés sur les pistes principales jusqu'à atteindre 25 % du total de la population de ce district. Ce serait une enquête partielle, mais on sent instinctivement qu'elle peut ne pas être représentative. L'hygiène et l'habitat peuvent être meilleurs sur les pistes, de plus les jeunes gens sont attirés plus facilement vers le chef-lieu et l'on observerait dans certaines tranches d'âge des déficits qui peuvent ne pas être valables pour le reste du district. Cette première méthode est à rejeter. On peut supposer alors que par des enquêtes antérieures on connaît le nombre moyen des habitants par village. On entreprendra un choix de telle manière que les villages, objets de l'enquête, constituent une espèce de réduction de l'ensemble de la population du district, eu égard à certaines caractéristiques connues à l'avance :

- vocation rurale ou urbaine des villages
- répartition ethnique
- rapport du nombre de garçons au nombre des filles
- taille moyenne du village, etc...

c'est là la méthode du choix raisonné, mais on s'aperçoit tout de suite de la difficulté, car la création de cet échantillon dépend de la connaissance préalable d'une certaine documentation qui peut être incomplète ou périmée et pratiquement il est difficile de se rendre compte de la qualité de ces informations préliminaires.

c) Enfin, dernière méthode directe, au lieu de mettre sur pied une organisation capable de photographier une situation donnée, on peut concevoir une organisation dont l'observation est continue et qui a un rôle d'enregistrement ; on a ce que l'on appelle un système de statistiques courantes. Dans ce cadre, rentre l'enregistrement de l'Etat Civil d'une population. - Les statistiques douanières constituent un autre exemple de statistiques courantes.

Déjà sur le plan de l'organisation, on s'aperçoit d'une différence fondamentale entre enquête instantanée et enquête continue.

Le premier cas, qui consiste en une observation photographique d'une situation, exigera un personnel important mais de brève utilisation. Dans le second cas, observation d'une évolution, d'un mouvement, on peut disposer d'un personnel moins nombreux, mais qui sera occupé la plus grande partie de son temps. Ce dilemme est particulièrement marqué dans le cas de la recherche des mouvememts naturels d'une population (enquête spéciale par interrogatoire ou état civil itinérant).

2.4.2- Les méthodes indirectes -

L'observation directe peut ne pas être possible, soit pour des raisons pratiques, soit pour des raisons particulières du moment (crédits budgétaires par exemple). Il faut recourir aux méthodes d'estimation.

L'estimation suppose qu'un caractère de l'unité statistique observé, que l'on ne peut mesurer, est lié étroitement avec un caractère déjà connu.

On connaît la population d'une région, on croît savoir la superficie cultivée en moyenne par personne et on estime la superficie cultivée totale.

Dans un pays du Sud-Est Asiatique, on croyait connaître la consommation de sel au total et par habitant, on pensait alors pouvoir calculer une estimation valable de la population. A une époque plus récente connaissant la consommation d'essence, le nombre de tonnes kilomètres transportées en était déduit. Les exemples pourraient être multipliés car cette méthode a été abusivement employée et elle doit être formellement déconseillée.

Les cas foisonnent pour lesquels cette méthode a conduit à des énormités :

- production totale de manioc au Gabon correspondant à une disponibilité de 20 Kgs par homme et par jour ;

- population de l'A.E.F. estimée à dix millions de personnes en 1910 et à quatre millions en 1920 ;

- rendement de l'igname estimé à Bouaké à cinq tonnes par hectare et s'élevant en réalité à douze tonnes ;

- superficie estimée de riz dans un cercle du Soudan.. ... supérieure à la superficie totale de ce cercle ;

- sous-estimation de 30 à 40 % des superficies plantées en café et cacao de la Côte d'Ivoire en 1956, amenant à des réévaluations de 400 à 600.000 hectares.

Un dernier exemple mettra en lumière le défaut de l'estimation qui est de ne pouvoir donner aucun renseignement sur l'évolution du milieu et sur sa variabilité ou son hétérogénéité interne.

En Haute-Guinée, on disposait en 1953 et 1954 de séries concernant la riziculture de plaine

		1953/54	1954/55
Superficie	hectares	22.000	22.000
Rendement	Kg/Ha.	800	800

Une enquête réelle réalisée en 1954/55 a donné les résultats suivants :

Superficie : 27.000

Rendement : 500

autrement dit, si le volume de la production de paddy, estimé par l'administration n'était pas grossièrement erroné, par contre les deux phénomènes de l'augmentation des superficies et de la médiocrité des rendements étaient passés totalement inaperçus (1).

o
o o

L'estimation est une solution de paresse et paralyse la recherche objective.

(1) La situation de la statistique agricole en France est identique, une enquête entreprise en 1956 sur les abricotiers a porté l'évaluation administrative de trois millions à près de quatre millions d'arbres.

Chapitre III

LE QUESTIONNAIRE

3.1 L'importance du questionnaire - Les types de questionnaires.

3.2 La préparation du questionnaire.

3.3 La rédaction du questionnaire.

3.4 La distribution des questionnaires et la collecte des renseignements par des enquêteurs spécialisés.

3.5 La vérification des bulletins. Le dépouillement sommaire.

" L'enquête doit avoir un but précis, qui
" puisse être expliqué aux enquêtés, et les
" questions posées doivent se rapporter à
" cette fin."

Frank YATES

3.1- L'importance du questionnaire - Les types de questionnaires.-

Une vingtaine d'étapes jalonne la vie d'un questionnaire depuis sa naissance, c'est-à-dire son élaboration, jusqu'à sa fin, c'est-à-dire la mise en archives ou au pilon. Au cours de chacune d'elles, les traquenards se multiplient sous le pas du chercheur. Aussi, ceux qui sont à la source même des informations doivent savoir que l'établissement d'un questionnaire soigné est une des conditions primordiales d'une enquête. Cette entreprise exige de la finesse, de la psychologie et évidemment du soin et de l'intelligence.

Le questionnaire est défini comme une demande écrite de renseignements adressée par un enquêteur à un informateur, (le questionnaire est aussi appelé : le document de base). Il s'agit donc d'établir une communication aussi bonne que possible entre l'informateur et l'enquêteur.

Le document de base joue en conséquence un rôle essentiel dans le déroulement de l'enquête. Il peut revêtir des formes très variées : On distingue principalement deux types de questionnaires :

- la liste collective
- la fiche unitaire ou individuelle.

- La liste collective est constituée par des lignes à la suite, sur chaque ligne, l'agent enquêteur (plus rarementl'intéressé) remplit les indications demandées. La liste collective est employée notamment pour les registres de population ou les cahiers de village.

- La fiche unitaire ne concerne qu'une unité statistique : bulletin de naissance, questionnaire d'établissement, fiche de paye, etc... Elle peut être remplie soit par un enquêteur, soit par l'enquêté lui-même.

Chacune de ces formes a ses avantages .... et ses inconvénients.

La première en général coûte moins cher, exige peu de place, mais elle se prête mal à l'étude d'une évolution, même à l'occasion de chaque recensement successif, on rebâtit la liste collective, au surplus ce document manipulé trop fréquemment se détériore vite. La seconde

(fiche unitaire) se prête plus facilement au classement et à la mise à jour, mais le prix de revient est plus élevé et dans certains cas, des renseignements qui seraient disponibles sur la liste collective disparaissent : structure de la famille.

Il peut être utile de prévoir un type intermédiaire analogue à celui utilisé au cours d'enquêtes démographiques récentes, qui permet d'obtenir des renseignements individuels détaillés et de posséder malgré tout des éléments d'information sur la cellule familiale élémentaire, c'est ce que l'on a tendance à appeler une fiche collective.

L'établissement d'un questionnaire pose chronologiquement différents problèmes :

- Préparation du questionnaire.
- Etablissement du questionnaire.
- Distribution ou diffusion du questionnaire.

3.6 - Préparation du questionnaire.-

L'étude préalable portera :

- sur le problème à traiter,
- sur les observations numériques recherchées,
- sur les moyens matériels disponibles,

car le volume des crédits et les effectifs de personnel conditionnent le nombre des questions à poser et le nombre d'unités statistiques à inventorier.

On rédigera à l'avance les brouillons des tableaux désirés afin de vérifier si les questions prévues permettent de les obtenir.

Pour déterminer les besoins des utilisateurs, on prendra contact avec les techniciens intéressés pour leur soumettre le projet pour critique. Il existe alors un danger qui est d'alourdir le questionnaire avec des questions supplémentaires. Ce traquenard est classique et des recommandations doivent être formulées au titre du volume du ques-

tionnaire : même en cas d'enquête monographique de village, il faut savoir se limiter et le questionnaire ne doit pas être un "fleuve".

Chaque question doit être pensée, pesée, et on ne doit pas inclure des questions seulement "pour voir" ce que cela peut donner : interrogatoire sur le nombre d'heures de travail d'homme par an, nombre d'heures passées au défrichement, etc...

C'est courir à l'échec d'une manière certaine. Il faut avoir à l'esprit qu'il vaut mieux poser mille fois dix questions que dix fois mille questions. Le responsable doit résister à toutes les pressions des éléments extérieurs qui souhaitent toujours l'insertion de nouvelles questions. Il y a un nombre optimum de questions à poser qu'il est fâcheux de dépasser (1).

On examinera aussi les questionnaires utilisés dans des enquêtes antérieures. Si ces questionnaires existent, même s'ils sont imparfaits, ils peuvent renseigner sur les besoins des utilisateurs, les lacunes ou les erreurs des relevés précédents.

3.3 - L'établissement du questionnaire.-

Une fois pris les contacts nécessaires, on dresse une première maquette du questionnaire. Son établissement, il faut le répéter, est d'une importance capitale car le questionnaire est le seul lien qui unisse l'informateur à l'enquêteur. Au point de vue rendement il ne faut pas hésiter à perdre plusieurs dizaines d'heures pour le préparer, car finalement l'économie réalisée pour le compléter sur le terrain et le dépouiller est infiniment plus importante que le temps passé au stade préparatoire.

Un questionnaire doit :

- être maniable,
- être correctement présenté,
- être clair,
- se suffire à lui-même,
- être adapté,

afin d'obtenir des renseignements sincères, exacts, précis.

(1) Cet optimum dépendant du genre d'enquête, de la forme du questionnaire, des facultés de l'enquêté, etc...

3.3.1- Il doit être maniable dans ce sens qu'un document de format trop important est à déconseiller. Le format 21 x 27 ou à la rigueur ses multiples est recommandé. Il faut éviter les documents comprenant de nombreuses feuilles, car en ce cas les risques de perte augmentent et l'abus des imprimés indispose l'enquêté et ralentit l'étude.

De même l'abus d'instructions auxiliaires est à proscrire.

3.3.2- La simple présentation a aussi son importance : la qualité du papier (couleur et forme) ne doit pas être négligé ; un espace suffisant est à prévoir pour chaque question, les lignes trop serrées sont à éviter. Enfin le document devra éviter de paraître trop rébarbatif ou trop administratif.

3.3.3- Il doit être clair. C'est l'étude préalable vue au paragraphe 3.2 qui a précisé les questions à poser, elles sont le fruit des confrontations des services intéressés. Aussi le questionnaire final est-il un compromis et des préoccupations divergentes peuvent apparaître dans le libellé des questions. Or, il faut se rappeler l'appréciation de BERTILLON "l'informateur est paresseux, bête et ignorant", aussi si le questionnaire doit, d'une part, mettre en confiance l'enquêté, d'autre part, le questionnaire doit répondre aux préoccupations suivantes :

Tenir compte de la qualité de l'informateur. Un questionnaire commun à toute une population, doit être simple, il faut éviter toute phrase susceptible d'être mal interprétée ; si on pose une question sur le nombre de mariages, le mariage actuel doit être compris dans ce nombre.

Demander des renseignements numériques, des faits positifs et non qualitatifs. Les questions doivent être énoncées de façon qu'il n'y ait pas de place pour des commentaires superflus.

Exiger de l'informateur le minimum de travail ; lui demander les renseignements sous une forme analogue à celle qui lui est habituelle pour d'autres besoins (dans une enquête en France sur les budgets de famille, les questions concernant le gaz et l'électricité reproduisaient les indications des quittances).

Poser des questions faciles à comprendre et suscepti-

bles d'une interprétation unique et immédiate. Une question qui revient à une alternative est de beaucoup préférable à deux questions différentes. En milieu africain ou malgache, on peut avoir intérêt à préciser dans les instructions une traduction en langue locale pour éviter toute déformation de la part des enquêteurs.

En ce qui concerne les emplacements réservés aux réponses, quand la réponse consiste en un choix entre un nombre réduit de réponses (Forme juridique d'une société, nature d'un établissement) on peut demander de souligner, de cocher un carré ou de barrer les mentions inutiles. L'idéal serait de prévoir des questions auxquelles on répondrait seulement par oui ou par non. Si le nombre des réponses possibles est considérable, on est forcé de demander à l'informateur de répondre d'une manière très explicite (profession individuelle).

Si le système de cocher des réponses toutes proposées ne peut être envisagé, on peut (questionnaires démographiques) prévoir en tête de colonne les diverses réponses. Ce système est valable quelque soit le mode de remplissage du document et il diminue le nombre de réponses inexploitables.

Nom	Prénom	Sexe	Age	Situation professionnelle
(1)	(2)	(3)	(4)	(Employeur (Isolé (Employé (Aide familial (5)

Ne pas abuser des transcriptions par initiales majuscules ou par abréviations.

Les rubriques "autres" ou "divers" sont à éliminer afin d'éviter de gonfler un poste passe-partout, apprécié particulièrement des enquêteurs embarrassés.

Si la réponse comporte une indication chiffrée, il ne faut pas oublier de spécifier l'unité (âge, nombre d'enfants, production, etc...). En matière de production agricole, on préfèrera l'indication de poids à l'indication du nombre d'unités.

Des <u>réponses de même type</u> appelle des transcriptions semblables. Dans une enquête démographique en Guinée étaient posées les questions suivantes :

durée de l'absence	si la valeur numérique ..... était	:	en mois à droite d'un pointillé
âge du recensé	inférieure à un an les ..... consignes	:	x à droite d'un pointillé
âge d'un décédé	étaient :	:	- 1.

Il est clair que ces notations auraient du être unifiées.

Le questionnaire enfin doit comporter des <u>recoupements internes</u>. Dans un questionnaire démographique, il sera intéressant d'être certain du nombre des enfants survivants d'une mère donnée ; si l'on fait suivre la ligne réservée à cette mère, des lignes de ses enfants non mariés, il sera utile dans la colonne observations, de préciser si les enfants vivants absents de la liste sont mariés, restés dans un village voisin, etc...

Au cours d'une enquête agricole, si l'on se trouve dans une zone d'itinérance des cultures, on devra veiller à ce que les défrichements de l'année soient bien précisés ainsi que les champs de femmes quand ils existent.

En matière d'enquêtes alimentaire et budgétaire, des questionnaires "articulés" sont indispensables : apports, achats, consommation familiale, ventes, dons, de façon à suivre le sort d'un article ou d'un produit quand il entre dans le circuit.

Sur la précision et la manière de poser les questions, on peut citer quelques exemples (1) :

- <u>Recensement de la population de la France en 1946</u>, le bulletin individuel portait les questions suivantes :

10 profession	on a obtenu des	fonctionnaire
11 situation	réponses de ce	fonctionnaire
12 lieu de travail	genre :	Préfecture de X

---

(1) Certains d'entre eux sont empruntés au cours encore inédit de Monsieur l'Inspecteur Général CHEVRY.

Il pouvait s'agir du garçon de bureau ou du chef de cabinet du Préfet.

- Questionnaire de la zone française d'occupation en Allemagne, relatif à la main-d'oeuvre dans l'industrie. On trouvait la question :

"Depuis quand êtes-vous en Z.F.O.?" Beaucoup d'enquêtés ont répondu : "depuis mai 1945", c'est-à-dire la date du début de l'occupation.

- Enquête française sur l'artisanat en 1942.-

"Exercez-vous votre métier seul ?" Certains enquêtés ont compris que "seul" était équivalent à indépendant sans associé, mais en employant des compagnons, alors qu'il fallait comprendre strictement seul.

Dans la même enquête, on demandait : "Quelles sont les différentes matières premières que vous employez ?".

Un boucher a signalé le papier et la ficelle, mais a omis la viande.

- Il y a lieu de se montrer très attentif sur la disposition même de l'imprimé afin d'éviter toute ambiguïté.

En 1946, on demandait lors d'un recensement outre-mer le lieu de naissance des personnes :

Lieu de naissance : Département .... (territoire d'outre-mer ou pays étranger).

Beaucoup de personnes nées en dehors de France, barraient simplement territoire d'outre-mer ou étranger, sans comprendre que dans ce cas il fallait compléter l'espace réservé au département, par la mention du territoire ou du pays.

En 1953, au cours des enquêtes pilotes précédant le recensement de la France la mention de la date de naissance était placée comme suit :

1. nom et prénom 2. Sexe
3. Date de naissance

L'expérience a montré la fréquente omission de la question 3.

Les questions doivent d'ailleurs être imprimées dans la forme même où elles seront posées, être de difficulté croissante et l'on doit avoir constante à l'esprit la préoccupation de la réaction de l'interlocuteur qu'il faut éviter de choquer dès le premier contact. Il faut au contraire, susciter la coopération des enquêtes et tenter même par une ou deux questions sans relation directe avec l'enquête de déclancher l'intérêt de l'informateur.

Sur l'ambiguïté de certaines questions, on peut encore citer : en 1942 encore lors de l'enquête sur l'artisanat, on demandait le métier et l'on donnait une liste :

Alimentation	Cuirs et peaux
Bâtiment	Métaux
Bois	Textiles
Céramique	Divers

or un métier pouvait être classé sous deux rubriques :

Serrurier à :	Bâtiments ou métaux
Charpentier, menusisier à	Bâtiments ou Bois
Sabotier à :	Bois ou cuirs et peaux
Encadreurs à :	Bois ou divers

Une enquête prévoyait en Nigéria une question sur la durée de la préparation et de la transformation de certains produits. On s'est aperçu au dépouillement que le temps de séchage au soleil était compris dans cette durée.

Deux derniers exemples pour terminer :

- Enquête agricole d'automne 1943 :

  Nombre d'exploitants agricoles
  Nombre de détenteurs d'animaux.

Il s'agissait de distinguer les exploitants agricoles qui sont tous en principe détenteurs d'animaux, des détenteurs d'animaux qui ne sont pas tous exploitants agricoles. Faute de l'avoir dit, on obtint souvent le même chiffre.

- Sur un bulletin individuel on demandait :

  "Combien avez-vous d'enfants ?

  Age en années révolues des enfants vivants ?"

Mais on avait oublié de demander "avez-vous eu des enfants ?" Ce qui ne permettait pas de distinguer les gens n'ayant pas d'enfants de ceux qui ne répondent pas à ces questions.

- On viemt de voir que l'on doit préférer les questions qui entraînent l'alternative oui ou non, toutefois, il faut s'assurer que l'alternative est bien posée ; il ne faut pas que les questions, sans être tendancieuses, semblent oriemter les réponses de l'enquêté.

Il faut donc éviter les tournures telles que :

Savez-vous que .....
Avez-vous songé que ...
Utilisez-vous telle ....

car l'enquêté peut répondre oui systématiquement pour ne pas paraître ignorant.

3.3.4- Le questionnaire doit se suffire à lui-même.-

C'est un point extrêmement important. Le questionnaire que remplit l'enquêteur ne doit pas être un tableau élaboré, mais un document élémentaire de travail, sur lequel toutes les informations de base ou primaires doivent être transcrites. On doit éviter les totalisations ou les multiplications sur le terrain, calculs dont on peut charger le service central de dépouillement. Si des totalisations doivent néanmoins être entreprises, il faut qu'elles soient verticales, les additions horizontales étant une source d'erreurs.

Si un document plus élaboré relève de la responsabilité de l'enquêteur, l'état ou les états élémentaires et le document élaboré comporteront des éléments numériques communs pour vérifications ultérieures. La non-observance de cette recommandation a eu pour conséquence que le questionnaire établi en 1950 pour le recensement mondial de l'agriculture, pour les pays africains a été très mal rempli : c'était un questionnaire global à l'échelon village au lieu de constituer un questionnaire d'exploitation agricole et il était pratiquement impossible de critiquer ou recouper les informations du document de base.

Il peut être utile, que les réponses soient indépendantes de la date d'établissement du questionnaire : si l'âge est demandé, il peut être préférable de noter l'année de naissance.

3.3.5- Le questionnaire doit être adapté.-

La technique de dépouillement influe sur la forme du questionnaire. Si le dépouillement est manuel, le nombre des renseignements à double entrée sera réduit au minimum. Si le dépouillement est mécanographique, il est indispensable que le service de statistique soit consulté sur le questionnaire (importance et emplacement des grilles de chiffrement).

3.3.6- On ne doit pas faire appel à la mémoire des personnes et les enquêtes rétrospectives présentent des inconvénients graves. L'exemple type est celui d'une enquête en Nigéria au cours de laquelle on demandait le coût de construction des cases depuis 1920. Il est classique de remarquer que lorsque l'on interroge les femmes sur le nombre total des enfants nés vivants, les vieilles femmes ont tendance à omettre certaines naissances.

Exemple d'oublis de la part de femmes âgées (1)

Age des femmes	Nombre d'enfants nés vivants par femme
14-19 ans	0,7
20-24 ans	2,1
25-29 ans	3,0
30-39 ans	5,4
40-49 ans	5,8
50-59 ans	5,5
60-69 ans	4,6
70 et plus	3,9

L'un des enseignements de l'enquête démographique de Guinée est que les informations les plus douteuses sont celles relatives aux naissances et décès intervenus dans les douze derniers mois. Les décès en particulier étaient souvent oubliés. En Cote d'Ivoire le phénomène est identique.(2)

Dans une enquête à Ghana (juin 1952) on suivait au jour le jour les budgets de recettes et dépenses courantes. On trouvait alors :

Recettes	5 L/ 15/IO
Dépenses	5 L/ I2/7

(1) Enquête faite à Nihien en Guinée forestière. 1955.
(2) Les Britanniques ont appelé "effet télescopique" la conséquence de cette erreur d'appréciation de temps, laquelle peut se compliquer parfois d'un phénomène d'oubli.

parallèlement par interrogatoire, on demandait aux habitants de donner une estimation des mêmes postes, mais sur la base annuelle cette fois, on a :

Recette	7 L/3/6
Dépenses	2 L/3/5

il y a là un écart difficilement imputable à l'épargne.

Une autre enquête à Ghana portait :

sur un interrogatoire global sur le nombre de sacs de cacao vendu depuis le début de la traite.

sur l'enregistrement jour par jour des ventes de cacao,

une fois la campagne agricole terminée on s'aperçoit que l'interrogatoire rétrospectif est entaché d'une sous-estimation de 40 %.

Cet écart systématique est d'ailleurs lié à l'importance de la récolte de l'exploitant :

Sacs vendus l'année précédente.	Déclarés par la famille D	Relevés hebdomadaires R	$\frac{D}{R}$
	nombre de sacs vendus		
- 25	10,3	25,3	0,41
25 à 54	33,0	41,7	0,79
55 à 124	68,6	66,7	1,03
125 et plus	183,5	116,3	1,58
Ensemble	24,5	35,3	0,69

Au cours de la même enquête une étude sur la consommation domestique démontre que les relevés au rythme hebdomadaire demeurent systématiquement inférieurs aux relevés journaliers.

	Sept.	Oct.	Nov.	Déc.	Janv.	Fév.	Mars
	Shillings par famille						
Relevas Hebdomad.	108,7	156,8	200,1	317,0	134,4	106,0	98,6
Relevés journaliers	111,9	176,8	236,2	396,6	144,3	109,1	101,5

Des exemples empruntés à des horizons très différents peuvent illustrer ce phénomène de l'oubli.

Divers échantillons d'individus sont interrogés sur les maladies qui les ont touché à un, deux, trois mois de distance. Les maladies sérieuses restent vivaces dans la mémoire des intéressés, le souvenir des maladies bénigmes s'estompe rapidement.

Type de maladie	Enquêtes conduites en	Mois déclaré de maladie		
		Octobre	Novembre	Décembre
		nombre de cas pour 100 personnes interrogées :		
maladie sérieuse :	Janvier	1,7	2,1	2,0
	Février		1,8	2,0
	Mars			2,0
Grippe	Janvier	1,5	8,5	14,2
	Février		5,2	10,7
	Mars			8,1
Rhumes	Janvier	4,9	9,8	19,3
	Février		8,3	14,0
	Mars			11,3

Source A.S.A. Journal Juin 1955, p. 349

En Inde, au cours du National Sampling Survey, en 1954 on demandait le nombre d'enfants nés vivants selon le sexe. Le taux de masculinité à la naissance croît régulièrement avec l'éloignement du mariage :

avant 1910	147
1910-1919	127
1920-1929	143
1930-1939	108
1940-1945	107
1946-1951	107

Pour en terminer avec ce problème des interrogatoires, on rapportera ici le résultat d'interrogatoires d'enfants des écoles de Dakar, auxquels on demandait le nombre de frères et de soeurs de leur famille (l'enfant interrogé étant compris dans ce nombre) :

Enfants interrogés :	63I	garçons : 5I2	filles : 119
nombre de frères		1.407	221
soeurs		85I	328

Source : communiqué par le docteur CANTRELLE de l'Institut Français d'Afrique Noire.

L'enfant se rappelle plus facilement les enfants du même sexe.

Toutes ces recommandations, et elles sont probablement incomplètes montrent que la tâche n'est pas facile, il faut se mettre dans la peau du personnage, or le statisticien ou le fonctionnaire à l'origine du questionnaire est saturé de son sujet. Il ne se trouve donc pas dans les conditions intellectuelles et psychologiques de l'informateur à qui est destiné le questionnaire.

3.3.7- Notes, définitions, explications.-

Un questionnaire, mises à part les questions, renferme d'autres indications : notes, définitions et explications ; elles peuvent être placées soit au début après le titre , soit sur une feuille séparée. S'il se trouve des questions particulières à préciser, les indications seront portées dans le corps du questionnaire. Il est primordial, que la personne soumise à l'enquête lise les instructions avant de commencer la rédaction des réponses.

La mention T.S.V.P. est nécessaire si le questionnaire est imprimé recto verso. Toutes les enquêtes et tous les tests montrent que rien ne doit être négligé pour que les instructions soient lues.

L'expérience montre d'ailleurs que c'est le travail réel sur le terrain et non une lecture intensive des instructions qui fait "rentrer le métier".

Ces instructions sont de trois ordres :

- Notes explicatives
- Définitions
- Directives.

Les notes explicatives préciseront l'autorité qui procède à l'enquête, éventuellement le but poursuivi et l'usage qui sera fait des documents recueillis.

Les définitions doivent porter sur les termes employés, l'unité statistique et le champ de l'enquête de telle manière que les différentes notions aient le même sens pour l'informateur et pour l'enquêteur.

Les directives donnent des renseignements sur le destinataire des réponses, le nombre d'exemplaires et les délais de retour. Dans l'ensemble il faudra intéresser l'enquêté, mais si la notice est trop longue, l'enquêté ne la lira pas. Si elle est trop courte, elle peut ne pas être comprise.

3.4- La distribution des questionnaires et la collecte des renseignements par des enquêteurs spécialisés.-

3.4.I- Organismes administratifs ou professionnels et distribution des questionnaires.

On peut passer par l'intermédiaire d'un organisme administratif se trouvant plus près des recensés : l'administration des communes par exemple. Mais si cette administration manque de moyens matériels, ou n'a pas bien saisi l'importance de l'enquête, cette méthode présente de graves inconvénients.

Si tout est cas d'espèce, on ne doit pas oublier que la publicité et la propagande auprès des organisations professionnelles est très importante. Il faut trouver des arguments simples et convaincants mais ... les résultats sont parfois bien décevants (recensement des entreprises du Sénégal en 1949).

Les questionnaires peuvent être envoyés par la voie postale et retournés par la même voie. C'est une solution de facilité, car en général, le lot des réponses n'est pas représentatif de l'ensemble interrogé. De plus, on risque que le questionnaire rejoigne la corbeille à papiers. Dans les pays d'Outre-Mer où le champ des enquêtes est parfois plus restreint, cette voie peut être employée, mais il faut à l'avance être assuré que la localisation des informateurs a bien été faite, que les adresses sont correctes et que le pourcentage des non-réponses ne sera pas important. Cette méthode est utilisée par exemple pour obtenir les renseignements trimestriels sur les salaires de la domesticité.

Si l'on passe par l'intermédiaire d'organisations syndicales ou professionnelles, il faut être renseigné sur leur importance, car seules, les organisations puissantes

rassemblent tous les assujettis. En général, les gens répondent plus facilement à leur organisation professionnelle ou à leur syndicat, mais il peut se poser la question du secret des affaires et les enquêtés préfèreront dans certains cas répondre directement à l'Administration ; il peut être avantageux de leur laisser le choix du canal de retour, il faut faire attention à ce que les voies de retour transitent par l'organisme expéditeur (une enquête sur la production industrielle a échoué en 1954 en A.O.F. parce que cette recommandation n'avait pas été respectée).

Il peut arriver que les entreprises ou les services mettent leur documentation à la disposition de l'organisme d'étude.

Toutes ces possibilités ont l'avantage d'être économiques, mais on est obligé dans la plupart des cas d'avoir recours à des enquêteurs spécialisés, dont l'intervention accroît la précision et l'homogénéité de l'enquête.

3.4.2.- L'emploi d'enquêteurs spécialisés.-

En 1952, la recommandation de l'emploi d'enquêteurs spécialisés au sein d'enquêtes ayant pour but une meilleure connaissance quantitative du milieu autochtone aurait pu sembler purement formelle.

Certains chercheurs avaient bien dans le passé, utilisé les services d'informateurs pour reprendre leur terminologie, mais seuls les territoires étrangers, tels les Rhodésies, le Ruanda Urundi avaient montré la voie. Dans les pays de la Communauté française, rien de systématique et d'important n'avait été tenté.

Aujourd'hui l'expérience est riche : Guinée, Cote d'Ivoire, Sénégal, Soudan, Togo, Cameroun, Moyen Congo, Oubangui, Madagascar, etc... ont été le théâtre d'opérations d'envergure, au cours desquelles des enquêteurs africains et malgaches ont été recrutés, formés et employés avec succès. Ils ont rendu les plus grands services et se sont revelés être des auxiliaires précieux : sans eux, en effet, des enquêtes de cette importance n'auraient pu être réalisées. Il faut se rendre compte d'ailleurs que pour entreprendre une investigation dans un certain milieu, ce sont des ressortissants de ce même milieu qu'il faut y envoyer pour obtenir des réponses sans erreur systématique.

Etant bien entendu que sélection, formation et encadrement de ces enquêteurs sont réalisés correctement.

3.4.3- La sélection des enquêteurs.-

Il n'existe pas de règle absolue, car les qualités requises sont liées au genre de travail auquel l'enquêteur est destiné : tel sera valable dans une enquête urbaine et sera inutilisable en brousse. Un autre sera capable de mesurer correctement des champs et non adapté pour participer à une enquête démographique.

Toutefois, on peut définir un certain nombre de qualités communes qui caractérisent un bon enquêteur :

- qualités sociales : la présentation, le tact, le contact humain sont des facteurs importants et parfois décisifs de succès. La surface personnelle de l'enquêteur doit lui permettre d'inspirer confiance et d'en imposer aux notabilités traditionnelles ou élues avec lesquelles il sera en relation. Il doit être persuasif, connaître les coutumes locales. On aura donc intérêt à retenir pour enquêter dans une zone donnée, des enquêteurs de la même race.

- qualités morales : il doit être consciencieux, tenace (on a vu des enquêteurs remplir "de chic" les questionnaires), ne pas se décourager au premier échec. Il ne doit pas modifier progressivement les instructions. Il doit être impartial afin de ne pas orienter les réponses. Enfin l'enquêteur doit être compétent, afin d'être capable de découvrir les réponses incomplètes, les incohérences et se tirer d'affaire, lorsqu'il se trouve en présence d'enquêtés évasifs ou malintentionnés.

Il doit connaître les buts de l'enquête afin de pouvoir les présenter convenablement. Avant de lancer les équipes sur le terrain, on a donc avantage à leur faire jouer un psychodrame, qui permet d'ailleurs de découvrir des arguments inattandus pour présenter les objectifs de l'étude.

Toutes ces qualités ne sont pas liées nécessairement aux diplômes acquis ou à l'éducation scolaire reçue et il n'existe guère de tests qui permettent de retenir à coup sûr les sujets répondant à toutes ces qualifications, d'autant qu'en même temps, ces enquêteurs ne devront pas être trop détribalisés afin qu'ils aient gardé le contact humain préconisé ci-dessus.

Pour sélectionner les candidats valables, il faut les mettre le plus rapidement possible sur le terrain. La sélection physique se fait sans beaucoup de retard, ensuite la

rencontre avec le village fait ressortir les éléments les plus sérieux et les plus efficaces. Il est difficile d'émettre un jugement définitif sur les mérites comparés des agents fonctionnaires, de ceux qui sont simplement recrutés sur place. Des fonctionnaires déjà anciens peuvent être extrêmement précieux, mais on s'aperçoit aussi que de jeunes auxiliaires ne possédant guère que le certificat d'études primaires, se révèlent être souvent de très bons enquêteurs.

3.4.4- La formation.-

Il faut répéter que la formation en classe n'est pas ce qui importe le plus, mais bien le travail sur le terrain. La répétition de l'opération réelle est l'exercice le plus utile, elle classe rapidement les candidats. Le travail matériel est source de perfectionnement du questionnaire, des instructions, de l'approche des habitants.

La formation débutera donc par des enseignements généraux :

- prise de contact
- présentation de l'enquête
- notation des observations
- manière de résoudre les difficultés
- etc...

Puis très rapidement les équipes seront implantées dans des villages pour entamer les travaux pratiques. Cette enquête préliminaire ou enquête pilote ("l'expérience du concierge") est indispensable. Elle met d'abord à jour des difficultés ignorées :

- problème des rationnaires dans une enquête alimentaire ;
- problème des ressources occultes dans une enquête budgétaire.

Cette enquête pilote permet aussi :

- de tester les réactions du milieu : on peut en rencontrer d'inattendues qui peuvent fausser systématiquement les résultats.

Dans certaines régions de la Guinée Forestière, il est interdit de mentionner le nom des morts, faute de connaître cette information, l'enquête pilote fit apparaître une nette sous-estimation des décès.

Un statisticien faisant personnellement une enquête sur les prix du marché de Potopoto (Brazzaville) remarqua que de nombreux acheteurs se pressaient derrière lui, les commerçants le confondant avec un contrôleur des prix, avaient baissé leurs barêmes de vente.

L'université d'Achimota (Department of Economics) avait organisé une enquête à un bac sur les migrations. Le but était d'étudier les mouvements de salariés, en conséquence les commerçants n'étaient pratiquement pas interrogés. Faute d'enquête pilote permettant de trouver un "filtre" efficace, 50 % des migrants se sont déclarés commerçants (40.000 personnes interrogées).

- de détecter les erreurs d'interprétation du questionnaire et de l'ajuster progressivement (le questionnaire sur le recensement métropolitain de 1954 a subi plus de dix modifications successives).

- d'améliorer la qualité du questionnaire.

- d'aguerrir et rôder le personnel, car l'assimilation des instructions est un long apprentissage et de longues heures sont nécessaires pour mener à bien cette formation. A la fin, le succès couronne les efforts : A Bouaké, le nombre de questionnaires mis entre les mains des enquêteurs dépassait la dizaine .... et ils étaient correctement employés et remplis. Certaines enquêtes dans la vallée du Sénégal comportaient des instructions relativement complexes. Il faut alors insister sur le fait que l'encadrement doit payer de sa personne et montrer l'exemple.

3.4.5- L'encadrement.-

La présence d'un personnel d'encadrement est nécessaire, il faut que le personnel enquêteur se sente surveillé, contrôlé et aidé. Si ce personnel n'existait pas, l'enquête pourrait peut-être démarrer convenablement, mais très vite, il y aurait mauvaise interprétation des instructions reçues et le travail sur le terrain deviendrait rapidement défectueux.

Le hasard même des contrôles peut faire découvrir l'erreur grave :

- carrés mesurés au pas et non à la chaîne ;
- pesées de riz en gerbes et non en paddy ;
- nombre de naissances totales confondu avec le nombre de naissances dans les douze derniers mois.

Par ailleurs, il est préférable que le contrôle ne soit pas entrepris directement par les fonctionnaires responsables de l'enquête : il n'est d'ailleurs pas possible pour un seul fonctionnaire de vérifier simultanément et avec efficacité le travail des cinq ou six équipes qui sont sous sa supervision. Il est recommandé de prévoir un échelon intermédiaire composé par exemple de certains fonctionnaires africains et des auxiliaires les plus compétents. Un contrôleur pour deux ou trois équipes semble être une norme à conseiller. Les expériences de Côte d'Ivoire et de Guinée sont à cet égard probantes. Il faut noter que les contrôleurs doivent être associés à l'élaboration des questionnaires et des instructions (stage en deux temps en Guinée, élaboration progressive des formulaires en Côte d'Ivoire).

On comprend alors que cette période préliminaire de formation et d'essais puisse être fort longue. Il est préférable de retarder une enquête de quelques semaines plutôt que de lancer prématurément des équipes incomplètement formées sur le terrain. En Guinée en 1954 la formation a duré plus de huit semaines pour douze semaines d'enquête principale. Cette proportion n'est pas à copier, mais elle met l'accent sur l'importance primordiale de la formation des enquêteurs.

3.5 - La vérification des bulletins de dépouillement sommaire.

Le dépouillement sommaire préalable à l'exploitation des documents de base : A l'échelon enquêteur, cette vérification portera sur les erreurs visibles, erreurs involontaires ou ..... volontaires.

Aux divers échelons de la centralisation, le contrôle portera sur l'exactitude matérielle des réponses, sur la comparaison des divers résultats et sur la vraisemblance. Certaines lacunes pourront être comblées : le sexe par exemple ou certains chiffres parus au premier abord illisibles. Mais le contrôle de vraisemblance a postériori, s'il permet de déceler les erreurs flagrantes ne fournit pas les raisons des discordances et des incohérences. Il est surtout trop tard pour recommencer l'enquête sur le terrain. Au Dahomey, en procédant au dépouillement d'un recensement agricole en 1951 dans une zone où la superficie cultivée par habitant était assez faible : 47, 18, 22 ares par personne, on relève subitement un village pour lequel la superficie moyenne par habitant s'élève à 341 ares. L'anomalie saute aux yeux, lors du dépouillement, mais il est trop tard.

Il est donc conseillé de tenter de dépouiller au fur et à mesure que l'enquête se déroule. A l'examen des résultats numériques provisoires, il est alors possible de redresser les erreurs, d'attirer l'attention des enquêteurs sur certains détails :

Enquête démographique : natalité, mortalité dans les douze mois,
recensés de un an révolu et de moins de deux ans,
répulsion ou attraction des âges ronds.

Enquête budgétaire : couverture des dépenses par les recettes

Enquête agricole : superficie moyenne par habitant ou adulte actif.

En même temps que se poursuivent contrôle et dépouillements provisoires, il y a lieu d'entreprendre le dépouillement définitif. Le premier travail consiste en un tri, une mise en ordre, un classement et le pointage des dossiers.

Le tri permettra l'élimination des documents inexploitables. On tentera de compléter les manques grâce à toutes les informations disponibles.

Les autres opérations consistent en une préparation générale des documents de base afin d'être certain que tous les documents ont bien été retournés au bureau central et peuvent être transmis à l'atelier de dépouillement.

BIBLIOGRAPHIE

Les ouvrages sur ce chapitre sont rares, en raison du caractère très pragmatique de la matière. On pourra consulter les documents suivants :

- Enquête démographique en Guinée : Instructions aux contrôleurs (voir en particulier les batteries de tests).

- Manuel pour la formation d'agents recenseurs dans le cadre d'une étude démographique par sondage dans un pays sous-développé.

- La formation du personnel d'enquête
- La statistique agricole africaine - Section 7 le questionnaire.

) Centre africain d'application de statistique agricole - Ihadan juillet -septembre 1953.

- Manuel d'enquêteur pour le recensement de l'agriculture en Turquie - 1950 (en anglais).

- Manuel de l'enquêteur "Current Population Survey". Monthly report on the labor force U.S. Department of Commerce.

- Colonial Agriculture Statistics. The organisation or field work. Part II Administrative work - par K.E. Hunt - Colonial Research Publication n° 22.

Chapitre IV

LA MISE EN ORDRE DES OBSERVATIONS - LE TABLEAU STATISTIQUE.-

4.1 Classements qualitatifs et nomenclatures.

4.2 Classements quantitatifs.

4.3 L'exécution du dépouillement.

4.4 Le tableau statistique

4.4.I- Ordre d'un tableau
4.4.2- La construction du tableau
4.4.3- Signes conventionnels - L'arrondissement.
4.4.4- Titre et notes.

4.5 La précision.

4.6 La vérification.

" Deux consignes permanentes règlent l'établisse-
" ment d'un tableau : avoir du bon sens, et se
" mettre à la place de l'utilisateur".

J.C. CAPT Directeur du
"Bureau of the Census"

Chapitre IV

LA MISE EN ORDRE DES OBSERVATIONS - LE TABLEAU STATISTIQUE.

La mise en ordre des données recueillies, constitue la seconde phase qui suit le relevé des observations. Elle a pour objet :

La répartition en groupes c'est-à-dire en classes, des unités statistiques, ces groupes étant constitués soit d'après la présence ou l'absence de certains caractères, soit d'après l'intensité de ces caractères. Le classement est qualitatif lorsque le caractère n'est pas mesurable, quantitatif dans le cas contraire.

4.I- Classements qualitatifs et nomenclatures.-

Certaines classifications peuvent être très sommaires, la répartition d'une population selon le sexe, n'exige que deux classes. Il existe, par contre, des classifications plus diversifiées, qui requièrent des milliers de rubriques. Il est alors nécessaire de faire appel à un classement systématique, c'est-à-dire à une nomenclature.

4.I.I- Une nomenclature se présente sous la forme d'une liste ordonnée : à chaque terme sont attachés un ou plusieurs caractères. Cette nomenclature peut être, soit alphabétique, soit systématique. Dans ce dernier cas, les termes de la liste sont groupés par catégories, on dit aussi que l'on a affaire à une nomenclature analytique.

Il n'existe pas de nomenclatures qui satisfassent ni aux besoins de tous les utilisateurs possibles, aussi représentent-elles souvent un compromis entre les différents utilisateurs.

Pour être valable, une nomenclature doit être :

a) complète : elle doit prévoir tous les cas possibles et comporter une rubrique finale pour les statistiques qui n'ont pu trouver de place ailleurs : d'où l'utilité de la rubrique : "autre". Une rubrique "non déclaré" doit être aussi prévue.

Il faut veiller à ce que l'importance relative de ces deux rubriques demeure faible. Si un grand nombre d'observations figure sous les rubriques "autres" ou "non déclaré", on peut être certain que la nomenclature est défectueuse.

b) adaptée aux besoins. Il est des distinctions administratives qui n'ont pas d'intérêt. Avant 1939 le Ministère des Finances publiait le produit de la taxe sur le revenu des valeurs mobilières en deux catégories :

valeurs abonnées au timbre
valeurs non abonnées,

il eut été plus utile pour l'étude de la balance des comptes de séparer les valeurs françaises des valeurs étrangères.

Dans le domaine de l'outre-mer, on peut citer d'autres cas de nomenclatures inadaptées à l'analyse économique : l'ancienne nomenclature budgétaire, issue du décret de 1912, bloquait dans certains de ses postes les travaux et la main-d'oeuvre journalière. La nomenclature utilisée pour le Plan de Développement Economique et Social répond au même reproche : la présentation est telle que l'on ne peut, dans l'exécution d'un marché, faire le partage dans le montant total de certaines dépenses de matières premières des payements de salaires. Jusqu'à une époque récente, les dépenses publiques étaient distinguées selon le chapitre budgétaire, mais la nature économique de l'opération demeurait inconnue.

En matière de commerce extérieur, la fiscalité a une influence profonde sur la structure de la nomenclature. Des détails nécessaires du point de vue commercial ou économique ne figureront pas, parce que le tarif ne les discrimine pas (exemple : papier Kraft bloqué dans papiers et applications).

4.I.2- Pour faciliter la désignation des différents postes d'une nomenclature, on adopte généralement une numérotation décimale. A chaque poste, on fait correspondre un nombre :

1- Recettes fiscales
  1.1. Impôts directs
    1.1.1- Impôts forfaitaires sur le revenu
    1.1.2- Impôts proportionnels et progressifs.
  1.2. Droits de douane
    etc...

2- Revenus du domaine
3- Recettes des exploitations
    etc...etc...

4.I.3- L'unification des nomenclatures :

Même sur le plan national, le problème de l'unification des nomenclatures de même espère n'est pas une oeuvre facile. Pour étudier les échanges intérieurs de l'A.E.F., on doit étudier en particulier la répartition des importations ; le transport d'une partie d'entre elles emprunte la voie ferrée (Pointe Noire - Brazzaville) puis une autre partie emprunte la voie fluviale (Brazzaville - Bangui), - or les trois nomenclatures douanière, ferroviaire et fluviale ne sont pas identiques et ce fait gêne considérablement les études économiques. Pendant longtemps les statistiques démographiques (professions) et les statistiques du travail (activités individuelles) n'ont pas été présentées selon la même nomenclature.

Sur le plan international, la difficulté est multipliée. Aussi, pour permettre les comparaisons des statistiques établies par les différentes nations, on a tenté de mettre sur pied des nomenclatures minima unifiées. Les nomenclatures actuellement en usage sont les suivantes :

a) sur le plan national

- nomenclature des produits (commerce extérieur)
- nomenclature des activités individuelles (professions)
- " " collectives
- code des catégories socio-professionnelles qui complète la nomenclature des professions. Cette dernière nomenclature mérite d'être reproduite en raison de son intérêt :

Groupes	Principaux sous-groupes
0 Agriculteurs	Propriétaires exploitants, fermiers, métayers
1 Salariés agricoles	Salariés agricoles
2 Patrons de l'industrie et du commerce	Industriels, Artisans, Patrons pêcheurs, Commerçants
3 Professions libérales et cadres supérieurs	
4 Cadres moyens	Instituteurs, techniciens, cadres administratifs et non administratifs moyens
5 Employés	Employés de bureau et de commerce
6 Ouvriers	Contremaîtres, ouvriers qualifiés, spécialisés, mineurs, marins, manoeuvres

7 Personnel de service
8 Autres catégories Artistes, clergé, police
9 Personnes non actives.

Elle présente en effet l'avantage de regroupements particulièrement homogènes.

La nomenclature de la situation dans la profession doit être aussi connue :

P	patron	I	isolé	A	aide familial
S	salarié	AS	associé	ST	sans travail
				SP	sans profession

b) <u>sur le plan international</u> :

O.M.S. nomenclature internationale des causes de décès

Bureau de Statistique Nations Unies { classification type pour le commerce international (1)
classification type des activités collectives

O.I.T. classification type des activités individuelles

O.A.A. nomenclature des produits agricoles

Avant toute enquête, le problème des nomenclatures qui seront utilisées au dépouillement doit être posé, car il faut connaître parmi les nomenclatures existantes, celles qui seront employées, et de plus on doit songer aux adaptations nécessaires. Il y a intérêt lorsque l'on utilise une nomenclature à quatre chiffres de conserver les positions principales (sections ou groupes) et d'adapter les sous-positions. On peut, ainsi par regroupement manuel ou mécanique, avoir des bases de comparaison avec des travaux semblables. De telles adaptations sont fréquentes pour des codes de professions, d'ethnies, etc...

---

(1) Il existe aussi une nomenclature douanière et tarifaire mise sur pied par le bureau d'étude pour l'Union Douanière Européenne (1.300 positions). Cette nomenclature peut être étendue pour les nomenclatures nationales est dite de Bruxelles. La Direction des Douanes française présente également la statistique du commerce extérieur de la France selon la nomenclature de Bruxelles.

4.2- Classements quantitatifs.-

Le classement quantitatif est fondé sur l'importance numérique du caractère observé. Le nombre qui tient lieu de repérage numérique peut être entier : classement de mères d'après le nombre des enfants vivants par exemple : il peut être décimal : classement de manoeuvres d'après le salaire horaire, nombre de lingots d'or classés d'après leur poids, etc... Les unités statistiques peuvent être classées par importance croissante ou décroissante. Lorsque l'on a affaire à un grand ensemble de statistique, on est amené à les regrouper et à les classer par tranche ou par classe.

4.2.I. Limites de classe.-

Dans ce cas, les limites de classe doivent être bien définies. Des personnes sont classées selon la catégorie des revenus :

présentation correcte :	présentation incorrecte :
moins de 100.000 francs	moins de 100.000 francs
100.000 à 199.999 francs	100.000 à 200.000 francs
200.000 à 299.999 francs	200.000 à 300.000 francs

Cette faute est très courante, et il convient d'éviter la moindre équivoque sur le classement des nombres ronds 200.000, 300.000, etc...

Par ailleurs, le libellé des limites de classe doit tenir compte de la façon dont les valeurs du caractère sont appréciées : si l'âge est exprimé en âges ronds, il faut savoir si 25 ans signifie :de 24 ans, 6 mois à 25 ans, 5 mois, 29 jours, ou de 25 ans à 25 ans 364 jours ; car le centre de classe selon l'un ou l'autre sera différent ; (ce problème est généralement résolu de la manière suivante: les âges sont exprimés en années révolues).

Il est indiqué, en général, de constituer des classes égales : dans le cas contraire l'oeil est attiré par les classes de grande étendue qui contiennent des observations plus nombreuses. La statistique reproduite ci-dessous est un exemple à ne pas suivre (pour diverses raisons que le lecteur trouvera sans difficulté) :

Tranche de revenus	Nombre relatif (%)
1 à 5.000	5,7
5.000 à 15.000	23,5
15.000 à 50.000	45,0
50.000 à 100.000	17,7
100.000 à 300.000	7,65
supérieur à 300.000	1,2

4.2.2- Nombres ronds.-

La détermination des classes doit tenir compte d'un autre phénomène : l'attraction des nombres ronds. Si des recensés sont classés d'après l'âge qu'ils déclarent, on constate l'importance du nombre des observations pour les âges : 10 ans, 20 ans, etc... Les classes devront être constituées de telle sorte que ces nombres ronds se trouvent placés approximativement au milieu de l'intervalle de classe 0 - 4 ans, 5 - 14 ans, 15 - 24 ans, etc...

Cette attraction des nombres ronds est générale, on trouvera ci-dessous trois illustrations empruntés à un recensement à Dakar en 1955, une étude dans les maternités à Madagascar en 1951-1952 et le recensement de la population du Brésil en 1940.

Dakar			Madagascar		Brésil	
âges	hommes	femmes	âges des mères		âges	
9	1253	I396	- 15	29	18	482.704
10	1391	1553	15	226	19	382.983
11	825	882	16	1099	20	503.939
...	....	....	....		2I	332.I47
14	784	10I4	19	1755	...	
15	1244	1899	20	4I36	29	247.749
16	96I	1377	...		30	44I.9I6
...	....	....	29	I009	3I	I8I.256
19	1323	I579	30	2760	...	
20	2068	3327	3I	498	39	I80.462
2I	1604	1779	...		40	397.254
...					41	113.946
59	2I6	152	34	584		
60	592	672	35	1292		
61	195	158	36	686		

Le même phénomène se remarque si l'année de naissance est demandée à la place de l'âge. Il ne faudrait pas croire que cette erreur frappe seulement les statistiques des pays en voie de développement :

L'exemple suivant est inspiré d'une enquête britannique : on demandait à des enquêteurs d'interroger des personnes dont les noms n'étaient pas spécifiés, mais dont les caractéristiques : lieu d'habitation, sexe, âge, etc... étaient strictement précisées. En dépouillant les résultats, on s'est aperçu, alors que les âges finissant par 0,1,2 devaient

figurer en proportion semblable (10%), que le chiffre 5 figurait pour près de 17 %.

Le phénomène de l'attraction des nombres ronds se manifeste aussi lorsque la réponse à un questionnaire est le résultat d'une estimation à partir d'un calcul à priori : consommation annuelle de viande : 180 Kgs, 360 Kgs,etc...

4.2.3- Création des classes.-

Les classes doivent être constituées en tenant compte de deux remarques contradictoires :

Subdiviser le plus possible pour accroître l'homogénéité de chaque classe, mais garder un nombre suffisamment grand d'observations dans chacune d'elles. On souhaite étudier la production agricole, on a intérêt à subdiviser les résultats selon le climat, les zones de végétation, l'ethnie dominante, etc... mais, si la classification est poussée trop loin, le nombre d'observation par classe devient faible, donc dépourvu d'intérêt. Si on étudie la mortalité de zones géographiques trop réduites, on risque, d'une année à la suivante, d'observer des variations trop brutales dénuées de toute signification. De même, il est absurde de présenter une statistique de personnes selon leur âge, année par année, lorsque les effectifs sont trop faibles.

Il faut donc examiner l'ampleur du phénomène à présenter de manière à réaliser l'équilibre entre ces deux exigences ; ce n'est pas toujours facile, car trop regrouper risque de faire manquer des observations intéressantes.

Le dépouillement provisoire d'une enquête sur des budgets familiaux ne comportait pour l'alimentation que la distinction : produits locaux et produits importés. Il est bien certain que dans ce cas, le dépouillement était trop sommaire.

Dans d'autres cas, une analyse très fine peut être utile pour décéler des anomalies : Le recensement de la France de 1954 permet de dresser une statistique des communes selon leur taille. On s'aperçoit que les limites de 10.000 et 110.000 habitants sont le siège d' "incidents" :

9.500- 9.999	16 communes	100.000-109.999	2 communes
10.000-10.499	30 communes	110.000-119.999	7 communes
10.500-10.999	19 communes		

Il est vraisemblable que la réglementation administrative (représentation politique, régime fiscal, degré d'autonomie financière) a conduit les maires à forcer l'effectif de leurs ressortissants. Seule, une statistique détaillée pouvait faire découvrir ce phénomène.

Il est enfin recommandé de grouper les fins de séries afin d'éliminer les suites interminables de nombres très faibles.

4.3- L'exécution du dépouillement.-

Les moyens matériels dont on dispose ont une influence sur le dépouillement. Les monographies mises à part, il sera rare qu'un dépouillement puisse être entrepris en brousse et il est préférable de s'adresser au service statistique le plus proche.

L'exécution du dépouillement s'effectue soit par voie manuelle, soit par voie mécanographique. Ce paragraphe ne traitera que du premier procédé.

Le dépouillement manuel combine l'état en liste de dépouillement sur lequel sont reportées résumées ou codifiées les données de l'enquête, avec les procédés de partage successifs et de pointages.

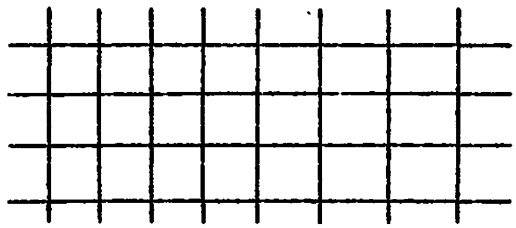
4.3.I- Etat en liste.-

Un exemple est donné ci-dessous d'un dépouillement d'une enquête agricole en Afrique Occidentale, il s'agit d'un simple état en liste sur lequel sont reportées les caractéristiques essentielles de chaque exploitation agricole.

N° village	N° exploitation	Composition de la population				N°champ ou plantation	Régime foncier	Superficie du champ ou de la plantation	Cultures portées		Année de défrichement	Durée antérieure de jachère
		Total	Actives dans l'agric. H/F	Actives en dehors de l'agric. H/F	Inactives H/F							

4.3.2- Partages successifs et pointages.-

- Dans le premier cas, on dispose de documents de base individuels : recensement par exemple ; on sépare les hommes les femmes, puis, si on souhaite ventiler les bulletins selon l'état matrimonial et par groupe d'âge, on place des casiers sur une grande table selon le principe suivant :



et chaque bulletin est selon les données déposé dans un casier.

On procède ensuite au comptage des bulletins de chaque casier.

- Dans le second cas, procédé du pointage, deux personnes sont mises à contribution. La première examine les documents et dicte les renseignements à une seconde personne qui côche sous forme de bâtonnets, on inscrit les données successives sur une grande feuille qui a l'allure du tableau définitif, mais qui comporte des cases plus grandes afin de permettre l'inscription de ces bâtonnets et leur totalisation ultérieure. Dans certains dépouillements, des nombres remplacent les bâtonnets.

Age de la mère	Nés vivants		Encore en vie	
	Masc.	Fém.	Masc.	Fém.
- 20	2,	2,1,	1,	1,0
2o à 24	3,1,2	2,	1,1,0,	0,

etc...

On peut craindre des erreurs de reports aussi à la place de bâtonnets, utilise-t-on parfois un signe rappelant un caractère de classement. Soit à classer des cases ou des habitations selon le nombre des habitants, L'enregistrement des données pourra être réalisé comme suit :

2 personnes	3 personnes	4 personnes
222	33	4.444

Remarque : Quand il n'y a pas seulement comptage de bulletins, mais sommation d'indications numériques précisée sur les documents de base, il faut reporter les renseignements primaires bruts sur des tableaux préparés à l'avance appelés tableaux de dépouillement ou tableaux de calculs qu'il ne faut pas confondre avec les tableaux statistiques proprement dits, appelés aussi tableaux de présentation ; si l'on conserve toutes les informations disponibles dans le tableau, on a un état dit de référence. S'il s'agit d'un tableau destiné à mettre en valeur certaines données d'une étude, il s'agira d'un tableau à but spécifié (documentaire ou descriptif).

Dans les tableaux de dépouillement, les renseignements de chaque bulletin sont inscrits sur une ligne, de façon à pouvoir totaliser les valeurs de caractères portés dans la même colonne.

Ces tableaux de dépouillement peuvent être établie "en cascade" selon la complexité de l'organisation administrative ou le volume du dépouillement. Le dépouillement manuel d'une enquête de budgets de famille portera sur la catégorie socio-professionnelle du chef de ménage, le nombre de personnes par budget et les diverses rubriques de dépenses. Le plan de dépouillement devra permettre à la fois l'obtention de résultats partiels, intermédiaires et finaux.

La présentation des documents statistiques.-

La dernière phase de l'élaboration des statistiques est abordée ici, il s'agit de la présentation des résultats. Cette présentation est très importante, car elle oblige l'auteur à définir, à serrer de plus près le problème ; quand ce souci est négligé, il n'est rien de plus décourageant pour l'utilisateur que de ne pas saisir totalement et rapidement la signification du chiffre publié. L'effort de l'auteur est rentable, car plus la présentation est soignée, plus cette statistique est lue, et ce, dans le minimum de temps. Cette présentation peut être réalisée sous la forme de tableaux statistiques et de graphiques. La présentation par le texte est à déconseiller car elle est lourde, même lorsqu'il s'agit de présenter un nombre réduit de chiffres. L'oeil saisit mal les chiffres dans le corps du texte.

4.4- Le tableau statistique.-

Un tableau statistique doit être clair, il doit permettre de se rendre compte s'il y a association entre les différents chiffres présentés ; il doit pouvoir être utilisé par les professionnels de la statistique et par le chercheur occasionnel. Pour ce dernier particulièrement on devra s'attacher à ce que le sens du tableau statistique soit sans ambiguïté.

4.4.1- Ordre d'un tableau.

Les tableaux de présentation ou de classification sont dits de différents ordres, selon le nombre de variables présentées :

1er ordre : population de Madagascar de 1921 à 1959.

2ème ordre : population de Brazzaville selon le sexe de 1900 à 1956.

3ème ordre : activité des équipes médicales du Cameroun en 1939 et 1949 selon le type de formations et le résultat des visites et dépistages.

4ème ordre : population de la Côte des Somalis selon le sexe, l'âge, la situation matrimoniale et la nationalité abrégée.

La complexité de la présentation et de l'interprétation s'accroît très rapidement avec le nombre des variables mises en jeu, et l'oeil a de plus en plus de difficultés à saisir les chiffres importants ou les liaisons principales. En général, on se limite à quatre variables.

4.4.2- La construction du tableau.-

Les buts des tableaux doivent être définis :

- nombre des tableaux ;
- forme générale (répartition des observations selon les classes choisies) ;
- titre ;
- notes annexes.

On précisera ici les principes fondamentaux d'un tableau à deux variables (appelé aussi tableau à double entrée) :

- l'entête de la marge à gauche correspond à la désignation du contenu des lignes.

- le haut de l'entête présente la désignation du contenu des colonnes. Nécessité de la mise en page et respect d'un certain équilibre dictent le choix en lignes ou en colonnes de l'une ou l'autre variable ; cependant, si l'on doit présenter une série d'observations dans le temps, les années seront précisées (pour chaque ligne) dans la première colonne. Il en est généralement de même pour des fréquences que l'on a intérêt à présenter les unes au-dessous des autres.

- On utilisera de préférence le format à la française et non à l'italienne, le format 21 x 27 et ses multiples.

- Ne pas oublier le libellé dans la cartouche en haut à gauche.

L'ordre des intitulés peut :

- être alphabétique. En cas de données goégraphiques, on peut adopter l'ordre Nord au Sud, Est à Ouest, etc... Mais un ordre étant déterminé il faut s'y tenir dans le cours de tout l'ouvrage.

- lié à un ordre de grandeur, dans le cas d'une variable prenant des valeurs numériques.

- être chronologique, si les données numériques sont associées à des époques.

- être analytique, dans le cas d'une nomenclature.

Dans l'élaboration du tableau, il faut prévoir les totaux marginaux. Les unités employées ne doivent pas être oubliées (ne jamais écrire "en milliers de francs", mais simplement "milliers de francs").

Les recommandations suivantes doivent être formulées au point de vue présentation :

utilisation de - caractères différents : on aura intérêt à adopter un détail de caractères typographiques (gras : totaux généraux, romain : totaux partiels, italique : pourcentages) ou des décalages, afin d'éviter des doubles emplois dans les totalisations.

- traits de séparation ou filets : gras, pointillés ou doubles.

- renvois et notes en cas de chiffres anormaux.

- espaces : toutes les cinq lignes en cas de tableau compact (série chronologique par exemple).

- Si les chiffres sont importants, les séparer par groupes de trois.

- Le tableau doit être lisible - Dans le cas de tableaux récapitulatifs, il peut être plus élégant de répartir les observations dans chaque case, avec une légende spéciale, précisant la signification de chaque nombre selon sa position.

Superficie (1.000 has)	
Rend.t kg/ha	Prod. T

- Si l'étude comprend plusieurs tableaux de contenus semblables, ils doivent être présentés selon le même modèle ; la publication gagnera en homogénéité et en logique : soit deux tableaux présentant l'évolution dans le temps d'indices boursiers et d'indices de prix, tous les deux devront spécifier les dates de la même manière.

La numérotation des lignes et des colonnes facilite les citations. Si un tableau exige l'emploi de plusieurs pages, on renouvelle les entêtes des colonnes.

Si un tableau nécessite un dépliant ou est publié sur deux pages, il est recommandé de répéter les titres des lignes et leur numéro à l'extrême gauche et à l'extrême droite.

4.3.3- Signes conventionnels. L'arrondissement.-

Il ne doit pas exister de case vide (sauf dans le cas de séries chronologiques pour les périodes récentes).

En 1936, l'Institut International de Statistique a étudié la normalisation des tableaux statistiques. En France, ce système, avec quelques variantes, a été adopté et les signes conventionnels utilisés par l'I.N.S.E.E. pour expliquer les absences de chiffres sont les suivants :

" : ne peut être obtenu ou n'existe pas, faute d'enquête.

/// : il n'est pas question d'inscrire un nombre en raison de la nature de la chose.

... : renseignement fait défaut parce qu'il n'est pas publié ou n'est pas encore parvenu.

\- : rigoureusement nul.

ε (epsilon) : nombre différent de zéro, mais inférieur à la moitié de l'unité choisie.

* : nombre provisoire (p).

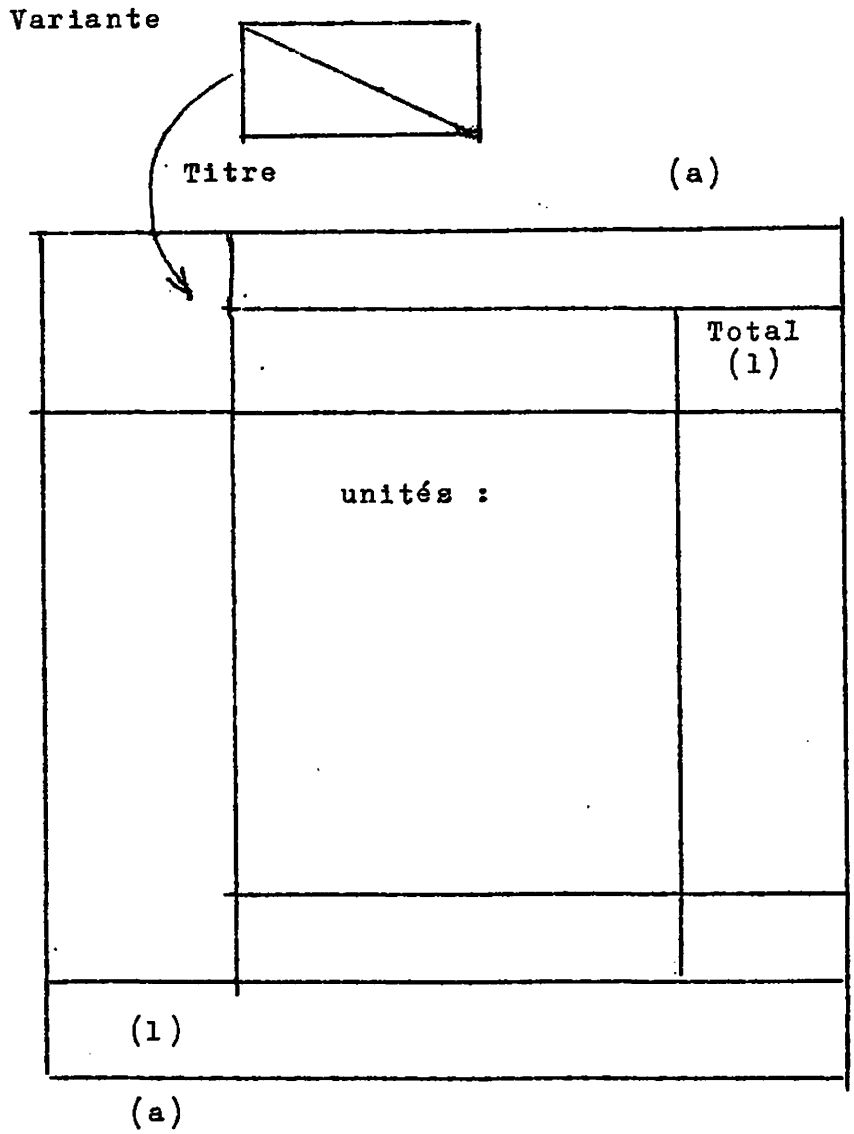
.. : nombre rectifié (r).

→ : valeur sans changement.

A cette occasion, il convient de rappeler la règle de l'arrondissement : arrondir en milliers, est supprimer les unités, les dizaines et les centaines. Si le chiffre des centaines est inférieur ou égal à 4, le chiffre des milliers est gardé ; s'il est égal ou supérieur à 5, on majore de 1 le chiffre des milliers.

4.4.4- Titre et notes.-

Le titre doit comporter le numéro du tableau, résumer le contenu du tableau ; on soulignera les termes importants afin que sautent aux yeux les éléments les plus intéressants ; ne pas utiliser dans le titre le terme "répartition", car un tableau est toujours une répartition ; la numérotation des notes doit se suivre de gauche à droite et de haut en bas ; les notes elles-mêmes se trouvent : dans le cadre, si elles se rapportent à un élément situé à l'intérieur du tableau, en dehors du cadre si les notes se rapportent au titre. La source du tableau doit être indiquée en bas et à droite à l'extérieur du tableau.



Les notes peuvent rappeler les <u>définitions</u> principales. Ce détail est utile pour le lecteur inattentif qui a négligé de lire ces précisions dans le texte. Si des tableaux différents sont publiés sur le même sujet, mais issus de <u>sources différentes</u>, on doit s'assurer que les <u>définitions</u> sont <u>identiques</u> ; dans la négative il est indispensable d'expliquer par avance les divergences éventuelles que pourrait constater le lecteur.

4.5 - La précision.-

Les chiffres publiés n'ont pas besoin d'être aussi précis que ceux d'un grand-livre de caisse. D'une part on cherche à connaître l'ordre de grandeur caractéristique d'un phénomène. D'autre part, l'erreur commise pour certaines enquêtes est largement supérieure à la précision du dernier chiffre caractéristique, aussi ne doit-on pas craindre d'arrondir largement les chiffres. Il y a, si l'on peut s'exprimer ainsi, un abus de la fausse précision. Lorsqu'un texte ou un tableau traîne des chiffres interminables, on peut dire que l'erreur est double :

- En premier lieu, les nombres sont marqués d'une approximation qui dépasse le degré de précision apparent que donne la présentation (Production de dizaines de milliers de tonnes exprimée en kgs).

- De plus, cette précision est nuisible à l'étude, car ce qui constitue le phénomène, ce n'est pas la tonne ou le franc en plus ou en moins, mais au contraire la masse élaguée des "bavures". Ce qu'il est intéressant de suivre, ce sont les grandes fluctuations atteignant un certain pourcentage de la moyenne, et non pas les petites fluctuations accidentelles. Enfin, dans tout tableau élaboré, on risque de voir figurer des données numériques de précision inégale, or c'est évidemment la plus mauvaise précision qui commande (1).

Il faut également se rendre compte de la précision qu'implique la publication d'une donnée numérique accompagnée d'un certain nombre de chiffres significatifs : en 1946, le gouverneur de la Guadeloupe publiait un indice du coût de la vie avec quatre chiffres après la virgule : 8I2,7255 ; autrement dit l'indice était donné au

$$\frac{1}{10.000.000} \text{ près !}$$

Si l'évolution du budget d'un pays est retracée, il ne faut pas hésiter à arrondir largement au million ou au milliard de francs, la lecture de l'étude en est facilitée. La même remarque est à formuler lorsqu'il s'agit de pourcentages : très souvent il est inutile de pousser après la virgule.

(1) Le tableau "Evolution en 1938 des quantités de brut traitées par les raffineries" (Etudes et Conjoncture, Economie Mondiale - n° 12 - 1947, page 6I), contient des quantités arrondies soit au millier de tonnes, soit à la tonne, le total général exprimé en tonnes est donc d'une précision très illusoire.

On rencontre certaines difficultés, dans l'application de ces principes à des tableaux à double entrée, car les arrondissements successifs risquent souvent de fausser le total général : aussi arrondit-on d'abord le total général, et modifie-t-on certains des chiffres des termes de l'addition, de telle manière que l'erreur relative soit la plus faible.

	Production	
	kgs	tonnes
A	14.540	14,6
B	3.587	3,6
C	800	0,8
Tot.	: 18.927	19,0

Il est à signaler que l'on a tendance de plus en plus à laisser tels quels les arrondissements partiels, au risque d'avoir une petite différence avec la totalisation générale. Le fait étant simplement signalé en tête du document.

4.6 La vérification.-

Le tableau achevé doit être vérifié. Cette vérification porte sur la répétition des calculs (l'idéal est la répétition systématique des premiers calculs par une seconde personne mais cette duplication est rarement possible). Cette recommandation n'est pas inutile, car trop souvent des documents officiels laissent passer des erreurs d'addition.

La vérification porte également sur :

- l'inversion des opérations.
- l'exactitude des additions des lignes et des colonnes.
- la vérification des pourcentages dont le total doit reconstituer 100.
- le contrôle avec des tableaux précédents du même genre.
- la vérification de certains éléments du tableau qui peuvent être répétés dans d'autres tableaux. Il faut éviter de trouver à quelques pages de distance, des chiffres différents relatifs au même sujet.

- la vérification indirecte : rapports entre salaires et nombre d'employés.

- le contrôle de vraisemblance.

Tranches de superficie	Nombre d'exploitations.	Superficies par tranche de superficie.
- de 3Has 3 à 4 Has 5 à 9 Has etc...		

Un tableau donne le nombre d'exploitations et leur superficie. Un contrôle de vraisemblance rapide portera sur la division de la troisième colonne par la seconde, et l'on vérifiera que le quotient est un chiffre compris entre les limites de la tranche.

- La dernière opération est la vérification des épreuves d'imprimerie et la suppression systématique des coquilles typographiques.

Remarque : Un certain nombre de termes doit être connu lorsque l'on est en relation avec des imprimeurs. On trouvera ci-dessous quelques indications utiles (1).

Format : Carré 45 x 56
Raisin 50 x 64
Jésus 55 x 72

Pratiquement, on utilise l'in-quarto dans le format carré :

Brut 22,5 x 28
Façonné 21 x 27

Force : Poids d'une feuille d'un mètre carré. Pratiquement la base des imprimeurs est le poids de la rame de 500 feuilles format coquille (44 x 56). Du Jésus 8 kilos correspond à un poids de 65 grammes pour une feuille de 1 mètre carré, la rame pesant $\frac{8 \times 55 \times 72}{44 \times 56} = 12,86$ Kgs.

Mesures typographiques : Un point égal = 0,376 mm.
12 points = 1 cicero = 1 douze = 4,5 mms.

(1) L'Association Française de Normalisation a systématisé les méthodes de classement des annuaires et la présentation des revues. Se reporter en particulier aux normes NF Z 44-001 et Z 4I-001.

BIBLIOGRAPHIE

Bureau of the census manual of tabular presentation
U.S. Department of Commerce - Washington D.C. 1949

Centre de Bingerville 1957 : conférence complémentaire sur les problèmes de dépouillement manuel (polycopié).

Enquête agricole sur la subdivision de M' Balmayo 1957 : annexes sur les états de dépouillement manuel.

Nomenclatures : Classification type commerce international M-10 N.U.

Classification type activités collectives M-4 N.U.

Chapitre V

LES MÉTHODES GRAPHIQUES

5.1 Caractère des graphiques.

5.2 Les diagrammes.

5.3 Les cartogrammes et les stéréogrammes.

5.4 Les graphiques triangulaires et carrés.

5.5 Les fautes à éviter.

" Il faut distinguer entre le graphique de
" propagande et le graphique destiné à l'étude,
" à la recherche scientifique".

Michel HUBER

L'étude méthodologique et critique des graphiques est pratiquement sans limite et les pièges qui guettent le rédacteur sont très variés. Ce chapitre sera consacré à quelques consignes simples relatives à leur établissement et à la revue des principales erreurs qu'il est conseillé d'éviter.

Les méthodes graphiques désignent les méthodes qui permettent de présenter par des grandeurs ou des figures géométriques, les observations statistiques.

5.I - Caractères des graphiques.-

Les graphiques ne remplacent pas les tableaux chiffrés, mais ils les complètent. Ils permettent, pour les gens pressés de saisir très vite l'ensemble du phénomène tout en négligeant certains détails. Autrement dit, les représentations graphiques répondent à trois préoccupations :

a) Souci de résumer d'un seul coup d'oeil un ensemble, et de faciliter sa compréhension.

b) Souci de contrôler, ou de mettre en évidence les régularités, de déceler les anomalies et les discontinuités. Un graphique, mieux qu'un tableau, permet de repérer sans hésitation une valeur pour laquelle existe un nombre anormal d'observations (1).

c) Souci de mettre en lumière des comparaisons.

Ce souci de faire apparaître certaines caractéristiques de séries numériques est aussi le défaut de ces représentations : un diagramme tendancieux ou mal établi est dangereux, car il peut fausser complètement un raisonnement.

Un graphique doit être simple, clair. Il doit aider à la découverte d'évolution, de différences ou de rapports, mais il ne doit pas être trop simple : on voit parfois un graphique composé d'une seule courbe ou de 4 ou 5 points,

(1) Voir exemples cités par Monsieur l'Inspecteur général E. MORICE dans "La méthode statistique dans les entreprises", Annales de la Faculté de Droit de Bordeaux, Série économique n° I-1955.

occupant une page grand format, on peut trouver à juste titre que c'est gâcher le papier et ... l'attention du lecteur (1).

Un graphique inutile peut être excusable dans un placard publicitaire ; il ne l'est plus dans un rapport administratif. Si l'on souhaite construire un graphique, il faut s'interroger sur son utilité et se demander si son insertion ajoute quelque chose au texte. Une fois sur deux, la critique personnelle démontrera que l'intention initiale ne doit pas avoir de suite.

On examinera successivement les diagrammes d'une part, les cartogrammes et les stéréogrammes d'autre part.

5.2 - Les diagrammes.-

Les divers types de diagrammes utilisés sont :

- Les diagrammes cartésiens, utilisant des échelles arithmétiques et des échelles logarithmiques.

- Les diagrammes à barres, secteurs, bandes concentriques, ou utilisant des surfaces.

- Les diagrammes polaires.

- Enfin, les diagrammes dits spéciaux.

5.2.I- Diagrammes cartésiens.-

On rappellera quelques notions élémentaires sur l'utilisation des systèmes de coordonnées.

On peut définir un point dans un plan par rapport à deux coordonnées rectangulaires par son abscisse Ox et son ordonnée Oy.

Si l'on désire représenter une série d'observations, il suffira que pour chacune d'elles, on puisse disposer des couples de valeurs abscisse-ordonnée (il est inutile de préciser à côté des points figuratifs les valeurs absolues qui ont servi à les construire).

---

(1) Certains rapports aux Nations Unies sont prodigues de telles représentations.

Cette figuration est particulièrement frappante lorsque l'on veut caractériser la variation d'un ensemble auquel sont attachées deux valeurs numériques (Commerce extérieur : importations et exportations en valeur).

Commerce extérieur du bloc africain tropical
en 1953 et 1955
(A.O.F., Togo, Cameroun, A.E.F.)

	1953	1955
	milliards F.M	
Importations	173	214
Exportations	145	174

Les figures V.I et V.2 précisent deux représentations possibles. On remarquera que dans le cas de la figure V.2, si on trace par l'origine un vecteur égal et parallèle au vecteur OF, on peut pour chacun des quatre secteurs du plan retrouver une propriété particulière de l'évolution du phénomène.

5.2.2- Diagrammes à barres.-

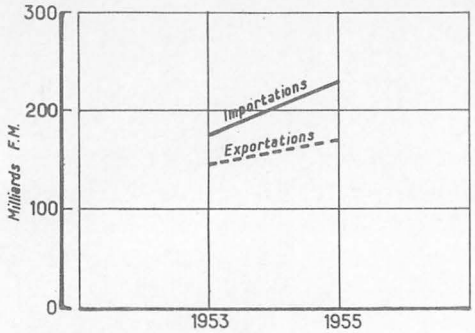
Dans cette représentation, chaque grandeur numérique est figurée par une surface rectangulaire proportionnelle à cette grandeur. Si l'on convient d'adopter pour chaque rectangle une largeur constante, à la grandeur numérique correspondra un rectangle dont la longueur lui sera proportionnelle (Fig. V.3 : Population de l'A.O.F. selon les pays constituant le groupe). Plusieurs diagrammes peuvent être placés côte à côte, soit dans le sens horizontal, soit dans le sens vertical (Fig. V.4).

On peut également combiner l'emploi de plusieurs barres, chacune d'entre elles étant la somme de plusieurs grandeurs intermédiaires (comparaisons de deux budgets de deux exercices différents - Commerce extérieur selon le groupe d'utilisation et la zone monétaire - Calendrier agricole Fig. V.5).

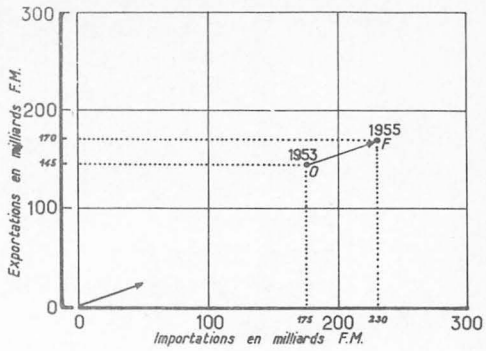
Les diagrammes à barres sont particulièrement utiles dans deux cas particuliers qui sont d'usage constant. Au lieu d'utiliser une échelle pour les abscisses et une autre pour les ordonnées, on emploie deux échelles issues d'une même origine, alignées, mais orientées de sens contraire.

COMMERCE EXTÉRIEUR DU BLOC AFRICAIN TROPICAL

Graphique V_1

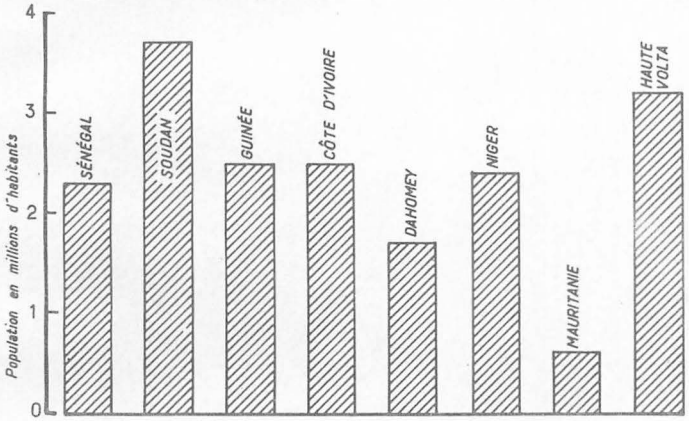


Graphique V_2



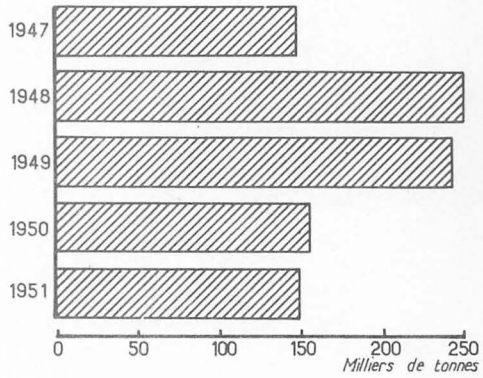
Graphique V_3

POPULATION DES PAYS DE L'A.O.F. EN 1957



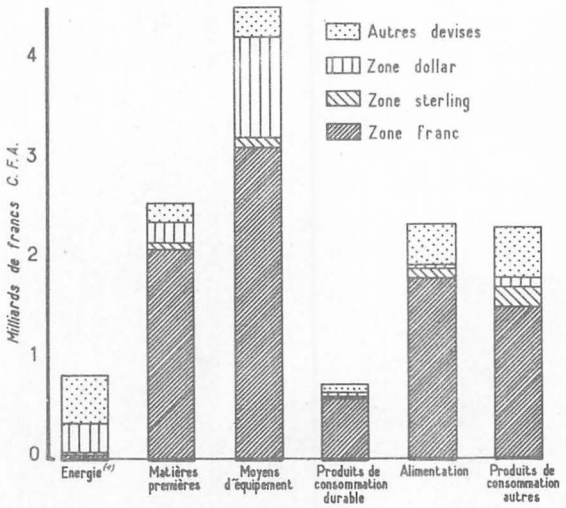
Graphique V_4

EXPORTATIONS DE MADAGASCAR (1947-1951)



Graphique V_5

A. E. F.
IMPORTATIONS EN 1950
PAR GROUPE D'UTILISATION ET ZONE MONÉTAIRE
(Valeurs)



(1)-Une part des produits pétroliers a pour origine les Antilles Néerlandaises, de ce fait, cette importation apparait dans la zone "autres devises". – En réalité, le devises qui servent au règlement financier de l'opération sont : soit le dollar, soit la livre sterling.

19.061_3_58

Dans un premier cas, où l'on emploie deux abscisses, on a une représentation particulièrement utilisée pour les pyramides d'âge (Fig. V.6). Une précaution essentielle, souvent négligée, est de s'assurer que les intervalles de classe sont bien égaux et de préciser si les effectifs sont ceux d'une tranche annuelle d'âge ou d'une tranche quinquennale, décennale, etc...

Le second cas est celui où sont mises en oeuvre deux ordonnées : commerce extérieur d'un pays année par année (Fig. V.7). Afin de faire apparaître le solde apparent de la balance commerciale, on dessine le symétrique et on ombre la différence pour faire apparaître la balance débitrice ou créditrice. Cette figuration est utile pour représenter la variation d'un budget.

La représentation par diagrammes à barres peut être lourde, on peut l'alléger en remplaçant le rectangle par un segment de droite égale longueur.

5.2.3- D'autres représentations dérivées des diagrammes à barres sont également utilisées :

- graphique à banderolle utilisé notamment par les journaux financiers pour représenter les valeurs extrêmes prises par la variable pour chaque période étudiée (fluctuation des changes, actions, etc...)

- diagramme cumulatif : on associe à une abscisse donnée, le nombre d'observations dont l'intensité est égale ou inférieure à cette abscisse. Ces diagrammes peuvent rendre de grands services lorsque l'on veut retracer la commercialisation d'un produit mois par mois depuis le début de la campagne agricole (1). Ce graphique peut être aussi utilisé pour suivre des rentrées fiscales (2). Dans ces exemples, seule la valeur représentée en ordonnée est cumulée, dans d'autres cas abscisse et ordonnée sont tous les deux cumulées : nombre d'exploitations dont la superficie unitaire est égale ou inférieure à une valeur donnée, et superficie totale de ces exploitations (Fig. V.8). Dans ce dernier cas, on peut porter en abscisse et en ordonnée

(1) Bulletin d'informations économiques et sociales de la statistique générale d'A.E.F. n° 31, janvier 1951, les impôts directs de 1948 à 1950, figures page 636.

(2) Commercialisation de l'arachide au Sénégal selon la semaine de traite.

- soit les valeurs absolues
- soit les valeurs relatives en pourcentage des totaux généraux (1).

5.2.4- Les diagrammes à échelle logarithmique.-

Les représentations graphiques jusqu'ici utilisées étaient telles qu'une grandeur double d'une autre était schématisée par un segment de droite d'une longueur double d'un autre. Etait appliquée la propriété de la graduation arithmétique : des nombres en progression arithmétique 3, 5, 7, 9 etc.... sont figurés graphiquement par des points équidistants.

Dans une graduation logarithmique, des points séparés par des distances égales représentent des valeurs numériques dont les rapports successifs sont égaux, ces points représentent des nombres en progression géométrique.

Si dans un système de coordonnées cartésiennes, seul un axe est gradué en échelle logarithmique, le graphique est dit semi-logarithmique. Le graphique V.9 montre que deux segments de droite de même pente sont représentatifs de variations relatives égales. En coordonnées cartésiennes arithmétiques, deux variations absolues égales seraient représentées par deux segments de droites parallèles entre eux (Figure V.9bis). Cette représentation permet de retracer sur un même graphique, une variable dont l'ampleur des variations peut être importante et elle permet aussi de faire figurer sur un même graphique des courbes dont les ordres de grandeur sont très différents (2). Elle permet aussi de se rendre compte directement de la valeur (ou de la variation) du taux d'accroissement. La pente de la droite (dans le cas où les points sont grossièrement alignés permet de mesurer précisément ce taux (3). L'erreur classique lorsque l'on a la mesure d'un phénomène à deux époques et que l'on souhaite calculer le taux annuel d'accroissement est de diviser la variation totale par le nombre d'années, ceci revient à surestimer parfois grossièrement le taux annuel réel d'accroissement :

(1) Voir mission socio-économique du Soudan rapport provisoire n° 2, page 16, $^{a}/_{s}$ distribution de la superficie des exploitations du delta du Niger.

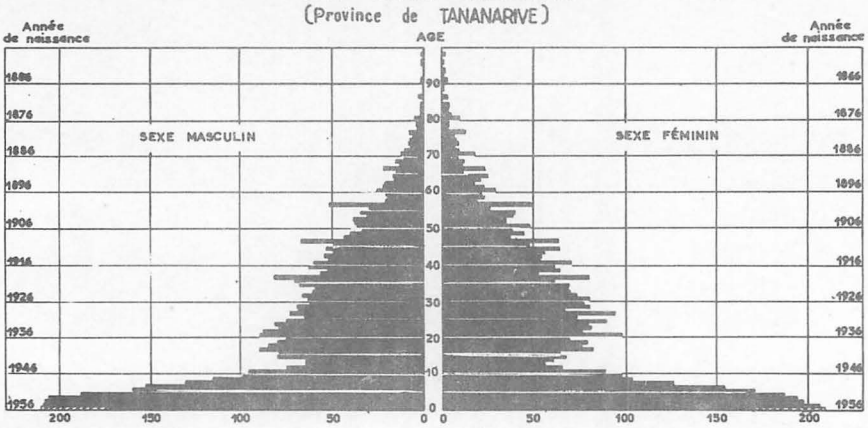
(2) On appelle module la distance entre deux puissances de 10. Le nombre de modules qu'il est nécessaire de réserver à la représentation graphique dépend des valeurs extrêmes prises par la variable.

(3) Ventes d'électricité en France de 1946 à 1958.

Graphique V_6

RÉPARTITION DE LA POPULATION POUR 10.000 HABITANTS

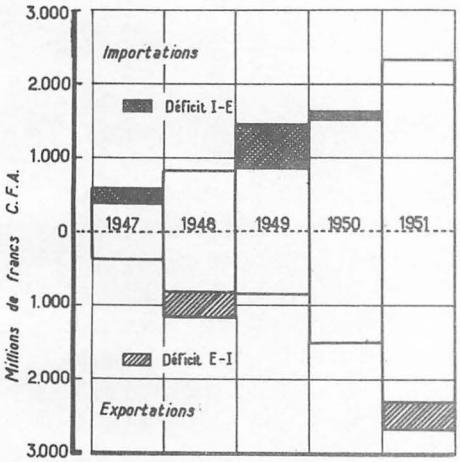
(Province de TANANARIVE)



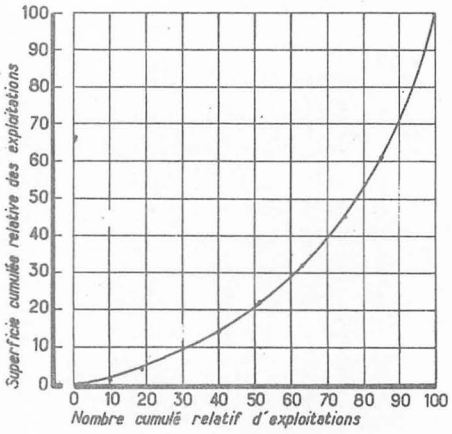
Bulletin mensuel statistique de Madagascar_ Août 1957

Graphique V_7

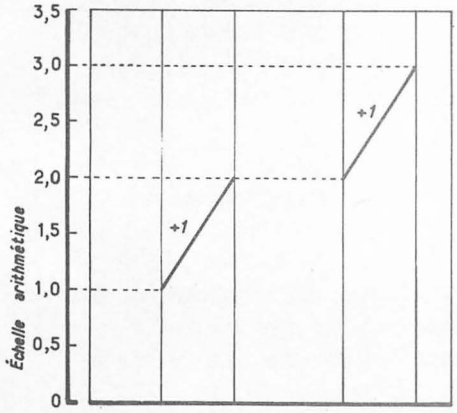
BALANCE COMMERCIALE DU TOGO



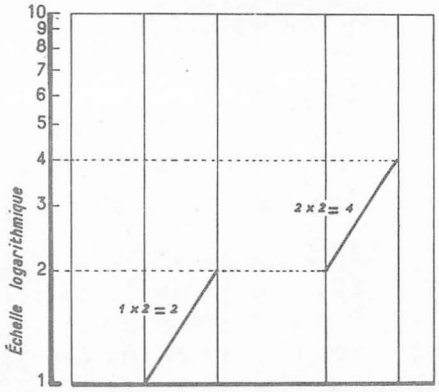
Graphique V_8



Graphique V_9 bis



Graphique V_9



19.061_4_58

La population de l'A.O.F. était en juillet 1926, de 13,5 millions d'habitants. Elle était évaluée en juillet 1956 à 18,7 millions d'habitants.

Le taux annuel d'accroissement n'est pas égal à

$$\frac{18,7 - 13,5}{30} \cdot \frac{1}{13,5}$$

En effet, si T est le taux d'accroissement moyen annuel au bout de la première année la population de l'A.O.F. est devenue :

$$13,5(1+T)$$

au bout de la seconde année : $13,5(1+T)\ (1+T)$ etc...

et au bout de la 30ème année : $13,5(1+T)^{30}$ qui est précisément égal à 18,7. Le taux d'accroissement sera donc défini par :

$$(1+T)^{30} = \frac{18,7}{13,5}$$

et en employant les logarithmes :

$$30(\log.1+T) = \text{Log}.18,7 - \text{Log}.13,5$$

(le même raisonnement s'applique à la recherche du taux d'intérêt tel qu'une somme S_o placée à intérêts composés, soit égale à une valeur S_1 donnée à l'avance au bout de n années).

La <u>représentation semi-logarithmique</u> présente l'inconvénient d'<u>aplatir les grandes valeurs</u>, aussi cette représentation employée à mauvais escient par un dessinateur malhonnête peut déguiser des augmentations en valeur absolue, parce que l'oeil ne sera influencé que par la <u>médiocrité</u> des augmentations en <u>valeur relative</u>.

Lorsque abscisse et ordonnée sont graduées en échelle logarithmique, on a un <u>graphique</u> dit <u>logarithmique</u> qui rend des services lorsque l'on a affaire à des courbes de la forme

$$y = ax^{\alpha}$$

Par une double transformation logarithmique

$$Ly = La + \alpha Lx ,$$

on est ramené en posant $Ly = Y$ et $Lx = X$ à une forme linéaire.

Cette représentation est fréquente en économie pour l'étude des lois de demande analysées par Alfred Marshall dans lesquelles les quantités Q demandées à un prix P sont liées à ce prix par une relation de la forme QP = Cte. (En coordonnées arithmétiques, la courbe a une allure hyperbolique qui est familière).

Une application importante du graphique cartésien logarithmique est celle où l'on représente la courbe cumulée du nombre de revenus dont le revenu est égal ou supérieur à une valeur x. La courbe représentative est de la forme

$$N = \frac{k}{x^{\alpha}}$$

et en coordonnées logarithmiques, on est ramené à

$$L\,N = \log k - \alpha\, L\,x.$$

Surtaxe progressive 1953 dans la métropole

Revenu net global annuel	Nombre (1000)	Montant de la tranche (milliards)	Revenu moyen par tranche (1000 frs)	Cumul du nombre des revenus (1000)
220.000 francs et moins	"	"	"	"
22I à 350.000 francs	729	20I	277	3098
35I à 600.000 francs	946	46I	490	2369
60I à 900.000 francs	805	586	730	1423
90I à 1.500.000 francs	425	474	1120	6I8
1.50I à 3.000.000 "	149	296	2000	193
3.00I à 6.000.000 "	33	131	4000	44
6.00I à 10.000.000 "	6	46	7500	11
plus de 10.000.000 "	5	51	10600	5
Total :	3098	2246		

Des répartitions justiciables de cette représentation ont été également observées pour des études de poids et d'âge de nourrissons et jeunes enfants (1), de population non primaire occupée et

(1) Economie Alimentaire du globe par CEPEDE, page 269.

de production industrielle (1) etc...

5.2.5.- Les diagrammes à secteurs -

Les valeurs numériques sont représentées par des surfaces proportionnelles aux grandeurs observées. Lorsqu'un seul cercle intervient, le paramètre est l'angle au centre. On peut figurer ainsi le commerce de la France à l'exportation avec ses différents pays clients. Ce procédé est un peu encombrant, car cette figuration, lorsqu'elle est claire fournit peu de renseignements, et si elle est surchargée devient vite illisible.

Un procédé dérivé du précédent est celui du diagramme à bandes concentriques. (Représentation de statistiques électorales : nombre de sièges par parti selon leur emplacement dans l'hémicycle).

Avant de passer aux graphiques spéciaux, il faut signaler enfin les graphiques qui utilisent des figurines (soldats, bateaux). Ils doivent être évités, car leur utilisation entraîne de lourdes erreurs.

5.2.6.- Deux graphiques spéciaux doivent être cités : graphiques de GANTT et de LANDAUER. Ils sont utilisés dans les ateliers pour contrôler l'utilisation des heures par tâche et la manière dont sont respectés les carnets de commande.

Le premier permet de comparer le rendement effectif au rendement théorique. Si un ouvrier a effectué les 3/4 de sa tâche journalière par rapport à la norme théorique, le trait ne couvrira que les 3/4 de la ligne totale.

Le graphique de LANDAUER nécessite l'emploi d'un "planning". Le haut du tableau est occupé par un emplacement millimétré, divisé en unité de temps. On affecte chaque bande horizontale à un atelier et on glisse des rectangles de carton dont la longueur est proportionnelle au travail réalisé ; grâce à des fils à plomb placés aux dates de livraison, on peut se rendre compte du respect de l'horaire, des retards et des avances. Ce tableau peut être utilisé d'une

(1) Etudes et conjoncture économie mondiale, mars-avril 1953, page 148. Voir aussi - Fichiers des établissements I.N.S.E.E. 1954, nombre d'entreprises selon l'effectif des salariés par entreprise.

autre manière : soient un métier à tisser et une commande, compte tenu du rendement du métier on coupe une bande de carton correspondant à la période d'utilisation du métier, si un retard est constaté, on pousse la bande vers la droite et la position de l'extrémité indique la nouvelle date théorique de fin de fabrication.

5.2.7- Les diagrammes polaires -

Il a été indiqué au paragraphe 5.2.I qu'un point du plan peut être défini par son abscisse et son ordonnée, il peut aussi, étant donnés une direction origine Ox et un point O appelé pôle, être défini par l'angle polaire et la distance OA, appelé rayon vecteur. Ce système est particulièrement utile lorsque l'on représente un phénomène périodique : la période étant couverte par une rotation complète du rayon vecteur.

Si on représente la circulation monétaire à Madagascar : dans un graphique normal à coordonnées arithmétiques, en période d'inflation, les années sont situées les unes au-dessus des autres, par contre en coordonnées polaires, on obtient une spirale qui se dilate progressivement autour du pôle. Cette représentation est utilisée dans le domaine des indices de prix. En météorologie on a une "rose" des températures.

5.3 - Les cartogrammes et les stéréogrammes.-

5.3.I- La carte statistique est familière : à une région géographique donnée, on associe un signe graphique, cette figuration diminue l'effort du lecteur et simplifie l'étude. On emploie les cartogrammes :

- à hachures ou à teintes
- à diagrammes de surfaces
- à épingles et rubans
- à images
- à courbes isométriques ou isogrammes.

a) La reproduction par hachures ou par teintes est de compréhension immédiate, mais lorsque cette méthode est employée, il ne faut pas abuser de la diversité des hachures, car on risque d'aboutir à une figuration illisible.

Sept figurations différentes semblent devoir constituer un maximum (en une couleur unique d'impression).

b) Le cartogramme à diagramme de surfaces attribue à chaque région géographique un rectangle ou un cercle dont la surface est proportionnelle à l'intensité du nombre représenté. Lorsque l'on représente des rectangles, il est recommandé d'utiliser des rectangles étroits de même base, la comparaison visuelle se basant automatiquement sur l'appréciation des hauteurs. Si l'on use de cercles, le rayon du cercle doit être proportionnel à la racine carrée de la grandeur à représenter.

Une variante est d'utiliser des petites surfaces circulaires de taille fixe mais plus ou moins découpées.

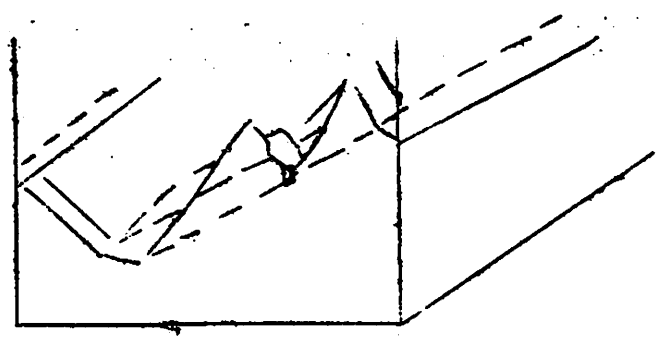
c) Les cartogrammes à épingles sont fréquents dans le commerce et la météorologie. Les cartogrammes à images sont parfois reproduits en géographie économique, mais ne sont pas à recommander.

d) Enfin les diagrammes à courbes isométriques sont constitués par des réseaux de lignes continues qui joignent des points de même intensité. Les cartes météorologiques constituent une bonne application de ces diagrammes.

5.3.2- Les stéréogrammes.-

Les graphiques dont il vient d'être fait mention font intervenir deux variables. Dans le cas où trois caractéristiques interviennent, on peut au-dessus d'un plan, matérialiser les points par des épingles d'inégales longueurs, on peut bâtir également des surfaces de fréquences :

Soit un tableau à double entrée, on peut constituer des prismes droits dont les hauteurs représenteront les valeurs portées dans les cases. Si les observations se multiplient, on peut diviser les classes en classes plus petites et la surface "en escalier" tend vers une surface continue.



Une application de cette représentation est fournie par la consommation d'électricité selon l'heure et le jour.

Une seconde application fréquente en démographie est celle des surfaces de mortalité en fonction de l'âge, du temps et du taux de mortalité par âge.

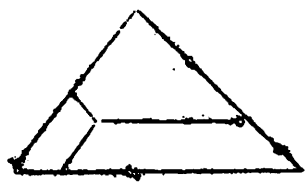
5.4 - Graphiques triangulaire et carré.-

Pour compléter ces généralités sur les graphiques, il faut citer deux graphiques particuliers : les graphiques triangulaires et les graphiques carrés.

a) Si on étudie des budgets de familles et qu'on divise les dépenses en trois groupes : Alimentation, Habillement, Autres dépenses, on peut calculer les pourcentages de dépenses de ces trois postes par rapport au total de la dépense familiale, on a l'identité :

% alimentation + % habit. + % autres = 100

En construisant un triangle équilatéral dont la longueur du côté est égale à 100, si on mène à partir d'un point intérieur M les trois parallèles aux côtés, la somme des trois segments est constamment égale à 100, inversement si on se donne trois pourcentages correspondant à chaque groupe de dépenses, le point M sera bien déterminé.



Ce système a de nombreuses applications lorsque la "trichotomie" est possible : élements d'un prix de revient : matières premières, main-d'oeuvre, frais généraux.

L'homogénéité des observations correspondant à diverses entreprises résultera de la répartition des points et de la forme du nuage.

Cette représentation peut caractériser des régimes alimentaires d'après la répartition des calories apportées par les lipides, les glucides et les protides.

b) Le graphique carré permet la représentation graphique d'une variable numérique qui est la somme de deux va-

leurs constituantes A et B et cela de deux manières différentes :

$$100 = A_1 + B_1 = A_2 + B_2$$

Cette méthode peut être utile dans l'examen de bilans, de budgets, etc...

Bilans	Actif	:	Immobilisation Fonds de roulement
	Passif	:	Fonds propre Passif à l'égard des tiers
Budgets	Recettes	:	Impôts directs Impôts indirects
	Dépenses	:	Fonctionnements Investissements

5.5 - Les erreurs à éviter.-

Après avoir vu ce qu'il convenait de respecter lorsqu'on a l'intention de complèter une étude ou un exposé par des graphiques, il faut passer en revue les principales fautes qu'il faut éviter dans la construction des graphiques.

5.5.I- Non représentation d'un phénomène réel.-

Un graphique doit être utile, c'est-à-dire qu'il ne doit pas être"creux" et il ne doit pas être trop touffu. Il faut aussi, dans un autre ordre d'idée, qu'il représente réellement un phénomène :

En 1947 l'Office Algérien d'Action Economique et Touristique publiait un graphique représentant l'exportation mensuelle de squamme de scille (plante bulbeuse utilisée en médecine) ; il est bien évident que cette exportation est sporadique et liée aux mouvements de bâteaux. Il en résulte que, d'une part, le fait même de construire un tel graphique est une erreur et que, d'autre part, la comparaison entre les exportations de chacune des deux années consécutives ne peut donner naissance à aucun renseignement utile.

5.5.2- Importance des échelles.-

Un graphique doit être honnête ; ainsi, en choisissant telles ou telles échelles pour l'abscisse et l'ordonnée on donne l'impression d'une progression brutale ou, au contraire, très lente. On doit faire très attention à cette remarque lorsqu'il s'agit de représenter la variation d'un indice des prix à la consommation familiale.

Si on utilise en ordonnée une échelle logarithmique, l'augmentation absolue d'un indice est mal perçue par le lecteur et ce dernier peut de bonne foi se tromper dans l'interprétation des données. L'indication du type d'échelle doit apparaître clairement.

5.5.3- Représentation par surface ou figures.

On utilise des surfaces circulaires pour représenter des grandeurs, et on oublie que deux cercles ont des surfaces dont le rapport est égal au rapport du carré de leur rayon. Or, on voit souvent figurer des valeurs numériques doubles par des cercles de rayon double, les surfaces sont dans le rapport 4 et le graphique est faux (1). Si l'on représente un secteur circulaire en perspective, il faut prendre soin de ne pas le décentrer.

Si on choisit des figurines (sacs de cacao) qui, en réalité, sont à trois dimensions, un sac dont la hauteur est double d'un autre sac signifie que les tonnages sont dans le rapport de 1 à 8 (2^3). Cette faute est très fréquente.

5.5.4- Axes interrompus.

Une autre erreur très répandue est celle qui consiste à couper et à remplacer par des pointillés les diagrammes à barres lorsque les valeurs à représenter deviennent trop grandes et que la figure représentative risque de sortir du cadre de la page. Le lecteur ne se rend finalement plus compte de l'ampleur du phénomène (2).

---

(1) Rapport de la conférence cacao 1951. Voir la carte relative au Cameroun intitulée : production sur les marchés contrôlés en 1950.
Se reporter également à la carte Agriculture du rapport aux Nations Unies de 1949.
(2) La série Etudes et Conjoncture présentait en mai 1947 (page 8) un graphique donnant par ordre décroissant une liste de pays selon la variation de leur circulation monétaire. Le graphique est inutile : un simple tableau aurait suffi. De plus certaines barres sont coupées et l'impression visuelle était faussée.

Une autre erreur est de faire partir l'échelle des ordonnées d'une origine différente de zéro (1).

Si on utilise des graphiques à colonnes, il faut prendre garde à ce que l'on appelle en dessin, la ligne de fuite ne sort pas trop inclinée, car, dans ce cas, les différences des niveaux s'apprécient moins bien.

5.5.5- Séries temporelles.

Il faut éviter de représenter des périodes inégales par des espaces égaux :

Représentation de la production de charbon en 1938, 1949, 1952 et 1958, les écarts des abscisses des points étant égaux.

Cours du coton année par année, puis mois par mois pour la dernière année : l'écart mensuel étant égal à l'écart annuel. Dans ce cas, il faut placer à part et dans une cartouche bien indépendante le graphique pour lequel les variations sont mensuelles (Voir aussi les graphiques publiés dans l' "Economist" du 10 septembre 1949).

Un cas particulier de représentation des séries temporelles est celui où deux séries se rapportent à des variations dans le temps de deux grandeurs de natures différentes : variation de prix et variation de tonnages échangés, variation de salaires et variation de production, d'effectifs, etc... On adoptera en ordonnée deux échelles distinctes et situées à chaque extrémité du segment qui porte la graduation des abscisses et l'on mettra autant que possible les ordonnées moyennes pour l'époque considérée en regard. Il convient d'être extrêmement consciencieux, car il est facile de fausser l'impression visuelle.

5.5.6- Titre -

Chaque graphique doit être numéroté et être accompagné d'un titre. Un graphique ne doit pas être une devinette. Dans le titre, éviter les termes : variation ou comparaison. La signification des unités en abscisse et en ordonnée doit être claire.

Les graphiques en coordonnées polaires ne seront pas trop chargés, car ils deviennent vite difficiles à comprendre.

---

(1) Inventaire du département de la Marne (annexe VIII sur la population active).

BIBLIOGRAPHIE

- Institut National de la Statistique et des Etudes Economiques.
  Elaboration des statistiques Figures et Graphiques.
  E. Morice.
  Cours à l'Ecole d'Application de l'I.N.S.E.E.

- Institut International de statistique. Centre international d'enseignement de la statistique à Beyrouth.
  Cours de statistique appliquée à l'économie. F. Chartier, chapitre VI, le graphique statistique.

- Robert Satet et Charles Voraz "les graphiques" Paris. Editions Georges frères, 1945.

- La statistique graphique et l'économique d'entreprise. F. Divisis - Hommes et Techniques, janvier et février 1947.

- How to lie with statistics - Darrell Hoff
  Norton et Co.
  (Voir en particulier, chapitres IV et V).

- Facts from figures - M.J. Moroney
  Penguin Books (Voir chapitre III).

Chapitre VI

LES CARACTERISTIQUES DES DISTRIBUTIONS

6.I Les distributions de fréquences.

6.2 L'intervalle de variation.

6.3 Les valeurs centrales.

6.4 Les paramètres de dispersion.

6.5 L'asymétrie.

Chapitre VI

Aux méthodes d'élaboration, succèdent les méthodes d'utilisation et d'interprétation des statistiques. Une seconde étape commence au cours de laquelle il s'agit d'examiner les résultats numériques, et de rechercher les valeurs caractéristiques des phénomènes observés. Ces méthodes descriptives seront divisées en trois parties :

- étude des distributions
- analyse des séries chronologiques
- calcul des nombres indices.

6.I - Les distributions de fréquences.-

Le dépouillement a donné lieu à la présentation de résultats numériques classés selon le caractère ou l'intensité d'un paramètre :

caractères	A	B	C	D	intensité	x_1	x_2	x_3	x_4
nombre d'observations	n_A	n_B	n_C	n_D	ou nombre d'observations	n_1	n_2	n_3	n_4

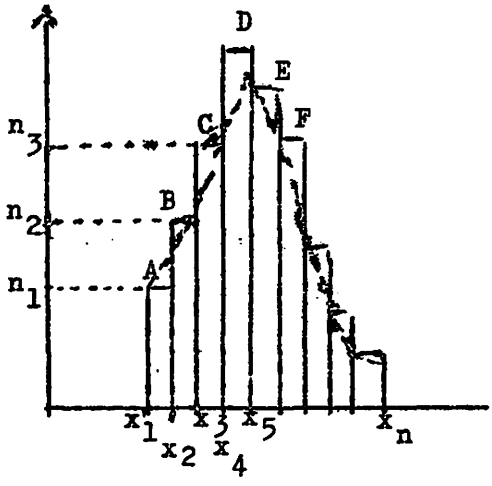
Dans le premier cas, une distribution qualitative est créée, dans le second cas on a une distribution quantitative = à chaque valeur isolée x_i correspond un nombre d'observations n_i. Dans le cas où le nombre des valeurs isolées prises par la variable x est important et le nombre total d'observations est suffisant, on est amené à grouper ces valeurs isolées en classes d'amplitude en général égale.

Exemple : nombre de personnes soumises à l'impôt sur le revenu par tranche de revenu (voir tableau ci-dessous).

Nombre de personnes soumises à l'impôt sur le revenu par tranche de revenu :

Limites de classes	Nombre d'observations
x_1 à - x_2	n_1
x_2 à - x_3	n_2
x_3 à - x_4	n_3
etc.....	.....

Il est alors possible de construire un diagramme à barres, en portant en abscisse les limites de classes de revenu et en ordonnée le nombre d'observations correspondant à chaque classe.



On obtient une figure appelée : histogramme de fréquence.

Si les classes ne sont pas d'une amplitude trop importante on peut repérer les points A, B, C, D, E, milieux des côtés supérieurs des rectangles, en les joignant on obtient une ligne brisée qui constitue un polygone de fréquence.

On conçoit intuitivement que si le nombre des observations s'accroît et, que dans le même temps, on divise les intervalles de classe, qui deviennent ainsi de plus en plus petits, cette ligne brisée finira par tendre vers une courbe appelée : courbe de fréquence.

Les histogrammes de fréquence peuvent revêtir diverses formes, aussi est-il nécessaire pour les caractériser de calculer divers coefficients ou valeurs numériques.

On est alors amené à définir :

a) les valeurs extrêmes prises par la variable et l'intervalle de variation ;

b) les valeurs centrales ;

c) la dispersion de la distribution ;

d) le degré d'asymétrie.

6.2 - L'intervalle de variation -

Une statistique est disponible sur le nombre de planteurs de cacao selon leur production au cours de la dernière campagne agricole. Le plus petit producteur a récolté 150 Kgs de dacao, le plus gros 8.500 Kgs. Tous les autres planteurs se répartissent entre ces deux valeurs extrêmes

$$x_m = 150 \qquad x_M = 8.500$$

L'intervalle de variation (le "range") est la différence entre l'abscisse maximum et l'abscisse minimum :

$$\Delta = x_M - x_m = 8.350$$

Cette notion est utilisée particulièrement en matière de contrôle de fabrication.

6.3 - Les valeurs centrales -

Il n'existe pas de valeur centrale unique pour caractériser une distribution. Pour être acceptable, une valeur centrale doit répondre à certaines conditions :

- être définie objectivement
- dépendre de toutes les observations
- avoir une signification concrète simple
- être d'un calcul aisé et rapide
- être peu sensible aux fluctuations de l'échantillon
- se prêter aux calculs algébriques ultérieurs.

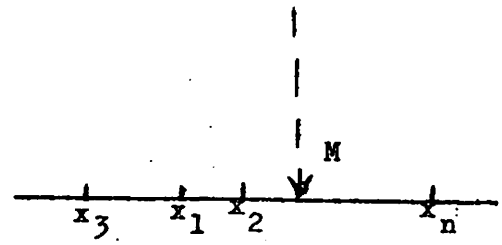
Les principales valeurs centrales sont :

les moyennes
la médiane
le mode.

6.3.I- Les moyennes dans le cas d'observations isolées.

a) Pour des observations isolées, donc non groupées, la moyenne la plus simple à calculer est la moyenne arithmétique ; on calcule la somme des valeurs observées que l'on divise par le nombre m des observations :

$$x = \frac{x_1 + x_2 + \ldots + x_n}{n} \quad (1)$$



b) Une autre moyenne est la <u>moyenne géométrique</u> : Elle est égale à la racine $n^{ième}$ du produit des n valeurs observées. La moyenne géométrique de deux termes est la racine carrée du produit de ces deux termes, de trois termes la racine cubique du produit de ces termes, etc...

Cette moyenne est utilisée dans le cas d'<u>indice de prix</u>. Les prix de deux articles sont connus à deux époques :

		1949	1953	indice
pain	le kg.	30	60	$200 = \frac{60}{30} \times 100$
vin	le litre	70	35	$50 = \frac{35}{70} \times 100$

pour le premier article, le prix a doublé, pour le second, il a baissé de moitié ; alors que la moyenne arithmétique est égale à

$$\frac{200 + 50}{2} = 125$$

la moyenne géométrique donne $\sqrt{200.50} = 100$ elle répond mieux à cette idée intuitive, qu'un prix dans son expansion se heurte à moins de résistance que dans sa tendance à la diminution.

<u>La moyenne géométrique est toujours plus petite que la moyenne arithmétique</u>.

Il existe une <u>moyenne dite harmonique</u> qui se calcule à partir de la formule

$$\frac{1}{H} = \frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_n}\right)$$

6.3.2 - <u>Les moyennes dans le cas d'observations groupées</u>.

Jusqu'ici, n'ont été considérés que des termes isolés, or, on peut (et c'est le cas le plus fréquent) avoir affaire

(1) On calcule également la <u>moyenne quadratique</u> : racine carrée de la moyenne arithmétique des carrés des observations.

à des répartitions par classes ; dans ce cas, en raison de ce groupement des observations, chaque valeur prise par la variable est répétée une ou plusieurs fois, on a affaire à des moyennes dites "pondérées". Dans ce type de moyenne, chaque terme intervient avec son importance propre, c'est-à-dire avec son poids.

En appelant x_1 x_2 .... x_n le centre de chaque intervalle de classe, les moyennes décrites précédemment sont calculées ainsi :

a) moyenne arithmétique :

$$\bar{x} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + n_n \cdot x_n}{n}$$

Des ouvriers sont classés selon leur salaire horaire :

12	ouvriers gagnent	40	francs	de l'heure	
15	" "	45	"	"	"
18	" "	47,5	"	"	"
22	" "	50	"	"	"
7	" "	52,5	"	"	"
3	" "	55	"	"	"
1	ouvrier gagne	60	"	"	"
78					

Le total des salaires horaires perçus par les 78 ouvriers sera égal à :

$$12 \times 40 + 15 \times 45 + 18 \times 47,5 + \text{etc...} = 3.702,5$$

et le salaire moyen :

$$\bar{x} = \frac{3.702,5}{78} = 47,5$$

Soit une population de personnes actives du sexe masculin classée par âge selon les deux secteurs du commerce et de l'armée :

Tranche d'âge	Commerce	Armée
15-19 ans	15	18
20-24 ans	127	205
25-29 ans	154	111
30-34 ans	112	122
35-39 ans	88	70
40-44 ans	73	50
45-49 ans	61	8
50-54 ans	40	7
55-59 ans	16	1
60-64 ans	7	-
65 ans et plus	3	-
Total :	696	592
Age moyen	34,5	29,6

Source : Population du Cameroun. Non originaires. 1951.

b) <u>moyenne géométrique</u>

Des calculs analogues peuvent être menés pour la moyenne géométrique pondérée

$$\bar{x} = \sqrt[n_1+n_2+.]{\underbrace{x_1 . x_1 . x_1}_{n_1} \cdot \underbrace{x_2 . x_2 . x_2}_{n_2} \ldots}$$

ou $$\bar{x} = \sqrt[n]{x_1^{n_1} \, x_2^{n_2} \ldots\ldots x_n^{n_n}}$$

<u>Remarque</u> : Soit une série de termes isolés, on calcule les moyennes issues de cette série, si l'on regroupe les observations individuelles, les moyennes obtenues à partir de la distribution par classes sont un peu différentes des premières.

6.3.3- <u>La médiane</u> -

a) <u>observations isolées.</u>

Une série de termes en nombre n est rangée par ordre croissant. Dans cette série, un terme est tel que le nombre de termes qui lui sont inférieurs est égal au nombre de termes qui lui sont supérieurs. Ce terme est la <u>médiane</u> de la série.

On a par exemple la série des températures maxima moyennes à Conakry :

Janv.	:	30,9	Août :	27,5	La moyenne arithmé-
Févr.	:	31,0	Sept.:	29,2	tique est égale à
Mars	:	31,8	Oct. :	31,1	30,4 tandis que la
Avr.	:	31,2	Nov. :	31,3	médiane se trouve
Mai	:	32,0	Déc. :	30,7	dans l'intervalle
Juin	:	30,0			30,9 - 31,0
Juil.	:	28,8			parce qu'il y a un nombre pair de termes.

b) <u>observations groupées.</u>

Le tableau de présentation, doit être ordonné de telle manière que les intervalles de classe se succèdent par ordre croissant.

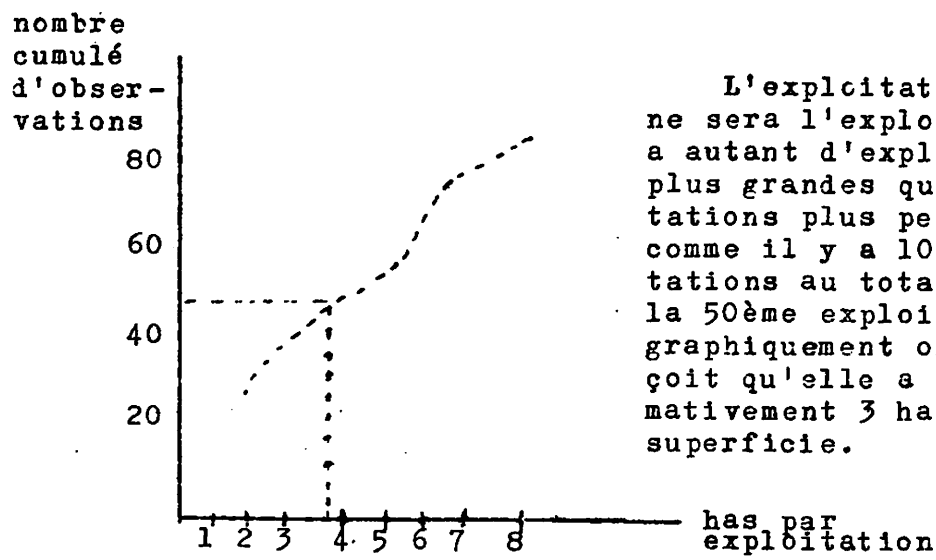
Ensuite on cherchera pour chaque limite de classe, le nombre de termes dont la valeur est égale ou inférieure à cette limite, on dit que l'on pratique le cumul de la série initiale.

Il faut distinguer deux cas :

b_1) Le caractère dont on mesure l'intensité est susceptible d'une variation continue. Dans ce cas, on construit le graphique dit cumulatif du polygone de fréquence initial et l'on joint les points représentatifs.

Cent exploitations agricoles sont classées selon leur superficie :

Superficie	Nombre d'exploitations	Nombre cumulé
- 1 ha	20	20
1 à 1,99 has	16	36
2 à 2,99 has	12	48
3 à 3,99 has	10	58
4 à 4,99 has	12	70
5 à 5,99 has	8	78
6 à 7,99 has	11	89
8 à 9,99 has	5	94
10 et plus	6	100



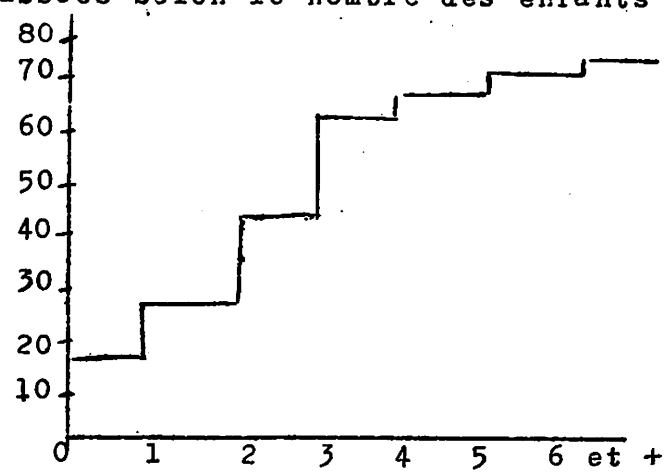
L'exploitation médiane sera l'exploitation qui a autant d'exploitations plus grandes que d'exploitations plus petites, comme il y a 100 exploitations au total ce sera la 50ème exploitation : graphiquement on s'aperçoit qu'elle a approximativement 3 has 3 de superficie.

Le même raisonnement s'applique pour calculer le budget médian, ou la ration calorique médiane ou encore le rendement médian d'une culture.

b_2) Un deuxième cas se présente lorsque l'intensité du caractère ne peut, de par sa nature même, varier d'une manière continue.

Des mères ont été classées selon le nombre des enfants encore vivants

sans enfant vivant	12
1 enfant	15
2 enfants vivants	18
3 " "	22
4 " "	7
5 " "	3
6 " et plus	1
	78



Il ne peut y avoir de mères avec des nombres entiers d'enfants suivis de décimales, on construit donc le graphique cumulatif dit "en escaliers". On dira simplement que la médiane se trouve dans la classe des mères de deux enfants. Il faut insister sur ce fait que ce n'est que dans le cas où la variable peut logiquement varier d'une manière continue, que l'on est en droit de calculer par une répartition proportionnelle la valeur de la médiane.

c) Intérêt de la médiane -

Dans le cas d'une série de valeurs isolées, on remarque que la valeur de la médiane n'est pas modifiée quand on change la valeur des termes de la série mais non leur ordre. La moyenne arithmétique par contre dépend de la valeur absolue des termes et, en particulier est sensible aux variations des termes extrêmes : si une observation accidentelle fournit une valeur très faible ou très forte, la moyenne arithmétique peut fluctuer dans une mesure appréciable.

C'est ainsi que, dans le domaine des prix en France pour des articles présentant une homogénéité de qualité suffisante, on calcule un prix moyen par une simple moyenne

arithmétique, alors que pour des échantillons d'informations instables ou peu importants (textiles, consultation médicale) on retient la valeur médiane des observations.

6.3.4 - Le mode -

Une valeur centrale citée dans certains cas, mais qui comme la médiane se prête mal aux calculs est le mode : c'est la valeur de la variable pour laquelle on a relevé le plus grand nombre d'observations. On étudie la durée de fonctionnement d'un lot d'ampoules électriques : le mode sera la durée pour laquelle on a constaté la survie du plus grand nombre d'ampoules. Cette valeur a donc toujours une signification concrète, avantage sur la moyenne arithmétique qui peut n'en revêtir aucune. Pour une rivière à régime torrentiel, le mode correspond à une notion réelle : hauteur d'eau qui a été le plus fréquemment relevée (ce peut être séro), tandis que le niveau moyen pour l'année n'a pu être enregistré que durant un bref instant.

o
o o

Le tableau ci-dessous résume les trois valeurs centrales qui viennent d'être étudiées :

salaires mensuels	nombre de salariés	
2.000	I4	← mode
3.000	1	← médiane
3.700	4	
5.000	3	
7.700	1	← moyenne arithmétique
10.000	2	
15.000	1	
45.000	3	
	29	

salaire modal : 2.000
salaire médian : 3.000
salaire moyen : 7.700

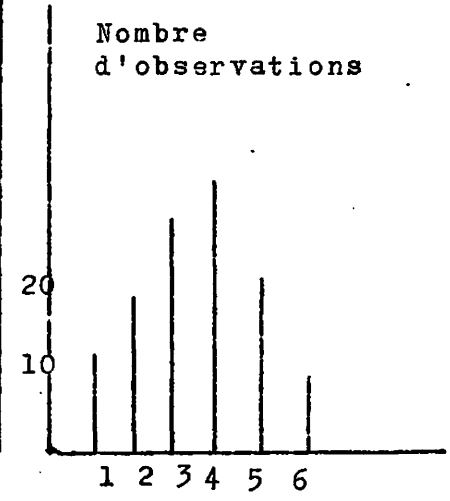
Cet exemple est évidemment scolaire, mais il illustre les différences parfois importantes des trois valeurs centrales que l'on peut calculer à partir d'une distribution de fréquences.

6.4. Les paramètres de dispersion.-

6.4.I- Utilité de la connaissance de la dispersion d'une distribution.-

Un recensement de population ou une enquête démographique permet de dresser une statistique des ménages. Il est alors intéressant de calculer la taille moyenne d'un ménage, en vue de connaître par exemple la structure des logements nécessaires à la population étudiée.

Taille du ménage.	Nombre d'observations	Nombre de personnes du ménage
1 personne	10	10
2 personnes	14	28
3 "	23	69
4 "	29	116
5 "	13	65
6 "	7	42
7 et + "	4	32
	---	---
	100	362



Cette taille moyenne est égale à 3,6 personnes autrement dit, connaissant ce seul renseignement on serait conduit à la construction d'appartements à deux chambres à coucher. L'examen de la distribution montre que si l'on ne se soucie pas de connaître la dispersion de la taille des différents ménages autour de la taille moyenne précédemment calculée 52 % des ménages seront logés normalement, mais 24 % seront trop au large et 24 % des logements seront surpeuplés.

La seule connaissance de la valeur centrale est insuffisante, il faut préciser la répartition des observations autour de cette valeur centrale. La dispersion peut être mesurée en utilisant comme valeurs centrales :

- soit la moyenne arithmétique
- soit la médiane.

6.4.2- Mesure de la dispersion dans le cas d'observations isolées par rapport à la moyenne arithmétique.

Les deux séries d'observations :

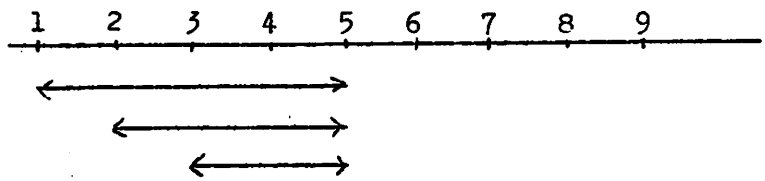
A :	1,	2,	3,	4,	5,	6,	7,	8,	9
B :	4,	4,25	4,50	4,75	5	5,25	5,50	5,75	6

admettent toutes les deux la même moyenne, cependant l'une est caractérisée par un intervalle de variation de 8 unités, tandis que l'autre ne l'est que par un intervalle de deux unités.

Pour caractériser la dispersion dans ce cas, on calcule les écarts de chaque observation à la moyenne arithmétique. Soit d'après l'exemple précédent

1 - 5 , 2 - 5 , 3 - 5 etc...

- on élève au carré chacune de ces différences
- on somme ces carrés
- on divise par le nombre des observations



et la racine carrée de cette valeur (la variance) appelée écart type de la distribution est une valeur caractéristique de la dispersion : plus l'écart type est important, plus les observations individuelles sont dispersées autour de la moyenne arithmétique. L'écart type peut être considéré comme une mesure de la variabilité du phénomène étudié, ou de la précision d'une mesure.

Pour prendre une comparaison concrète, on peut imaginer que l'on mesure une petite distance avec un instrument peu précis et que l'on répète la même opération cinquante fois. On trouve en moyenne que cette longueur est égale à 10 mètres et que l'écart type de ces cinquante observations est égal à 10 centimètres.

Dans une seconde série d'observations, au cours de laquelle on mesure une grande distance, on trouve une distance moyenne de 10 kilomètres avec un écart type de 1 mètre. Le rapport de l'écart type à la moyenne dans le

premier cas est égal à $\frac{1}{100}$ et dans le second $\frac{1}{10.000}$ (1).

Ce résultat enseigne que la précision est meilleure dans le second cas que dans le premier.

Pour exposer un autre exemple pratique, on recherche le rendement en riz de deux plaines : l'une est amenagée, bien planée, la seconde au contraire est bosselée, sujette à des inondations irrégulières car le plan d'eau n'est pas contrôlé.

On pratique des coupes d'un are. Dans le premier cas, puisque la rizière est homogène les rendements obtenus varieront peu et les points représentatifs des prélèvements seront peu dispersés, dans le second cas les rendements seront très variables et la moyenne calculée sera issue d'un ensemble hétérogène.

Dans chacun des deux cas on pourra calculer la précision.

Calcul de l'écart type

La formule générale est $\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

si les x_i sont de petits nombres, on a intérêt à utiliser une table de carrés et on calcule $\frac{\sum x_i^2}{n}$

dont on retranche $\bar{x}^2$ c'est-à-dire $\left(\frac{\sum x_i}{n}\right)^2$

Lorsqu'on a affaire à des chiffres importants ou décimaux, on prend comme origine de travail une valeur ronde x_o voisine de $\bar{x}$ et on calcule les écarts à cette origine et les carrés des écarts.

Le carré de l'écart type s'obtient alors par la formule :

$$\text{variance} = \frac{\sum \left(x_i - x_o\right)^2}{n} - d^2$$

$$\text{avec} \quad d = \bar{x} - x_o$$

(1) Ce rapport de l'écart type à la moyenne multiplié par 100 est appelé coefficient de variation.

Application numérique.-

Vingt coupes échantillons de riz ont été prélevées dans des champs. Calculer la dispersion (rendement exprimé en kg par are - paddy humide).

2,4	8,8	18,5
3,0	9,1	19,3
3,6	9,7	
4,1	11,7	
4,6	11,9	
4,8	12,4	
5,0	12,5	
5,6	12,6	
6,2	13,8	

$\bar{x} = 8,98 \qquad x_0 = 9$

$$\sum(x_i - x_0)^2 = 477,52 \qquad \sigma^2 = \frac{477.52}{20} - \overline{0,02}^2 = 23.9$$

$$\sigma = 4,9$$

6.4.3.- Mesure de la dispersion dans le cas d'observations groupées, dispersion calculée à partir de la moyenne arithmétique.

Un histogramme de fréquences étant donné, il s'agit d'en calculer la dispersion selon la méthode précédente des écarts à la moyenne. Le principe est identique mais si le centre d'une classe i a pour valeur x_i, l'écart $x_i - \bar{x}$ est répété n_i fois s'il y a n_i observations dans la classe i.

On adopte toujours une origine arbitraire de travail (1) et l'on établit le tableau numérique suivant :

(1) N'est pas traité ici le cas particulier où les classes sont d'amplitude égale, pour lequel une variable auxiliaire est très utile. La correction dite de Sheppard ne sera pas évoquée dans ce cours.

Modèle de feuille de calcul de l'écart type d'une distribution -

Valeur du caractère (1)	Nombre d'observations (2)	(1)x(2) (3)	Ecart entre origine arbitraire et valeur du caractère (4)	Ecart élevé au carré (5)	Ecart carré pondéré (5)x(2)
	Total N	Total			Total : T^2

Somme de (3) divisée par N = moyenne
(moyenne - origine arbitraire) au carré = d^2

$$(\text{écart type})^2 = \frac{T^2}{N} - d^2$$

Comme dans le cas des observations individuelles, le coefficient de variation

$$C\ V = \frac{\sigma}{\bar{x}} \times 100$$

caractérise la dispersion relative de la distribution. Deux histogrammes peuvent avoir des moyennes sensiblement égales, mais le premier peut à premier examen apparaître plus "aplati" que le second. Le calcul montrera alors si V_1 est bien supérieur à V_2. On verra dans le chapitre réservé aux tirages au hasard, l'utilité pratique de cette notion de dispersion.

<u>Application numérique</u> : Calcul du coefficient de variation de la distribution des ménages selon leur taille :

Valeur du caractère (1)	Nombre d'observations (2)	(1)x(2)	Ecart à x_o	Ecart au carré	Ecart au carré pondéré
1	10	10	- 3	9	90
2	14	28	- 2	4	56
3	23	69	- 1	1	23
4	29	116	0	0	0
5	13	65	1	1	13
6	7	42	2	4	28
7	4	28	3	9	36
		358			246

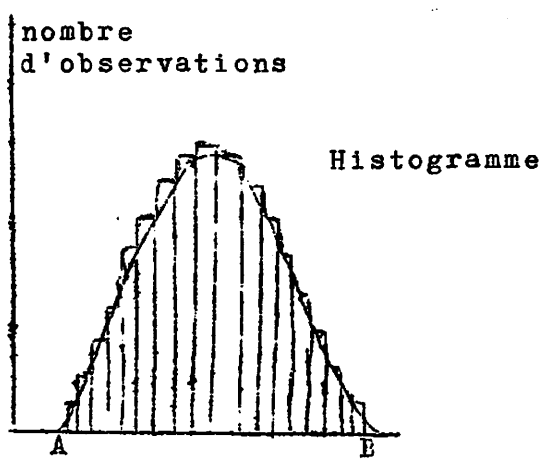
$\bar{x} = 3,58$ origine de travail $x_o = 4$

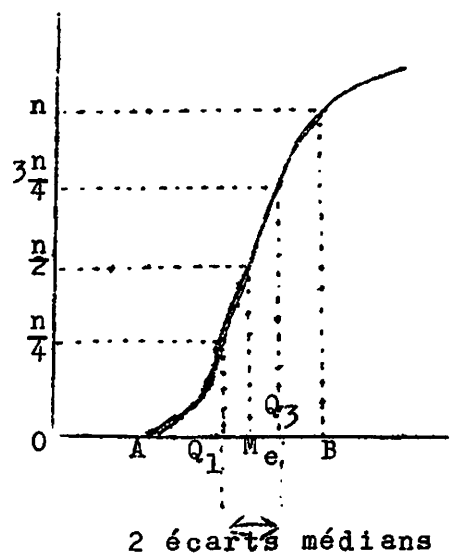
$$\frac{246}{100} - \overline{0,42}^2 = (\text{écart type})^2$$

$$\sigma = \sqrt{2,28} = 1,50 \qquad V = \frac{1,50}{3,58} \times 100 = 42$$

6.4.4.- Mesure de la dispersion d'une distribution à partir de la médiane.-

Une seconde valeur centrale est utilisée comme origine des écarts : la médiane. Cette valeur centrale sépare l'effectif total des observations classées par intensité croissante de la variable en deux groupes égaux. On peut, comme on le fait pour l'ensemble de la distribution rechercher pour chacun de ces deux groupes la valeur de la variable qui à nouveau a autant d'observations qui lui sont soit inférieures soit supérieures.





On obtient deux nouvelles valeurs qui fixent le 1er et le 3ème quartiles.

La moitié de la valeur $Q_3 - Q_1$ s'appelle l'écart médian $= \frac{Q_3 - Q_1}{2}$ (1).

On vérifiera graphiquement que si l'histogramme H est aplati Q_3 est éloigné de Q_1 et l'écart médian sera important, au contraire si H est pointu, la dispersion donc peu marquée $Q_3 - Q_1$ sera faible. La notion de quartile peut être étendue en considérant les valeurs de la variable telles que 10 % des observations lui soient inférieur, puis 20 %, 30 %, ces valeurs sont appelées déciles.

Le bulletin de la statistique générale de France (avril juin 1952) montre l'utilisation graphique de la notion de décile en matière de statistique de poids et de taille d'écoliers selon leur âge.

En psychotechnie, ces grandeurs sont extrêmement utiles pour repérer l'aptitude d'un individu selon ses réactions à un test donné, on peut alors par rapport à une population standard le classer et lui attribuer une note comprise entre 0 et 10 (2).

L'exemple ci-dessous est une illustration du comportement différent à deux tests, d'une même population.

Un même groupe de deux cents candidats a été examiné pour deux tests différents : réactions auditive et optique à un signal acoustique ou optique (le candidat en entendant ou en voyant le signal actionne un poussoir). Pour l'ensemble des candidats, les temps de réaction moyens selon l'un ou l'autre test sont voisins (valeurs centrales égales) mais les graphiques construits à partir des données numériques sont fort différents de forme :

---

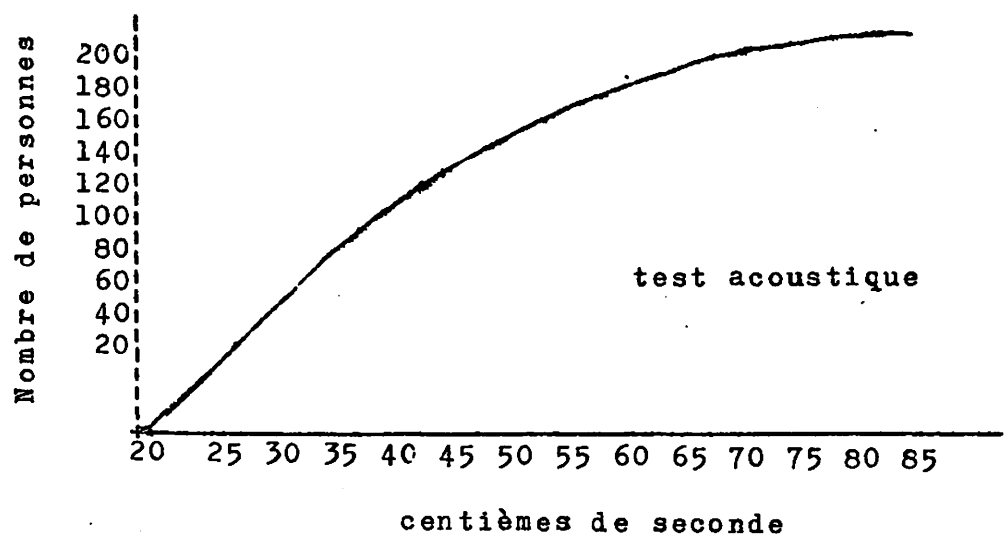
(1) Lorsque l'on a des distributions différentes à comparer, on peut être amené à calculer un coefficient de dispersion relative, dans ce cas, on définit l'écart interquartil relatif par le rapport de l'écart médian à la médiane :

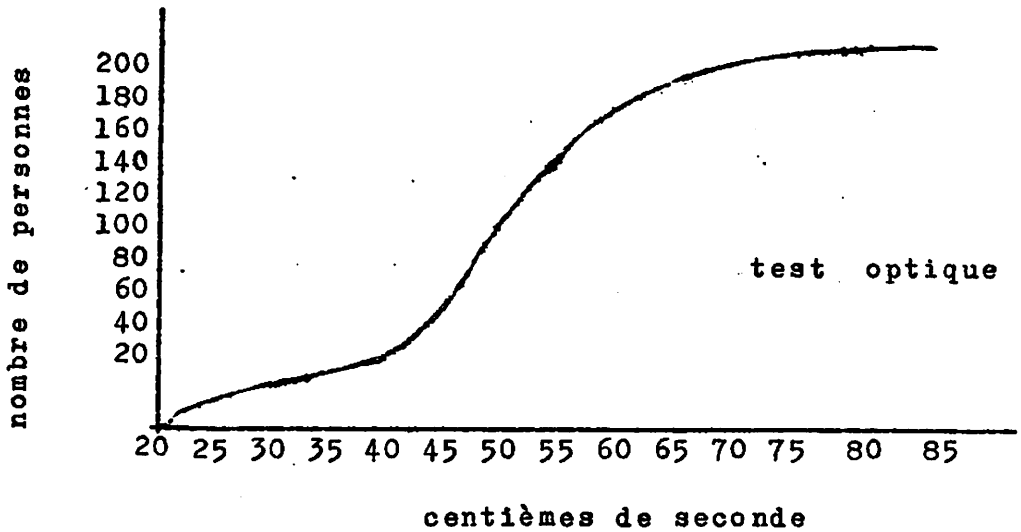
$$\frac{Q_3 - Q_1}{2\ M_e} \times 100.$$

(2) Test de pointillage par exemple. L'ensemble des résultats d'une batterie de tests constitue un profil.

Temps de Réaction	Test n° 1 acoustique		Test n° 2 optique	
	Nombre d'individus correspondant	Nombre cumulé	Nombre d'individus correspondant	Nombre cumulé
25 centièmes de seconde	11	11	1	1
30 "	18	29	3	4
35 "	20	49	9	13
40 "	24	73	11	24
45 "	28	101	24	48
50 "	29	130	72	120
55 "	21	150	55	175
60 "	17	168	12	187
65 "	11	179	6	193
70 "	9	188	3	196
75 "	6	194	2	198
80 "	4	198	1	199
85 "	2	200	1	200
Total :	200		200	

On s'aperçoit que le test acoustique fournit une distribution étalée, tandis que le test optique donne un graphique beaucoup plus ramassé. Si l'on trace les polygones cumulés à partir de ces deux distributions on obtient :





1° quartile, médiane et 3° quartile (50° individu, 100° individu et 150° individu) ont des temps de réaction fort différents selon l'un et l'autre tests.

		Test acoustique	Test optique
Q_1	1° quartile	35 cent. de seconde	45 cent. seconde
M_e	médiane	45 " " "	48 " "
Q_3	3° quartile	55 " " "	53 " "

Dans le premier cas (distribution aplatie)

$Q_3 - Q_1 =$ 20 centièmes,

dans le second cas (distribution pointue)

$Q_3 - Q_1 =$ 8 centièmes,

la différence entre 3° et 1° quartiles est bien caractéristique de la dispersion d'une distribution statistique étudiée.

Ces deux tests étant étalonnés, on peut alors situer les résultats obtenus par un candidat. Pour que le test soit sélectif, il faut que la distribution de référence soit suffisamment étalée de telle manière que l'on puisse placer sans trop d'erreur les résultats obtenus par le candidat dans les déciles correspondants.

6.5 L'asymétrie.-

L'asymétrie d'une distribution se définit par l'expression $A = \frac{M_o - \bar{m}}{\sigma}$ à partir de la valeur du mode de la moyenne arithmétique et de l'écart type.

o
o o

Les notions essentielles sur les distributions se ramènent donc à :

- la notion de fréquences par classe
- l'existence de moyennes arithmétiques et géométriques, pondérées ou non
- la connaissance du calcul d'une valeur numérique caractéristique de la dispersion appelée écart type, le coefficient de variation étant la mesure de la dispersion relative.

BIBLIOGRAPHIE

- Méthode statistique - E. Morice et F. Chartier
Chapitre I.

- Eléments de technique statistique - Michel Huber
Deuxième Partie - Chapitres I et II.

Chapitre VII

LES SÉRIES CHRONOLOGIQUES

7.I Définitions - Eléments constituants.

7.2 La tendance - Les variations saisonnières.

7.3 Comparaison de séries chronologiques.

7.4 Corrélation et contingence.

Chapitre VII

LES SÉRIES CHRONOLOGIQUES

7.I Définition - Eléments constituants.

Les notions exposées précédemment visaient à la description d'un ensemble statistique, à un instant donné, l'observation était en quelque sorte statique. Il est bien évident que cet ensemble se déforme, varie selon les époques et que l'introduction de la variable temps ouvre le champ à d'autres recherches. L'analyse de la variation des données numériques en fonction du temps constitue l'étude des séries chronologiques.

7.I.I- La représentation graphique d'une série chronologique ressemble à l'enregistrement d'un baromètre enregistreur : l'abscisse repère le temps, l'ordonnée mesure la valeur prise par la variable. Les observations sont effectuées généralement à intervalles réguliers.

L'observation peut être définie - soit à un instant donné (température à une certaine époque, prix à une date déterminée) - soit par sa valeur pendant l'intervalle de temps séparant l'observation actuelle de la précédente (chutes de pluies durant un mois, chiffre d'affaires durant un exercice annuel).

L'établissement d'un graphique à l'occasion d'une série chronologique favorise l'exposition du phénomène et son analyse : l'étude visuelle permet en effet de découvrir certaines anomalies ou certaines tendances particulières.

7.I.2.- Une série chronologique est le résultat de la combinaison des variations de quatre éléments :

a) la tendance de longue durée
b) les oscillations cycliques
c) les variations saisonnières
d) les variations résiduelles.

a) La tendance de longue durée comme l'indique son nom est le mouvement "de fond" qui traduit une évolution sociale ou économique à très long terme : les statistiques des pays en voie de développement sont encore trop récentes pour illustrer cette composante. On peut citer les tendances de la natalité et de la mortalité en France, depuis 1870 (1).

b) L'oscillation cyclique a été comparée par certains auteurs à un mouvement pendulaire autour d'une position d'équilibre. On peut citer la variation des cours du coton à New-York de 1876 à nos jours (2).

c) Les variations saisonnières : une économie non diversifiée est très influencée par le rythme de la production agricole. Il n'est donc pas étonnant de trouver de nombreuses séries pour lesquelles on constate dans le cadre de l'année une pulsation régulière : circulation monétaire, perceptions fiscales (taxe sur les transactions en A.O.F.), commerce extérieur, etc... (voir graphiques VII.I et VII.2).

d) Les variations résiduelles sont les petits accidents aléatoires ou conjoncturels à très court terme qui se traduisent par les irrégularités inévitables de la courbe représentative de la série étudiée.

7.2 La tendance - Les variations saisonnières.-

Deux observations sont fondamentales dans les pays sous-développés : la recherche de la tendance et du mouvement cyclique, l'étude des variations saisonnières. L'élimination des variations accidentelles est nécessaire avant tout calcul.

7.2.I - Les variations accidentelles sont en partie éliminées par un procédé simple appelé "méthode des moyennes mobiles". Chaque observation est remplacée par une combinaison d'un certain nombre d'observations voisines.

---

(1) Annuaire statistique abrégé de la France, deuxième volume, 1949, page 12 - graphique 3, et mise à jour dans Tableaux de l'économie française, page 68bis.

(2) I.N.S.E.E. Etudes et documents. Le marché mondial du coton - page 128.

On substituera par exemple à une observation la moyenne arithmétique de cette observation et des deux observations encadrantes.

Si on veut attribuer plus d'importance à certaines mesures, on peut pondérer différemment les observations étudiées, attribuer le coefficient 2 à l'observation centrale et le coefficient 1 aux observations encadrantes. On obtient ainsi un graphique que l'on peut qualifier de raboté, lissé.

7.2.2. Seuls les petits accidents de la courbe sont jusqu'ici égalisés, mais les différentes composantes de la série ne sont toujours pas connues. La suppression des variations saisonnières afin de déterminer la tendance de longue durée à laquelle s'ajoute le mouvement cyclique peut être opérée par deux voies : numérique et graphique.

a) Le calcul numérique est lié à une hypothèse de travail sur la période de répétition des phénomènes : on suppose en général que la période est annuelle. Il est donc fondamental de s'assurer avant tout calcul que cette hypothèse est vraisemblable. A cette condition, on remplacera chaque observation isolée par la moyenne arithmétique de celle-ci et des douze observations encadrantes. Ceci se justifie intuitivement et peut se vérifier en considérant la série :

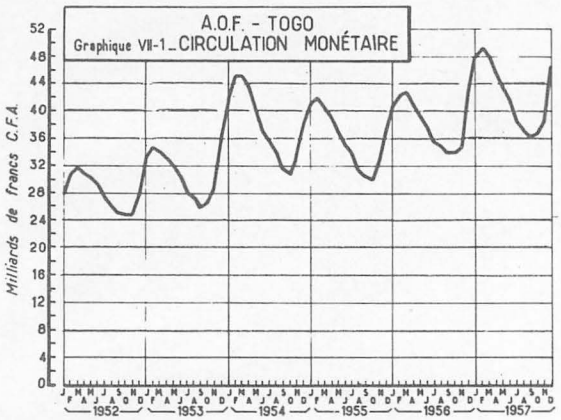
<- - - - - - - - - - - - année n - - - - ->												<- année n+1 ->	
J	F	M	A	M	J	J	A	S	O	N	I	J	F
+2	+3,5	+5	+7	+5	+2,5	0	-3	-5,5	-7	-5	-2	+2	+3,5

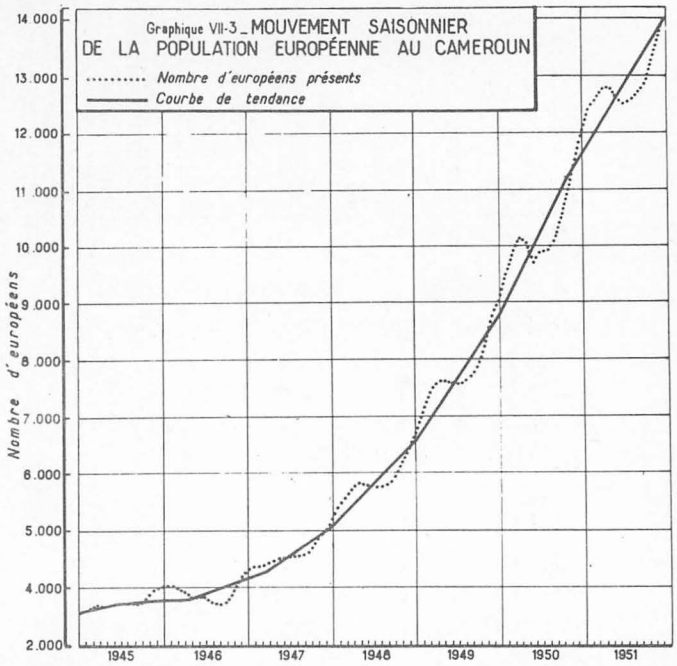
La figure représentative est une courbe qui serpente autour de l'axe des abscisses.

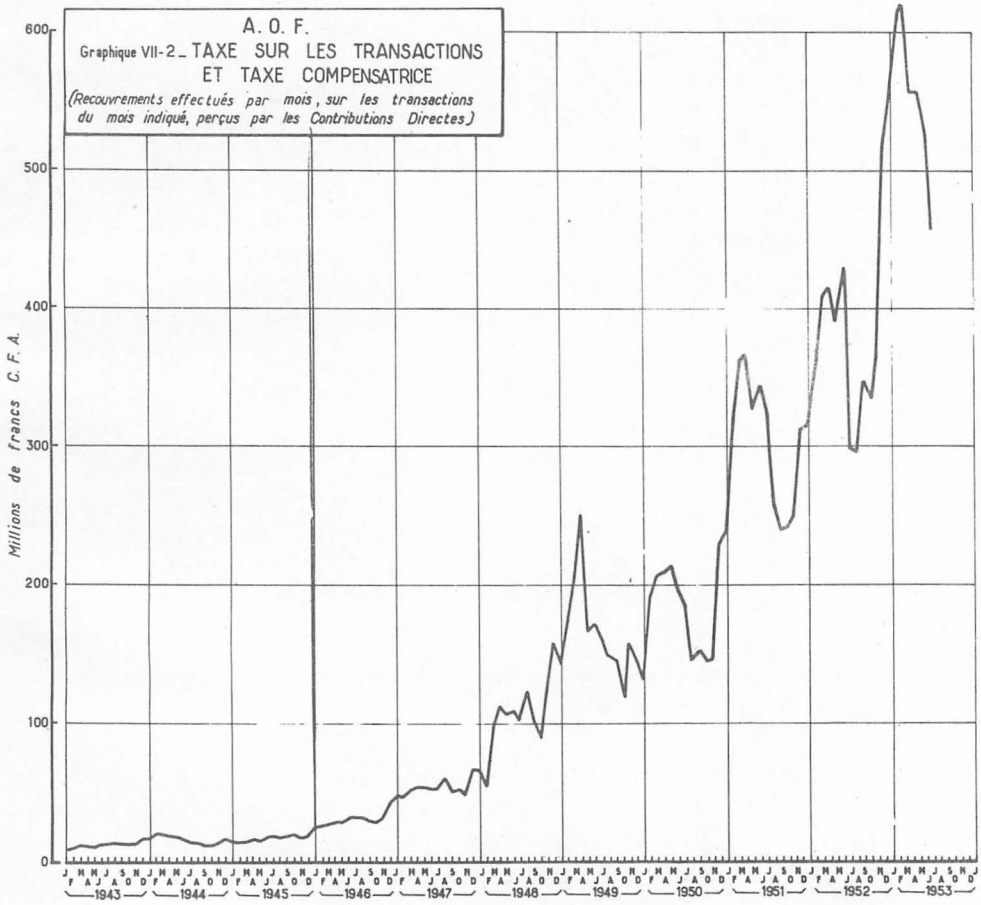
En appliquant la méthode des moyennes mobiles sur 13 mois, chaque observation est dans cet exemple remplacée par un nombre très proche de zéro. (En Août l'ordonnée initiale -3 serait remplacée par une ordonnée égale à 0,5, etc...).

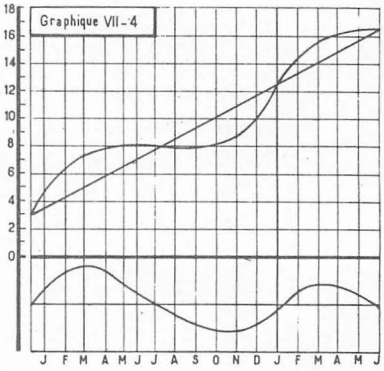
Les nouveaux points obtenus sont joints (1) et on

(1) Il faut remarquer que l'on perd de l'information sur les six premiers et les six derniers mois de la série totale.

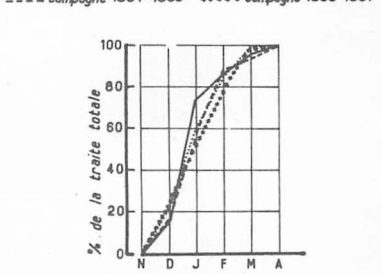








Graphique VII-5_VARIATIONS, MOIS PAR MOIS, DU TONNAGE RELATIF CUMULÉ DE COTON-GRAINES EN A. E. F.



19.061_1_58

obtient la courbe recherchée débarassée des variations saisonnières : la courbe est dite "ajustée" (1).

b) On peut arriver à un résultat semblable par une méthode graphique.

On trace à main levée la courbe ajustée de telle manière que le tracé laisse autant de points au-dessus et au-dessous : plus exactement, en utilisant un papier quadrillé, il est nécessaire que les deux nombres des carrés compris entre le contour polygonal et la courbe ajustée soit au-dessus, soit au-dessous soient égaux.

La courbe représentative de la tendance et du cycle étant connue, il est alors possible de préciser les coefficients saisonniers.

On donnera ci-dessous une solution générale, qui peut être simplifiée lorsque la moyenne annuelle des observations est stable durant la période étudiée.

Cas général : La courbe ajustée étant connue, on mesure successivement pour chaque mois l'écart entre la valeur initiale de ce mois et la valeur ajustée. (Graphique VII.4). Si les données existent pour n années successives, on a pour un mois m donné n écarts. Pour éliminer l'influence des écarts extrêmes, on retient pour chaque mois l'écart médian (et non la moyenne arithmétique des n écarts mesurés). Chaque écart médian mensuel fait ensuite l'objet d'une correction de telle manière que la somme algébrique des douze écarts mensuels soit nulle.

Remarque : Dans certains cas, on calcule non les écarts à la courbe ajustée, mais les rapports. Dans ce cas, la somme des douze coefficients mensuels doit être égale à 12.

Cas particulier : méthode des rapports à la moyenne mensuelle.

Dans ce cas la tendance est sensiblement parallèle aux abscisses ou présente une évolution peu marquée. On calcule pour chaque année d'observations (en nombre n) la moyenne mensuelle et on calcule par année les écarts des

---

(1) Cette méthode a été appliquée dans une étude du service de la statistique générale du Cameroun sur les fluctuations de l'effectif de la population non-originaire présente, de 1945 à 1951. Voir graphique VII.3.

valeurs mensuelles à cette moyenne. Pour un mois donné m, on a n écarts dont on retient encore la médiane. Lorsque la tendance n'est pas en réalité rigoureusement stable, cette seconde méthode a pour désavantage d'exagérer certains coefficients saisonniers.

Autre application -

On étudie l'allure de certaines séries cumulées et cette préoccupation se rattache à la recherche de coefficients saisonniers ; pour cela on calcule en fonction du cumul final les pourcentages atteints à la fin des périodes successives. Au bout de n années d'observations, on peut encore pour un mois donné rechercher le pourcentage médian et être ainsi en mesure à partir d'un étalonnage préalable d'avancer certaines prévisions. Un exemple est donné graphique VII.5 d'un tel étalonnage de la commercialisation du coton-graines en A.E.F.

Tous les procédés de calcul qui viennent d'être exposés ne doivent pas faire oublier que la recherche de variations saisonnières se justifie si ces mouvements périodiques existent en réalité, autrement dit tout calcul peut mener à un résultat numérique sans pour cela correspondre à une réalité concrète (1).

7.3 - Comparaison de séries chronologiques :

7.3.I. Après l'étude individuelle d'une série chronologique, il convient de passer à la comparaison de deux séries chronologiques entre elles.

Il arrive en effet que l'on soupçonne une liaison entre deux variables, qui font l'objet de mesures ou d'observations régulières dans le temps, simplement par l'examen des graphiques représentatifs.

Le trafic d'un chemin de fer et le volume de la manutention d'un port situé à une extrémité du même réseau

(1) Voir l'étude parue en Côte d'Ivoire en 1950 dans "Statistiques Côte d'Ivoire" sur les variations saisonnières du prix de cacao, variations qui en définitive résultent de seuls artifices de calculs.

constituent un exemple d'une telle liaison (1). La liaison commune est le trafic des marchandises embarquées et débarquées.

Un cas plus complexe est celui de l'évolution de l'emploi des ouvriers dans le bâtiment et de l'évolution de la consommation du ciment. Le lien est plus souple, car la productivité intervient, nouveau paramètre qui peut avoir une influence négative sur le volume de l'emploi. Un nouveau type de matériau peut être aussi la cause d'une disparité dans les évolutions des deux séries.

7.3.2 - Pour se rendre compte de l'allure de deux évolutions chronologiques, on est amené d'abord à présenter sur le même graphique les deux séries observées.

Si les grandeurs mises en jeu en ordonnées sont identiques, il n'y a pas de difficulté sauf si leurs ordres de grandeurs sont très différents, dans ce cas on est obligé de décaler, de faire glisser l'un des graphiques par rapport à l'autre.

Dans le cas de grandeurs d'essences différentes sans commune mesure, en abscisse, il n'y a pas de difficulté, car l'échelle des temps est commune, en ordonnée, par contre, le problème se complique car, si l'on adopte des échelles très dissemblables, l'impression visuelle risque d'être faussée (2).

On peut remédier à cet inconvénient en retenant une année pour période de base, commune aux deux séries et rapporter toutes les variations ultérieures en indices par rapport à cette année de base.

Mais, là encore, on n'est pas à l'abri de la critique car le fait de choisir une période donnée pour période de base laisse supposer, soit qu'à ce moment existait une situation normale et que les fluctuations postérieures peuvent être considérées comme s'écartant d'une position

---

(1) Bulletin d'informations économiques et sociales, juillet 1950, page 486. Trafics du chemin de fer Congo Océan, du port de Pointe Noire et de la navigation sur le Congo.

(2) On a vu chapitre V comment pouvait être résolue graphiquement cette difficulté.

d'équilibre, soit que pour la période de base commune, les situations enregistrées par l'une et l'autre série étaient comparables.

7.3.3 - La liaison entre deux séries chronologiques peut être définie par le calcul d'<u>indices de covariation</u>.

Si l'on considère deux séries U et V se composant chacune de n observations, on calcule pour chacune de ces deux séries la moyenne générale de ces observations $\bar{u}$ et $\bar{v}$ et les tendances (u) et (v). Graphique VII.6

Pour déterminer la liaison entre les deux séries, on considérera les <u>écarts</u> de chaque observation, soit par rapport à la <u>moyenne générale correspondante</u> (covariation tendancielle),
soit par rapport à la <u>valeur ajustée</u> de la courbe de tendance (covariation différentielle).

Si les écarts sont de même sens, la covariation est dite positive, s'ils sont de sens contraire la covariation est négative.

Autrement dit, la liaison de deux séries chronologiques sera caractérisée numériquement en étudiant la concordance ou la discordance des variations enregistrées durant les mêmes périodes.

a) Si l'on étudie la <u>covariation tendancielle</u> et si U et V sont les deux courbes, on calculera $\bar{u}$ et $\bar{v}$, moyennes arithmétiques des observations, durant la période considérée et on formera les différences :

$$u_i : \quad u_1 - \bar{u} \qquad\qquad v_i : \quad v_1 - \bar{v}$$
$$\qquad\quad u_2 - \bar{u} \qquad\qquad\qquad v_2 - \bar{v}$$

Soit "C" le nombre de ces concordances, et "D" le nombre des discordances.

On calcule l'<u>indice simple de covariation tendancielle</u> par la formule :

$$i = \frac{C - D}{C + D} = \frac{C - D}{N}$$

S'il y a concordance ou discordance complète, on a :

$$i = \pm 1$$

s'il n'y a pas covariation : $C = D$ et $i = 0$.

Cet indice ne tient pas compte de la grandeur des différences, aussi on peut calculer un <u>indice pondéré de covariation tendancielle</u> :

on calculera la somme des $u_i \cdot v_i$ positifs, soit S_c

la somme des $u_i \cdot v_i$ négatifs, soit S_d

et l'indice pondéré de covariation tendancielle sera égal à :

$$I = \frac{S_c - |S_d|}{S_c + |S_d|}$$

b) Lorsque l'on veut étudier la <u>covariation différencielle</u>, on compare chaque terme U et V aux valeurs U_t et V_t pour la même période des deux tendances de longue durée, et l'on calcule les deux indices simple et pondéré définis ci-dessus.

o
o o

Une variante de cette méthode est de comparer l'année "n" d'une série à l'année "n+1" ou "n-1" de l'autre. En effet, dans le domaine agricole, il arrive souvent que deux séries aient des <u>variations</u> identiques mais <u>décalées</u> l'une par rapport à l'autre.

Soient la production d'arachides du Sénégal et le prix payé au producteur. Le tableau ci-dessous précise pour chaque campagne la production base coques et le rapport du prix de l'arachide au prix du riz. La production est certes liée aux facteurs climatiques, toutefois elle dépend des superficies ensemencées, elles-mêmes fonction du <u>pouvoir d'achat</u> de l'arachide au cours de la campagne précédente.

Campagnes	Production 1.000 T.	Prix (100 k.) Arach. Coques	Prix (100 k.) Riz	Rapport Arachide / Riz
1934/35	364	78,5	70	112,1
1935/36	466	72,5	62,5	116,0
1936/37	574	81,18	95	85,5
1937/38	483	92,72	165	56,2
1938/39	569	94,20	165	57,1
1939/40	560	90	175	51,4
1940/41	419	95	280	33,9
1941/42	199	140	320	43,8
1942/43	114	185	450	41,1
1943/44	274	188	1000	18,8
1944/45	234	300	1400	21,4
1945/46	354	350	1700	20,6
1946/47	365	600	1800	33,3
1947/48	452	700	2000	35,0
1948/49	434	1450	2400	60,4
1949/50	429	1450	3400	42,6
1950/51	342	2300	3500	65,7
1951/52	441	1925	3600	53,0
1952/53	449	1850	4500	41,0

Le décalage des variations d'une série sur une autre série est offert par l'étude de la variation de l'indice des prix de détail à Brazzaville en 1951-1952 qui suit avec un décalage de 6 à 9 mois la variation de l'indice des prix à Paris.

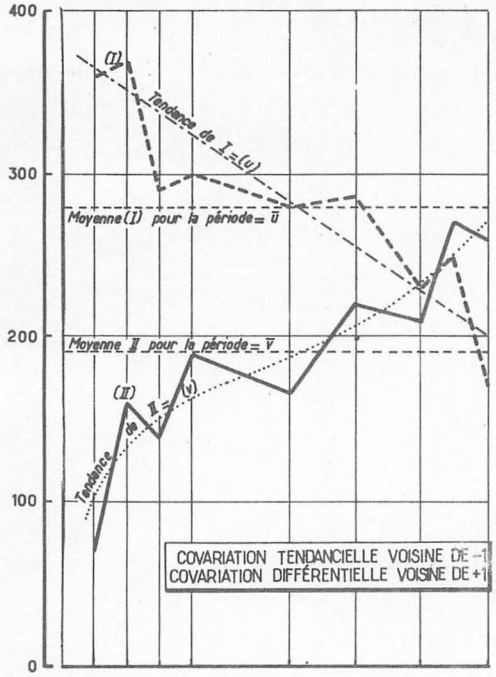
7.3.4 - Variations à court terme. Coefficient de dépendance.

La liaison entre deux séries chronologiques peut être étudiée en calculant très simplement les différences successives (avec leurs signes) des termes de chaque série, soit pour reprendre les notations précédentes Δu et Δv. On effectue les produits $\Delta u \cdot \Delta v$ et on définit un nouveau coefficient dit de dépendance par :

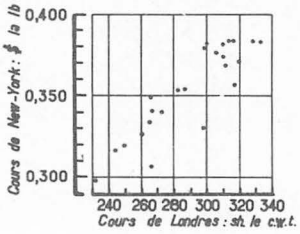
$$\frac{\text{somme algébrique des produits } \Delta u \, \Delta v}{\text{somme des valeurs absolues des produits } \Delta u \, \Delta v}$$

L'exemple ci-dessous est emprunté à deux séries publiées dans la publication " Mouvement économique en France de 1929 à 1939 " éditée par le Service National des

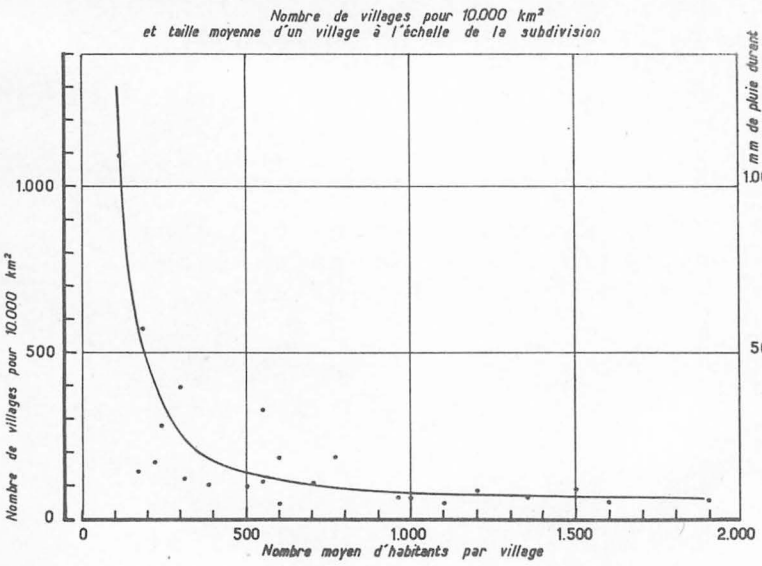
Graphique VII_6



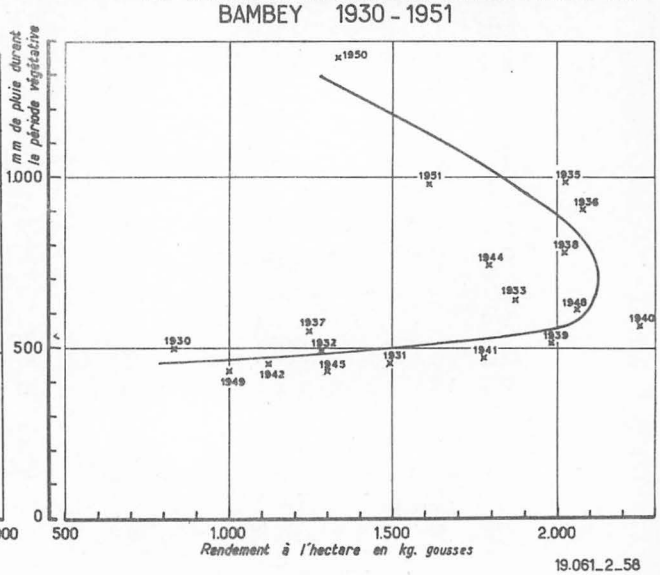
Graphique VII-7



Graphique VII-8_GUINÉE_RECENSEMENTS ADMINISTRATIFS 1954-1955



Graphique VII-9_RENDEMENT À L'HECTARE DE L'ARACHIDE AU CENTRE DE RECHERCHES AGRONOMIQUES DE BAMBEY 1930-1951



19.061_2_58

statistiques - pages 160 et Annuaire statistique de la France 1949, page 161.

	Montant des mandats ordinaires et bons de poste émis millions f.	Δu	Recettes hebdomadaires de la S.N.C.F. millions f.	Δv	$\Delta u \cdot \Delta v$ +	$\Delta u \cdot \Delta v$ -
1928	5.252		282			
1929	5.866	+ 6I4	298	+ 16	9.824	
1930	7.067	+ I20I	297	- 1		1.201
1931	6.591	- 476	270	- 27	12.852	
1932	6.902	+ 3II	23I	- 39		12.I29
1933	6.58I	- 32I	2I8	- 13	4.I73	
1934	5.9I4	- 667	208	- 10	6.670	
1935	6.835	+ 92I	189	- 19		17.499
1936	7.800	+ 965	192	+ 3	2.895	
1937	8.7I3	+ 9I3	242	+ 50	45.650	
1938	9.954	+ 1241	285	+ 43	53.363	
					135.427	30.829

$$C = \frac{135.427 - 30.829}{135.427 + 30.829} = 0,63$$

7.4 Corrélation et contingence.-

7.4.I- L'étude de la corrélation entre deux séries chronologiques, facilite l'étude de la corrélation entre deux grandeurs quelconques.

Les cours mensuels du cacao sur les marchés de Londres et New-York sont suivis régulièrement. Le tableau ci-dessous précise ces cours mensuellement en 1951 et 1952.

	1951		1952	
	Londres	New-York	Londres	New-York
Janvier	311-2	0,370	298-8	0,331
Février	309-6	0,376	316-10	0,358
Mars	316-1	0,384	328-6	0,384
Avril	315-0	0,384	319-11	0,381
Mai	310-10	0,382	333-13	0,384
Juin	299-6	0,383	305-7	0,378
Juillet	263-8	0,350	297-8	0,381
Août	281-8	0,355	285-2	0,354
Septembre	276-7	0,341	263-6	0,333
Octobre	250-10	0,320	263-4	0,341
Novembre	231-4	0,295	242-3	0,318
Décembre	260-6	0,326	264-7	0,308

Londres : Good Fermented Accra CIF Londres Sh le cwt
New-York : Accra disponible en vrac FOB New-York $ la lb.

Depuis la fin de la guerre, on a pu noter sur les marchés mondiaux un certain alignement (compte tenu des taux de change), des cours des principales matières premières. C'est la raison des évolutions très parallèles de chaque série de cotations.

Au lieu de chercher simplement la covariation entre ces deux séries, c'est-à-dire s'il y a concordance ou non des variations des deux courbes, on peut tenter de caractériser numériquement la liaison entre ces deux séries d'indices.

On construit un graphique, en abscisse on porte le prix du cacao à Londres et en ordonnée le prix du cacao à New-York. A chaque mois correspond un point.

On vérifie que les points sont grossièrement alignés de plus si les cours augmentent de 40 % à Lonires, ils augmentent également de 40 % à New-York en raison de l'interdépendance des deux bourses. En fait la liaison n'est pas rigoureuse et on obtient un nuage de points, nommé diagramme de dispersion (Graphique VII.7). Plus le nuage sera allongé et filiforme, plus on dira que la corrélation est forte.

Au contraire, si on a affaire à un nuage dispersé de densité peu variable, on en déduit que la corrélation est faible.

Le raisonnement qui vient d'être précisé à partir de deux séries chronologiques s'applique à deux mesures d'une unité statistique quelconque, et la corrélation dépasse là la simple comparaison de deux séries chronologiques : pour une même unité statistique, on mesurera deux caractères différents et l'on recherchera la liaison qui peut exister entre les deux variables étudiées. On vient de donner un exemple de corrélation positive : les deux variables variant en moyenne dans le même sens. La variation des superficies en cultures arbustives en fonction de l'âge de l'exploitant en constitue une autre illustration :

	Age de l'exploitant				
	15/29	30/39	40/49	50 et +	Ensemble
	hectares par exploitation				
Cacao	1,38	2,57	4,32	4,60	3,19
Café	1,14	2,03	2,57	2,88	2,15
Total	2,52	4,60	6,89	7,48	5,34

La liaison entre les âges des époux lorsque l'on étudie les statistiques de mariages est classique.

Mais le cas de corrélation négative est également fréquent.

Diminution de la natalité avec l'augmentation du niveau de vie :

Tranche de dépenses par personne (roupies/mois)	zone rurale	zone urbaine
	nombre d'enfants nés vivants par ménage	
0 - 10	3,40	3,44
11 - 20	3,02	3,38
21 - 30	2,95	3,10
31 et au-dessus	2,70	2,78

Enquête nationale indienne de 1954

Proportion d'accidents d'enfants selon l'importance relative d'espaces non bâtis

Districts	Proportion d'espaces non bâtis	Proportion d'accidents d'enfants
A	5,0	46,3
B	2,2	43,4
C	1,3	42,9
D	4,2	42,2
E	1,4	40,0
F	2,0	38,8
G	7,0	38,2
H	2,5	37,4
I	4,5	37,0
J	3,1	33,3
K	5,2	33,6
L	7,2	33,6
M	6,3	30,8
N	12,2	28,3
O	14,6	23,8
P	23,6	17,8
Q	14,8	17,1
R	27,5	10,8

Nombre de décès par jour férié occasionnés par le trafic automobile selon la durée du pont aux Etats-Unis de 1946 à 1955.

Durée du pont	Nombre de ponts	Nombre de jours d'observation	Nombre de morts	Moyenne journalière de décès
1 jour	6	6	821	137
2 jours	6	12	1.674	139
3 jours	30	90	10.321	120
4 jours	8	32	3.IOI	97
	--	---	------	---
	50	140	16.4I7	117

7.4.2 Pour caractériser numériquement la corrélation, il faut distinguer d'abord le cas où le nuage de points est sensiblement filiforme. Dans ce cas on calcule la position d'une droite dite de régression ayant la propriété suivante : la somme des carrés des écarts, comptés parallèlement à l'axe des ordonnées, entre le point observé et le point correspondant situé sur la droite de régression, est minimum. Toute autre position de la droite donnerait lieu à une somme supérieure, elle est appelée aussi droite des moindres carrés.

Cette droite passe par le point G dont l'abscisse est la moyenne arithmétique des valeurs de la variable portée en abscisse $\bar{x}$ et l'ordonnée la moyenne arithmétique des valeurs de la variable portée en ordonnée $\bar{y}$. Si les écarts de chacune des deux variables par rapport à leurs moyennes sont désignés par X_i et Y_i

$$\text{avec} : \quad X_i = x_i - \bar{x}$$

$$Y_i = y_i - \bar{y}$$

la pente de la droite de régression est égale à :

$$a = \frac{\sum X_i \; Y_i}{\sum X_i^2}$$

et le coefficient de corrélation linéaire est égale à :

$$r = \frac{\sum X_i \; Y_i}{\sqrt{\sum X_i^2 \; . \sum Y_i^2}}$$

Ce rapport varie de -1 à +1.

Le tableau ci-dessous donne un exemple numérique de calcul à partir des données du tableau sur les accidents d'enfants en fonction de la proportion d'espaces verts.

	Δx	Δy	$(\Delta x)^2$	$(\Delta y)^2$	$\Delta x \times \Delta y$ +	$\Delta x \times \Delta y$ −
A	- 3,0	+ 13,2	9,0	174,24		39,60
B	- 5,8	+ 10,3	33,64	106,09		59,74
C	- 6,7	+ 9,8	44,89	96,04		65,66
D	- 3,8	+ 9,1	14,44	82,81		34,58
E	- 6,6	+ 6,9	43,56	47,61		45,54
F	- 6,0	+ 5,7	36,00	32,49		34,20
G	- 1,0	+ 5,1	1,00	26,01		5,10
H	- 5,5	+ 4,3	30,25	18,49		23,65
I	- 3,5	+ 3,9	12,25	15,21		13,65
J	- 4,9	+ 0,2	24,09	00,04		0,98
K	- 2,8	+ 0,5	7,84	00,25		1,40
L	- 0,8	+ 0,5	0,64	00,25		0,40
M	- 1,7	- 2,3	2,89	5,29	3,91	
N	+ 4,2	- 4,8	17,64	23,04		20,16
O	+ 6,6	- 9,3	43,56	86,49		61,38
P	+ 15,6	- 15,3	243,36	234,09		238,68
Q	+ 6,8	- 16,0	46,24	256,00		108,80
R	+ 19,5	- 22,3	380,25	497,29		434,85
			991,54	1.701,73	+ 3,91	1188,37
						3,91
						1184,46

$$\sum x = 144,6 \qquad \sum y = 595,3$$

$$\bar{x} = \frac{144,6}{18} = 8,0 \qquad \bar{y} = \frac{595,3}{18} = 33,1$$

$$r = \frac{-1184,46}{\sqrt{991,54 \times 1701,73}} \# -0,90$$

$$a = \frac{-1184,46}{991,54} \# -1,2$$

7.4.3.- Le nuage de points n'affecte pas toujours une forme linéaire, dans le cas d'une étude d'une loi de demande, on observera une répartition de forme hyperbo-

lique. Il en est de même si l'on porte sur un graphique par subdivision la population moyenne des villages de Guinée et la densité par km2 de ces villages (Graphique VII.8). Un autre exemple est celui du rendement de l'arachide en fonction de la pluviométrie utile (Graphique VII.9).

Les formules données précédemment ne doivent pas être employées dans ces cas où la régression n'est pas linéaire.

Remarque importante : Il faut faire attention au fait que l'on peut calculer une corrélation très voisine de l'unité sans pour cela qu'il y ait aucune liaison entre les deux variables.

On a eu l'idée de comparer le trafic des chemins de fer français pendant huit ans et le nombre de décès pendant le même laps de temps. On trouve : C = + 0,46. En Angleterre la corrélation entre le nombre de postes de TSF déclarés et le nombre de cas de maladies mentales détectées est très élevé C = 0,98. Une corrélation peut être trouvée entre l'âge des femmes et l'angle des empreintes des pieds au cours de la marche, une autre entre les salaires des ministres du culte presbytérien et le prix du rhum à la Havane ....

Toutes ces liaisons sont évidemment absurdes et illustrent le danger des calculs menés d'une manière mécanique.

7.4.4.- La contingence : Les méthodes ci-dessus supposent que l'on peut mesurer quantativement les caractères observés. Il est des cas où les caractères présentent plusieurs possibilités mais non susceptibles d'une mesure quantitative.

On a classé dans le recensement de la population non originaire du Cameroun en 1951 les recensés du sexe masculin selon le lieu de naissance de leur père et le lieu de naissance de leur mère.

Tableau de contingence relatif à la liaison qui existe entre le lieu de naissance des parents des recensés européens et assimilés en 1951 au Cameroun.

Lieu de nais. du père \ Lieu de naissance de la mère	France	Afrique du Nord	Départements d'O.M.	Territoires d'O.M.	Europe	Asie	Afrique	Amérique	Océanie	Etrangers s.a.i.	Non déclaré	Total
	recensés de sexe masculin											
France	5064	57	13	71	117	3	11	14	1	3	22	5376
Afrique du Nord	43	47	-	7	4	-	-	-	-	-	-	101
Départements d'Outre-Mer	18	-	74	2	1	-	-	1	-	-	1	97
Territoires d'O.M.	26	-	-	123	-	2	1	-	-	-	3	155
Europe	157	6	-	13	1063	7	10	12	-	1	3	1272
Asie	7	1	-	6	7	128	4	-	-	-	-	153
Afrique	8	-	1	3	2	3	11	-	-	-	-	28
Amérique	10	-	-	-	3	2	1	66	-	-	-	82
Océanie	-	-	-	-	-	-	-	-	-	-	-	-
Etrangers s.a.i.	3	-	-	-	2	-	-	-	-	30	-	35
Non déclaré	37	1	-	8	8	-	1	1	-	-	225	281
Total	5373	112	88	233	1 207	145	39	94	1	34	254	7580

L'examen de ce tableau laisse pressentir une liaison, puisque le lieu de naissance de la mère n'est pas indépendant du lieu de naissance du père.

Dans une enquête sur les résultats de la vaccination contre la fièvre typhoïde, on observe les résultats suivants :

		Manifestation de la maladie		
		oui	non	total
Vaccinations	oui	56	6759	68I5
	non	272	II396	11668
Total		328	18I55	18483

JSS Paris, avril mai juin 1957. Etude statistique de l'influence de la vaccination sur la manifestation de la fièvre typhoïde.
Travaux de Greenwood et Yule
(suite à campagne de presse de l'Association des parents d'élèves contre la vaccination)

S'il y a indépendance entre les deux variables : existence de vaccination - manifestation de la maladie, on doit enregistrer, que les enfants aient été ou non vaccinés, la même proportion de malades (Cette proportion pour l'ensemble du lot observé est de $\frac{328}{18483}$). Par conséquent on peut bâtir un tableau théorique qui résume les observations en supposant l'inefficacité totale de la vaccination.

		Manifestation de la maladie		
		oui	non	total
Vaccinations	oui	122	6.693	6.8I5
	non	206	II.462	II.668
	total	328	I8.I55	I8.483

Pour caractériser numériquement la dépendance on calcule l'expression :

$$\chi^2 = \frac{(56-122)^2}{56} + \frac{(6759-6693)^2}{6759} + \frac{(272-206)^2}{272} + \frac{(11396-11462)^2}{11391}$$

$$= 66^2 \left[\frac{1}{56} + \frac{1}{6759} + \frac{1}{272} + \frac{1}{11.396} \right]$$

Les tables construites à cet effet montrent que l'indépendance est peu vraisemblable et que la vaccination a, dans ce cas, un effet de protection. Plus le nombre obtenu appelé χ^2 est important plus les écarts avec la répartition théorique sont importants et l'hypothèse d'indépendance est à rejeter.

Ces calculs sont fréquemment utilisés dans les enquêtes d'opinion publique ou les enquêtes sociologiques,

soit pour tester les réactions et comportements d'un même lot d'individus à différentes questions :

Enquête en Belgique - 1948

		oui	non	sans opinion	total
Croyez-vous que le gouvernement indiqué désire sincèrement la paix :	Etats-Unis	685	139	176	1.000
	U.R.S.S.	180	602	2I8	1.000
		865	74I	394	2.000

soit pour comparer les réponses de deux groupes à un même questionnaire (l'un des groupes pouvant éventuellement être retenu comme groupe-témoin) :

L'exemple ci-dessous est tiré d'une étude sociologique récente (1).

Opinion de la population lettrée sur l'attribution des enfants en cas de divorce :

En cas de séparation	Pour cent personnes interrogées de chaque sexe Hommes	Femmes
les fils doivent rester		
avec le père	91	48
avec la mère	0	25
être partagés	1	5
pas d'opinion	8	22
Total :	100	100

(1) Source : PORTO-NOVO par C. Tardits (Série Etudes de l'Ecole pratique des Hautes Etudes, n° VII).

En cas de séparation	Pour cent personnes interrogées de chaque sexe	
	Hommes	Femmes
les filles doivent rester		
avec le père	88	26
avec la mère	3	48
être partagés	1	3
pas d'opinion	8	23
Total :	100	100

Le test précédent du χ^2 confirme la différence significative de réaction aux questions posées aux personnes de chaque sexe.

BIBLIOGRAPHIE

En dehors des livres déjà cités dans les chapitres précédents, on consultera avec utilité :

- L'entreprise et la statistique - R. Dumas
  (Chapitre III, Sections 3 et 4) chez Dunod.

- Initiation pratique à la statistique - A. Liorzou
  (Chapitres V et VI) chez Eyrolles.

- Economie politique et technique statistique -
  (3ème partie, chapitre II) A. Marchal
  Librairie générale de droit et de jurisprudence.

- Statistical Methods - Herbert Arkin et Raymond R-Colton
  (Chapitres V à XI). Barnes and Noble Inc.

- Trafic postal et cycle économique - Henri Anglade
  I.N.S.E.E. - Etudes et documents.

- Documents et Statistiques, Ministère de la France
  d'Outre-Mer : production d'arachide et circulation
  monétaire au Sénégal.

Chapitre VIII

LES INDICES

8.I Définitions.

8.2 Les problèmes.

8.2.I- Les difficultés
8.2.2- La base
8.2.3- Les formules de calcul

8.3 Signification des coefficients de pondération.

8.4 Les indices de volume.

8.5 L'utilisation des indices. Indices bruts et corrigés.

Chapitre VIII

8.I Définitions.-

8.I.I- Dans le domaine statistique, un indice n'est pas une indication permettant d'asseoir une prévision, au risque de décevoir, un indice n'est pas non plus un renseignement qui facilite l'arrestation d'un coupable. En statistique, un indice est un nombre qui caractérise, qui résume, la variation relative de la valeur d'une certaine grandeur ou l'évolution d'un certain ensemble, variation entre une époque de référence 0 et une autre époque 1 ; la situation 0 est dite la situation de base, et la situation 1 est la situation actuelle. Si on considère une seule grandeur mesurable telle que l'effectif d'une promotion d'étudiants, pour caractériser à l'aide d'un indice la variation de cet effectif en 1958 par rapport à 1938, on calcule le nombre I tel que :

Indice de l'effectif $^1/_0$: $\frac{NI \quad 1958}{NO \quad 1938}$

D'une manière plus générale la formule est :

$$I = \frac{1}{0} = 100 \frac{GI}{GO}$$

Cette notation signifie que 100 repère la valeur de la grandeur à l'époque de base.

Un indice simple sera dono le rapport (multiplié par cent) de la valeur du nombre, mesure de la grandeur à son niveau actuel, à un premier nombre qui mesure la grandeur à l'époque de base.

Pour reprendre l'exemple précédent,

si $N_1 = 40$ et $N_o = 45$, l'indice en 1958 sur la base 100 en 1938 est égal à 89, on peut dire aussi que l'effectif a fléchi de 11 %, il ne faut donc pas confondre indice et pourcentage d'accroissement ou de diminution. Il n'est d'ailleurs pas inutile de rappeler que les pourcentages ne s'ajoutent pas et ne sont pas réversibles.

Prix kg. café	1949 : 200			
	1950 : 250	+ 25 %	100 %	et non
	1953 : 400	+ 60 %		25 + 60

De 200 frs à 400 frs, l'augmentation est de 100 %.

De 400 à 200 frs, la diminution est de $\frac{200-400}{400}$ soit 50 %.

Un indice n'est pas un nombre absolu ou une mesure, mais une <u>grandeur relative</u>, il s'agit d'un simple repérage par rapport à une <u>origine arbitraire</u> : <u>la base</u>.

8.I.2- On distingue trois catégories d'indices :

- les indices <u>partiels</u>
- les indices <u>synthétiques</u> et
- les indices <u>composites</u>.

a) Un <u>indice partiel</u> analogue à celui qui vient d'être calculé est le rapport multiplié par cent de deux nombres relatifs à un même fait. On dit que l'indice partiel du prix du café sur le marché français en Janvier 1953, (la moyenne annuelle en 1949 étant retenue pour base et les cotations concernant le Robusta courant au stade C.A.F.) est :

moyenne 1949 = 152.6 francs par kilog.
janvier 1953 = 36I

I # 237

De même, on peut calculer l'indice partiel de la production commercialisée d'arachides au Sénégal pour la période 1951-1954, la période 1935-1938 étant choisie pour base.

Campagnes	1.000 tonnes (coques)	
1935-1936	530	moyenne avant guerre : 539
1936-1937	574	
1937-1938	482	
1938-1939	568	
1951-1952	442	moyenne après guerre : 456
1952-1953	449	
1953-1954	552	
1954-1955	383	

Indice de la production $\frac{\text{après guerre}}{\text{avant guerre}}$ $\frac{456}{539} \times 100 = 84.$

b) Un indice synthétique est un groupement d'indices partiels relatifs à des grandeurs de même espèce. Un indice général des prix ou un indice de production répond à cette définition : l'un comprend le groupement d'indices partiels de prix, et l'autre le groupement d'indices partiels de tonnages. Les indices synthétiques répondent au souci de représenter par un nombre unique un phénomène complexe.

c) Un indice composite est un groupement d'indices partiels ou d'indices synthétiques, mais ces indices se rapportent à des grandeurs d'essence différente : un indice de productivité est le rapport d'un indice synthétique de production à un indice d'emploi ou d'heures travaillées.

On insistera ici sur les indices synthétiques, car les indices composites font appel beaucoup plus à l'interprétation des statistiques qu'à leur élaboration.

d) Du point de vue de la classification des indices, on distingue les indices chronologiques : les comparaisons visant les variations dans le temps et les indices régionaux ou géographiques qui ont pour objet l'étude à une même époque d'éléments semblables mais observés en des lieux différents (parité de change).

8.2 - Les problèmes.-

Trois problèmes fondamentaux sont soulevés par le calcul des indices :

les pondérations
le choix de la période de base
la formule de calcul.

8.2.I- Les pondérations. Un indice synthétique résulte de la combinaison d'indices simples, si dans cette combinaison chaque indice partiel a une importance égal aux autres l'indice est dit non pondéré si au contraire chaque indice partiel intervient avec son importance propre, l'indice est pondéré. Soient les prix du lait et du vin à deux époques :

Situation	Prix du lait (francs par litre)	Prix du vin (francs par litre)
0	25	50
1	40	60

Les indices partiels sont :

pour le lait ..... 160 pour le vin 120

L'indice synthétique non pondéré est égal à :

$$\frac{160 + 120}{2} = 140$$

Par contre, si l'on estime qu'il ne faut pas accorder la même importance aux variations des prix des deux articles considérés, différentes solutions peuvent être apportées :

a) on multipliera chaque indice partiel par un coefficient appelé poids et la somme des indices partiels (multipliés par ces coefficients) dits alors pondérés sera divisée par la somme des poids.

Si on adopte une optique antialcoolique, on peut attrivuer le poids 4 au lait et le poids 1 au vin ; le calcul donne :

$$\frac{160 \times 4 + 120 \times 1}{4 + 1} = 152$$

Au contraire, si on juge la consommation de vin essentielle, on adoptera d'autres coefficients pour caractériser l'élévation des prix, compte tenu des habitudes de consommation par la formule :

$$\frac{160 \times 1 + 120 \times 4}{1 + 4} = 128 \quad (1)$$

b) On aurait pu également mettre en oeuvre une autre méthode : additionner 40 et 60, 25 et 50, et comparer les deux dépenses. On aurait obtenu un nouvel indice égal à :

$$100 \cdot \frac{100}{75} = 133,3$$

---

(1). Note. Sans même pondérer on peut arriver à des résultats fort paradoxaux selon le choix de l'année de base

	1949	1956
lait	50	25
vin	50	100
Indices base 100 en 1949	50 / 200	 125
Indices base 100 en 1956	200 / 50	 125

Ce phénomène sera évoqué au paragraphe 8.2.3.I, à propos de la réversibilité des indices établis à l'aide de formule arithmétique.

Ce nouvel indice appelé indice agrégatif comporte un inconvénient grave car il dépend des unités employées : si on maintient la cotation du lait au litre mais que l'on enregistre le prix du vin à l'hectolitre, on a :

$$\frac{40 + 6.000}{25 + 5.000}$$

qui est évidemment différent du premier indice ; le résultat est d'autant plus différent que l'unité utilisée pour l'un des produits est plus grande.

On s'aperçoit que les résultats sont fort dissemblables selon les diverses hypothèses, les formules de calculs employées et que l'époque retenue pour période de base joue aussi son rôle propre.

8.2.2.- La base.- Un indice mesure une variation, il en résulte que si, à l'origine la base correspond à un niveau absolu, peu élevé, la variation ultérieure peut être importante et s'exprimer par un indice élevé, si au contraire, la base est établie à une position haute de la conjoncture, les indices seront faibles. Des controverses passionnées ont visé dans le passé l'évolution de la production industrielle en France. Cet indice est un indice de quantité ou de tonnage.

Si l'on retient l'année 1928 pour base 100, la série suivante pour les années d'avant-guerre pour la production (y compris bâtiment) est disponible :

1928	.....	100	1934	.....	82
1929	.....	I09	I935	.....	80
I930	.....	I08	I936	.....	85
I93I	.....	94	I937	.....	89
I932	.....	79	I938	.....	83
I933	.....	88			

L'année 1938 par rapport à l'année 1928 se place à un niveau très inférieur à 100. Or, depuis la guerre, un indice de la production industrielle plus complet que le précédent, a été calculé sur la base 100 en 1938, pour les années récentes :

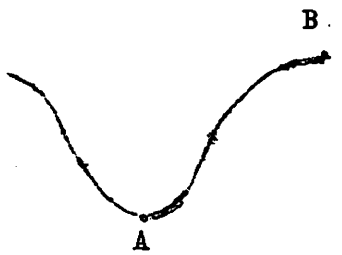
1946	.....	84	I950	.....	I23
1947	.....	99	I95I	.....	I38
I948	.....	II3	I952	.....	I44
I949	.....	I23	I953	.....	I40
			I954	.....	I52

Pour un lecteur non prévenu, il semble que la production est en plein essor ; mais, si l'on compare 1954 non pas à 1938, mais à 1928 (ceci par la simple multiplication

$$\frac{152 \times 83}{100} \quad (1) \quad)$$

on trouve que 1954 dépasse seulement de 25 % le niveau de 1928, et de 14 % celui atteint en 1929 juste avant la crise.

On juge ainsi d'une manière concrète, combien le choix de la période de base est important, notamment dans le domaine des études de prix car si la production physique présente une certaine inertie, les prix par contre, sont susceptibles de subir des fluctuations importantes et rapides, si l'on choisit au cours d'un cycle, une période de base, qui se place dans une phase de dépression des prix, on calcule des indices bien différents et très supérieurs aux indices calculés sur la base B. Un cas maintenant classique est celui des indices de prix à l'exportation des pays d'outre-mer de la Communauté Française : ces indices sont calculés sur la base 100 en 1949. Cette année est la dernière durant laquelle les prix des principaux produits à l'exportation des pays français d'outre-mer n'étaient pas encore alignés sur les cours sur les marchés mondiaux des mêmes produits, il en résulte que les indices calculés sur cette base sont artificiellement "soufflés". On se rend compte de ce phénomène en calculant les indices partiels du café libéré début 1950 et du cacao libéré début 1949 pour les trois zones monétaires en 1952 sur la base 100 en 1949.



	zone franc	zone sterling	zone dollar
café	243	218	166
cacao	165	154	164

Il faut donc que la période de base corresponde soit à une année normale de l'économie, soit à une moyenne d'années suffisamment nombreuses (dix années par exemple) pour atténuer l'effet du cycle.

(1) Sous certaines réserves que l'on verra au paragraphe 8.3.3.

Toutefois il est difficile de séparer le normal de l'anormal et la documentation chiffrée ne permet pas toujours le retour en arrière en vue du calcul de moyennes s'étendant sur plusieurs années. On se trouve alors dans l'obligation de retenir ce qui est disponible et le matériel statistique que l'on retrouve dans les archives n'est pas toujours, ni le meilleur ni le plus indiqué (1).

Il convient donc d'être prudent lorsque l'on aborde le calcul des nombres indices.

On vient d'en voir successivement les deux raisons principales :

a)- selon les hypothèses initiales sur les coefficients de pondération, on peut en toute bonne foi parvenir à des résultats parfois très différents les uns des autres.

b)- un indice repère une situation par rapport à une situation choisie comme terme de comparaison.

8.2.3. Les formules de calcul.

On peut combiner diversement les indices partiels, pour obtenir un indice synthétique, mais ces différentes combinaisons, autrement dit les formules de calculs, présentent des avantages et des inconvénients.

8.2.3.1- Qualités d'un indice. Irving Fisher a défini les propriétés d'un indice synthétique parfait :

a)- Si les valeurs à l'époque I sont égales à celles de l'époque 0, $I^{1}/_{0}$ doit être égal à 1 (ou 100) ; principe d'identité.

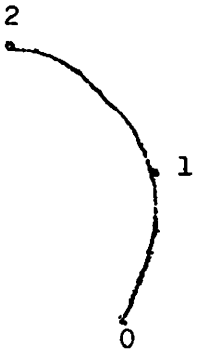
b)- Si l'on change d'époque de base en retenant les valeurs de l'époque 1 comme valeurs de référence, le calcul de $I^{0}/_{1}$ doit donner pour résultat :

$$\frac{1}{I^{1}/_{0}}$$: principe de réversibilité.

(1) Un exemple du caractère malencontreux de certains choix est la prise en considération de l'année 1938 pour le commerce extérieur de l'A.O.F., car cette année fut assez médiocre.

c)- Si l'on calcule un premier indice synthétique sur la base 0 pour l'époque I, puis un second indice pour l'époque 2 sur la base 1, on doit trouver pour l'indice de l'époque 2 sur la base 0 (à un facteur 100 près) :

$$I\ 2L0 = I\ 1/0 \times I\ 2/1$$: principe de <u>transitivité</u>.

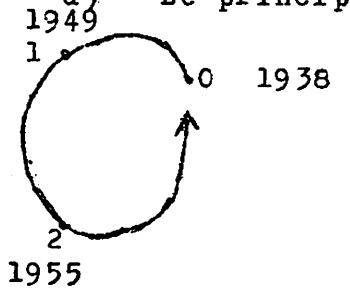


Soit un indice 1949 base 1938 = 500
et l'indice 1952 base 1949, = 150
l'indice 1952 base 1938 doit fournir la valeur :

$$\frac{500}{100} \times 150 = 750$$

(C'est le calcul entrepris pour l'indice de la production industrielle au paragraphe 8.2.2.).

d)- Le principe de transitivité entraîne la propriété de <u>circularité</u>, c'est-à-dire



$$I\ ^{1}/_{0} \times I\ ^{2}/_{1} \times I\ ^{0}/_{2} = 1$$

e)- Enfin l'indice doit être <u>indépendant</u> des unités choisies et avoir une valeur <u>finie et déterminée</u>, ces deux conditions sont dites d'<u>homogénéité</u> et de <u>détermination</u>.

8.2.3.2.- <u>Etude de la moyenne arithmétique simple</u>.

Pour préciser ces propriétés, on va étudier si un indice synthétique obtenu à partir d'une moyenne arithmétique simple non pondérée, répond à ces conditions :

		Lait francs par litre	Vin francs par litre
0	situation 1950	25	50
1	situation 1951	100	60

Une formule arithmétique non pondérée donne :

$$I\ {}^{1}/_{0} = \frac{\frac{100}{25} \times 100 + \frac{60}{50} \times 100}{2} = \frac{400 + 120}{2} = 260$$

Si à l'époque I, les prix sont égaux à l'époque de base, on a bien I = 100
mais si on calcule :

$$I\ {}^{0}/_{1} = \frac{\frac{25}{100} \times 100 + \frac{50}{60} \times 100}{2}$$ on trouve 55 qui n'est

pas l'inverse de

$$\frac{1}{I\ {}^{1}/_{0}} \times 100 = 39$$

$I\ {}^{1}/_{0}$ est donc différent de $\frac{1}{I\ {}^{0}/_{1}}$

De même, les conditions de transitivité et de circularité ne sont pas remplies. Une formule arithmétique pondérée présente les mêmes inconvénients.

8.2.3.2.- <u>L'indice géométrique</u>.-

Aussi a-t-on cherché des formules qui satisfassent mieux aux conditions ci-dessus.

On a déjà défini la moyenne géométrique d'une série de nombres, elle est égale à la racine Nème du produit de N nombres lorsque la moyenne géométrique n'est pas pondérée. Lorsqu'elle est pondérée on a la formule :

$$G = \sqrt[n_1+n_2+\ldots]{a_1^{n_1} \cdot a_2^{n_2} \cdot}$$

pour un indice synthétique, on calcule de même :
un <u>indice géométrique simple</u> donné par la formule

$$I = \sqrt[n]{I_1 \cdot I_2 \ldots I_n}$$

un indice géométrique pondéré étant défini par la formule :

$$\sqrt[a_1+a_2+\ldots+a_n]{I_1^{a_1}\, I_2^{a_2} \ldots I_n^{a_n}}$$

Ce deuxième type de moyenne est particulièrement utilisée pour les indices de prix, car la moyenne géométrique jouit de la propriété de réversibilité et de transitivité.

Par exemple : $I^{1}/_{0} = \sqrt{400 \times 120}$

et $I^{0}/_{1} = \sqrt{\frac{1}{400} \cdot \frac{1}{120}} = \sqrt{\frac{1}{400 \times 120}}$

$I^{1}/_{0}$ et $I^{0}/_{1}$ sont bien inverses.

Du point de vue des calculs, il faut passer par les logarithmes des indices partiels, logarithmes qui sont ensuite pondérés.

8.3 - Signification des coefficients de pondération.-

Un indice synthétique résulte de la combinaison d'indices partiels affectés de coefficients. En admettant pour l'instant que l'année prise pour référence demeure fixe, on peut distinguer les indices à coefficients de pondération fixes et les indices à coefficients de pondération variables. Il serait inexact de penser que le choix de ces coefficients est arbitraire, ils correspondent au contraire à une réalité bien concrète.

8.3.I- Recherche de la signification des coefficients de pondération.-

L'objectif est d'entreprendre le calcul d'un indice pondéré de prix, on constitue un panier de provisions comprenant des articles en différentes quantités mais sa composition reste fixe. On calcule aux deux époques : de base et actuelle (la composition de ce panier de provisions demeurant inchangée) les coûts de cet ensemble de marchan-

dises. Le rapport de ces coûts définit l'indice synthétique des prix pour les marchandises considérées.

On prendra un exemple numérique :

	prix			coût	
	époque de base	époque actuelle	consommation	époque de base	époque actuelle
Artide 1 - Lait	25	40	20 l.	500	800
Article 2-Vin	50	60	15 l.	750	900

ou

$$I = \frac{20x40 + 15x60}{20x25 + 15x50} \times 100$$

$$= \frac{\frac{40}{25} \times 25 \times 20 + \frac{60}{50} \times 15 \times 50}{20x25 + 15x50} \times 100$$

$$= \frac{40}{25} \times 100 \times \frac{25x20}{20x25+15x50} + \frac{60}{50} \times 100 \frac{15x50}{20x25+15x50}$$

$$= I_1 \times \alpha_1 + I_2 \alpha_2$$

or $\alpha_1 = \frac{25x20}{20x25+15x50}$ représente l'importance de la dépense de lait à l'époque de base par rapport à la dépense totale à l'époque de base.

On saisit immédiatement la réalité des coefficients de pondération dont la notion avait été introduite très empiriquement au début de ce chapitre. <u>Le coefficient de pondération d'un indice partiel I_j est égal au rapport de la dépense pour l'article j à l'époque de base, à la dépense totale pour l'ensemble des articles durant la même époque</u>.

Le tableau ci-dessous vérifie cette propriété dans le cas du calcul d'un indice des prix à l'exportation.

Produits	Q_o Quantités exportées à l'époque de base (T)	p_o Prix unitaire à l'époque de base (Milliers de Frs à la T.)	P.- Prix unitaire à l'époque actuelle (Miers Frs à la T.)	Q_o p - Valeur globale à l'époque de base (M.ons de francs)	α Coefficient de pondération.	Q_o P. - Valeur actuelle à tonnage constant	Indice partiel de prix. $\frac{p}{p_o}$	Indice partiel pondéré. $\frac{p}{p_o}$ x
Cacao	3.000	150	200	450	46	600	133	6.II8
Café	1.500	200	250	300	3I	375	I25	3.875
Palmiste	1.000	15	10	15	2	10	67	134
Huile de palme	400	25	20	10	1	8	80	80
Coton	2.000	100	125	200	20	250	125	2.500
Ensemble	.....	.....	.....	975	I00	I243		I27,07

$$\text{Indice : } \frac{1.243}{975} \times 100 = 127$$

8.3.2- Base fixe, base variable - Coefficients de pondération fixes, variables.

a) Il résulte du raisonnement précédent que l'indice à base fixe suppose que l'on compare une dépense actuelle à une dépense prise pour base dans l'hypothèse que les habitudes de consommation n'ont pas changé, c'est-à-dire à quantités constantes. C'est l'indice de Laspeyres, moyenne arithmétique pondérée, à base et coefficients de pondération fixes.

Périodiquement, il sera nécessaire de réviser les coefficients de pondération et de changer de base ; d'une part

les habitudes de consommation évoluent avec le temps, d'autre part il peut apparaître de nouveaux articles sur le marché. C'est ainsi qu'à Brazzaville, un indice des prix à la consommation familiale a été calculé sur la base 100 en 1945 jusqu'en 1952 et que depuis cette date un nouvel indice base 100 en 1949 a été établi pour tenir compte des modifications des habitudes de consommation.

Dans ce cas, on aura affaire à un indice chaîne. Cet indice chaîne sera calculé à partir d'indices synthétiques partiels.

$I^{1/0}$ système 0 de poids (habitudes de consommation 0)

$I^{2/I}$ système I de poids (habitudes de consommation I)

$I^{3/2}$ système 2 de poids (habitudes de consommation 2)

et on effectue :

$$J^{2/0} = I^{I/0} \times I^{2/I}$$

$$J^{3/0} \quad I^{I/0} \times I^{2/I} \times I^{3/2} \quad \text{etc...}$$

Cette méthode permet de tenir compte de la modification des structures économiques, de la disparition de certains articles, de l'apparition de nouveaux, mais il est évident qu'elle exige des enquêtes périodiques et qu'une erreur de calcul se répercute tout le long de la série, ce qui ne se produit pas dans le cas d'un indice à base fixe.

b) Il existe des indices dont les coefficients de pondération réflètent, non pas la structure de l'ensemble statistique considéré à l'époque de base, mais celle existante à l'époque actuelle ; la période de base demeure inchangée, mais les coefficients de pondération varient selon les modifications des conditions de la situation actuelle, cet indice à base fixe mais de pondération variable, qu'il ne faut pas confondre avec l'indice chaîne, est appelé indice de Paasche.

L'exemple à citer est celui de l'indice trimestriel à la construction des immeubles à usage d'habitation en France : une enquête périodique permet de connaître les caractéristiques de la construction par examen d'un

échantillon de devis : on décompose les devis en un certain nombre d'ouvrages et on calcule outre le coût actuel de la construction selon sa description actuelle, le coût de la même construction à l'époque de référence. L'indice est obtenu à partir de ces deux estimations de coût.

8.3.3- Défauts des indices de Laspeyres et de Paasche.-

Les deux indices de Laspeyres et de Paasche présentent un défaut commun : si on choisit une année intermédiaire, année 0', la relation suivante ne se vérifie pas en toute rigueur.

$$I^{I/0'} = \frac{I^{I/0}}{I^{0'/0}}$$

Lorsque les indices partiels calculés sont très différents de 100 et très dispersés autour de l'indice général, cet inconvénient peut être grave.

		Lait (poids 2)	Vin (poids 1)
0	1945	25	50
0'	1949	40	60
I	1955	80	200

on a : $$I^{55}_{I/0^{45}} = 100 \times \frac{\left(2 \times \frac{80}{25} + 1 \times \frac{200}{50}\right)}{3} = 346,7$$

$$I^{49}_{0'/0^{45}} = 100 \times \frac{\left(2 \times \frac{40}{25} + 1 \times \frac{60}{50}\right)}{3} = 146,7$$

et $$I^{1/0'} = 100 \times \frac{\left(2 \times \frac{80}{40} + 1 \times \frac{200}{60}\right)}{3} = 244,4$$

qui est fort différent de :

$$\frac{346,7}{146,7} \times 100 = 209,1$$

Cette difficulté doit être toujours présente à l'esprit lorsque, pour certaines études on a besoin de procéder à des changements de base, ou à des raccordements.

8.3.4.- Modalités de calcul d'un indice de prix.-

On peut assimiler un indice de prix,

- soit selon une conception dite "budgétaire" (un "panier de provisions" correspondant à un type de consommation étant donnée) au rapport du coût d'un panier de composition fixe aux deux époques considérées :

$$I = \frac{\text{Somme de } q_o\ p}{\text{Somme de } q_o\ p_o} \times 100$$

- soit à une moyenne pondérée des indices partiels calculés pour chaque article :

$$I = \text{Somme de } \alpha_i\ I_i$$

avec $\alpha_i = \dfrac{(q_o\ p_o)i}{\text{Somme de } q_o\ p_o}$ et $I_i = \left(\dfrac{p}{p_o}\right) i$

On a vu sur un exemple numérique que ces deux relations sont équivalentes.

Pour construire un indice de prix, on entreprendra une enquête de budgets familiaux pour le ou les groupes sociaux intéressés, on chiffrera la dépense familiale moyenne totale et les dépenses moyennes par groupe de marchandises et articles, on déterminera par groupe de consommation les dépenses relatives correspondantes (en % ou en ‰ du total).

pain	α_1
biscuits	α_2
farine	α_3
produits laitiers	α_4
etc...	---
	1000

puis pour chaque groupe élémentaire, on retiendra un article dont les variations de prix sont représentatives

des variations des prix des articles qui rentrent dans ce groupe de dépense, on calculera ensuite l'indice partiel des prix de cet article auquel on appliquera le coefficient de pondération correspondant au groupe de marchandises que cet article est censé représenter.

8.3.2- Précautions dans les manipulations d'indice.-

En résumé, si on doit entreprendre des calculs d'indices, d'une part, on se rappelera toujours la relativité des indices, car un indice suppose obligatoirement une comparaison et ne préjuge en rien du niveau absolu des prix ou des quantités ou de la grandeur quelconque entrant en jeu, à l'époque retenue pour référence.

a) Si on examine l'évolution simultanée de deux indices durant une même période, on ne compare que des variations relatives : si un premier indice augmente plus vite que l'autre, cela ne signifie pas que la première grandeur mise en jeu devient en valeur absolue supérieure à la seconde : l'indice du coût de la vie à Brazzaville en 1952 est au niveau 500 base 100 en 1945 et celui de Dakar sur la même base est au niveau 300, on ne peut assurer que la vie est plus chère à Brazzaville, on peut seulement affirmer qu'en sept ans, le coût de la vie à Brazzaville a augmenté plus rapidement que le coût de la vie à Dakar, mais on ignorera toujours si l'on n'a pas d'autres renseignements si le coût de la vie lui-même est plus élevé dans l'un ou l'autre centre.

b) Ce problème de la comparaison géographique des niveaux de prix donne lieu à des développements théoriques très importants. On en donnera seulement ici une solution concrète :

On dresse une série commune d'articles entrant pour une part importante dans la consommation des ménages des deux localités. Dans chaque localité, on dispose de la série des coefficients de pondération α_i et α'_i. On calcule les indices partiels de prix, le prix dans une localité servant de base ; les indices partiels I_i sont pondérés par α_i d'où après sommation des indices partiels pondérés il résulte un premier indice I_1, on pondère ensuite par α'_i d'où I_2 après sommation. L'indice de disparité des niveaux de prix est égal à $I = \sqrt{I_1 I_2}$

Ces calculs exigent un examen préalable très détaillé qui portera sur :

- le champ de l'indice et les classes sociales à étudier ;

- l'étude de la formation des prix et des habitudes commerciales dans l'une et l'autre localité ;

- choix des rubriques de dépenses retenues, choix des articles représentatifs et comparabilité des cotations ;

- qualité égale des relevés de prix.

8.4 - Les Indices de volume.-

Il existe également des indices de quantité et de volume qui interviennent dans les cas suivants :

On suppose qu'un district commercialise une année donnée (1949) 500 tonnes de palmistes, 2.000 tonnes de cacao, 100 tonnes de café, 200 tonnes de manioc ... selon les différents prix pratiqués à cette époque la valeur totale de cette production est de 44I millions de francs C.F.A. Quatre ans plus tard, en 1953, le même district exporte des tonnages différents, de plus les prix de chacun des produits ont varié, la valeur totale commercialisée s'élève à 408 millions.

Commercialisation du district de .....

	1949 réel			1953 réel			1953 conventionnel.
	quantités tonnes	valeur 1.000 frs	p.u. frs le kg.	quant. tonnes	val. 1.000 frs	p.u. frs le kg.	valeur 1.000 frs
manioc	200	1.000	5	300	1.200	4	1.500
palmistes	500	10.000	20	800	12.000	15	I6.000
cacao	2000	400.000	200	2I00	3I5.000	I50	420.000
café	100	30.000	300	200	80.000	400	60.000
	2800	44I.000		3400	408.000		497.500

$$I = \frac{497.500}{44I.000} \times 100 = \underline{113}$$

On ne peut, si l'on recherche un indice représentatif de l'évolution de la production physique, calculer brutalement le rapport :

$$\frac{408}{441} \times 100 = 93, \quad \text{rapport des valeurs totales}$$

car les prix des différents produits ayant évolué différemment, l'augmentation des prix globaux peut être purement fictive et cacher une diminution de la production.

La comparaison des tonnages est également trompeuse :

$$\frac{3.400}{2.800} \times 100 = 121$$

Le prix moyen du kilogramme est également un renseignement insuffisant = 157 en 1949 ; 120 en 1953.

Il convient chaque année de multiplier les tonnages commercialisés par les <u>prix unitaires de l'époque de base</u> <u>prix unitaires</u> qui, <u>par hypothèse sont conservés fixes</u>. On additionne les valeurs obtenues pour chaque produit et on rapporte la valeur totale obtenue à la valeur observée à l'époque de base. D'une manière plus précise, l'indice est défini au facteur 100 près par la formule :

$$I = \frac{\text{Somme de } p_o\, q}{\text{Somme de } p_o\, q_o}$$

par une transformation simple, on trouve que cette formule est équivalente à :

$$I = \text{Somme de } {}_i \cdot I_i \quad \text{avec} \quad I_i = \frac{iq}{iq_o}$$

comme pour les indices de prix, étant égal à l'importance relative en valeur du produit i par rapport à la valeur de la production totale à l'époque de base.

<u>La symétrie des formules</u> apparaît clairement. Dans le cas d'un indice de volume, <u>ce ne sont pas les quantités qui restent fixes</u>, <u>mais les prix unitaires observés durant l'année de base</u>.

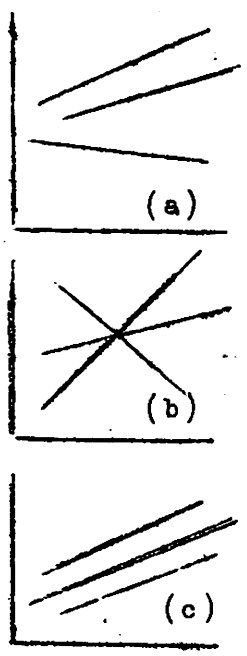
Il ne faut donc pas confondre indice de volume et variation générale de tonnages. Cette précaution doit être rappelée car la confusion existe entre ces deux notions.

Dans un article datant déjà de 1953 (1), l'auteur indique que l'A.O.F. n'a exporté en 1952 que :

944.964 tonnes contre 987.770 en 1938
et que le commerce de l'A.E.F. reste stationnaire car on observe
en 1952 : 355.000 tonnes, contre 333.000 tonnes en 1938.

Il est bien certain qu'en A.O.F. une tonne d'arachides décortiquées est plus avantageuse qu'une tonne d'arachides-coques, et qu'en A.E.F. il vaut mieux exporter une tonne de coton-fibres qu'une tonne de bois en grumes.

8.5 - L'utilisation des indices - Indices bruts et corrigés.



Les indices sont certainement parmi tous les outils disponibles en statistique, ceux qui sont employés le plus fréquemment, on ne peut donc s'étonner du nombre de pièges qui guettent l'utilisateur.

8.5.I.- Il faut d'abord noter que des valeurs égales d'un indice général synthétique à deux instants différents ne caractérisent pas forcément deux situations identiques ; en effet la dispersion autour de la valeur centrale peut être très différente et il peut exister des distorsions génératrices de troubles : si l'indice des prix de gros révèle un écart entre les prix industriels et agricoles pour un indice général donné, il est évident que cette situation sera moins favorable que la période où l'on constatait le même niveau général des prix mais au cours de laquelle ce niveau était inexistant (cas b et c).

(1) Marchés Coloniaux - 31 Octobre 1953, pages 3053 et suivantes.

Une deuxième difficulté se révèle lorsqu'on souhaite comparer des indices de même genre mais pour des pays et localités différents.

Ainsi, la comparaison des variations des indices de prix de détail à Paris et à Abidjan soulève les objections suivantes :

- Les habitudes de consommation sont très différentes et les articles concourant à l'approvisionnement dans l'une et l'autre capitale pourront être très variés.

- Les coefficients de pondération sont donc très dissemblables.

- Pour un même article, la "satisfaction" n'est pas identique.

- Les classes sociales considérées ne sont pas les mêmes.

De plus implicitement la comparaison des valeurs atteintes par les indices suggère l'égalité des niveaux absolus des prix à l'époque commune de base et cette comparaison escamote sans critique des problèmes sérieux.

8.5.2- La combinaison de plusieurs indices conduit parfois à des erreurs :

En raison de la hausse des prix, on est amené à recalculer la valeur de commerce des pays d'outre-mer de la Communauté (tant à l'importation qu'à l'exportation) non pas en francs courants, mais en "francs constants" valeur 1938 par exemple. On cherche dans ce but une série annuelle représentative de la hausse des prix en France (indice des prix de gros), et on a l'habitude de diviser pour une année donnée, le montant des échanges exprimé en francs courants par l'indice des prix de gros en France pour l'année considérée. On élimine ainsi l'effet de la hausse des prix dans l'appréciation de l'évolution du commerce. A l'importation dans les pays étudiés, le principe du calcul est grossièrement valable, car on peut admettre que les marchandises intervenant dans le commerce à l'importation sont celles retenues dans l'indice des prix de gros en France.

Rubriques	1938	1946	1947	1948	1949	1950	1951	1952
Commerce à l'importation	3220	20390	39722	80489	146698	177860	258491	272580
I) dans les pays de la Communauté (Millons FM.) (n.c. Indochine)								
2)Indice prix de gros en France.	100	606	819	1711	2114	2289	2924	3063
(1)divisé par : (2)=(3)	3220	3365	4850	4704	6939	7770	8840	8899
Variation du (3) pour 100 en 1938	100	105	151	146,1	216,5	241,3	274,5	276,4

Mais à l'exportation, un raisonnement identique ne peut s'appliquer : on constate en effet que la liste des matières premières d'Outre-Mer est bien originale et que l'évolution relative des prix de ces produits est particulière : le niveau moyen des prix des produits d'Outre-Mer était en 1952 à peu près au niveau 45, tandis que le niveau des prix des marchandises en France atteignait seulement le niveau 30 ; par suite si on adopte la même série que pour les importations, on obtient une valeur corrigée de beaucoup supérieure à la réalité. En fait il faut calculer un indice spécial des prix à l'exportation des pays d'Outre-Mer de la Communauté.

Rubriques	1938	1946	1947	1948	1949	1950	1951	1952
Commerce (1) à l'exportation des pays de la Communauté	3II6	I8745	29893	70055	I063I3	I30I25	I649I7	I6988I
(2) Indice des prix de gros en France	100	606	8I9	1711	2114	2289	2924	3063
(3) Indice des prix de gros des produits d'Outre-Mer	100	692	998	2090	2790	3733	4523	4210
(I) divisé par (2) = (4)	3II6	3093	3650	4094	5029	5665	5640	5546
(I) divisé par (3) = (4')	3116	2709	2995	3352	3811	3486	3646	4035
Variation de (4), pour cent en 1938	100	99,3	117,2	131,4	161,4	182,5	181,0	178,0
Variation de (4'), pour cent en 1938	100	87,0	96,1	107,6	122,3	111,9	117,0	129,5

On voit qu'à partir du moment où l'on s'écarte des chemins battus, l'utilisation des indices pose des problèmes et des difficultés particulières qui n'apparaissent pas toujours au premier examen.

8.5.3- Pour en terminer avec ces notions sur les indices, il faut signaler le cas de la correction des indices chronologiques lorsque la grandeur observée concerne des périodes de temps inégales. La production industrielle fournit une illustration de ces problèmes.

On pratique deux corrections distinctes selon qu'il s'agit d'entreprises qui fonctionnent d'une manière discontinue (avec fermeture complète les dimanches et jours fériés) ou d'entreprises marchant en continu (hauts-fourneaux).

Dans la première situation, on multiplie pour chaque article considéré la production du mois courant par :

$$\frac{\text{nombre mensuel moyen de jours ouvrables durant l'époque de base}}{\text{nombre de jours ouvrables du mois courant.}}$$

Dans la seconde situation de la marche continue, le coefficient est :

$$\frac{\text{nombre mensuel moyen de jours durant l'époque de base}}{\text{nombre de jours du mois courant}}$$

Une fois cette correction effectuée, on compare pour le calcul de l'indice partiel, et pour l'article en cause, la production corrigée du mois considéré à la production mensuelle moyenne de l'année de base.

BIBLIOGRAPHIE

- Supplément méthodologique au bulletin mensuel de statistique de l'I.N.S.E.E. - avril 1958.

  (voir en particulier - partie mensuelle n° 44 à 54 et partie trimestrielle, chapitres II.A, VI et VII).

- Livre blanc de l'I.N.S.E.E. - 1953 -
  Les statistiques de prix en France et en divers pays étrangers.

- Bulletin mensuel de statistique d'outre-mer.
  Supplément série études n° 21 - Les indices de prix dans l'U.F.O.M.

- Documents et statistiques . Ministère de la France d'Outre-Mer, n° VIII.

- C.E.C.A. - Informations statistiques.
  Numéro spécial - Juillet-août 1957.

Chapitre IX

HASARD ET PROBABILITÉ

9.I Le jeu de pile ou face.
Notion de probabilité d'apparition d'un phénomène.

9.2 Propriétés du graphique de Laplace.

9.3 La loi des grands nombres.
Le théorème de Bernouilli.

9.4 Loi de probabilité et variable aléatoire.

Chapitre IX

HASARD ET PROBABILITE

Etudier toutes les unités statistiques d'un ensemble est une solution à laquelle il est difficile de se rallier, lorsque les moyens en personnel et en matériel sont réduits. En réalité au cours des enquêtes réelles sur le terrain l'univers n'est pas mesuré dans son intégralité et seul un échantillon fournit une estimation des caractéristiques de l'ensemble inconnu. Cette technique opératoire est fréquente mais peut-on être sûr que les renseignements issus d'un échantillon sont utilisables et ne risque-t-on pas d'aboutir à des conclusions erronées en étudiant ou en mesurant les unités d'un seul échantillon ?

C'est le dessein de ce chapitre que de justifier intuitivement la procédure de l'échantillonnage.

9.I Le jeu de pile ou face. Notion de probabilité d'apparition d'un phénomène.

Un opérateur jette une fois une pièce de monnaie sur une table. Il existe deux possibilités d'apparition : pile ou face. Pile apparaît soit 0 fois, soit une seule fois, autrement la proportion de pile pour un seul jet est de 0 ou 1.

On peut se poser deux problèmes :

a) Que se passe-t-il si l'on répète un grand nombre de fois cet jet élémentaire de pièce ? Comment va varier la proportion de pile au fur et à mesure que la série de jets va devenir plus longue ?

b) Si l'on imagine diverses séries de 100 jets pour chacune de ces séries, la composition en pile et face varie, peut-on prévoir la répartition de ces divers échantillons ?

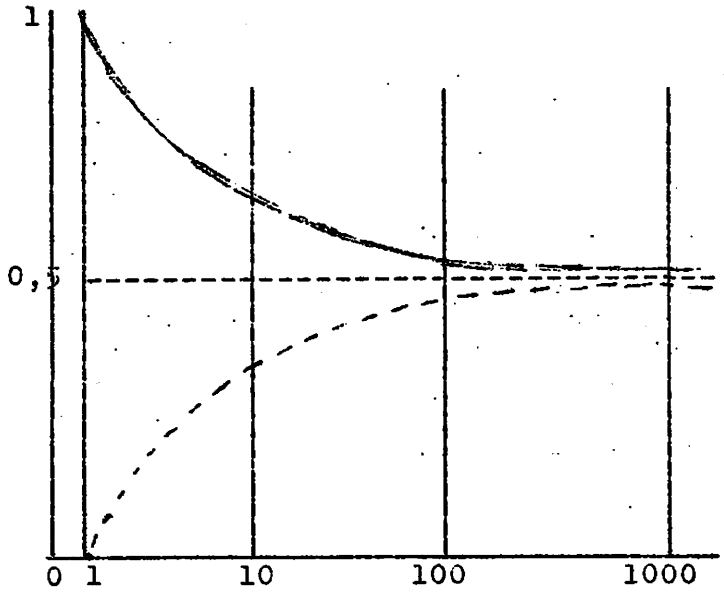
9.I.I- Soit donc une série de 100 jets : on compte 47 faces et 53 piles qui constituent l'échantillon de 100 observations, la proportion de piles est de 0,53.

Côté d'apparition	Fréquences absolues	Fréquences relatives
Pile	53	0,53
Face	47	0,47
	100	

On recommence une autre série de 100 jets encore, on obtient cette fois 52 faces, 48 piles, la proportion de pile est pour cette nouvelle expérience de 0,48.

En répétant cette opération plusieurs fois, toujours pour des séries de 100 jets, on trouve des proportions en face et en pile voisines de $\frac{1}{2}$, mais à chaque expérience, cette proportion varie, et sa valeur oscille autour de cette valeur centrale. On remarque toutefois que des proportions assez éloignées de la valeur de $\frac{1}{2}$ se rencontrent peu. La série d'expériences se continue, mais par série de mille jets, on trouve par exemple quatre cent quatre-vingt onze (49I) faces et cinq cent neuf (509) piles, la proportion de piles est de 0,509.

La série s'allonge encore soit dix mille jets, on trouve 5.025 faces et 4.975 piles, la proportion de piles est de 0,4975. En augmentant les tirages à chaque expérience, on s'aperçoit que la proportion de piles et par voie de conséquence la proportion de faces se rapproche de plus en plus de 0,5. Intuitivement, en admettant la série infinie, la proportion de piles qui serait observée serait égale à 0,5 - Et on remarque qu'au fur et à mesure que la série augmente de longueur, la fluctuation de la proportion calculée à mesure que la série s'allonge



devient de plus en plus faible, autrement dit les proportions s'écartant largement de 0,5 en pratique disparaissent, pour un nombre suffisant de jets.

Une expérience concrète du même genre peut être réalisée en remplissant une urne d'un grand nombre de billes de deux couleurs, rouge et blanche chaque couleur étant représentée également. On dispose d'une palette pour effectuer des prélèvements par 5, 10, 15, 20 et 25. L'opérateur tire des séries de 5 billes, de 10, etc... finalement puise par 25 billes, chaque opération est répétée dix fois pour chaque type de série. L'exemple montre que la proportion de billes blanches varie de 0 à 100 % dans des tirages de 5 billes, tandis que cette proportion ne varie plus par exemple que de 30 à 70 % dans les tirages de 25 billes.

On peut inversement passer à une autre série d'observations.

Soit une urne composée de boules noires et de boules blanches très nombreuses, mais en proportions inconnues, pour découvrir ces proportions on tire une série de 100 boules, puis une série de 1.000, 10.000 boules. Pour ce faire une boule est tirée, sa couleur notée et elle est remise dans l'urne, une seconde fois on tire une boule, etc... et expérimentalement on observe une proportion de noires qui tend, avec quelques fluctuations, vers la valeur 0,5. De cette expérience on sera amené à penser que le nombre des boules noires est égal au nombre des boules blanches.

Dans le premier cas, on se contente d'établir une statistique à partir d'observations répétées : la pièce de monnaie n'étant pas truquée, on admet qu'elle n'a pas plus de raisons de tomber sur un côté plutôt que sur un autre (a), et en la lançant en l'air n'importe comment, on s'aperçoit que pour une série de jets assez longue, la proportion des piles tend vers 0,5. Dans le deuxième cas, on ignore la composition de l'urne, mais on s'aperçoit que la proportion des noires tend à la longue vers 0,5, on en conclu qu'il est vraisemblable que les deux couleurs figurent en parties égales dans cette urne (b). De plus, l'expérience montre que la proportion observée oscille autour de cette valeur, mais de plus en plus faiblement avec l'importance des tirages.

Autrement dit, en entreprenant l'inventaire d'un seul échantillon, on commet une certaine erreur par rapport à la valeur réelle (en plus ou en moins) mais cette erreur dans un nombre très grand de cas devient de plus en plus faible à mesure que l'échantillon devient important (1).

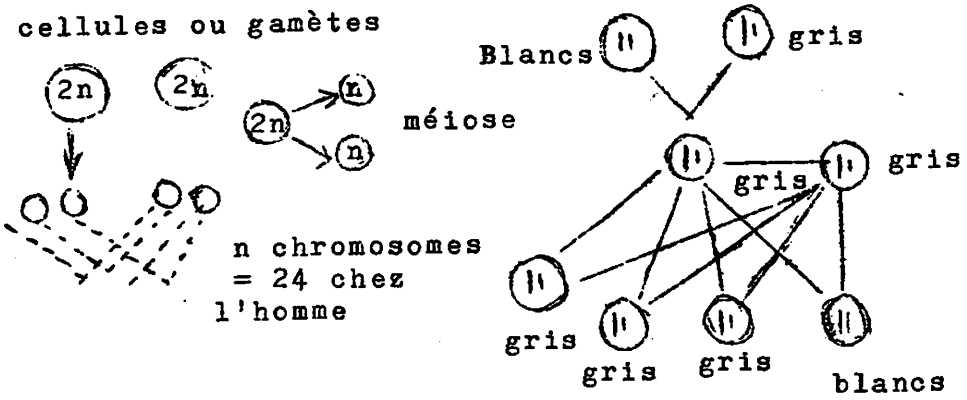
9.I.2 - On peut donc assurer que l'analyse d'un échantillon permet de déduire avec une certaine approximation les caractéristiques de l'ensemble de la population.

Un exemple, emprunté à la génétique peut être cité et illustrer la richesse de cette propriété : un élevage renferme des souris blanches et des souris grises, ces bêtes sont croisées, on note que la première génération donne des produits gris en totalité, un premier lot de cette génération croisée entre elle, donne 74,5 % de souris grises, 25,5 % de souris blanches, un second lot donne 76 % de souris grises, 24 % de souris blanches, etc... finalement, on observe toujours un pourcentage voisin de 3/4 de souris grises et 1/4 de souris blanches. Cette observation statistique qui constitue une simple expérimentation conduit alors à certaines hypothèses et à la construction d'un schéma explicatif. Série d'observations féconde car elle est la base des théories de Mendel : en admettant l'existence de gênes dominants et de gênes récessifs. Mendel inféra :

- la pureté des caractères génétiques,

- la disjonction des gênes au moment de la reproduction selon le schéma ci-dessous :

(1) Cet écart par rapport à la valeur réelle qui demeurera inconnue s'appelle un écart aléatoire.
(a) c'est le principe d'indifférence.
(b) la conclusion est basée ici sur l'observation d'une fréquence expérimentale.



Différence entre mitose et méiose :

disjonction générale des cellules non reproductives → 2n 2n mitose

disjonction des cellules reproductives → n n méiose

9.I.3- Le graphique binomial.-

On s'est posé un autre problème qui est l'analyse des séries comprenant le même nombre de jets.

En revenant en effet à la partie de pile ou face, on peut chercher à savoir, en se fixant à l'avance le nombre de jets, quatre par exemple, quelle va être la composition de chaque série :

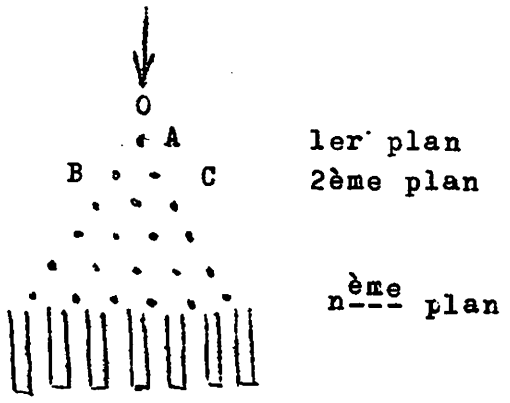
- toujours pile
- toujours face
- ou mélangés.

Auparavant, on va reproduire l'appareil suivant, imaginé par Francis Galton.

9.I.3.I- Des clous sont disposés en quinconces, sur une planchette, verticale, au-dessus du clou supérieur un entonnoir contient des billes d'un diamètre très légèrement inférieur à l'écartement des clous. En bas de la planchette sont disposées des éprouvettes pour recevoir les billes.

Une première bille tombe sur le premier clou. L'alternative est gauche ou droite, et on peut penser, puisque l'entonnoir est bien à l'aplomb du premier clou, qu'en moyenne si on lâche plusieurs billes, une fois sur deux, la bille tombera à droite, autrement dit le flux sera divisé en deux parties sensiblement égales, le flux de droite rencontre ensuite un clou de la seconde rangée soit C nouvelle alternative une fois sur deux gauche ou droite ; par suite, si on laisse se vider l'entonnoir, la moitié seulement des billes qui sont tombées à droite du clou de la première rangée A, donc un quart, fileront à droite du clou C et un quart à gauche du même clou. Or

$$\frac{1}{4} \quad \text{s'écrit} \quad \frac{1}{2} \times \frac{1}{2} \quad \text{ou} \quad \left(\frac{1}{2}\right)^2$$

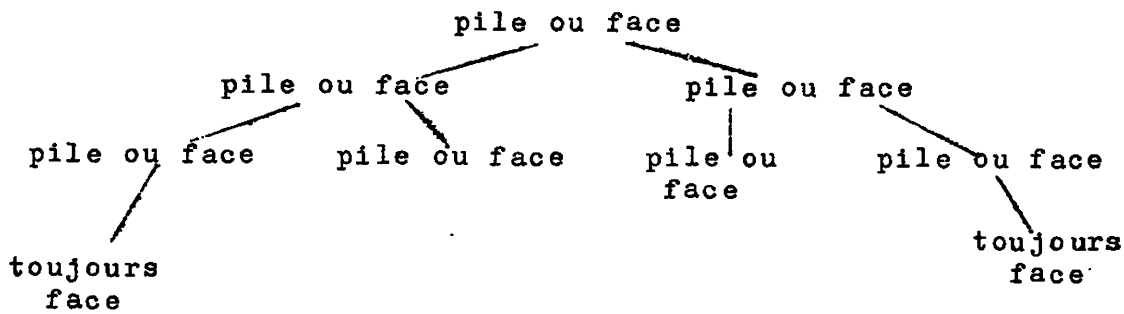


En répétant le raisonnement, on s'aperçoit que si l'on se fixe d'avance un trajet et que si ce trajet suppose la traversée de n plans horizontaux, il y aura en moyenne une proportion de $\left(\frac{1}{2}\right)^n$ qui auront frappé ces trous. Ce trajet aboutit à une éprouvette et on peut imaginer que pour tomber dans cette éprouvette les billes puissent emprunter plusieurs "circuits". Autrement dit si l'on ne se préoccupe plus d'un trajet donné (en se fixant à l'avance les clous) mais seulement de l'éprouvette d'arrivée, on conçoit qu'il y aura d'autant plus de billes dans une éprouvette, qu'il y aura plus de trajets différents pour y parvenir. En particulier pour les éprouvettes situées aux extrêmes il n'y a qu'une manière d'y parvenir : la bille doit tomber systématiquement toujours à droite ou toujours à gauche du clou.

Ces divers trajets peuvent être décomptés et leur nombre diminue à mesure que l'on s'éloigne du centre vers les extrémités.

9.I.3.2- Si on appelle probabilité, le rapport du nombre des éventualités favorables, c'est-à-dire le nombre des billes tombées à l'endroit voulu, au nombre des éventualités possibles, nombre total des billes lâchées, on se rend compte visuellement que la probabilité qu'une bille atteigne l'éprouvette située à la verticale de l'entonnoir est beaucoup plus élevée que celle de tomber dans les éprouvettes des extrémités. Est-il possible de préciser numériquement ces diverses probabilités ?

Si on revient à la partie de pile ou face, et que l'on entreprenne une série de quatre jets de pièce, on peut avoir les possibilités suivantes selon qu'à chaque jet, la pièce fait apparaître pile ou face (les éventualités d'apparition de l'un ou l'autre évènement étant considérées comme également vraisemblables)(1).



La similitude apparaît avec le jeu précédent : la probabilité est faible d'avoir continuellement soit pile ou face. Comme il y a seize cas différents et une seule possibilité favorable, la probabilité d'avoir une série totalement pile ou totalement face est égale à 1/16. Pour les autres combinaisons on peut les compter directement :

Combinaisons	Nombre de manières d'y parvenir		Probabilité
4 piles	PPPP	1 possibilité	$\frac{1}{16}$

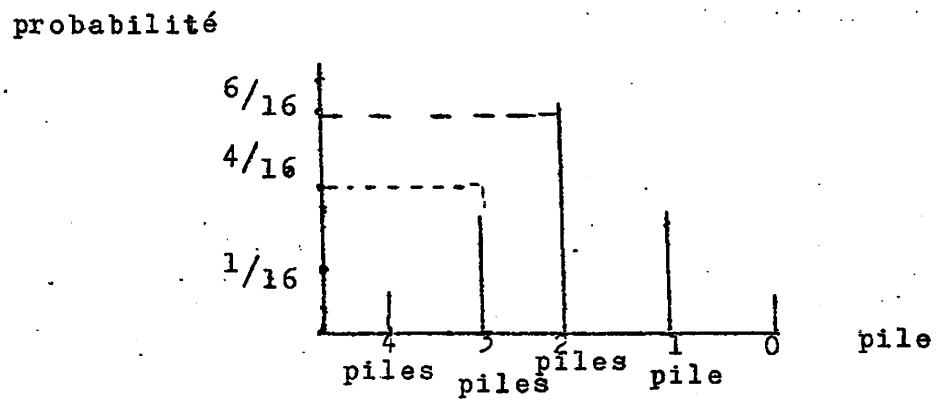
.../..

(1) On se trouve dans le premier cas précisé à la note (a) de la page 171, sur l'équiprobabilité définie à priori.

Combinaisons	Nombre de manières d'y parvenir		Probabilité
3 piles 1 face	PPPF PPFP PFPP FPPP	4 possibilités	$\frac{4}{16}$
2 piles 2 faces	PPFF FFPP FPPF PFPF FPFP PFFP	6 possibilités	$\frac{6}{16}$
1 pile 3 faces	FFFP FFPF FPFF PFFF	4 possibilités	$\frac{4}{16}$
4 faces ...	FFFF	1 possibilité	$\frac{1}{16}$

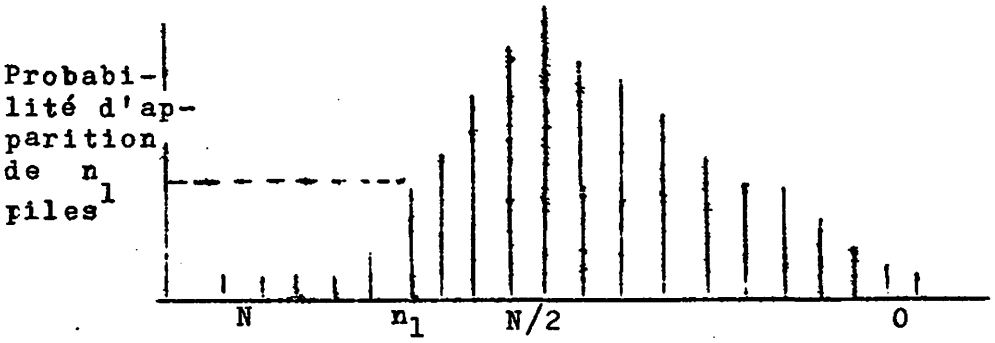
On remarque que la somme de ces probabilités reconstitue l'unité.

Si on représente graphiquement ces probabilités, la figure est symétrique :



Au lieu d'observer une série de 4 jets, on pourrait augmenter la longueur de la série : la probabilité dans une série nombreuse de jets d'obtenir continuellement pile ou face serait toujours infiniment plus faible que d'avoir une série composée d'un nombre à peu près égal de pile ou face. En représentant graphiquement les diverses probabilités, on obtiendrait encore un graphique symétrique rappelant la figure représentée par les billes entassées dans les éprouvettes de l'expérience de Galton. Ce graphique toujours dans l'hypothèse de l'<u>équi-probabilité élémentaire</u> d'apparition de pile ou face permet d'obtenir la probabilité pour que dans

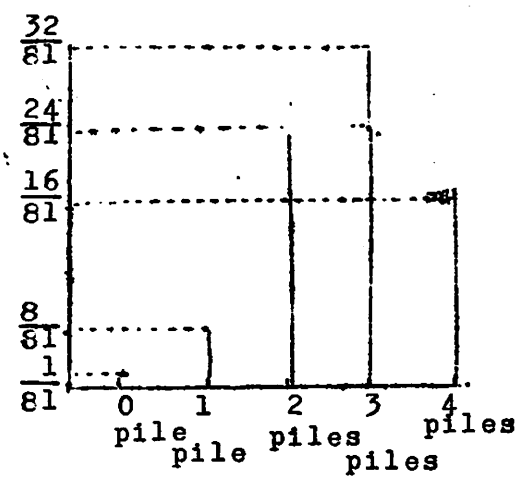
une série de N jets on observe n_1 piles (graphique binomial).



Ce résultat ne préjuge pas du résultat d'une partie unique, mais signifie seulement qu'en répétant de nombreuses fois la même expérience comprenant N jets, on enregistrera l'apparition de n_1 piles dans une proportion sensiblement égale à la probabilité calculée à priori. Pour prendre un exemple si, chaque matin, sur un même trajet, une voiture doit franchir quatre <u>feux automatiques indépendants</u> (de durées rouge et verte égales), en moyenne 6 fois sur 16, on se trouvera arrêté deux fois sur quatre, et en moyenne une fois sur 16, on sera arrêté quatre fois, ou .... on passera sans difficultés.

La forme du graphique varie si la probabilité élémentaire d'apparition du phénomène n'est plus égale à $\frac{1}{2}$; il n'est plus symétrique, ce cas se présente dans le cas où sur l'une des faces de la pièce on colle un petit poids supplémentaire de manière à ce que pile sorte deux fois plus fréquemment que face. Dans ce cas en ne considérant que des séries d'un seul jet, face apparaît en moyenne une fois sur trois. Pour une série de quatre jets de pièces on a les probabilités suivantes d'apparition :

4 faces	$\frac{1}{81}$
3 faces 1 pile	$\frac{8}{81}$
2 faces 2 piles	$\frac{24}{81}$
1 face 3 piles ...	$\frac{32}{81}$
4 piles	$\frac{16}{81}$



Cependant, si la série de jets au cours d'une expérience augmente indéfiniment, la figure tend à devenir symétrique et à la limite, le diagramme représentatif tend vers le diagramme dit normal de Laplace Gauss, d'une forme semblable à celle obtenue dans le diagramme de la loi binomiale lorsque $p = 1/2$.

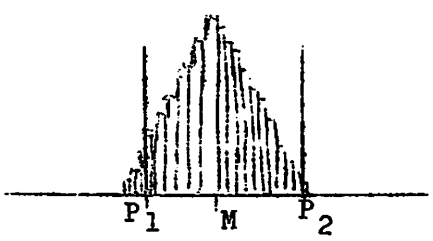
Le diagramme cumulé à partir du diagramme de Laplace est appelé diagramme de Galton. Ce diagramme trouve de nombreuses applications en psychotechnie.

9.2 - Propriétés du graphique de Laplace.-

Le graphique de Laplace représente donc la probabilité élémentaire d'apparition d'un phénomène étant connue, soit p, la probabilité pour que dans une série de N évènements consécutifs, on en enregistre n évènements favorables. Le mode du graphique est toujours atteint pour la valeur

$$n = N p.$$

Cette distribution numérique : nombre d'évènements favorables et probabilité correspondante d'apparition, est caractérisée par une certaine dispersion et on en peut en calculer l'écart-type. Le graphique de Laplace fournit une information importante qui se démontre algébriquement.



A partir de la valeur moyenne M, on porte de part et d'autre deux longueurs MP_1 et MP_2 telles que $MP_1 = MP_2$ = deux écarts-types.

Si on élève en P_1 et P_2 deux perpendiculaires l'histogramme se trouve séparé en trois zones. On démontre que la surface comprise entre ces perpendiculaires représente 95 % de la superficie totale de l'histogramme. Autrement dit : au cours d'une seule épreuve un évènement favorable a une probabilité p de se produire, si on répète N fois cette épreuve, l'évènement favorable dans cette série de N répétitions peut apparaître 0, 1... N-1, N fois. Si l'on considère un nombre important de séries de N éprauves, le nombre d'évènements favorables sera compris dans 95 cas sur 100 entre

Np $\pm$ deux écarts-types. L'écart-type étant

$$\sqrt{Np(1-p)}\ .$$

La probabilité de 95 % suppose que p ne soit pas trop petit si N n'est pas grand, ni trop voisin de l'unité (voir paragraphe 9.3.2).

Exemple : Un dé à six faces est jeté à cent reprises, on considère l'apparition de l'as comme un évènement favorable. Si l'on répète cette série de cent jets, dans 95 % des cas, l'as apparaîtra de 9 à 24 fois.

Il résulte de cette remarque que si au cours d'une expérience, on trouve une valeur extérieure à cet intervalle P_1 P_2 on peut légitimement tenir pour invraisemblable l'hypothèse envisagée. Si, dans une partie de cent piles et faces (95 % des observations se trouvent dans le cas de séries de cent jets dans l'intervalle

$$50 \pm 2\sqrt{\frac{100}{4}} = 50 \pm 10\ ),$$

on trouve 30 piles et 70 faces, c'est-à-dire P = 0,3, on peut conclure que la pièce est truquée, parce qu'il y a moins de 1 chance sur cent que cette éventualité puisse arriver.

9.2.I- Le même raisonnement fait conclure que, comme sur 10.000 naissances, on observe 5.250 garçons, il est hautement improbable qu'à la naissance, il y ait égalité des naissances masculines et féminines. Au contraire, si l'on suppose d'après la seule expérience, qu'il naît 105 garçons pour 100 filles, on trouve (sans qu'on ait encore pu le justifier) que les données numériques tirées de l'observation concernant d'autres pays et d'autres peuples confirment cette moyenne expérimentale.

Pour prendre un autre exemple soit une statistique de naissances de jumeaux :

- Naissances jumelées de deux garçons	2.933
- " " de deux filles	2.797
- " " un garçon, une fille ..	3.I88
	8.9I8 naissances.

Sur 8.9I8 naissances on pourrait s'attendre aux chiffres suivants : 2.229, 2.229 et 4.460. On explique la différence avec les résultats expérimentaux, par le fait

qu'il existe des vrais jumeaux, ces jumeaux sont obligatoirement de même sexe, tandis que les faux jumeaux peuvent être de sexes différents.

La question se pose de savoir, à partir du moment où le premier jumeau est né, si cette première information permet de prévoir le sexe du second enfant, ou si, au contraire on a toujours 105 chances sur 205 de voir apparaître un garçon.

Le calcul montre que 28,5 fois sur 100, le deuxième enfant est du même sexe que le premier, et que 71,5 fois sur 100, le sexe du deuxième enfant est indépendant du sexe du premier. Par suite au lieu d'une chance sur deux, on est un peu plus avance : 64/IOO.

9.3 - La loi des grands nombres - Le théorème de Bernouilli.

9.3.I- Lorsqu'une étude théorique préalable ou une analyse antérieure exhaustive montre que, dans une population nombreuse d'unités statistiques, un caractère donné prend une valeur moyenne m, on peut donc s'attendre, en étudiant divers échantillons, à voir les valeurs m' observées se disperser autour de la valeur m théorique, cependant ces valeurs m' ne se dispersent pas n'importe comment : plus les échantillons seront importants, plus m' sera voisin de m ; sous une forme littéraire, c'est la loi des grands nombres qui vient d'être énoncée.

D'autre part et inversement, si un échantillon fournit une valeur de m' très différente de m, on peut en conclure avec une marge d'erreur très faible, que cet échantillon n'a pas été tiré de la population totale, ou encore que les mesures et relevés sont incorrects.

Sous une autre forme on peut dire que si un phénomène A a une probabilité théorique p de se produire, si on observe r répétitions du phénomène A au cours de n épreuves, la probabilité pour que $\left(\frac{r}{n}\right) - p$, soit en valeur absolue, inférieure à un nombre positif très petit ε tend vers l'unité, quand le nombre des épreuves croît indéfiniment.

Ce théorème est le théorème de Bernouilli.

On démontre d'autre part que, pour qu'une distribution se répartisse comme une distribution de Laplace Gauss (on dit aussi une distribution normale), il suffit que :

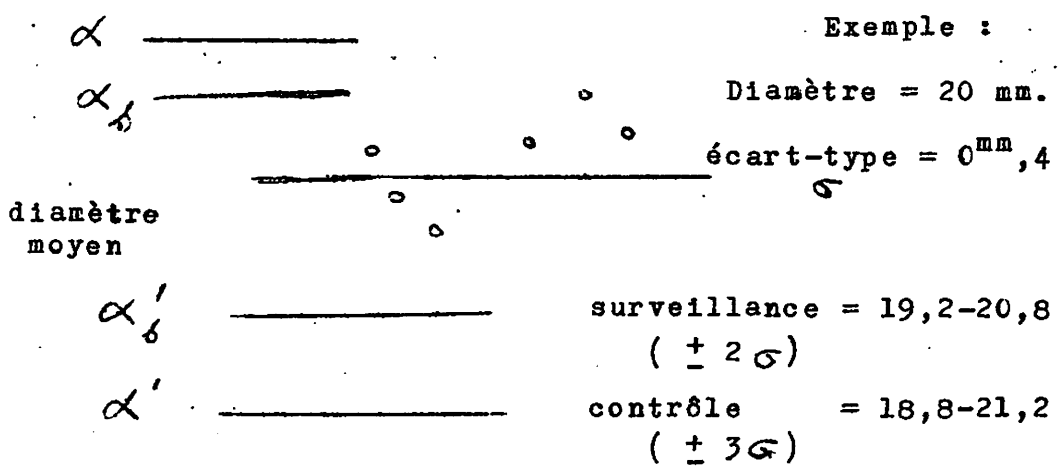
a) les causes accidentelles pouvant influencer sur la valeur du phénomène soient nombreuses,

b) l'effet d'une de ces causes doit être faible vis à vis de la somme de toutes ces causes,

c) elles doivent être indépendantes entre elles, et leurs effets sans relation avec la valeur de la variable (1).

Cette forme de courbe se rencontre fréquemment : courbe de survie de lampes à incandescence, vitesse de molécules dans un gaz, etc...

Dans le domaine du contrôle des fabrications, cette notion est capitale lorsque l'on étudie l'homogénéité d'une fabrication ; le principe est de pratiquer un prélèvement à intervalle périodique au cours de la fabrication d'une machine donnée, soit par exemple le diamètre d'un tube. Une étude antérieure fait connaître les limites inférieures et supérieures de deux grandeurs appelées limites de tolérance, car il est normal que même convenablement réglée, la machine donne lieu à une production non rigoureusement identique. Cette production devra être telle que le diamètre des tubes fabriqués soit compris entre ces limites de tolérance (l'une est appelée limite de surveillance (95 %), l'autre limite de contrôle (99 %).

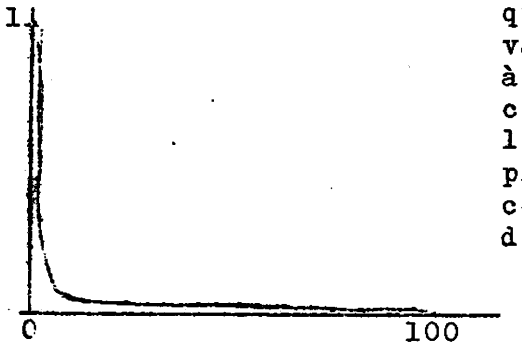


On reporte sur une carte de contrôle les mesures effectuées lorsqu'à un moment donné, un prélèvement sort des limites de contrôle, on arrête la machine.

(1) On rappellera pour mémoire la distribution de Gibrat pour laquelle l'importance des effets d'une cause est proportionnelle à la valeur de la variable.

9.3.2. Il existe d'autres distributions dont les diagrammes représentatifs sont différents du diagramme de Laplace Gauss.

Lorsqu'un évènement intervient très rarement au cours d'une certaine série de répétition, on obtient un graphique très dissymétrique dont la valeur modale peut correspondre à zéro. Ainsi si l'on a une chance sur 10.000 de gagner un billet à la Loterie Nationale, en prenant 100 billets, on a une courbe représentée ci-contre, de gagner 0 fois,
1 fois,
2 fois,
etc...



Cette courbe est la loi de Poisson ou dite des évènements rares. (1)

Autres exemples de la loi de Poisson :

- Nombre de décès du sexe féminin par jour de 85 ans et plus publiés par le Times :

nombre de décès par jour	fréquence
0	364
1	376
2	218
3	89
4	33
5	13
6	2
7	1

- Durant la guerre 1939-1945, Londres fut l'objet de bombardements par projectiles V_1. La zone fut divisée en 576 parcelles de $\frac{1}{4}$ Km2. Si le bombardement était dirigé sur le centre de Londres, on aurait dû trouver des zones où les points de chute étaient les plus fréquents. Par contre si l'arrosage est aléatoire, la répartition des 535 impacts

(1) Il convient de se rendre compte que la répétition de l'épreuve conduit à des probabilités de sortie de l'évènement non négligeables : soit un automobiliste qui chaque jour emprunte un croisement de rue fréquenté. La probabilité d'accident est de $1/1.000$. Au bout de mille passages, la probabilité de rencontre est de 0,63.

doit suivre une loi de Poisson.

Nombre de V_1 par parcelle	Nombre d'observations	Nombre théorique
0	229	226,74
1	211	211,39
2	93	98,54
3	35	30,62
4	7	7,14
5 et plus	1	1,57
	576	576,00

On peut en conclure à un arrosage aléatoire de la zone londonienne.

9.4 Loi de Probabilité et Variable aléatoire.

Ces notions seront complétées par deux définitions : la première concerne la définition d'une loi de probabilité : une loi de probabilité est constituée par une série de nombres $p_1, p_2, \ldots p_n$ auxquels sont attachés des valeurs $x_1, x_2, \ldots x_n$; x_i représente une valeur qui peut apparaître au cours d'un tirage et p_i est la probabilité attachée à la sortie de cette valeur.

On rejoint ici la notion de distribution de fréquences. On fait jouer aux fréquences d'apparition le même rôle que les probabilités mathématiques.

x dans ce cas est appelé une variable aléatoire.

La deuxième définition concerne l'espérance mathématique d'une variable aléatoire, elle est donnée par l'expression :

$$E(x) = \frac{x_1 p_1 + \ldots x_n p_n}{p_1 + p_2 + \ldots + p_n} = \frac{\sum x_i p_i}{1}$$

C'est la valeur pondérée des valeurs de x, les poids étant égaux aux probabilités partielles. Si on prend l'exemple d'un dé, chaque valeur du dé = 1,2,3,4,5,6, a une probabilité $\frac{1}{6}$ d'apparaître (si le dé n'est pas truqué) par conséquent, si on désire calculer l'espérance mathématique attachée à l'apparition d'un point du dé, on calcule

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} \ldots\ldots = 3,5.$$

Au cours d'un grand nombre de jets de dés, la valeur moyenne des points tirés sera précisément égale à cette espérance mathématique.

Avec un dé pipé, où le 5 apparaît en moyenne 10 % plus fréquemment que les autres valeurs :

valeur de la variable :	1	2	3	4	5	6
probabilité d'apparition :	0,164	0,164	0,164	0,164	0,18	0,164

$$\sum p_i \, x_i = 0,164 + 0,328 + 0,492 + 0,656 + 0,900 + 0,984$$
$$= 3,524$$

Au bout de cent jets, le joueur au dé pipé gagnera 2F,4 de plus qu'un joueur doté d'un dé normal.

Ces rudiments d'analyse statistique sont terminés et il faut en retenir les deux notions précises suivantes :

Notion de probabilité élémentaire : c'est le rapport du nombre des cas favorables à l'apparition d'un évènement donné, au nombre total des différents cas possibles.

Notion que les faits statistiques sont hasardeux : c'est-à-dire que ces faits sont variés et complexes : il en découle qu'on ne peut prévoir à coup sûr leur distribution, car les causes d'erreur sont multiples ; cependant, si le nombre des observations est suffisamment grand, il se dégagera certaines permanences, on pourra construire certaines courbes, et à l'examen des résultats, des régularités apparaîtront, là où l'on apercevait aucun ordre.

Souvent on pourra rattacher alors les courbes expérimentales à des courbes théoriques, et on pourra assimiler un phénomène observé à un phénomène ou à un schéma théorique.

Pour reprendre une phrase de Joseph Bertrand, on dira "que le hasard a des caprices, mais jamais on ne lui vit des habitudes".

Au contraire, si l'on constate des divergences systématiques on pourra conclure et rejeter le schéma initial.

De plus, et cela sera examiné en détail pour les statistiques agricoles et démographiques, l'analyse statistique viendra en aide lorsque les moyens matériels dont on dispose ne permettent pas d'entamer une enquête exhaustive.

En émettant certaines réserves, et en prenant certaines précautions, on peut alors étudier une partie seulement du domaine statistique : d'observations apparemment fragmentaires, on peut déduire des données numériques extrapolables de l'échantillon à l'ensemble de cette population inconnue. Cette assimilation de la partie au tout constitue la méthode des sondages, méthode qui a déjà fait l'objet de réalisations pratiques sur le terrain.

Lorsque dans le passé, on a parlé du hasard, bien souvent, ce mot n'a évoqué qu'incohérence ou incertitude ; le jeu de pile ou face, ou le retournement d'un roi dans un jeu de cartes, sont les types même du jeu hasardeux, cependant les probabilités permettent de prévoir pour un grand nombre d'observations les permanences de ce hasard et ce n'est pas le moindre paradoxe que ce soit précisément le caractère hasardeux de ces phénomènes qui permette de déduire les règles auxquelles ces phénomènes obéissent (1).

Il est enfin une remarque de bon sens qu'il faut formuler et elle constitue la conclusion de ce chapitre : lorsque l'on donne un chiffre, il faut délimiter l'erreur qui lui est attachée, on insiste généralement peu sur ce point. En fait, c'est une question d'honnêteté intellectuelle et même d'honnêteté tout court.

Le chiffre qui est diffusé, que les statisticiens publient, n'est jamais rigoureusement exact ; il ne l'est pas pour deux raisons principales : - la première est que l'observateur est homme, et que malgré le soin que l'on peut apporter à son travail, des erreurs, même monumentales, peuvent toujours se glisser, (mais il n'y a que ceux qui ne font jamais rien qui ne commettent jamais d'erreur),

---

(1) Mais ce caractère hasardeux ne doit pas faire oublier la constance de la probabilité élémentaire d'apparition du phénomène et c'est de cette connaissance positive dont on se sert pour déduire la distribution résultant d'épreuves répétées.

- la seconde raison, est que les observations ne sont jamais exhaustives, et qu'une marge d'intermination, une frontière un peu floue demeure : en conséquence une erreur en plus ou en moins, mais une erreur quand même, subsiste, même dans l'hypothèse où l'observation est sans défaut. Il est donc important, et c'est dans ce domaine que la technique statistique peut être utile, que les limites d'erreur soient bien délimitées afin que le lecteur ne soit pas abusé.

On peut donner une conséquence de cette remarque : on admet généralement en statistique qu'une erreur relative de 3 à 5 % constitue une très bonne approximation de la mesure d'un phénomène donné, or, bien souvent, on demande au statisticien de confirmer ou d'infirmer des chiffres fournis par un autre service, mais le demandeur exige parfois une précision encore plus grande, c'est-à-dire une marge d'erreur plus faible ; intellectuellement cette demande n'est pas recevable, et les utilisateurs sont choqués des doutes légitimes, et de ce que l'on pourrait appeler les réticences, les restrictions mentales ou les appréhensions du statisticien lorsqu'on le sollicite de publier des statistiques d'une précision outrancière. (Exemple : variation de la production agricole d'une année à l'autre)

Il ne faut donc pas oublier cette remarque et ne pas exiger plus que le calcul et le soin le plus grand ne peuvent donner. Une enquête en A.E.F. amenant à penser que le pouvoir d'achat du manoeuvre dans une localité était supérieure de 7 % à celui mesuré dans une seconde localité. Mais le calcul d'erreur montre que la mesure est entachée d'une erreur de 7 % en plus ou en moins, l'écart des pouvoirs d'achat n'est donc pas caractéristique.

IMPRIMEUR : Direction Générale de l'INSEE, 29, Quai Branly - PARIS 7e