



**BULLETIN DES SEANCES
MEDEDELINGEN DER ZITTINGEN**

51 (3)

EXTRAIT - OVERDRUK

**ACADEMIE ROYALE
DES SCIENCES D'OUTRE-MER**

Sous la Haute Protection du Roi

**KONINKLIJKE ACADEMIE
VOOR OVERZEESE WETENSCHAPPEN**

Onder de Hoge Bescherming van de Koning

La statistique dans les pays en développement: regard et perspective*

par

Jean-Jacques CLAUSTRIAUX**

MOTS-CLES. — Statistique; Biométrie; Agronomie; Outre-mer.

RESUME. — Dans le cadre des pays en développement, notamment dans le domaine de la recherche agronomique, la présente communication tente d'apporter quelques réponses aux questions suivantes:

- La statistique est-elle spécifique?
- Quel est son rayonnement?
- Comment augmenter sa pénétration?

TREFWOORDEN. — Statistiek; Biometrie; Landbouwkunde; Overzee.

SAMENVATTING. — *Statistiek in ontwikkelingslanden: overzicht en vooruitzicht.* — Wat de ontwikkelingslanden betreft, en meer bepaald het landbouwkundig onderzoek, wordt in deze voordracht gepoogd een antwoord te geven op volgende vragen:

- Is statistiek specifiek?
- Wat is de uitstraling ervan?
- Hoe kan de impact ervan vergroot worden?

KEYWORDS. — Statistics; Biometry; Agronomy; Overseas.

SUMMARY. — *Statistics in Developing Countries: View and Tendency.* — In the context of developing countries, notably in the field of agronomical research, the author suggests some answers to the following questions:

- Is there a specificity in statistics?
- What is its influence?
- How to increase its field impact?

1. A propos de la statistique

Pour introduire ces quelques réflexions sur la statistique, particulièrement son impact dans les pays en développement, sujet choisi dans l'optique d'introduire

* Communication présentée à la séance de la Classe des Sciences techniques tenue le 25 novembre 2004. Texte reçu le 1^{er} décembre 2004. Décision de publication prise le 24 novembre 2005.

** Unité de Statistique et Informatique, Faculté universitaire des Sciences agronomiques, av. de la Faculté 8, B-5030 Gembloux (Belgique).

un large échange de vues à la suite de l'exposé entre des personnalités scientifiques particulièrement instruites sur les conditions spécifiques de la recherche outre-mer, je ne peux que m'inspirer des propos cités par MEYER (2004) au début de son article sur «Les leçons de la nature», pour vous convaincre, si besoin est, que le statisticien n'est pas nécessairement un individu insensible aux charmes de celle-ci.

Hostile ou accueillante, la nature cache aux hommes ses mystères, ne leur offrant à admirer que sa beauté. Mais les scientifiques, insatiables et fous du désir d'en violer les secrets, nous dévoilent sans cesse de nouvelles raisons d'être séduits.

Autour de nous, il n'y a pas de plus simples, de plus parfaites leçons d'harmonie qu'une rose, le vol d'un oiseau ou le mouvement des vagues, mais c'est en **statisticien** que j'aimerais vous parler d'autres accords de la nature, aux charmes plus austères.

Comme le signale DAGNELIE (1998a), le mot statistique est dérivé du substantif latin *status* (état) et il possède deux significations distinctes.

Utilisé le plus souvent au pluriel, le terme désigne tout ensemble cohérent de données, comme les statistiques de naissances ou les statistiques des productions agricoles. Employé au singulier, il concerne toutes les méthodes qui permettent de rassembler et d'analyser des données codées sous forme numérique.

Dans un cas comme dans l'autre, le quidam a tendance à retenir les aspects strictement mathématiques de la statistique, le calcul des probabilités en particulier, ce qui la rend souvent antipathique et difficile *a priori* (NELDER 1999). De plus, même pour certains scientifiques, elle est considérée comme une science inaccessible, difficile et nécessitant des prédispositions particulières pour sa maîtrise.

Pour ce qui nous concerne, rejoignant ainsi, notamment, les propos de WEGMAN (2000), si les aspects théoriques sont évidemment essentiels, nous préférons regarder la statistique sous un aspect beaucoup plus pratique, c'est-à-dire en la considérant davantage comme une discipline dont le commun dénominateur est la **donnée**, plutôt que la méthode, y associant aussi tout ce qui la lie à sa collecte, soit lors d'un échantillonnage, soit lors d'une expérimentation.

De cette manière, on mesure davantage, non seulement l'intérêt de la statistique, mais aussi son évolution (DROESBEKE & TASSI 1990), en mettant également mieux en évidence son indissociable deuxième outil, lui beaucoup plus récent, à savoir l'informatique, le premier étant la mathématique.

Ainsi, de nouveaux sujets qui concernent la gestion de grandes quantités de données comme le *data mining*, l'analyse d'image, les modèles de reconnaissance, l'intelligence artificielle, la simulation aléatoire, etc., deviennent des traits fondamentaux de la discipline.

Signalons encore que le statisticien exerce une discipline assez singulière dans la mesure où il peut développer son savoir au service des autres scientifiques et qu'il n'est donc pas tenu de consacrer tout son temps à sa propre science pour

mieux encore la mettre en évidence. C'est d'ailleurs en considérant le premier aspect que nous aborderons essentiellement la suite de l'exposé.

Un domaine de la recherche appliquée qui traite du vivant, ou de tout ce qui en est dérivé, et pour lequel la donnée a toujours fait l'objet d'attentions particulières est certainement l'agronomie. Dans ce contexte, la statistique est souvent appelée biométrie ou discipline visant à élaborer des modèles statistiques permettant une compréhension plus profonde des phénomènes biologiques.

Cette définition est cependant trop restrictive à mon point de vue car elle n'intègre pas suffisamment l'ensemble des règles pratiques qui conduisent à une action ou à une décision dans des conditions déterminées, c'est-à-dire qui lui associe ce que certains auteurs ont appelé la praximétrie (MARTIN & HENRY 1960).

Enfin, vu la pénétration importante de la statistique pour traiter aujourd'hui des questions pharmaceutiques et médicales, le terme biométrie est de moins en moins utilisé: il est remplacé par l'expression biostatistique et même parfois par celle de biomathématique (RASCH *et al.* 1994).

2. La statistique outre-mer

Dans les pays en développement, le secteur économique que représente traditionnellement l'agriculture est toujours essentiel; dès lors, les activités de recherches associées sont indispensables et importantes. Plus récemment, à côté de tout ce qui concerne la production, la transformation et la valorisation des denrées alimentaires, la gestion de l'environnement, notamment celle des grands espaces boisés, et sa préservation sont devenues également une préoccupation majeure complémentaire.

Certes, que ce soit dans nos pays développés ou dans les pays en développement, la recherche en statistique, tout comme l'application des méthodes statistiques, sont analogues.

Toutefois, ce qui pour moi est très différent, ce sont, d'une part, les conditions dans lesquelles le statisticien travaille dans les pays d'outre-mer et, d'autre part, son isolement.

2.1. LES CONDITIONS DE TRAVAIL

Tout d'abord, dans ces pays, les données présentent davantage une variabilité élevée, souvent en raison des conditions du milieu particulièrement hétérogènes et mal contrôlées. A cette forte variabilité ou hétérogénéité vient s'ajouter l'existence fréquente d'accidents de parcours dans l'échantillonnage ou l'expérimentation.

Pour illustrer ces propos, considérons la forêt tunisienne qui a perdu un quart de sa superficie en quarante ans, soit plus de 250 000 hectares. Pour restaurer et

reconstituer ce couvert végétal, des recherches à long terme sur une période de trente-six ans ont été entreprises dans quatre sites, notamment en matière de choix des provenances de pin d'Alep (*Pinus halepensis* Mill.). En résumé, la question essentielle qui se pose est de savoir comment la statistique peut contribuer à atteindre l'objectif fixé, compte tenu des conditions essentiellement pédologiques très variables et des taux élevés de mortalité des sujets?

Sans entrer dans tous les détails (SCHAIER *et al.* 2004), des méthodes moins classiques, aux fondements théoriques mal connus, comme celles basées sur les modèles des plus proches voisins, peuvent, par approximations successives, contribuer largement à une meilleure prise en compte du contrôle local de l'hétérogénéité.

A titre d'exemple de modèle, celui repris ci-après identifie les voisins notés $X_{i,c}$, d'une provenance $Y_{i,c}$, située à la ligne i et à la colonne c d'un dispositif expérimental, et pour lequel il convient d'estimer le contrôle local de l'hétérogénéité $X_{i,c}$, b_1 , b_2 et $\epsilon_{i,c}$ étant respectivement les coefficients et le résidu aléatoire du modèle:

$$X_{i,c} = b_1 (X_{i+1,c} + X_{i-1,c} + X_{i,c+1} + X_{i,c-1}) + b_2 (X_{i+1,c+1} + X_{i+1,c-1} + X_{i-1,c+1} + X_{i-1,c-1}) + \epsilon_{i,c}$$

Grâce à ce type d'outil, il est donc possible d'extraire des observations, des quantités d'informations qui masquent la qualité recherchée d'une provenance en termes de hauteur, du diamètre, de la forme, du couvert végétal ou de la qualité du bois.

Ensuite, un second exemple va mettre en évidence une autre caractéristique des pays en développement, montrant à nouveau combien l'approche classique du traitement statistique des données est parfois bien difficile.

Cette fois, il se rapporte au domaine des enquêtes par échantillonnage dans des régions qui n'ont rien de commun avec les nôtres, dans la mesure où l'environnement économique est difficile et les informations préalables nécessaires à la mise en place d'enquêtes sont inexistantes, pauvres, erronées, obsolètes, inadéquates, volontairement biaisées, etc.; les moyens humains affectés à l'étude sont réduits ou peu qualifiés; les facteurs historiques ou politiques freinent la recherche d'informations objectives et, enfin, les délais impartis à l'enquête et à son analyse sont faibles, tout comme bien souvent les budgets.

Dans ce cas précis, l'objectif est d'évaluer la situation technico-économique des exploitations agricoles privées de deux régions de la République du Kirghizistan (LABE & PALM 1999), sachant qu'il n'est donc pas possible de mettre en place des méthodes statistiques classiques d'enquêtes en utilisant des bases de sondages pour ensuite calculer, notamment, des estimations et des marges d'erreur.

Dans une telle situation, il ne reste comme outil que les méthodes d'enquêtes dites informelles (CASLEY & KUMAR 1988). L'objectif recherché est alors de

comprendre davantage des systèmes (l'exploitation agricole, la structure familiale, l'organisation villageoise, etc.), en collectant des données qualitatives et en considérant l'enquêté, non pas comme une simple unité d'échantillonnage, «ignorante, bête et paresseuse» à observer, mais comme le principal partenaire avec lequel il convient de dialoguer pour obtenir la précieuse donnée. Le commun dénominateur pour la réussite de la mise en place de ces méthodes est donc de trouver des prétextes pour faciliter la discussion. Il est évident aussi que les questionnaires sont peu structurés et les observations sont des phrases, des schémas, des dessins, des comptes rendus de visites, etc.

Le travail d'analyse des données et l'interprétation des résultats ne peuvent évidemment se réaliser en principe que par des statisticiens qualifiés, ayant une connaissance suffisante du terrain et non par des «formulateurs» de principes purement théoriques.

J'ajouterai encore que dans certaines circonstances, la situation est telle que la seule méthode statistique disponible n'est hélas et tout simplement plus que l'intuition, sur base d'observations évidemment subjectives.

2.2. L'ISOLEMENT

Au niveau individuel, la place occupée par le statisticien dans les pays en développement est tout à fait différente de celle rencontrée dans les pays développés.

Des études à long terme (DAGNELIE 1998b) se référant aux membres de la Société internationale de Biométrie (*IBS*) ont montré que, quel que soit le secteur d'activité du statisticien, par rapport aux effectifs de la population, on pouvait considérer qu'en moyenne il y avait de l'ordre d'un statisticien par million d'habitants dans le monde, mais avec des disparités énormes entre les continents puisque, s'il y avait un statisticien pour 125 000 habitants en Amérique du Nord et au Canada, et un statisticien pour 500 000 habitants en Europe, en Australie, au Japon et en Nouvelle-Zélande, il n'y avait qu'un statisticien pour 5 000 000 habitants pour les autres pays du continent américain, l'Afrique, les autres pays d'Asie et l'Océanie, soit un rapport de un à quarante entre le Sud et le Nord. Par comparaison, la Société belge de Statistique compte de l'ordre d'un membre pour 100 000 habitants en Belgique.

A ces constatations, il faut cependant ajouter quelques éléments d'analyse, même si ces chiffres sont intéressants et qu'ils mettent parfaitement en évidence un déséquilibre permanent entre le Sud et le Nord au cours du temps.

En effet, les données ne sont que le reflet de la répartition géographique des membres d'une seule société scientifique internationale et elles ne précisent pas ce qu'est un membre: statisticien isolé ou non, enseignant de la statistique sans pratique de celle-ci, statisticien théoricien ou praticien, etc. Par ailleurs, rapporter les résultats à la population est certes nécessaire, mais il conviendrait d'évaluer également la présence relative du statisticien par rapport aux autres scientifiques, chiffres beaucoup plus difficiles à obtenir.

Néanmoins, cette plus faible présence de statisticiens dans les pays du Sud est hélas aussi accentuée par le fait que nous avons pu constater que la statistique et le statisticien sont relativement « ignorés » dans l'enseignement et la recherche. Dès lors, le chercheur en sciences biologiques ne peut se rendre compte, ni résoudre des questions essentielles, qui se posent à lui en termes de planification expérimentale et d'interprétation des résultats. De plus, trop souvent, bien informé de la théorie, mais maîtrisant très peu des notions biologiques fondamentales, le mathématicien prend la place du biométricien, sans pour autant résoudre les questions plus appliquées. Aussi, les articles scientifiques proposés par les chercheurs de ces pays sont souvent refusés par les éditeurs du Nord faute d'une planification expérimentale adéquate ou d'une inadéquation entre les objectifs de la recherche et les méthodes statistiques utilisées (NELDER 1999).

Dans cet isolement, les statisticiens du Sud ont aussi leur part de responsabilité pour une question de mentalité, de perspectives professionnelles, etc.

Cependant, reconnaissons que la situation d'isolement relatif du statisticien outre-mer s'est améliorée, grâce, notamment, à l'arrivée massive des ordinateurs, des logiciels libres ou non et de l'accès au réseau internet, facilitant ainsi la réalisation des calculs et la mise à disposition de nouvelles méthodologies.

3. Quelques suggestions

Face à ce constat, spécificité des conditions de travail et présence moindre du statisticien dans les pays du Sud, que faire pour favoriser une amélioration de la situation?

Si réduire la variabilité n'est évidemment pas possible, il ne reste qu'à proposer des actions au niveau des hommes eux-mêmes.

Mais, avant d'apporter des éléments de réponse, ne faut-il pas s'interroger sur l'opportunité d'accroître la présence du statisticien outre-mer?

Par une simple phrase, Wilson (*in* NELDER 1999) répond à la question: «the presence of regulatory statisticians in the pharmaceutical field has surely been the major reason that the industry has become such a major employer of statisticians».

Le rôle joué par le statisticien en matière de recherche-développement dans ce secteur, certes particulier, a donc certainement un impact économique positif, tout comme pour l'entreprise publique ou privée, si la sécurité coûte, cela coûte davantage de ne pas s'en préoccuper. Pourquoi en serait-il autrement dans d'autres secteurs économiques, dont celui du développement?

En conséquence, quelles pourraient être les actions à entreprendre pour encourager une meilleure pénétration de la statistique?

Dans le contexte des décideurs publics ou privés, il est évident que toute institution universitaire ou de recherche scientifique ne devrait porter ces qualificatifs que si la présence de statisticiens expérimentés est effective, si besoin en association.

Par ailleurs, ces décideurs devraient aussi mieux se rendre compte que les bailleurs de fonds se préoccupent de cet indicateur puisque la qualité des résultats des travaux est souvent jugée par des experts statisticiens *a posteriori*, mettant ainsi éventuellement en cause la crédibilité de l'institution de recherche elle-même.

Mettre en valeur nos collègues statisticiens étrangers dans nos enseignements, même de premier ou de deuxième cycle, au titre de professeur visiteur, est une action qui devrait aussi être mieux soutenue, par exemple par la création de Chaires du Sud ou mieux encore de Chaires d'Excellence.

En effet, il ne faut pas cacher que des problèmes de carrière professionnelle se posent pour nos collègues étrangers, surtout si celle-ci est davantage valorisée par des publications plutôt que par des conseils judicieux à d'autres collègues.

Par ailleurs, puisque le statisticien est souvent seul, isolé dans son institution et qu'il agit parfois sur de vastes territoires, toute action qui favorisera la communication avec le monde extérieur, celui du Nord en particulier, pour lutter contre l'isolement scientifique rapide en statistique et encore plus en informatique doit être soutenue (séjours d'études, de recyclages, établissement de réseaux thématiques de statisticiens, via internet, etc.).

Un étudiant est très rarement enclin à étudier la statistique ou à se spécialiser en la matière. Il faut donc créer des filières d'enseignement spécifiques, davantage orientées vers le traitement statistique et informatique des données, et implémenter des incitants ou des encouragements à ces formations, tels que des bourses de formation, des prix pour récompenser des travaux réalisés dans le secteur, des subsides pour aider les jeunes statisticiens d'outre-mer à communiquer dans les colloques internationaux, pour autant que les organisateurs de ces derniers restent ouverts à des problématiques plus traditionnelles ou plus pragmatiques, ce qui est hélas de moins en moins le cas.

Si l'éthique en expérimentation humaine est une préoccupation certes essentielle et si elle concerne aujourd'hui toute la recherche animale, ne devrait-elle pas aussi concerner le secteur du végétal, si ce n'est pas pour la défense de l'organisme en tant que tel, tout au moins pour une question liée au coût de la recherche?

Dans ce cadre, tout chercheur pourrait rapidement se rendre compte que le statisticien est loin d'être un gêneur, mais qu'il est davantage un conseiller ou un allié pour, par exemple, peser sur le nombre de sujets à observer, le statisticien ne devant plus être ainsi cantonné au simple rôle d'analyseur des données.

En résumé, sans l'aval du statisticien, pas de recherche et donc plus d'institution de recherche, au Nord comme au Sud! Certaines entreprises ont parfaitement adopté ce principe depuis longtemps.

Le bon usage de la statistique doit encore être encouragé par la rédaction et la diffusion auprès des auteurs d'articles et des éditeurs de revues non statistiques de guides des bonnes pratiques statistiques en matière d'expérimentation, d'échantillonnage et de publication de résultats. En particulier, par exemple, ces

derniers ne devraient plus être considérés comme ayant une caution statistique suffisante par la simple présence dans les articles d'astérisques assimilant ainsi automatiquement le rejet d'une hypothèse à un progrès scientifique.

A l'heure de l'image et de la communication, si les mots ont leur importance pour une meilleure perception d'un domaine au Nord comme au Sud, ne faut-il pas donner à la statistique la place qui lui revient en identifiant un maximum de structures organisationnelles par les termes «science statistique» dans l'appellation des instituts, des départements ou des unités en lieu et place du simple mot «statistique»?

Dans cet isolement, les sociétés scientifiques, c'est-à-dire les statisticiens associés, ont aussi à apporter leur contribution comme nous l'avons déjà évoqué. Ne devraient-elles pas davantage considérer également comme dignes d'intérêt des publications à caractère moins théorique et organiser systématiquement, durant les réunions scientifiques, des séances plus appliquées sur des thèmes mettant en valeur le bon usage de la pratique statistique outre-mer, certes en prévoyant également des modalités plus avantageuses de participation à des congrès pour des ressortissants du Tiers-Monde?

Nous-mêmes, individuellement, statisticiens des pays dits développés, ne devrions-nous pas consacrer davantage de temps à ces pays en développement:

- Par des formations locales pour les chercheurs du Sud ou la participation à des séminaires régionaux, éventuellement à distance;
- Grâce aux moyens actuels et rapides de communications en initiant des centres-relais de consultations statistiques aussi pour les chercheurs du Sud, de manière à mieux définir les objectifs de leurs travaux et à mieux mettre en valeur les résultats obtenus;
- En influençant nos autorités, au sein de nos institutions, à mieux valoriser nos jeunes collègues qui consacrent une partie de leur temps de recherche en faveur des pays en développement pour leurs promotions académiques;
- Etc.

4. En guise de conclusion

La statistique est certainement un domaine de la science qui a discrètement contribué au développement socio-économique des pays du Sud, souvent sans nécessairement faire appel à des méthodologies sophistiquées. Mais, son taux de pénétration est moins important qu'au Nord et sa progression est lente, malgré plusieurs signaux d'alarme déjà anciens (DAGNELIE 1984, OTOUL 1994), alors que d'aucuns affirment que la maîtrise des nouveaux savoirs et la culture technologique sont devenues les seuls garants du développement durable du Sud.

Plus encore dans ces régions du monde aux conditions de milieux spécifiques, il nous semble que la qualité des recherches entreprises va de concert avec une présence plus effective de la statistique sur le terrain et qu'il faut l'encourager.

Sans cela, les travaux de recherches et les publications qui en découleront n'auront généralement plus de véritable portée et ne contribueront plus à un véritable développement, sauf si tout chercheur fait, tel Monsieur Jourdain, de la statistique sans le savoir...

BIBLIOGRAPHIE

- CASLEY, D. J. & KUMAR, K. 1988. The collection analysis, and use of monitoring and evaluation data. — Baltimore, Johns Hopkins University Press, 174 pp.
- DAGNELIE, P. 1984. La situation de la biométrie dans les pays du tiers monde. — *Biom. Praxim*, **24**: 3-21.
- DAGNELIE, P. 1998a. Statistique théorique et appliquée (tome 1: Statistique descriptive et bases de l'inférence statistique). — Bruxelles, De Boeck & Larcier, 508 pp.
- DAGNELIE, P. 1998b. A quantitative vision of biometry in developing countries. — *Biom. Bull.*, **15**: 6-9.
- DROESBEKE, J.-J. & TASSI, P. 1990. Histoire de la statistique. — Paris, Presses Universitaires de France, 128 pp.
- LABE, V. & PALM, R. 1999. Statistique, empirique, informelle: quelle enquête pour la collecte d'informations sur les exploitations agricoles? — *Cahiers Agricultures*, **8**: 397-404.
- MARTIN, L. & HENRY, J. 1960. Editorial. — *Biom. Praxim*, **1**: 2-4.
- MEYER, D. 2004. Les leçons de la nature. — *Le Monde*, **18585**: 20.
- NELDER, J. A. 1999. Statistics for the Millennium. — *The Statistician*, **48**: 257-269.
- OTOUL, C. 1994. La statistique et la recherche agronomique dans les pays du tiers-monde: son passé et les tendances nouvelles. — *Biom. Praxim*, **34**: 147-161.
- RASCH, D., TIKU, M. L. & SUMPF, D. 1994. Elsevier's dictionary of biometry. — Amsterdam, Elsevier, 887 pp.
- SGHAIER, T., CLAUSTRIAUX, J.-J. & BEJI, M. A. 2004. Intérêt des modèles des plus proches voisins pour le contrôle de l'hétérogénéité spatiale: application à un essai de provenances de Pin d'Alcp (*Pinus halpensis* Mill.) en Tunisie. — *Revue de l'INAT*, **19** (2): 5-22.
- WEGMAN, E. J. 2000. On the eve of the 21st century: statistical science at a crossroads. — *Comp. Stat. Data Anal.*, **32**: 239-243.