

Du Calcul des Probabilités considéré comme base de l'action

par W. Edwards Deming
The American Statistician - Novembre 1975
Traduction et adaptation : Jean-Marie Gogue

Résumé

L'auteur a pour but l'amélioration des pratiques statistiques. Il fait la distinction entre les études énumératives et les études analytiques. Une étude énumérative a pour but une estimation du nombre d'unités qui appartiennent à une classe spécifiée dans un cadre donné. Une étude analytique a pour but une base de l'action sur le système de causes ou le processus, afin d'améliorer la production future. Déterminer un prix raisonnable pour un stock de marchandises est un exemple d'étude énumérative. Des tests sur des variétés de céréales, des médicaments ou des procédés de fabrication sont des exemples d'études analytiques. Elles déterminent le choix de traitements qui affecteront la qualité des céréales, la santé des gens, de futurs produits. Les techniques et les méthodes de déduction qui s'appliquent à des études énumératives conduisent à des plans et des déductions fausses s'il s'agit de problèmes analytiques.

Il est possible, dans un problème énumératif, de réduire les erreurs d'échantillonnage à un niveau spécifié. Au contraire, dans un problème analytique, il est impossible de calculer le risque de prendre une mauvaise décision. L'auteur donne de nombreux exemples et plaide pour qu'on prenne un plus grand soin dans la rédaction et l'enseignement de la théorie statistique et de la déduction.

But et champ d'application de cet article

Nous avons ici pour but de contribuer quelque peu à l'amélioration des pratiques statistiques. Nos hypothèses fondamentales sont que toute investigation statistique est tournée vers l'action et que toute connaissance nouvelle modifie l'ensemble des connaissances existantes.

Le travail statistique est un besoin urgent

Les statisticiens n'avaient jamais eu à relever de tels défis. Tout le monde parle de la sécurité des automobiles, du danger des médicaments, de la pollution, de l'amélioration des techniques agricoles, du besoin d'une plus grande productivité dans l'industrie, de l'enrichissement des tâches. Le consommateur exige tous les jours une meilleure sécurité et une meilleure performance des produits. Le producteur rencontre lui aussi des problèmes avec les matériaux, les machines et le personnel. Il doit connaître de mieux en mieux ses produits. Quelles sont les obligations d'un fabricant ? Qu'est-ce qu'une faute professionnelle pour un médecin ? Les études statistiques sur la consommation sont dans un état lamentable, bien que les sommes dépensées dans ce domaine soient en augmentation chaque année.

On ne peut pas comprendre ces problèmes ni même les énoncer correctement sans recourir à une théorie statistique. On ne peut pas évaluer l'effet d'une nouvelle solution sans recourir à des méthodes statistiques. Quand des expressions telles que fiable, polluant, demandeur d'emploi, égal, rond, aléatoire, rouge, vert... sont utilisés dans les affaires ou dans l'administration, elles ne peuvent être définies qu'en termes statistiques. Une norme (de sécurité, de performance, etc.) ne peut avoir un sens dans le commerce, ou devant un tribunal, que si elle est définie en termes statistiques.

Si l'étiquette d'une couverture indique « 50 % laine », qu'est-ce que cela signifie ? Est-ce la moitié en moyenne dans la couverture, ou la moitié sur la production mensuelle ? Est-ce la moitié en poids ? dans ce cas, quelles sont les conditions d'humidité ? Quelles méthodes d'analyse chimique sont utilisées ? Combien d'analyses sont faites ? Si la moitié supérieure de la couverture est en laine et la partie inférieure en autre chose, la couverture est-elle 50 % laine ? Cette indication veut-elle dire qu'il doit y avoir de la laine dans n'importe quel échantillon grand comme une pièce de monnaie, découpé au hasard ? Dans ce cas, combien d'échantillons faut-il mettre à l'essai ? Comment les sélectionner ? À quel critère la moyenne doit-elle satisfaire ? Quelles variations entre les échantillons sont-elles

tolérables ? Il est évident que la signification de 50 % laine ne peut s'énoncer qu'en termes statistiques. Des mots simples en anglais, en français ou en japonais ne suffisent pas.

Nous avons besoin de changements énergiques dans notre manière d'agir, d'écrire et d'enseigner. Comme l'a dit Walter Shewhart, les normes de la connaissance de ce qui se fait dans l'industrie et les services publics sont plus sévères que ce qui est demandé dans la science pure. Il aurait pu ajouter que les conditions nécessaires pour la pratique des statistiques sont encore beaucoup plus strictes que les conditions nécessaires pour l'enseignement des statistiques. C'est regrettable mais c'est un fait (j'aborderai plus loin les problèmes de l'enseignement).

Le cadre, l'univers, l'environnement

Une étude statistique procède par l'étude de matériaux contenus dans un cadre. Le cadre est un assemblage d'unités physiques d'un même type, identifiables et tangibles, dont chacune peut être sélectionnée et étudiée. Le cadre peut être constitué par des listes de personnes, de zones, d'établissements, de matériaux ou d'autres unités identifiables qui devraient donner des résultats utiles si tout le contenu était étudié. Ce peut être un lot de pièces fabriquées. Tout aussi importante dans un problème analytique est la description des conditions d'environnement qui peuvent affecter les résultats.

Pour faciliter la présentation, nous utilisons un cadre de N unités échantillonnées, numérotées de 1 à N . On rencontre cependant des circonstances où la dimension du cadre est indéfinie et où chaque unité échantillonnée se voit attribuer une probabilité P d'échantillonnage. Un échantillon de 100 % (recensement complet) ne peut pas être défini en l'absence d'un cadre de N unités échantillonnées.

Distinction entre études énumératives et analytiques

UNE ETUDE ENUMERATIVE est faite quand une action doit être prise sur les matériaux contenus dans le cadre étudié. Cette action dépend entièrement des estimations ou des dénombrements complets d'une population du cadre. Une étude énumérative a pour but de décrire une situation. Combien de personnes appartiennent à telle ou telle catégorie ? Combien de pièces défectueuses trouve-t-on dans le lot ? Le but n'est pas de savoir pourquoi un certain nombre d'unités se trouvent dans une certaine catégorie.

UNE ETUDE ANALYTIQUE est faite quand une action doit être prise sur le processus ou le système de causes qui a produit le cadre étudié. Elle a pour but d'améliorer la pratique dans l'avenir. L'objet de l'étude n'est pas l'ensemble des matériaux étudiés mais la production future. L'action consiste à adopter une méthode B plutôt qu'une méthode A, ou à conserver une méthode A, ou à continuer l'expérience.

CRITERE - Il existe un moyen très simple de faire la distinction entre les études énumératives et les études analytiques. Un échantillon de 100 % du cadre donnera entièrement la réponse à la question posée dans un problème énumératif, sauf évidemment si la méthode d'investigation ne le permet pas. Au contraire, dans un problème analytique un échantillon de 100 % d'un groupe de personnes, ou d'un territoire, ou de la production de la semaine dernière, ne donnera pas la réponse. Bien que cette distinction soit fondamentale dans l'information statistique, c'est un point qui a échappé à de nombreux auteurs.

Les deux types de problèmes font appel à des procédures différentes pour sélectionner et calculer les estimateurs. Par exemple, dans un problème énumératif dont le but est d'estimer le nombre d'individus du sexe féminin d'âge compris entre 20 et 29 ans dans un cadre donné, nous n'avons pas besoin de reconnaître des strates au début ou à la fin, même si la stratification améliore la précision de l'estimateur sans augmenter le coût.

Au contraire, dans un problème analytique où la question est de découvrir dans quelles conditions deux traitements A et B donnent par exemple sur le rendement d'une culture de céréales un écart supérieur ou égal à D , il est recommandé de faire les comparaisons initiales avec des strates dans lesquelles les différences de température, d'humidité, etc. sont poussées à l'extrême. Si quelqu'un, faute d'une expérience suffisante, ne prévoit pas quelles strates pourront réagir de façon différente aux traitements A et B, le succès de son étude analytique sera fortement compromis.

Deux sortes d'erreurs dans un problème

Il y a deux sortes d'erreurs dans un problème énumératif. Par exemple dans une transaction commerciale, il est possible de :

1. Payer trop cher la différence d'un type de minerai à un autre
2. Vendre trop bon marché la différence d'un type de minerai à un autre

Il y a aussi deux sortes d'erreurs dans un problème analytique. Nous pouvons par exemple :

1. Remplacer le processus A par le processus B et le regretter plus tard
2. Garder le processus A et regretter plus tard de ne pas avoir adopté le processus B

La fonction du statisticien est d'essayer de minimiser la perte qui résulte des deux types d'erreurs, dans un problème analytique comme dans un problème énumératif. Dans un problème énumératif il y a des formules précises, mais dans un problème analytique il n'y a que des hypothèses incertaines.

Les statisticiens doivent prendre en compte le fait que, dans un problème analytique, il est impossible de formuler une fonction de perte et d'obtenir le minimum de perte. On ne peut donc pas, même avec des méthodes statistiques, éliminer ces erreurs en toute connaissance de cause. C'est parce qu'il est impossible de rassembler une somme de données empiriques assez grande pour prédire les futures conditions de l'expérience et la performance qui en résultera.

La comparaison de deux traitements n'est pas un test de régularité ayant pour but d'estimer la variance entre des individus soumis au même traitement. Ce n'est pas une investigation pour savoir si certaines données expérimentales sont conformes à la loi de l'hérédité ou à tout autre loi.

Limites de la déduction statistique

Tous les résultats des études analytiques, en agronomie, en médecine, etc. varient selon les conditions expérimentales : (a) le cadre dont proviennent les unités à tester ; (b) la méthode d'investigation (questionnaire ou méthode de test) ; (c) les personnes qui font l'enquête ou les mesures ; (d) les conditions d'environnement (par exemple le lieu, la date et la durée de l'expérience).

Dès qu'une expérience a été effectuée, on sait que les conditions d'environnement ne se reproduiront jamais exactement. Deux traitements qui n'ont pas fait apparaître de grandes différences dans certaines conditions d'environnement pourront donner des résultats très différents dans d'autres conditions (sols, climats, etc.).

Il n'existe pas de méthode statistique permettant d'extrapoler les conséquences de l'usage d'un médicament au-delà de la période de test et à d'autres patients. Des effets secondaires peuvent apparaître. Il est également impossible d'extrapoler les résultats des ventes d'un produit. Un concurrent peut lancer un autre produit sur le marché ou faire une grosse campagne de publicité. Les prévisions ne sont plus valables quand les conditions économiques ont changé. Telles sont quelques-unes des raisons qui font qu'un problème analytique ne peut jamais être résolu complètement et que les calculs utilisant une "fonction de perte" ne sont que conditionnels. Le fossé qui sépare la déduction statistique de la réalité ne peut être comblé que par la connaissance du sujet (en économie, en médecine, en psychologie, dans l'industrie etc.) Ces conseils sont passés sous silence dans tous les livres sur les méthodes de décision, un domaine qui en aurait le plus grand besoin.

Pour que la présentation des résultats ait l'efficacité optimale et qu'elle soit de bonne science, il faut qu'elle se conforme à la règle de Shewhart, c'est-à-dire qu'elle préserve tous les témoignages contenus dans les données originelles. Les données d'une expérience se composent de beaucoup plus de choses qu'une moyenne et un écart-type. En fait, l'ensemble des données d'un problème ne se limite pas aux données originelles. Pour comprendre les résultats d'une expérience, l'utilisateur doit obtenir la description de la méthode d'investigation, le lieu, la date et la durée des essais, le relevé des défauts mis en évidence par le contrôle statistique, le taux de non-réponse, les noms des observateurs.

En tant qu'architecte d'un projet, le statisticien a l'obligation d'aider son client à percevoir d'avance les limitations de l'étude envisagée et de modifier le projet si c'est nécessaire pour satisfaire à ses exigences. Une importante question à poser avant que l'étude soit très avancée est la suivante : À quoi les résultats vont-ils servir ? La réponse conduit parfois à modifier profondément les plans.

De quoi avons-nous besoin ?

Ce dont nous avons besoin, dans une étude comparative, c'est de savoir si la différence entre deux traitements A et B présentera une importance matérielle, économique ou scientifique, dans les futures conditions d'emploi. C'est un problème d'estimation. Quel est l'ordre de grandeur de la différence ?

Il est important de se rappeler que la moyenne, la variance et bien d'autres fonctions mathématiques d'un ensemble de nombres sont des fonctions symétriques. On ne change pas leur valeur en interchangeant deux nombres. Évidemment, la variance et les méthodes d'estimation élaborées font disparaître l'information contenue dans l'ordre des données originelles. Il faut donc considérer qu'elles sont inefficaces jusqu'à plus ample informé.

Les histogrammes, les diagrammes de corrélation, les graphiques de tendance tracés à la main sont des méthodes d'estimation plus efficaces que la déduction statistique car elles préservent l'information contenue dans les données originelles. Dans certains cas, lorsque l'étude s'est faite dans les conditions prévues, on peut résumer les résultats par des intervalles de confiance ou des écarts-type. Mais il faut d'abord démontrer le bien-fondé de ces méthodes.

Il faut reconnaître, avec Sir Winston Churchill, qu'il est parfois payant d'admettre une évidence. On ne conduit pas une expérience pour savoir si deux variétés de céréales sont égales. On sait à l'avance, sans dépenser un dollar, qu'elles sont inégales. La différence entre deux traitements ou deux groupes d'individus finira par être une « différence significative » pourvu que l'expérience s'étende assez loin, même si la différence est si faible qu'elle est sans intérêt sur le plan scientifique ou économique.

De même, la question de savoir si les résultats d'une enquête ou d'une expérience sont en accord avec une courbe théorique est sans intérêt sur le plan scientifique ou économique. En effet, si les résultats sont assez nombreux, aucune courbe ne s'ajuste aux résultats d'une expérience. La question intéressante quand on utilise une courbe est la suivante : Quelle est la robustesse des conclusions ? Est-ce qu'une autre courbe conduirait à des prévisions plus sûres ?

La signification statistique ne conduit à aucune connaissance, aucune base de l'action.

Utilisation des échantillons statistiques

1. L'utilisation d'un échantillon pour une étude énumérative (par exemple pour estimer la fréquence d'un certain défaut) n'a pas plus de valeur que la réputation de la personne qui signe le rapport. C'est parce que son jugement est le seul moyen de fixer les limites d'incertitude de l'estimateur. Les échantillons ont l'avantage d'éclairer un peu la question. Ils permettent d'évaluer l'incertitude sur des données qui proviennent d'une multitude de causes de variations, par exemple de la variance entre les observateurs, mais ils sont rarement nécessaires.

2. L'utilisation d'échantillons n'est pas toujours nécessaire dans un problème énumératif. Il semble qu'il y ait quelques exceptions, par exemple lorsqu'il s'agit d'un tas de charbon ou d'un chargement de minerai. (Dans ce cas, le problème de l'échantillonnage peut être résolu si les matériaux passent sur un convoyeur à bande).

3. En revanche, la majeure partie de nos connaissances scientifiques provient de jugements portés sur des échantillons dans des études analytiques.

4. Bien que l'on puisse s'arranger pour ne faire des comparaisons de traitements que sur des sujets extrêmement anormaux, ou dans des instituts spécialisés tels que le Centre d'Etudes agronomiques de Rothamsted, il est réconfortant de noter que les expériences faites sur les patients d'un hôpital pour comparer deux traitements nous apprennent aussi quelque chose. Si des expériences (organisées convenablement) montrent une différence appréciable, on peut affirmer, avec une probabilité d'erreur qu'il est possible de calculer, que deux traitements sont matériellement différents. On peut l'affirmer, même si les traitements ne sont plus appliqués ensuite aux mêmes personnes et dans les mêmes conditions. Le fait de constater une différence intéressante pour l'économie ou la science est une contribution incomplète, mais il peut être aussi une contribution importante pour nos connaissances.

5. Le dernier paragraphe met en évidence l'importance de la randomisation et de la théorie des plans d'expérience dans le jugement sur échantillon. La randomisation à l'intérieur de blocs dans un secteur que l'on a choisi convenablement (notamment en agronomie et en médecine) éclaire

considérablement la question. Si les conditions sont favorables, la randomisation des traitements dans une strate sélectionnée justifie l'usage des probabilités pour une déduction conditionnelle. La théorie des plans d'expérience aide à réduire les variances autant que possible pour un certain coût. Mais toute déduction fondée sur les résultats des expériences est conditionnelle, quelle que soit l'efficacité du plan d'expérience.

6. Nous avons déjà observé que la sélection de strates très différentes peut être l'approche la plus efficace pour une comparaison de traitements. Quand les résultats tendent à prouver que l'on a pris assez de strates pour identifier le domaine où il existe une différence égale ou supérieure à D entre des traitements A et B, on peut étudier les strates une par une.

Il est assez facile de comprendre qu'un échantillon statistique pris sur un cadre entier est inefficace pour une étude analytique. Par exemple, si une étude agronomique prétendait tester deux traitements en prenant au hasard des échantillons de terre arable sur toute la planète, le résultat serait inexploitable parce que chaque échantillon serait soumis à de trop nombreuses conditions de terrain, de climat etc. L'estimation de la différence serait une moyenne valable pour le monde entier, mais aucune déduction pratique ne pourrait être faite.

Critique de l'enseignement

Le moment est venu de faire quelques critiques sur l'enseignement des statistiques, en espérant qu'elles aideront à améliorer la pratique des statistiques dans l'avenir.

Exemple 1. Lu dans un livre de statistique mathématique. On a testé le temps de combustion sur 19 parachutes : 10 échantillons pour le lot A, 9 pour le lot B. Étant donné que $t(0,25) = 1,96$, il en résulte que l'hypothèse nulle (les deux populations sont identiques) ne peut pas être rejetée en faveur de l'alternative (les deux populations sont différentes), avec un niveau de signification de 0,05.

Commentaires : À quoi sont destinés les résultats ? Qu'est-ce qu'une population ? Qu'est-ce que deux populations identiques ? L'auteur du livre s'intéresse-t-il à deux lots déjà fabriqués et vérifiés, ou bien à deux processus de production de parachutes ? On ne s'intéresse probablement qu'à des processus de production car les parachutes testés (et brûlés) ne serviront plus. Mais quelle est la méthode d'essai ? on n'en parle pas. On pourrait exposer le problème comme étant un problème analytique dans lequel la question est de savoir si le coût de production, le marketing et les préférences des clients justifient un choix entre les processus A et B.

Si un statisticien travaillant pour une entreprise présentait son rapport de la même façon que ce problème est énoncé, il se ferait certainement mettre à la porte.

Pour extraire l'information d'une expérience si coûteuse, je souhaiterais avoir en main les données originelles, afin de pouvoir graphiquer le temps de combustion de chaque parachute testé, dans l'ordre des essais. Les articles de A et B ont-ils été testés alternativement ? Sinon, pourquoi n'a-t-on pas procédé ainsi ? Les étudiants ont besoin d'être encouragés à réfléchir et à poser des questions sur les données, mais comment peuvent-ils le faire dans ces conditions ? Ce qui manque dans le texte, ce sont les données originelles et la description de la façon dont les tests ont été faits, ce que Shewhart appelait les « données de l'expérience ».

Dans l'enseignement traditionnel, un étudiant consciencieux qui veut réussir son examen doit abandonner tout l'esprit scientifique qu'il a accumulé jusqu'alors et apprendre le livre, les erreurs avec tout le reste.

Exemple 2. Un jour, à l'Université du Wyoming, une personne du département de biologie annonça à mon ami le Dr. Bryant une découverte stupéfiante : les fibres de laine du côté droit et du côté gauche du mouton n'avaient pas le même diamètre. Le Dr. Bryant lui demanda combien de fibres il y avait dans l'échantillon. Réponse : 50.000. Ce nombre était assez grand pour établir la signification. Et alors ? Tout le monde sait à l'avance, sans dépenser un dollar, qu'il y a une différence entre les fibres de laine du côté droit et du côté gauche d'un mouton, ou d'un groupe de n moutons. La question est de savoir si la différence a une importance scientifique.

Exemple 3. Au cours d'une conférence de l'*American Statistical Association*, à Montréal en août 1972, on a distribué une brochure dans laquelle on pouvait lire la phrase suivante : « Une difficulté

fondamentale pour la plupart des étudiants est la bonne formulation des alternatives H_0 et H_1 pour un problème donné et la détermination des régions critiques qui en résultent (queues de distribution). Ici, l'hypothèse H_0 est que $\mu_1 = \mu_0$. »

Commentaire : Il ne faut pas s'étonner que les étudiants soient angoissés. Ils cherchent de temps en temps à réfléchir.

Exemple 4. Lu dans une revue de statistique. Les garçons de 12 à 17 ans ont généralement une meilleure acuité visuelle (binoculaire, de loin, sans correction) que les filles du même âge dans les quatre régions du pays. Cependant, c'est seulement dans le Midwest et le Sud que l'on a trouvé des différences assez grandes pour avoir une signification statistique.

Commentaires :

a) Les différences entre les garçons et les filles semblent persistantes dans les régions et présentent une véritable importance scientifique. Une nouvelle étude serait utile.

b) L'étude des chiffres détaillés (publiés par ailleurs) me semble en contradiction avec la conclusion énoncée plus haut. Pour 1.000 individus, dans cette tranche d'âge, il y a plus de filles que de garçons dont la vision est 20/17 et même 20/20, mais plus de garçons que de filles dont la vision est 20/15 et 20/12 ou même mieux.

c) Ce qui apparaît comme un pourcentage plus élevé de garçons dans le tableau relatif à une vision supérieure ou égale à 20/20 provient de la consolidation fortuite de quantités disproportionnées de différents âges et de différents degrés de vision.

d) Cette disproportion entre différents âges et de différents degrés de vision est probablement le résultat le plus important de l'étude, mais le texte ne retient pas cette possibilité.

e) La forte proportion de garçons et de filles avec une vision supérieure ou égale à 20/20 dans le Sud, comparée au reste du pays, a peut-être pour origine des différences entre la vision des noirs et des blancs, garçons et filles. Les tableaux ne donnent pas les chiffres par couleur, probablement parce que les échantillons de noirs sont faibles.

f) D'après mon expérience, les différences entre les examinateurs mériteraient d'être étudiées, mais le texte ne donne aucune indication sur la façon dont les garçons et les filles ont été répartis vers les examinateurs ni sur les différences entre les examinateurs.

g) Les écarts-type qui figurent sur le tableau n'ont aucun sens. Il me semble qu'ils ne font qu'obscurcir la compréhension du lecteur.

Autres idées sur l'enseignement des statistiques

Il y a peu de progrès à attendre dans l'enseignement des statistiques, et peu d'espoir pour que les méthodes statistiques soient utiles dans les problèmes redoutables auxquels l'homme est confronté aujourd'hui, tant que la littérature et les salles de classe ne se débarrasseront pas de termes aussi nuisibles pour la recherche scientifique que : « hypothèse nulle », « population » (au lieu de cadre), « valeur vraie », « niveau de signification » (pour comparer des traitements), « échantillon représentatif ». Il n'existe pas de valeur vraie d'une caractéristique mesurée. Il peut y avoir, bien sûr, une définition opérationnelle qui est acceptée et une valeur expérimentale qui est acceptée tant qu'elle n'est pas remplacée par une autre valeur jugée plus acceptable par les experts du domaine considéré.

Voici quelques suggestions pour remplacer les sujets qu'il faut éliminer. (1) Faire détecter et mesurer par des contrôles statistiques les erreurs qui ne sont pas dues à l'échantillonnage et évaluer leurs effets possibles sur l'utilisation des résultats. (2) Insister sur les différences entre les études énumératives et les études analytiques, les différences dans la préparation et l'analyse. Ceci entraînera automatiquement un usage plus efficace de l'échantillonnage. (3) Chaque étudiant devrait s'exercer à la déduction statistique à partir d'un ensemble de données originelles mélangées aux informations non-statistiques nécessaires sur les conditions d'environnement et les conditions d'expérience. Des contrôles croisés entre les étudiants d'une même classe les inciteraient à une certaine circonspection en la matière.

Les étudiants en statistiques ont besoin de professeurs engagés dans la pratique. Que se passerait-il en médecine si les étudiants apprenaient la chirurgie auprès de médecins qui ne la pratiquent pas ?

Un professeur qui est concerné par des problèmes statistiques a des bases solides pour faire un choix sur la théorie qu'il va enseigner. Il comprend parfaitement bien les exemples qu'il donne dans ses cours et dans ses livres, puisque ce sont les siens. Son enseignement porte les étudiants à réfléchir.

Nous avons aussi besoin de professeurs de théorie qui ne sont pas engagés dans la pratique. Un étudiant peut parfaitement n'étudier que la théorie. Cependant, les étudiants qui étudient la théorie ont besoin d'exemples de bonnes et de mauvaises pratiques avec des explications - en termes de théorie - de ce qui était juste et de ce qui était faux dans la façon de procéder. Il est dommage qu'un bon enseignement des statistiques soit trop souvent ruiné par des exemples de plans et de déductions qui égarent les étudiants, comme j'ai essayé de l'expliquer ici.

Le fait d'être impliqué dans un problème donne des responsabilités. Le statisticien qui travaille sur un problème pratique doit écrire un rapport sur la fiabilité statistique des résultats, en disant ce qu'ils signifient et ce qu'ils ne signifient pas. Son rapport est destiné au management d'une entreprise ou à l'instruction d'une affaire. Il doit indiquer les marges d'incertitude provenant de l'échantillonnage, les maladresses détectées par les contrôles, les chiffres illisibles, les chiffres manquants, les incohérences trouvées dans le codage. Il doit expliquer avec précision les méthodes de mesure, les questionnaires utilisés, et les difficultés rencontrées. Celui qui est impliqué dans un problème peut avoir affaire à un conseil d'administration ou à une audition contradictoire. Il fera des travaux fastidieux ; il étudiera des données sous différentes formes ; il fera et refera des tableaux et des graphiques ; il tentera de préserver les témoignages dans les résultats tout en étant assez clair pour le lecteur ; il connaîtra des phases de déception et de découragement. Mais quelques conseils donnés par-ci par-là sur la façon d'aborder une intégrale double ne constituent pas une implication dans un problème.