

REVUE DE STATISTIQUE APPLIQUÉE

A. VESSEREAU

Statistique et langage Dacosta

Revue de statistique appliquée, tome 25, n° 2 (1977), p. 83-87

http://www.numdam.org/item?id=RSA_1977__25_2_83_0

© Société française de statistique, 1977, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

STATISTIQUE ET LANGAGE DACOSTA

A. VESSEREAU

A Isidore IZOU
inventeur du lettrisme

S'agit-il d'un nouveau langage pour informaticiens — comme ALGOL, BASIC, FORTRAN, . . . ? Pas tout à fait, malgré quelques analogies.

DACOSTA. — Définition Abrégée d'un Concept Statistique.

J'ai fait allusion dans un précédent article (1), à ces curieux néologismes — pas toujours prononçables — qu'on voit apparaître de plus en plus dans la littérature statistique (anglo-saxonne surtout), et qui résultent de la compression dans un même vocable de l'ensemble des termes qui constituent une définition complète — opération un peu comparable à celle que réalise le sculpteur César à partir de plusieurs automobiles hors d'usage. Un exemple typique — choisi parmi les plus anciens de ces "digests" — est ANOVA : analysis of variance.

Le sujet mérite-t-il qu'on lui consacre un article ? Nous avons pensé qu'il méritait une courte note. De toute façon, et bien que cela ne soit pas explicitement écrit dans la R.S.A., (parce que cela va de soi), les opinions exprimées n'engagent que la responsabilité de leur auteur.

Imitant le dialogue du Juste et de l'Injuste, nous présenterons d'abord l'accusation, puis la défense. Un petit lexique (bien incomplet) de DACOSTA suivra.

L'ACCUSATION

Comme dans beaucoup de maladies, les premiers symptômes n'ont pas été très alarmants. L'un des premiers DACOSTA (dans la suite, je donnerai à ce terme le genre neutre) a sans doute été BIB (balanced incomplete block) : cela se prononce aisément et le résultat est assez amusant. ANOVA, qui fait penser à quelque personnage de la mythologie indoue, figure aussi parmi les ancêtres. Mais certains auteurs ont pensé que le taux de compression pouvait encore être amélioré. Au sommaire d'un récent numéro d'une très sérieuse revue statistique, on trouve deux articles dont les titres comportent :

(1) Statistique et Français. Revue de Statistique Appliquée XXV, n° 1, 1977.

..... random one-way ANOVA model

..... procedures for the two-way AOV model

On pourrait encore aller plus loin dans l'économie des moyens : AV suffirait.

Economie des moyens . . . Un écologiste pourra calculer que la réduction à deux ou trois caractères des 20 signes de "analysis of variance", et la généralisation de ce procédé, économiserait des tonnes de papier, donc protégerait des hectares de forêts. Un économiste évaluera l'économie dans les importations d'énergie qui résulteraient de la réduction des surfaces imprimées. Ce sont, de toute évidence, de bien médiocres arguments. Protection de la nature et économie d'énergie peuvent être réalisées sans qu'il soit porté atteinte au langage, sans que le signifié se présente sous la forme d'un rebus par rapport au signifiant.

Car le DACOSTA a pour conséquence une imprécision qui peut entraîner de véritables contresens. En voici un exemple, emprunté au contrôle statistique de la qualité. Dans AQL (acceptable quality level), la lettre L désigne un niveau (*level*), dans AOQL (average outgoing quality limit) une *limite*, de même que dans LTPD (limit tolerance percent defective) ; dans AQL, la lettre A veut dire *acceptable*, dans AOQL elle représente une moyenne (*average*). Si de telles confusions peuvent se produire dans un domaine étroit des applications de la statistique, le risque nous paraît encore plus grand dans la désignation de concepts généraux. La lettre M représente tantôt une moyenne (MS = mean square), tantôt un maximum (ML = maximum of likelihood), parfois un minimum.

Le DACOSTA étant conçu dans une langue particulière (l'anglais en général) comment lui trouver un correspondant dans une autre langue : le français en l'occurrence ? La solution la plus simple est la transcription telle quelle. Mais elle peut conduire à un curieux mélange, tel que le suivant, où l'expression française complète voisine avec un condensé anglais : "... on procède à l'estimation des paramètres par la méthode du maximum de vraisemblance, on trouve facilement, comme expression des MLE ...". L'autre solution consiste à créer un DACOSTA équivalent, que l'on pourrait trouver dans un dictionnaire bilingue, ou multilingue, des DACOSTA. Dans le domaine des institutions internationales, il y a de tels exemples : NATO et OTAN, WHO et OMS, ... — mais le problème n'a pas été résolu pour la FAO(1). En matière de statistique il y a aussi des exemples. Le fascicule de documentation X 05-030 de l'Association Française de Normalisation donne quelques équivalents français de DACOSTA anglais utilisés en contrôle statistique de la qualité : ainsi AQL (acceptable quality level) devient NQA (niveau de qualité acceptable). Les spécialistes de la fiabilité ont réalisé le tour de force d'imaginer un DACOSTA parfaitement bilingue : MTBF (mean time between failure) = MTBF (moyenne des temps de bon fonctionnement). Sans autant d'effort les statisticiens peuvent traduire test UMP (uniformly most powerful) par test UMP (uniformément de meilleure puissance).

Si le "principe DACO" est admis, pourquoi ne pas l'étendre à n'importe quel domaine ? L'imagination peut se donner libre cours.

(1) De même, l'Institut International de Statistique s'abrège IIS en français et ISI en anglais, tandis que l'organisation internationale de normalisation conserve en français l'abréviation anglaise ISO.

DACOMU (musique). Un disciple de Mozart, Haydn ou Beethoven intitulera les quatre mouvements d'une symphonie : ACOMO (andante con moto) – ALVI (allegro vivace) – AMATROPO (adagio ma non troppo) – FICOSPI (finale con spirito).

DACOLI (littérature). Un universitaire écrira un ouvrage, un candidat au doctorat préparera une thèse sur TITEBE (Corneille, Tite et Bérénice), ou le DUBONUMA (Pascal, Du bon usage des maladies).

Il est facile de poursuivre ce jeu de société jusqu'à la limite du mauvais goût. Imaginons par exemple le DACOSTA résumant la "minimale erreur relative des rapports estimés" ? De quoi remplir d'aise le père UBU.

LA DEFENSE

Elle reprendra point par point – en les réfutant – les arguments de l'accusation.

D'une façon très générale, DACOSTA est une abréviation, réduite souvent aux initiales de l'expression complète. Peut-on lui refuser le droit, lorsqu'elle est aisément prononçable, de devenir un nouveau mot du vocabulaire statistique ? Dans un autre domaine, RADAR et LASER n'ont pas d'autre origine.

Est-il une science, une technique, qui puisse se passer de conventions analogues ? Essayons d'imaginer ce que serait un traité de chimie si l'on n'avait pas imaginé de représenter le phosphore par P, le sodium par Na, . . . Et dans un traité de mathématiques, il est également fait appel à des signes conventionnels . . . ce qui implique évidemment l'obligation de les prononcer avec des mots du vocabulaire usuel. Il faut reconnaître que, dans ces exemples, symbole ou signe ne présentent, pour aucun lecteur, aucune ambiguïté d'interprétation.

Une même lettre, parfois un même groupe de lettres, peut avoir des sens très différents. Cela est inévitable, même si l'on fait appel à tous les alphabets existants. Mais, le contexte aidant, qui confondra la C.G.T., confédération générale du travail, et la C.G.T., compagnie générale transatlantique ? Dans la plupart des cas les auteurs prennent soin, dès le début d'un article, de définir les abréviations qu'ils utilisent par la suite : . . . "method of maximum likelihood (ML)". S'ils ne le font pas, ils ont incontestablement tort.

Le DACOSTA présente des difficultés lors de la traduction dans une autre langue ? Comme dans bien d'autres domaines, cela est inévitable, mais, en définitive, c'est l'usage qui fait la loi des parties. Lorsqu'on parle de l'UNESCO, on oublie qu'il s'agit d'initiales anglaises, et, si l'on n'a pas essayé de les transposer, sans doute est-ce parce que "cela sonne bien" – de même que sonnent également bien OTAN et NATO. Après tout, si les lecteurs français se plaignent de l'abondance des DACOSTA anglais, ils n'ont qu'à représenter par des DACOSTA français les nouveaux concepts qu'ils auront imaginés.

La généralisation du DACO ? Elle existe déjà, et nous avons cité plus haut le cas de RADAR et LASER. Faut-il citer encore, à titre d'exemple : etc. (ce qui n'empêche pas de prononcer et caetera), op. cit (opere citato) . . . etc. !

Et si les initiales conduisent en français à un résultat incongru, nous laisserons à nos amis anglais, utilisant un vestige ancien de notre langue, le soin de répondre "Honni soit qui mal y pense".

LE JUGEMENT

Il n'y aura pas ici de jugement : chaque lecteur de cet article pourra prononcer le sien. L'auteur se bornera à formuler un double souhait :

- que, dans toute publication, les abréviations (de même que les symboles) soient clairement définies, dès qu'il en est fait usage,
- que l'abus des DACOSTA ne conduise pas à transformer les textes en véritables rebus.

PETIT LEXIQUE (provisoire) des DACOSTA

ANOVA	Analysis of variance
AOV	Analysis of variance
AQL	Acceptable quality level
AR model	Autoregressive model
ASN	Average sample number
BE	Backward elimination
BIB	Incomplete balanced block
BIBD	Balanced incomplete block design
BLUE	Best linear unbiased estimate
CSP	Continuous sampling plan
ECB	Extended complete block
EPRB	Equal percent bias reducing
FE	Forward elimination
IQL	Indifference quality level
LMER	Least mean error regression
LQ	Limit quality
LTPD	Lot tolerance percent defective
MANOVA	Multivariate analysis of variance
MAR	Maximum average range
MINEMS	Minimum expected MSE
MINEV	Minimum expected variance
MINQUE	Minimum quadratic unbiased estimator
MIVQUE	Minimum variance quadratic unbiased estimator
ML	Maximum likelihood
MLE	Maximum likelihood estimator
MLUE	Maximum efficient linear unbiased estimator
MOLS	Mutually orthogonal latin square
MSD	Maximum standard deviation
MSE	Mean square error
MTBF	Mean time between failure

MV model	Moving average model
OLUMV	Ordered linear unbiased minimum variance
PBIBD	Partially balanced incomplete block design
QUE	Quadratic unbiased estimator
REML	Restricted maximum likelihood
RRM	Randomised response model
RRT	Randomised response technique
RSS	Residual sum of square (ou Royal Statistical Society suivant contexte)
SOS	Sum of squares
SPR	Sequential probability ratio
SSD	Sum of squared deviations
STP	Simultaneous test procedures
UMP	Uniformly most powerful

* * *

On trouvera une liste assez impressionnante de DACOSTA (sous le nom de "lettre-code") dans la Revue Internationale de Statistique.

Vol. 43 n° 1 pages 41-42
 " " n° 2 " 189-190