REVUE DE STATISTIQUE APPLIQUÉE

W. J. YOUDEN

Programmes d’essai pour l’évaluation statistique des matériaux

Revue de statistique appliquée, tome 10, nº 4 (1962), p. 71-75
<http://www.numdam.org/item?id=RSA_1962__10_4_71_0>

PROGRAMMES D'ESSAI
POUR L'ÉVALUATION STATISTIQUE DES MATÉRIAUX (1)

W. J. YOUDEN

RESUME

Les modestes prétentions statistiques de cet exposé et les simples exemples énoncés constituent une tentative voulue de réparer les dommages causés inconsciemment par des statisticiens amateurs par trop enthousiastes.

Quel que soit le domaine exploré, les chercheurs qui appliquent pour la première fois les techniques statistiques sont enclins à se laisser déborder par leur enthousiasme. On voit paraître des articles dans lesquels le problème étudié est noyé dans un traitement statistique touffu des données. Pis encore, des techniques statistiques alambiquées sont appliquées à des données qui s'énonceraient de façon convaincante au moyen de tableaux ou de graphiques. Il est probable que l'abus des statistiques n'incite pas les profanes à s'informer de ces question. La diffusion de quelques méthodes statistiques de base est plus bénéfique que l'usage confidentiel de techniques compliquées. Dans cet esprit, je propose que l'on s'en tienne à des statistiques aussi simples que possible et que l'on en fasse un usage modéré.

Un des problèmes auxquels le chercheur se heurte fréquemment est celui de l'évaluation de résultats préliminaires. Le chercheur se demande si ces résultats sont assez prometteurs pour justifier la poursuite du travail. Considérons une matière à laquelle une modification a été apportée dans l'espoir d'améliorer une de ses propriétés, par exemple la force, ou la résistance à l'usure. Supposons que trois échantillons ont été fabriqués avec la matière habituelle (C), et trois avec la matière modifiée (T). Quelle valeur faut-il accorder, en tant que preuve d'amélioration, au fait que chacun des trois échantillons $\underline{T}$ est supérieur aux trois échantillons $\underline{C}$? Ecrivons T sur trois petits papiers et C sur trois autres, et mélangeons le tout dans un chapeau. Combien de chances y-a-t-il pour que les trois premiers papiers soient tous des T ? Une chance sur vingt (3/6 ; 2/5 ; 1/4 = 1/20). Tous les papiers ont une chance égale d'être tirés. Il y a donc une chance sur vingt pour que l'on obtienne les résultats observés lors de l'expérimentation si la matière $\underline{T}$ n'est pas meilleure que la matière $\underline{C}$. Cette chance est relative-

----

(1) Communication présentée au Séminaire sur les applications industrielles de la Statistique - Paris - Septembre 1961.

ment faible et de nature à encourager l'expérimentateur à poursuivre ses essais.

Une autre situation se présente lors de la comparaison simultanée de plusieurs matières par rapport à un témoin. Soient les matières A, B, D, E et F et le témoin C. Très souvent, une telle situation est de nature à rendre particulièrement avantageuse une comparaison "par paires", (par exemple entre deux dents situées à des emplacements analogues). La tradition veut que l'expérimentateur fasse des essais sur les cinq paires CA, CB, CD, CE et CF. Supposons que chaque paire soit soumise à trois essais. Est-ce là le meilleur usage que l'on puisse faire de 15 comparaisons par paires ? Je ne le crois pas. Mieux vaut essayer une fois chacune des 15 paires qui peuvent être formées avec les six lettres. Ces paires sont :

AB	AC	AD	AE	AF
BC	BD	BE	BF	CD
CE	CF	DE	DF	EF

Plusieurs raisons peuvent être mises en avant en faveur de ce dernier choix. Avec cinq paires, la moitié de l'effort expérimental est consacré à C, et un dixième à chacune des matières essayées. Avec 15 paires, les six matières interviennent de façon parfaitement symétrique et un sixième de l'effort porte sur chacune. Chaque matière d'essai est présente cinq fois au lieu de trois. Il n'y a pas de raison valable pour que l'on connaisse le témoin avec plus de précision qu'aucune des matières d'essai. De plus, les comparaisons entre matières sont effectuées sur un pied d'égalité avec celles du témoin à une autre matière, tandis qu'avec cinq paires seulement, le témoin seul intervient comme base de comparaison. L'analyse des données est facile.

D'ailleurs il n'est pas nécessaire d'utiliser la totalité des 15 paires. Les paires AF, BE et CD, par exemple peuvent être laissées de côté. Les neuf paires AB, BC, CD, DE, EF, FA, AD, BE, CF peuvent aussi être utilisées pour un essai préliminaire. Chaque matière apparaît trois fois dans ces neuf paires, tandis que dans le schéma habituel, avec dix paires, chacune des matières d'essai n'apparaît que deux fois.

Il arrive que les données obtenues ne soient pas quantitatives. Malgré cela, une discrimination préférentielle peut être basée sur des comparaisons par paires du type qui vient d'être décrit. On procédera à quinze choix, un pour chacune des quinze paires dans le cas où six matières sont à l'étude. Le problème de la classification des matières se complique de contradictions telles que celle-ci :

A est choisi de préférence à B

B " " " C

C " " " A, et non A de préférence à C.

Inscrivons les six lettres dans n'importe quel ordre de préférence arbitraire tel que D, A, C, E, F, B. Cet ordre détermine quinze préférences telles que D par rapport à A, D par rapport à C, F par rapport à B, etc.. Si l'on compare ces choix aux quinze choix découlant réellement des essais, on constatera un certain nombre de désacords entre les choix résultant de l'ordre arbitraire et les choix effectivement faits. Il existe certainement un ordre d'inscription des lettres qui minimise le nombre de ces désaccords; cet ordre représente la meilleure façon de classer les matières à la suite des expériences.

Il existe une technique qui permet de trouver lequel des $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ arrangements possibles des six lettres entraîne le minimum de désaccords avec les 15 choix effectués. Cette technique a été publiée par M.G. Kendall dans Biometrics, vol. XI, page 43 en 1955. Les tableaux 1 à 3 représentent les opérations auxquelles les 15 choix sont soumis dans les résultats ci-après, où la première lettre de chaque paire représente la la matière victorieuse dans la comparaison correspondante :

AB AC DA AE AF CB BD BE FB CD CE CF ED FD EF.

Le premier tableau montre que l'on attribue 2 à une lettre par victoire, et 0 par défaite ; la cote est de 1 pour les "comparaisons identiques" comme AA, BB, etc.. Les totaux des cotes par rangée fournissent une classification provisoire. Cette classification est révisée à partir de l'idée qu'une victoire sur une matière hautement cotée a plus de valeur que sur une matière n'ayant que peu de points. Si A l'emporte sur C, côté 9 points, cela doit avoir plus de valeur que la victoire remportée sur B qui n'a que 5 points. Le second tableau est donc établi en multipliant les cotes inscrites au premier par les cotes des lettres qui accompagnent les perdantes. Ceci amène à une nouvelle série de totaux par rangée et à une classification plus exacte des matières concurrentes. Le processus est répété à partir de ces nouvelles cotes. Il peut en résulter une classification stable avec un minimum de désaccords.

Nous donnerons un autre exemple de technique rapide applicable à un petit nombre de résultats. Devant une situation entièrement nouvelle, les premiers essais ont naturellement un intérêt immédiat. La raison qui conduit habituellement à effectuer un nombre considérable d'essais est que l'on veut obtenir une moyenne qui subirait très peu de changement si l'on poursuivait les essais. Il est quelquefois utile d'établir, à partir des données disponibles, des limites qui ont une probabilité donnée de contenir la moyenne de nombreux essais. Un procédé simple consiste à prendre la différence $\underline{w}$ entre le plus grand résultat et le plus petit. Cette différence, multipliée par un coefficient convenable, est ajoutée au résultat le plus grand et soustraite du plus petit. Le coefficient dépend du nombre de résultats, et celui que nous indiquons ici correspond à un niveau de probabilité de 95 %, pour 2, 3, 4 et 5 résultats.

Nombre de résultats	2	3	4	5
Coefficient de $\underline{w}$	6,0	1,0	0,25	0

Exemple - On a obtenu trois résultats, 1,05 ; 1,16 et 1,08. La valeur de $\underline{w}$ est 1,16 - 1,05 = 0,11. Puisqu'il n'y a que trois résultats, le coefficient est l'unité. Le résultat le plus important plus 0,11 donne 1,27 ; le plus petit moins 0,11 donne 0,94. Dix-neuf fois sur vingt, les limites calculées de cette manière comprendront la moyenne qui aurait été obtenue sur un grand nombre de résultats. L'intervalle est large, parce qu'il n'y a qu'un petit nombre d'essais ; il se rétrecit rapidement à mesure que le nombre d'essais augmente.

En conclusion, il faut noter que certaines des idées les plus intéressantes en matière de statistique concernent le plan du programme d'essais. C'est dans cette intention qu'ont été mentionnées les comparaisons par paires et le choix des paires à essayer. Un programme bien conçu peut n'exiger qu'un minimum d'analyse statistique. Une analyse statistique élaborée montre parfois que certaines différences très petites sont "statistiquement significatives". Il ne s'ensuit pas que ces différences minimes aient une importance pratique : c'est là encore au chercheur qu'incombe la responsabilité de la conclusion.

Tableau 1

(Pour trouver la classification la moins illogique par rapport aux données ; Gagnant : 2 ; Perdant : 0 ; Egal : 1).

Gagnants \ Perdants	A	B	C	D	E	F	Total	Rang
A	1	2	2	0	2	2	9	I
B	0	1	0	2	2	0	5	II
C	0	2	1	2	2	2	9	I
D	2	0	0	1	0	0	3	III
E	0	0	0	2	1	2	5	II
F	0	2	0	2	0	1	5	II

Tableau 2

(Répéter le processus en se servant des totaux qui viennent d'être obtenus).

Gagnants \ Perdants	A	B	C	D	E	F	Total	Rang
A	9	10	18	0	10	10	57	I
B	0	5	0	6	10	0	21	III
C	0	10	9	6	10	10	45	II
D	18	0	0	3	0	0	21	III
E	0	0	0	6	5	10	21	III
F	0	10	0	6	0	5	21	III

Tableau 3

(Rang après chaque répétition)

Matière	1er	2e	3e	Final
A	I	I	I	I
B	II	III	IV	IV
C	I	II	II	II
D	III	III	III	III
E	II	III	IV	IV
F	II	III	IV	IV