REVUE DE STATISTIQUE APPLIQUÉE

F. BASTENAIRE

Sur l'efficacité des plans statistiques d'expérimentation

Revue de statistique appliquée, tome 3, nº 4 (1955), p. 81-86
<http://www.numdam.org/item?id=RSA_1955__3_4_81_0>

SUR L'EFFICACITÉ DES PLANS STATISTIQUES D'EXPÉRIMENTATION

par

F. BASTENAIRE

Ingénieur à l'Institut de Recherches de la Sidérurgie Française

Des controverses s'élèvent assez fréquemment entre techniciens et statisticiens, au sujet du nombre des essais qu'il faudrait effectuer pour comparer valablement deux populations de mesures.

Les techniciens connaissent intuitivement, par l'expérience, les variabilités des grandeurs qu'ils utilisent, mais la mémoire humaine, apte à ne retenir que les événements les plus fréquents, les conduit souvent à des évaluations trop optimistes.

Les statisticiens, au contraire, gens prudents par nature, quoi qu'on en dise, inspectent d'abord minutieusement leurs données, mais, cela fait, ne font confiance qu'à leurs tables.

Au moment d'établir un plan d'expériences, d'assez fortes divergences peuvent donc apparaître entre les uns et les autres, qu'il n'est pas toujours facile de résoudre.

Pour comprendre les deux points de vue et en faire, si possible, la synthèse, nous avons tenté de les analyser. Nous croyons avoir mené notre recherche sans idée préconçue.

Nous aurions pu raisonner d'un bout à l'autre dans l'abstrait, mais il nous a paru préférable de nous appuyer sur l'exemple même qui nous avait conduit à ces réflexions : la comparaison des « mises au mille » de coke dans les hauts fourneaux.

Nous nous appuyons sur un cas où existent à la fois des erreurs intrinsèques et des erreurs de mesure, mais cette condition n'est pas nécessaire pour nos conclusions.

I. - GÉNÉRALITÉS SUR LES FLUCTUATIONS DES CARACTÉRISTIQUES DE MARCHE DES HAUTS FOURNEAUX

Les compositions chimiques et les températures des produits qui entrent et qui sortent d'un haut-fourneau subissent en pratique de telles fluctuations que les réactions de cet appareil à des variations des conditions à l'entrée ne nous sont encore qu'imparfaitement connues, surtout quantitativement.

Sur de longues périodes, ces fluctuations se compensent en grande partie et les caractéristiques moyennes que l'on a coutume de calculer sont beaucoup plus stables bien que sujettes à des variations non négligeables.

Rien n'est plus simple que de déterminer une mise au mille (*) de coke sec ou de carbone mais les erreurs qui les affectent sont de trois sortes :

1° Les erreurs de pesée.

Lorsque les pesées sont bien faites avec contrôle périodique du zéro de la bascule, ces erreurs sont faibles mais il faut se souvenir que ces précautions sont rarement observées et que le coke est souvent chargé au volume.

2° Les erreurs de détermination d'humidité pour le coke sec et du taux de cendres pour le carbone.

(*) Le haut-fourneau est l'appareil sidérurgique destiné à produire la fonte. La "mise au mille" est le terme technique par lequel on désigne la quantité de coke (en kilogrammes) qu'il faut pour produire une tonne de fonte.

Elles résultent à la fois de l'échantillonnage et de l'analyse proprement dite, la part de l'échantillonnage étant souvent très grande.

3° Les erreurs de définition sur les entrées et les sorties en début et fin de période.

En effet, le coke qui a été enfourné ne correspond pas exactement à la fonte produite du fait du temps de passage et il faut décaler les entrées par rapport aux sorties d'une quantité qui n'est pas exactement connue.

L'importance de ces trois types d'erreurs décroît lorsque croît la longueur de la période sur laquelle on calcule la mise au mille à condition que ces erreurs soient purement aléatoires. Au contraire, les erreurs systématiques d'analyse et de pesée ne décroissent pas quel que soit le nombre des mesures effectuées.

Il serait relativement facile de déterminer a priori les variances des erreurs accidentelles dues aux pesées et aux analyses. Il suffirait pour cela d'effectuer des mesures en double (pesées, analyses, échantillonnage, etc...). En négligeant l'erreur de définition, on pourrait alors calculer l'intervalle de confiance d'une mise au mille et déclarer par exemple qu'elle se trouve avec une probabilité de 0,95 dans l'intervalle 900 - 980 kilogs.

Bien que digne d'intérêt en soi, cette variance d'erreur de la mise au mille n'est pourtant d'aucune utilité pour comparer des marches avec des minerais différents ou des cokes différents parce que l'affirmation que cette mise au mille est comprise entre 900 et 980 kgs avec une probabilité 0,95 n'est valable **que pour la période d'essai révolue seulement.**

En effet, les mises au mille relatives à des périodes successives et différentes ne varient pas seulement à cause des erreurs de mesure et d'échantillonnage mais, entre autres raisons :

1° Parce que les irrégularités de la marche ne se compensent qu'en partie. La preuve en est que les techniciens proposent toujours de plus longues périodes d'essais.

2° Parce que le minerai et le coke ont pu évoluer au point de vue composition et caractéristiques physiques d'une période à une autre (Exploitation de nouvelles couches, etc...).

3° Parce que les conditions météorologiques ont changé (changements de saison).

4° Parce que les erreurs systématiques d'analyse et de pesée ont également pu varier.

Etc....

Des conditions de marche différentes correspondant nécessairement à des périodes d'essais distinctes, les performances enregistrées seront nécessairement entachées d'erreurs de ce type.

Il n'y a donc pas d'autre moyen pour estimer la variance des erreurs **réellement commises** que de se baser sur des résultats relatifs à des périodes **différentes.**

II. - DÉTERMINATION DE LA VARIANCE D'ERREUR RÉELLE.

L'ordre de grandeur de cette variance pourrait être calculé d'après des résultats de marche antérieurs, établis sur des périodes de même longueur, pour le ou les fourneaux qui doivent servir à faire une expérience.

Une telle façon de procéder repose le problème de l'usage de ce que l'on appelle en statistique : l'information a priori.

On aborde rarement une expérience sans avoir quelques notions préalables sur les ordres de grandeur des résultats et sur leurs précisions. Cependant, ces connaissances sont vagues et presque toujours difficiles à exprimer en termes objectifs.

Depuis que Th. BAYES a énoncé son célèbre théorème sur la probabilité des causes, c'est-à-dire depuis cent ans environ, statisticiens et probabilistes sont, en gros, partagés en deux camps. Les uns persistent à employer des lois de probabilité a priori auxquelles ils attribuent en fait les formes mathématiques qui conviennent le mieux à leurs calculs. Les autres sautent ces difficultés en supprimant l'emploi de l'information à priori.

C'est R.A. FISHER qui s'est l'un des premiers (après STUDENT) engagé dans cette voie, cherchant à tirer en termes de probabilité des conclusions ne reposant que sur les résultats de **l'expérience faite** et négligeant l'information préalable presque toujours impossible à exprimer mathématiquement.

La méthode de FISHER consiste à comparer l'importance des effets obtenus dans une expérience à la variance de l'erreur telle qu'elle peut être estimée **d'après les résultats de la même expérience.**

En opérant ainsi, on perd le bénéfice de l'information a priori qui n'est jamais nulle.

Cette perte est en réalité minime dans la majorité des cas car l'expérience, dès qu'elle porte sur un nombre suffisant d'individus, apporte une information beaucoup plus considérable. En d'autres termes, les conclusions que l'on pourrait tirer en utilisant en plus l'information provenant de lois de probabilité a priori (d'ailleurs subjectives en général), ne seraient guère plus précises que celles que l'on obtient par la méthode fisherienne purement objective.

Cela n'est toutefois pas vrai lorsque le nombre des observations est très petit car l'estimation de la variance de l'erreur est, dans ce cas là, si mauvaise que l'information a priori peut être supérieure à celle qui provient de l'expérience.

Ainsi, en désignant par s^2 une estimation de variance basée sur un seul "degré de liberté" (ce qui équivaut à deux observations répétées dans les mêmes conditions), l'intervalle de confiance à 0,95 de la vraie variance σ^2 est donné par :

$$\frac{s^2}{\chi^2_{0,975}} < \sigma^2 < \frac{s^2}{\chi^2_{0,025}}$$

c'est-à-dire :

$$\boxed{0{,}44\ s < \sigma < 32\ s}$$

Si, par exemple, on trouvait pour le s d'une mise au mille la valeur de 25 kgs, les limites de l'intervalle de confiance de σ seraient dans ce cas : 11 kgs, 800 kgs. Celui qui est au courant de la technique, quel que soit ce qu'il pense de l'imprécision des mesures sait qu'elle ne peut pas atteindre 800 kgs.

L'information fournie par une telle expérience interprétée par les méthodes classiques de la statistique est donc très faible. En général, on est déjà a priori mieux renseigné sur l'erreur que par une telle estimation.

Il est facile de comprendre maintenant pourquoi la statistique nous donne, sur la validité des expériences, des résultats si contraires à l'intuition et même au bon sens dans des cas extrêmes : c'est que la méthode fisherienne classique veut ignorer toute information préalable et ne prendre en considération que celle qui provient de la seule expérience que l'on se propose d'interpréter.

Pour en revenir à l'expérimentation sur le haut fourneau, cela signifie que cette méthode ne peut donner de résultats intéressants que si le nombre des observations est suffisant pour que l'information apportée par l'expérience elle-même dépasse largement celle que l'on possède déjà.

III. - EFFICACITÉ DES PLANS STATISTIQUES.

En supposant que l'on essaye deux cokes, leur différence fournit dans un plan statistique une somme de carrés qui, avec un degré de liberté, peut être comparée à la variance résiduelle.

Il est intéressant de se demander avec quel nombre de degrés de liberté il faudrait connaître l'erreur pour que l'information a priori soit négligeable.

Nous envisagerons deux cas :

1° Nous supposerons la variance de l'erreur connue de façon certaine d'après des résultats antérieurs. C'est ce que l'on peut imaginer de plus parfait quant à l'information a priori. Cette variance n'est connue, en pratique, que de façon approximative et dans la majorité des cas, on ne sait pas si elle sera, dans l'expérience que l'on va faire, égale à celle que l'on a trouvée antérieurement (1).

2° Nous comparerons à ce premier cas celui où la variance est connue d'après l'expérience elle-même.

Les propriétés des tests statistiques sont, on le sait, définies par les valeurs de deux risques :

a) Le risque de trouver une différence significative alors qu'il n'y a pas de différence réelle. On le fixe assez souvent à 5 %.

b) Le risque, s'il y a une différence et pour un premier risque donné, de ne pas détecter un effet qui existe réellement.

Considérons une expérience visant à comparer deux cokes et pouvant éventuellement servir à étudier en même temps les effets d'autres facteurs (Expérience factorielle par exemple).

Soit r le nombre d'observations faites avec chacun des deux cokes dans des conditions qui peuvent être variables vis-à-vis des autres facteurs.

Soit d le nombre des degrés de liberté de la variance d'erreur; d peut varier selon le plan d'expérience et le nombre des facteurs étudiés.

Dans une expérience à blocs randomisés, d atteint la valeur maximum $2(r-1)$. Dans une expérience factorielle 2^n, d est fonction du nombre des interactions non significatives. Sauf pour des raisons de géométrie, d peut être considéré comme indépendant de r et l'on peut étudier l'efficacité du plan statistique en fonction de r et d.

Si Δ désigne la différence réelle des mises au mille moyennes des deux cokes et σ l'écart-type de l'erreur réelle, on démontre que le second risque ne dépend dans ce cas (2) que :

a) du premier risque
b) de d
c) de la quantité $\Phi = \frac{\Delta}{\sigma}\sqrt{\frac{r}{2}}$

Inversement, pour un premier et un second risque donnés, le choix de d détermine Φ.

Si l'on se fixe en plus la valeur de r, c'est alors $\frac{\Delta}{\sigma}$ qui se trouve déterminé. Ce rapport représente la différence des deux moyennes à comparer, mesurée en écarts-type.

Le tableau à double entrée ci-contre contient les valeurs de $\frac{\Delta}{\sigma}$ associées à diverses valeurs de d, r de telle façon que l'on ait :

- un premier risque égal à 5 %
- un second risque égal à 20 %

Exemples d'utilisation du tableau.

Supposons qu'un plan d'expérience soit tel que quatre observations soient effectuées sur chacun de deux cokes et que la variance résiduelle soit basée sur six degrés de liberté.

---

(1) C'est pour cette raison que cette façon de procéder est abandonnée des statisticiens d'aujourd'hui.

(2) Cf. O. Kempthorne. The design and analysis of experiments. Wiley and Sons. 1952, p. 217.

Pour r = 4 et d = 6, le tableau indique pour $\frac{\Delta}{\sigma}$ la valeur 1,68.

Cela signifie que, sous l'hypothèse où la différence moyenne réelle entre les deux cokes s'élève à 1,68 fois l'écart type σ de l'erreur, il y a encore 20 chances sur 100 (valeur du second risque) de ne pas conclure à l'existence de cette différence quand le test employé est tel que le premier risque soit fixé à 5 % (Risque de conclure à l'existence d'une différence alors qu'il n'y en a pas).

Les chiffres du tableau, multipliés par la valeur de l'écart type σ, sont donc des indices de la sensibilité du plan d'expérience.

d \ r	2	4	6	8	10	20	60	∞
1	11,5	8,14	6,63	5,74	5,13	3,63	2,1	0
2	3,90	2,76	2,25	1,95	1,74	1,23	0,71	"
4		1,88	1,53	1,33	1,19	0,84	0,49	"
6		1,68	1,37	1,19	1,07	0,75	0,44	"
8			1,31	1,13	1,01	0,72	0,41	"
10			1,27	1,10	0,98	0,69	0,40	"
14				1,07	0,95	0,67	0,39	"
18					0,94	0,66	0,38	"
38						0,64	0,37	"
∞	1,98	1,40	1,15	0,99	0,89	0,63	0,36	0

La dernière ligne de ce tableau correspond au cas où la variance d'erreur est supposée rigoureusement connue a priori (ce qui équivaut à d = ∞).

Si on la compare aux autres lignes, on voit que l'avantage de connaître a priori la variance n'est grand que pour les petites valeurs de r et d.

Par exemple, pour r = 2, d = 1 : Δ / σ = 11,5
tandis que pour r = 2, d = ∞ : Δ / σ = 1,98

la différence de sensibilité est donc considérable.

Par contre, pour r = 4, d = 6 : Δ / σ = 1,68
et pour r = 4, d = ∞ : Δ / σ = 1,40

la différence est peu sensible.

Ceci montre que le point de vue fisherien se justifie d'autant mieux que d est plus grand, c'est-à-dire que les observations sont plus nombreuses.

IV. - APPLICATION NUMÉRIQUE.

On est, en fait, assez mal renseigné sur la valeur de σ pour les mises au mille mais, pour apprécier la validité d'un plan statistique, il suffit d'en connaître l'ordre de grandeur.

D'après quelques données obtenues à l'IRSID, nous avons trouvé :

$$s = 20{,}6 \text{ kg/tf.}$$

Si l'on adopte ce chiffre, on voit en se référant au tableau précédent, qu'avec une expérience où r = 4, d = 6, la différence de mise au mille que l'on a des chances de pouvoir mettre en évidence par la méthode statistique classique doit atteindre 35 kg de coke par tonne de fonte.

Il n'est pas moins important de noter que si l'on supposait la variance d'erreur parfaitement connue a priori ce qui, répétons-le, est contraire aux principes essentiels de l'expérimentation statistique, on trouverait quand même $\Delta \simeq$ 28 kg (Dernière ligne du tableau : r = 4, d = ∞).

V. - RÉSUMÉ ET CONCLUSIONS.

Dans le paragraphe I, nous avons insisté sur le fait que les erreurs qui affectent les caractéristiques de marche d'un fourneau proviennent de sources diverses et ne peuvent pas être calculées a priori.

Dans le second paragraphe, nous avons confronté deux méthodes de calcul de l'erreur réelle.

Dans le troisième paragraphe, nous avons présenté un tableau de l'efficacité de différents plans statistiques.

Deux méthodes d'analyse statistique des résultats des plans d'expériences peuvent être envisagées :

1° La méthode classique rigoureuse.

2° Une méthode plus critiquable supposant la variance d'erreur connue a priori; on trouve que l'avantage de cette méthode par rapport à la première décroît lorsque croît le nombre des essais, au point de devenir négligeable pour quatre à six répétitions et à peu près autant de degrés de liberté pour estimer la variance d'erreur.

Les expériences doivent donc être envisagées de telle sorte que le nombre des unités expérimentales soit adapté à l'amplitude des effets à déceler.

Enfin, il n'est pas inutile de rappeler qu'il est plus économique d'étudier simultanément les effets de plusieurs facteurs en un seul plan d'expérience que d'étudier successivement les effets de chacun de ces facteurs dans des expériences séparées.