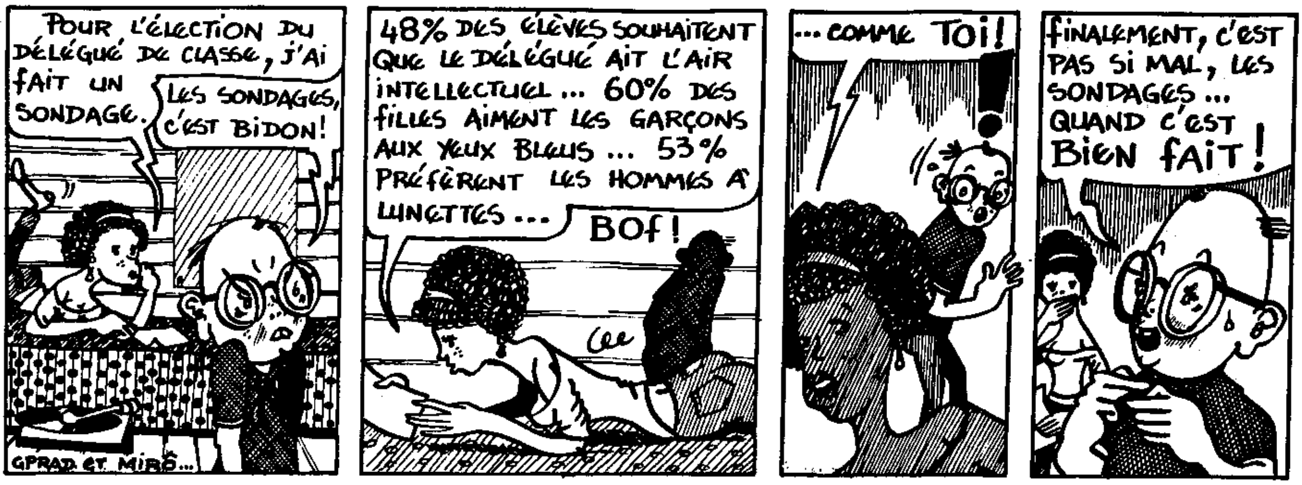

Quand un concept mathématique sert à MANIPULER L'OPINION

Albert JACQUARD

Pour les fervents de l'analyse factorielle, l'analyse de correspondance, l'analyse multivariée, ... , chères aux spécialistes de l'évaluation voici un texte d'Albert JACQUARD, directeur du département génétique de l'Institut National d'Etudes Démographiques, trouvé dans un livre publié par les éditions complexes : "Le genre humain".

Quel psychologue, quel sociologue, quel biologiste n'a pas, un jour ou l'autre, calculé un "coefficient de corrélation" ? Deux séries de mesures sont "corrélées" lorsqu'une certaine ressemblance apparaît entre elles, c'est-à-dire lorsque la connaissance d'une mesure apporte une information sur l'autre. Ainsi, mesurons sur un ensemble de squelettes les longueurs du fémur et de l'humérus ; il est fort probable que nous constaterons que les humérus les plus longs sont, en moyenne, associés aux fémurs les plus longs ; ces deux mesures sont corrélées positivement. Au contraire, si nous notons, pour un grand nombre de voitures, leur année de fabrication et leur consommation d'essence, nous observerons fort probablement une corrélation négative.

Mais ces notions sont floues. Pour préciser le concept de "ressemblance", les mathématiciens, jamais à court d'imagination, ont proposé divers paramètres ; citons les "coefficients de détermination", tombés on ne sait pourquoi dans l'oubli, et le "coefficient de corrélation", dont la fortune a au contraire été si prodigieuse qu'elle en est inquiétante. Epargnons au lecteur la définition arithmétique de ce dernier coefficient que la moindre calculette de poche sait maintenant déterminer. Retenons que c'est, par définition, un nombre compris entre -1 et + 1 : lorsqu'il est égal à + 1, les deux séries de mesures sont si bien ressemblantes, que la connaissance de l'une permet de calculer exactement l'autre.

Mais lorsque ce coefficient, représenté classiquement par la lettre r, est égal à une valeur inférieure à 1, quelle est sa signification ? C'est là que le bât blesse, aucune interprétation ne peut être fournie sans faire des hypothèses sur la répartition des deux séries de mesures. Même la valeur r=0 n'a, en soi, aucun sens précis.

> Ce fut le mérite insigne de l'école anglaise de psychologie, depuis Sir Francis Galton, d'avoir transformé les tests mentaux ; jadis manigances de charlatan, ils sont devenus, grâce à cette méthode d'analyse mathématique, un instrument scientifique de précision, reconnu de tous.
>
> *Cyril Burt, 1921*

Pour avancer, les mathématiciens se placent en général dans un cas particulier, celui où les mesures de chacune des séries sont, en moyenne, des fonctions du premier degré des mesures de l'autre série. Bien souvent, en fait, les observations ne sont pas trop éloignées de cette hypothèse, ce qui rend valide l'interprétation proposée.

Mais celle-ci est assez différente de ce que la plupart des utilisateurs imaginent de bonne foi. Supposons, par exemple, que nous trouvions entre deux séries de mesures A et B un coefficient de corrélation r = 0,80. Ce nombre étant beaucoup plus proche de 1 que de 0, ont est tenté de croire que ces deux séries sont "fortement corrélées", que par conséquent la connaissance de l'une permet d'estimer correctement l'autre. Mais regardons de plus près.

En fait, l'ensemble des mesures de la série B a une certaine dispersion que l'on peut caractériser par son "écart-type" ; sans entrer dans les détails techniques, disons que ce paramètre est tel que (en gros) 95% des mesures se situent dans une plage dont l'amplitude est de quatre écarts-types. Ainsi, le célèbre quotient intellectuel est construit de telle façon que son écart-type est égal à 15 points : 95% des QI d'une population se situent donc dans la plage 70-130, dont l'amplitude est bien de 4 fois 15=60.

La connaissance de la valeur de la caractéristique A apporte une information importante sur B, si lorsque A est connu, la dispersion des mesures possibles pour B est nettement réduite.

A la limite, la connaissance de A entraîne celle de B si l'écart-type de B est réduit à zéro lorsque A est connu. On peut montrer que le rapport

$$\frac{\text{Ecart-type de B lorsque A est connu}}{\text{Ecart-type lorsque A est inconnu}}$$

est égal à $\sqrt{1-r^2}$.

Ce résultat permet de répondre à la question posée précédemment : que signifie un coefficient de corrélation égal à 0,80 ? La connaissance de A réduit la plage de dispersion des mesures de B dans le rapport $\sqrt{1-0{,}80^2} = 0{,}60$. Cette plage reste donc très importante puisqu'elle n'est réduite que d'un peu plus de 1/3.

On constate ici combien les définitions qui fondent le coefficient de corrélation ont pour effet de jeter de la poudre aux yeux : r = 0,80 peut être, de bonne foi interprété comme la marque d'une liaison élevée ; en fait, cette liaison est très lâche.

> Si vous voulez en savoir plus sur l'utilisation des statistiques dans les Sciences Humaines un excellent livre est paru aux éditions Ramsay (198) :
> "La mal mesure de l'homme" *de Stephen Gould.*

Un article écrit par un psychologue mondialement connu H. Eysenck, illustre cette erreur d'interprétation.

H. Eysenck veut montrer que les QI est un paramètre stable qui mesure une caractéristique intrinsèque des individus. Pour cela, il a déterminé le QI de nombreux enfants à 5 ans, puis onze années plus tard, des mêmes enfants âgés donc de 16 ans. Il a constaté, dit-il, une corrélation de 0,80 entre ces deux séries. Il en conclut que la connaissance de la note à 5 ans permet de prévoir "avec une précision remarquable" la note à 16 ans. On imagine les conséquences qu'il en tire pour le système éducatif.

Ne chicanons pas ses chiffres, mais tirons-en les véritables conclusions.

Lorsque la note à 5 ans est connue, l'écart-type des notes à 16 ans est égal, selon le résultat que nous avons obtenu, à 60% de l'écart-type initial, soit 15x0,60=9 Autrement dit, pour une note à 5 ans donnée, 95% des notes à 16 ans se répartissent dans une plage dont l'amplitude est égale à 36 points.

Supposons que le QI d'un enfant soit, à 5 ans, de 110 ; que pouvons-nous dire de son QI à 16 ans ? Simplement qu'il est compris, avec une probabilité de 95%, dans la plage 90-126.

Le moins que l'on puisse dire, est qu'une telle conclusion n'apporte guère d'information.

59 % des français jugent utile la publication des sondages.

Un sondage sur *les* sondages, publié dans Le Matin de mercredi révèle que 59 % des français jugent utile la publication des sondages et que 56 % d'entre eux font confiance à leurs résultats.

Réalisé par l'institut Harris-France, entre le 30 janvier et le 2 février, auprès d'un échantillon de 1000 personnes âgées de dix-huit ans et plus, le sondage fait apparaître que, notamment sur les problèmes politiques, 13 % des personnes interrogées estiment que les sondages sont très utiles et que 46 % d'entre elles les jugent assez utiles. 24 % des personnes les considèrent peu utiles et 12 % pas du tout.

Sur la crédibilité des sondages, 56 % des personnes interrogées font confiance au résultat du sondage (7 % leur font très confiance, 49 % leur font assez confiance). 28 % ne leur accordent que peu de confiance, tandis que 11 % des personnes interrogées ne leur font pas confiance du tout.

(La Montagne, Février 1980)

Il faut un certain aplomb, ou une certaine inconscience, pour voir là une prédiction d'une "remarquable précision". Une cartomancienne n'obtenant que de si piètres performances risquerait de faire faillite.

Nous sommes ici face à un cas typique d'utilisation de l'argument d'autorité camouflé par un langage d'apparence mathématique. Le pouvoir intimidant des formules et des nombres sert à assener des affirmations idéologiques en les faisant passer pour des résultats rigoureusement démontrés.

Les mathématiciens inventent des outils toujours plus efficaces pour disséquer les résultats des observations. Que ne prennent-ils la précaution d'insister sur les limites de leur emploi ! Certains chercheurs en sciences sociales s'imaginent que leurs conclusions sont "scientifiques" lorsqu'ils ont employé pour y parvenir des outils mathématiques sophistiqués. Mais il ne suffit pas de prendre en main un bistouri même ultra moderne, pour se transformer en chirurgien.

Suite de la page 22 : de GALION à BERNOUILLI

par Michel CLINARD - Olivet

Face à ceux qui vous feront remarquer que la formule générale est bien plus compliquée que la somme de départ même si elle a moins de termes, qu'il faut connaître les valeurs des nombres de BERNOUILLI et que de toutes manières on trouve par itération très facilement et très rapidement les sommes souhaitées avec n'importe quel micro-ordinateur, voire n'importe quelle calculette programmable si on n'est pas trop exigeant, ...

Face à ces élèves franchement vicieux,

Ne mollissez pas, soyez imaginatif et rapide, entraînez-vous par avance si c'est nécessaire, il y va peut être de l'avenir de l'enseignement des mathématiques.

BIBLIOGRAPHIE

The American Mathematical Monthly
Volume 87, n° 6, june-july 80.

Histoire des Mathématiques T2
J.P. COLLETTE, Vuibert ERPI

L'intégrale
Paul DEHEUVELS, PUF

Analyse Combinatoire, T1 et 2
L. COMTET, PUF Sup.

Theory of Numbers
DICKSON, Volume 1, Chelsea

Les nombres remarquables
F. Le Lionnais, Hermann

Tous ces livres sont à la bibliothèque de l'IREM d'ORLEANS.