PROF. O. N. ANDERSON

Corrélation et causalité

I.

Dans son article „L'essence de la statistique et de la méthode statistique", inséré dans le fascicule IV, année I, de la présente Revue et constituant l'introduction de l'ouvrage paru en tchèque et intitulé „Les bases de la théorie de la méthode statistique", M. le Professeur *S. Kohn* fait remarquer à juste titre qu'un des principaux problèmes de la théorie de la méthode statistique est la recherche des liaisons causales dites „libres„ ou „stochastiques". L'essentiel de ces liaisons consiste en ce „qu'un phénomène A est suivi par un autre phénomène B qui se trouve indubitablement lié au premier et arrive non pas par nécessité, mais seulement avec certaine probabilité" (page 384). Comme exemple on peut citer la liaison entre la taille des parents et celle des enfants, la dépendance de la conjoncture économique du nombre des établissements industriels déclarés en état de faillite, etc. L'existence de telles liaisons libres n'est point en contradiction avec le principe de la dépendance générale de causalité, mais s'explique par l'impossibilité pour nous (ou du moins par notre inhabileté) de délimiter assez complètement et exactement le cercle de ces causes que nous mettons en liaison avec l'effet qui nous intéresse. Particulièrement dans le domaine de la statistique sociale, nous sommes très rarement en état de ramener notre recherche à la détermination de la probabilité mathématique pour l'arrivée d'un certain phénomène (à condition qu'il existe un groupe donné de causes). Les cas sont très nombreux où nos observations statistiques, appliquées aux différentes masses, nous amènent à des séries de caractéristiques numériques générales de ces masses, et notre tâche est réduite à éclaircir la question de savoir si les variations des chiffres d'une série se réfléchissent (et si elles se réfléchissent, comment cela se fait) sur celles des chiffres correspondants de l'autre série. L'un des modes d'appréciation de la concordance mutuelle des variations de 2 ou plusieurs séries statistiques, mode le plus généralement utilisé dans les pays occidentaux (particulièrement dans les pays anglo-saxons), est actuellement le calcul du *coefficient de corrélation.*

On arrive au calcul de ce coefficient par trois côtés différents. D'abord, on peut passer par l'étude des „lois de répartition", en se posant pour but de construire un système de „paramètres" qui puisse exprimer ces lois de la manière la plus simple. Ce moyen est très perfectionné au point de vue de la théorie des probabilités, mathématiquement très élégant, mais il est en même temps très difficile pour la non-mathématique. Il est traité de la manière la plus avantageuse dans les ouvrages de feu le professeur *A. A. Tchouproff* *).

Le second procédé qui a été, grâce à *G. U. Yule*, adopté dans presque tous les manuels de statistique, se distingue par sa simplicité comparative, et pour cette raison on le considère comme très convenable pour l'instruction des statisticiens débutants. Il existe quelques variantes différentes de ce procédé: les unes s'appuyant principalement sur son interprétation géométrique, les autres sur la forme algébrique de l'exposition.

La forme la plus simple de la variante algébrique est, à notre sens, la suivante:

on a deux séries statistiques:

la première: $X_1, X_2, X_3, \ldots . X_N$

la seconde: $Y_1, Y_2, Y_3, \ldots . Y_N$

On suppose qu'entre X_1 et Y_1, entre X_2 et Y_2, entre X_3 et Y_3, etc. il existe une certaine dépendance „libre"et on veut mesurer son intensité. On calcule d'abord la moyenne arithmétique pour la première série (soit Mx) et celle pour la seconde série (égale à My); on déduit Mx de chacun des termes de la première série, et My de chacun des termes de la seconde série. Désignons les différences obtenues par les lettres minuscules x et y et par les mêmes indices que ceux des lettres majuscules. De cette façon on obtient deux nouvelles séries transformées, exprimant les écarts consécutifs des moyennes:

la première série: $x_1, x_2, x_3, \ldots . x_n$

la seconde série: $y_1, y_2, y_3, \ldots . y_n$.

*) A. A. Tchouproff, Grundbegriffe und Grundprobleme der Korrelationstheorie, Leipzig 1925. Voir de même: Oskar Anderson, Die Korrelationsrechnung in der Konjunkturforschung. Ein Beitrag zur Analyse von zeitreihen-Bonn 1929 (Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung, 4) Pages 24—32.