BULLETIN OF THE INTERNATIONAL
STATISTICAL INSTITUTE

BULLETIN DE L'INSTITUT
INTERNATIONAL DE STATISTIQUE

CONTRIBUTED PAPERS

COMMUNICATIONS LIBRES

TWO BOOKS . DEUX LIVRAISONS

BOOK 1 LIVRAISON 1

48th SESSION - CAIRO 1991 - 48ème SESSION

DIVERSITY INDICES AND ANALYSIS OF DIET-CANCER RELATIONSHIP

Decarli A. and Ferraroni M.

Istituto di Statistica Medica e Biometria, Università degli Studi di Milano

In epidemiological studies concerning chronic diseases, more often it is common to evaluate the relationship between dietary habits, expressed as absolute intake of different nutrients/foods, and risk of disease. For each subject, let Y be the response variable taking value 1 if the disease is present, 0 otherwise; $\underline{x} = (x_1, x_2, \ldots, x_k)$ the vector of k non dietary variables related to disease (confounding and risk factors); $\underline{z} = (z_1, \ldots, z_h)$ the vector of other variables quantifying the absolute intake of h nutrients. Generally speaking, the relationship between disease and explanatory variables is usually investigated using models as:

$$pr(y = 1 \mid \underline{x}, \underline{z}) = f(\underline{x}\underline{\beta}' \cdot \underline{z}\underline{\gamma}') \qquad (1)$$

where $\underline{\beta}'$ and $\underline{\gamma}'$ are unknown vectors of parameters, which are related to some measure of risk depending on the definition of the function f. In addition, these models, allowing for the evaluation of the dose-response relationship between nutrients and risk, usefully contribute to the study of the role of nutrients in the etiology of disease. However, the total intake of a nutrient is usually scattered among many food items and the modalities with which a nutrient is assimilated in the diet may act as a modifier of its impact on the risk . For this reason it seems restrictive to consider in model (1) only variables measuring the absolute intake of each nutrient. It would also be of interest to quantify the different patterns of nutrients intake. This could be done by characterizing each nutrient z_j assimilated in the diet with a diversity index d_j to measure its degree of dispersion. Consider a nutrient distributed among a countable set of foods, labelled i=1,2 ..I and let p be the proportionate share of this nutrient raising from the i-th food. The measures or indices of diversity which are considered in the present work are those defined according to Patil and Taillie (1982) as:

$$D = (1 - \sum p_i^{\beta-1}) / \beta \qquad (2)$$

As β varies in the real field, with $\beta \geq -1$, expression (2) defines different forms of diversity indices with desirable mathematical properties. Among those, the index E known as Shannon's Entropy is obtained for $\beta = 0$:

$$E = -\sum p_i \log p_i$$

The properties of some diversity indices of the family are investigated in relation to their use in dietary data analysis. Some interpretations of these indices, when applied to this kind of data, are given. Exemplifications using real data obtained from a case-control study recently carried out in Italy are also shown. This data-set includes information on dietary habits on 1059 cases of gastric cancer and 1116 population controls from 7 different areas in Italy. Dietary data were collected by means of diet history questionnaires with 181 questions on 146 different food items.

References:

Patil G.P., Taillie C.
Diversity as a concept and its measurement.
JASA, 1982, 77, 548-561.