

15th European Meeting of Statisticians

Palermo 13-18 September 1982



ES

ABSTRACTS

ENTROPY CORRELATION COEFFICIENT, A MEASURE OF STATISTICAL DEPENDENCE
FOR CATEGORIZED DATA

by
Ilkka Virtanen
University of Vaasa
School of Business Studies
Vaasa, Finland

ENTROPY CORRELATION COEFFICIENT, A MEASURE OF STATISTICAL DEPENDENCE FOR CATEGORIZED DATA

by
Ilkka Virtanen
University of Vaasa
School of Business Studies
Vaasa, Finland

Abstract

The concept of *entropy* has been widely used in physics and information theory. Over the years the idea has been borrowed by other disciplines and has then been applied in several problem areas within the social sciences, especially in statistics, economics, business, geography and operational research. Entropy has become an important tool for planning purposes in the wide and fast developing area of system modeling.

The use of entropy in statistics has its origin in information theory. Shannon's measure for uncertainty, for example, has been introduced as a measure of dispersion for qualitative (or categorized) data. It is, however, possible to calculate entropy also for a two-dimensional distribution of two qualitative variables, i.e. for a bivariate distribution given as a frequency table. This two-dimensional entropy has been shown to reveal both the dispersion of the marginal distributions and the dependence of the two variables. The analysis of entropy as a measure of dependence has remained, however, quite slight.

The aim of the paper is to carry out a detailed analysis of the concept of entropy defined for two-way frequency tables (called now *coentropy*). We also give *coentropy* an interpretation as the mean uncertainty appearing in the table and demonstrate its definitional analogy with the covariance of two quantitative variables. Further, we construct an entropy-based measure, called *mean dependence information*, for the degree of dependence and analyze and interpret this measure (e.g. we show its definitional analogy with product moment correlation coefficient of two quantitative variables).

And finally, we scale this measure in order to get such a measure of dependence that fulfills the requirements set for a correlation coefficient. We call our measure *entropy correlation coefficient*. The paper ends with an analysis and numerical illustration of this coefficient.

Reference:

Astola, J. and Virtanen, I.: Entropy correlation coefficient, a measure of statistical dependence for categorized data, Lappeenranta University of Technology, Department of Physics and Mathematics, Report 4/1981

Preuss, L.G.: A class of statistics based on the information concept, Communications in Statistics, Theory and Methods, Vol. A 9 (1980), No 15, pp. 1563-1586