



Fitting a bivariate normal distribution to a two-way contingency table using maximum likelihood estimation

Gretel Crafford*

University of Pretoria, Pretoria, South Africa – gretel.crafford@up.ac.za

Nico Crowther

University of Pretoria, Pretoria, South Africa – nico.crowther@up.ac.za

The historic height data set (1885) of Sir Francis Galton which considered the relationship between the height of parents and their children is revisited. The data set consists of 928 cases and is categorised in a two-way contingency table. Although the data is only available in a grouped format, an underlying bivariate normal distribution is evident. Employing an iterative maximum likelihood (ML) procedure, the estimated bivariate cumulative relative frequencies are fitted to follow a cumulative bivariate normal distribution curve at the intersections of the upper class boundaries of the two-way contingency table. These estimates are referred to as the ML estimates under constraints of the bivariate normal probabilities. The five parameters of the bivariate normal distribution, namely the mean and variance of the two variables and the correlation coefficient can now be solved uniquely from these ML estimates under constraints. In the iterative procedure use is made of the fact that the marginal distributions are normally distributed and that the correlation coefficient can be uniquely expressed as a function of the total probability of the positive quadrant of the bivariate normal distribution. The conditions of the marginal and joint distributions are incorporated simultaneously in the ML estimation procedure. By obtaining the ML estimates of the underlying bivariate normal distribution the complete relationship between the two variables can be investigated. The slope of the estimated regression line is 0.63 which corresponds closely to that of Galton, who suggested a slope of two thirds.

Keywords: bivariate grouped data; maximum likelihood estimates under constraints.