Prof. John Rayner, University of Newcastle and **CSSM**,

Authors: *John C. W. RAYNER and Eric J. BEH*

Title: Towards a Better Understanding of Correlation

Abstract: In most introductions to correlation the student is taught that for bivariate random variables (X, Y) two important properties of the correlation ρ_{XY}, defined by

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}},$$

are that

- independence implies zero correlation, but the converse is not true, and
- the correlation is ±1 if and only if X and Y are linearly related.

In fact many ‘correlations’ may be defined. Here *generalised* correlations for bivariate discrete distributions are defined and used to show that independence holds if and only if *all* the generalised correlations are zero. The correlation ρ_{XY} above is the simplest generalised correlation, that subsequently we will refer to as the order (1, 1) correlation. The reason why this $\rho_{XY} = 0$ is necessary but not sufficient for independence is that there are many other generalised correlations, and all of them must be zero for independence to hold. We next consider generalisation the notion of correlation to higher dimensions and discuss issues concerning the interpretation of independence and its lack.

Finally we consider linearity to two and more dimensions. It is not difficult to give results, and they do not seem to be well-known. Thus in two dimensions linearity between X and Y means that for some constants a, b and c, $aX + bY = c$. That this holds if and only if $\rho_{XY}^2 = 1$ is shown in many standard textbooks. Subsequently we ask first, how does this extend to more than two dimensions, and second, is it important to do so?

€

References:

RAYNER, J.C.W. and BEH, Eric J. (2009). Towards a Better Understanding of Correlation. *Statistica Neerlandica*.