Statistical inference for direction of dependence in linear models

Yadolah Dodge
Professor Emeritus, University of Neuchâtel, Switzerland
yadolah.dodge@unine.ch

Valentin Rousson
Associate Professor, Division of Biostatistics, Institute for Social and Preventive Medicine, University Hospital Lausanne, Switzerland
valentin.rousson@chuv.ch

Abstract

This paper considers the problematic of the detection of the direction of a regression line as originally proposed by Dodge and Rousson (2000, 2001). Their approach uses the fact that the response should be closer to a normal distribution than the predictor in a linear regression model with normal residuals, such that one can select as predictor that variable which is more skewed, respectively has more excess kurtosis than the other. We propose a model framework to formally test for the significance of the direction of a regression line. We identify five mutually exclusive situations, three of which defining the null hypothesis. We then use a simplification of the testing procedure proposed by Pornprasertmanit and Little (2012), which maintains the Type I error defined by our null hypothesis, as checked via simulations. The testing procedure is found more powerful to detect the true direction of dependence when the predictor is skewed than when it is symmetric with an excess of kurtosis. We also argue that our model framework, which implies a linear model with non-normal variables and normal residuals if the null hypothesis is wrong, protects us to some extent against the existence of a lurking variable in case of a significant result, unless one selects an artificial non-normal subpopulation from a normal population. Finally, we illustrate the fact that a few outlying observations may have a high influence on the testing procedure, and that removing these observations may bring us back towards the normality which we wish to avoid, such that proving statistically the direction of a regression line remains a ma jor challenge.

Key words: Causality, direction of dependence, kurtosis, linear regression model, non-normality, null hypothesis, outliers, skewness, Type I error.

Dodge Y, Rousson V (2000). Direction dependence in a regression line. Communications in Statistics: Theory and Methods, 29, 1957-1972.

Dodge Y, Rousson V (2001). On asymmetric properties of the correlation coefficient in the regression setting. The American Statistician, 55, 51-54.

Dodge Y, Yadegari I (2010). On direction of dependence. Metrika, 72, 139-

150.

Pornprasertmanit S, Little TD (2012). Determining directional dependency in causal associations. International Journal of Behavioral Development, 36, 313-322.