

How important are the assumptions of statistical techniques?

Peter O'Donoghue

Cardiff School of Sport, University of Wales Institute Cardiff, Cyncoed Campus, Cardiff, Wales, CF23 6XD, UK.

The statisticians tell us that our research is invalid if our data violate the assumptions of the statistical techniques that we have used. However, many in sports psychology continue to use advanced multivariate procedures with likert scale data that is not normally distributed. There is little evidence to show that violating the assumptions of such tests increases the risks of sampling error in research studies. Therefore, the purpose of the current investigation was to compare 2 sets of 4 models of international soccer performance; each set consisted of (a) multiple linear regression, (b) multiple linear regression with proportional outcomes, (c) simulation based on multiple linear regression and (d) discriminant function analysis and binary logistic regression. One set of models used data where transformations were made to ensure the data satisfied the assumptions of the modelling techniques while the other set of models used raw untransformed data that violated the assumptions of the techniques. A total of 8 predictive models of the 2010 World cup were made and tested using actual match outcomes from the tournament. After the 48 pool matches, there was little difference in the accuracy of the 4 models that satisfied the assumptions (27.125 points out of 48) and those that violated the assumptions (27.375). It will be interesting to see which models are most accurate after the knockout stages are completed, but at this stage it is difficult to justify the effort of transforming data to satisfy the assumptions of the modelling techniques used.