DO WE TEACH REGRESSION CORRECTLY

Paul Hewson
Plymouth University, UK
paul.hewson@plymouth.ac.uk

BACKGROUND

Many interesting social phenomena are innately multidimensional and require suitable data modelling tools. Regression modelling (which includes log linear modelling for contingency tables as a special case) is often the „go-to"‘ tool. However, much of the math theory was developed for designed experiments (where explanatory variables **X** are orthogonal and fixed). Conversely, societal data is often observational with random non-orthogonal **X**.

The pedagogic route to data modelling usually starts with linear models before the introduction of the generalised linear models that can address contingency tables. Whilst good textbooks do feature caveats, the search for a parsimonious model is often carried out in a manner that may promote unsafe interpretation of observational data. This poster tries to present a case for a reform in the teaching of regression for observational data.

CONTENT

A non-modelled example on social attitudes to leaving the EU, as well as text book examples on characteristics of teenage gamblers and restaurant quality in New York will be used to illustrate the need for careful modelling, and the pitfalls present within regression modelling.

CONCLUSION

Understanding data about society requires a framework for dealing with more dimensions that we can visualise or handle intuitively. Regression seems stretched beyond its intended purpose when applied to observational data. It seems time to consider how we teach students to understand multidimensional data.

In: J. Engel (Ed.), *Promoting understanding of statistics about society.*
Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE),
July 2016, Berlin, Germany. iase-web.org/Conference_Proceedings.php