CrossMark

Strong model dependence in statistical analysis: goodness of fit is not enough for model choice

John Copas[1] · Shinto Eguchi[2]

Abstract

Most statistical methods are based on models, but most practical applications ignore the fact that the results depend on the model as well as on the data. This paper examines the size of this model dependence, and finds that there can be very considerable variation between the results of fitting different models to the same data, even if the models being considered are restricted to those which give an acceptable fit to the data. Under reasonable regularity conditions, we show that different empirically acceptable models can give rise to non-overlapping confidence intervals for the same parameter. Application papers need to recognize that the validity of conventional statistical results rests on the assumption that the underlying model is known to be correct, and that this is a much stronger requirement than merely confirming that the model gives a good fit to the data. The problem of model dependence is only partially resolved by using formal methods of model selection or model averaging.

Keywords Goodness-of-fit · Model choice · Model uncertainty · Subset selection

---

✉ John Copas
jbc@stats.warwick.ac.uk

[1] Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

[2] Institute of Statistical Mathematics, Midori-cho 10-3, Tachikawa, Tokyo 190-8562, Japan