COMMENTS

CHOOSING MODELS FOR CROSS-CLASSIFICATIONS*

(Comment on Grusky and Hauser, *ASR*, February 1984)

ADRIAN E. RAFTERY
Trinity College, Dublin, Ireland

Grusky and Hauser (1984)—hereafter GH—reanalysed three-stratum intergenerational mobility classifications for men in sixteen countries. They rejected the Lipset-Zetterberg hypothesis that observed mobility rates are much the same in western industrialized societies on the basis of a highly significant likelihood ratio chi-square test statistic, $L^2 = 3201$ with 64 degrees of freedom (df) for the nine most industrialized nonsocialist nations in their sample.

They then considered the quasi-perfect mobility model, which yields a highly significant $L^2 = 150$ with 16 df, corresponding to a P-value of about 10^{-120}. They nevertheless adopted this model on the grounds that it "fits extremely well, accounting for 99.7% of the association under the baseline model of independence." This sounds sensible, yet the procedure of rejecting one model on the basis of the likelihood-ratio test (LRT), and accepting another in spite of it, seems unsatisfactory.

My purpose here is to argue that it is not GH who are illogical, but that the LRT is ill-suited to the task of model selection, especially in large-sample situations such as theirs. An alternative statistical procedure is proposed which, unlike the LRT, leads to the same conclusions as GH.

WHAT IS WRONG WITH THE LIKELIHOOD RATIO TEST?

The aim of much social research is to describe the main features of selected aspects of social reality; such a description is often called a model, and is necessarily to some extent approximate. The LRT, however, in common with other significance tests, is designed to detect *any* discrepancies between model and reality. Such discrepancies do exist, by definition, although if the model is satisfactory, they should be small. With a large enough sample, the LRT will find them and reject even a good model.

In the contingency table case, the LRT tests a model M_0 say, against the saturated model M_1. Assume for the moment that no other models are being considered. Rejection of M_0 then implies acceptance of M_1, which says that each cell is a special case. This does constitute a statement about the underlying social reality, and may, indeed, itself be a model of interest (Featherman and Hauser, 1978:161–66). Rejection of M_0 does not imply that M_1 provides a better description. The point is that we should be *comparing* the models, not just looking for possibly minor discrepancies between one of them and the data.

AN ALTERNATIVE: POSTERIOR ODDS

The question to which we really want an answer can perhaps often best be expressed as follows: which model better describes the main features of social reality as reflected in the data? A closely related and more precise question is: given the data, which of M_0 and M_1 is more likely to be the true model?

The latter question can be answered by calculating the posterior odds for M_0 against M_1, defined by

$$B = \frac{\text{Prob } [M_0 \text{ is the true model given the data}]}{\text{Prob } [M_1 \text{ is the true model given the data}]}$$

B is defined in the context of the Bayesian approach to statistical inference, which views probability as a measure of the opinion of a coherent individual, and which is well described by Edwards et al. (1963). In principle, to find a value for B one needs to specify prior beliefs about the models and their parameters, but in the large-sample situation considered here, the effect of the prior beliefs on the conclusions drawn would be very small.

It is shown in Raftery (1985), drawing on results of Spiegelhalter and Smith (1982), that approximately, in large samples,

$$-2 \log B = L^2 - (\text{df}) \log N$$

where N is the total sample size. This quantity, denoted BIC, can be easily calculated from the output of standard contingency table analysis programs such as GLIM. If BIC is negative, therefore, we should accept M_0 in the sense of

* Address all correspondence to Adrian E. Raftery, Department of Statistics, University of Washington, Seattle, WA 98195.

I am grateful to Michael Hout and an anonymous referee for helpful comments.

preferring it to the saturated model. If we are comparing several models, we should prefer the one with the lowest BIC value. BIC provides a consistent model selection procedure in the sense that in large samples it chooses the correct model with high probability, and it is thus valid beyond the Bayesian context; see Schwarz (1978). It provides an automatic way of making the often difficult and subjective trade-off between L^2 and df which is inherent in the conventional LRT model selection procedure.

CONCLUSION

For the Lipset-Zetterberg hypothesis, BIC = 2528, so that it provides a poor description of the data relative to the saturated model and should be rejected. This is GH's conclusion also. For the quasi-perfect mobility model, BIC = −36, indicating that it should be accepted. This agrees with GH's "commonsense" conclusion, and disagrees with the conclusion based on the LRT.

The implications of this comment go beyond the GH study and suggest that posterior odds provide a more sensible basis than likelihood-ratio testing for choosing between models for contingency tables in many social science applications.

REFERENCES

Edwards, Ward, Harold Lindman and Leonard J. Savage. 1963. "Bayesian Statistical Inference for Psychological Research." *Psychological Review* 70:193–242.

Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.

Grusky, David B. and Robert M. Hauser. 1984. "Comparative Social Mobility Revisited: Models of Convergence and Divergence in 16 Countries." *American Sociological Review* 49:19–38.

Raftery, Adrian E. 1985. "A Note on Bayes Factors for Log-Linear Contingency Table Models with Vague Prior Information." *Journal of the Royal Statistical Society, Series B*.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6:461–64.

Spiegelhalter, David J. and Adrian F. M. Smith. 1982. "Bayes Factors for Linear and Log-Linear Models with Vague Prior Information." *Journal of the Royal Statistical Society, series B* 44:377–87.