Model Selection and Criticism

In many applications of statistics, little prior knowledge or relevant theory is available, and so model choice becomes an entirely empirical, exploratory process. Three different approaches to model selection are described in the first three sections of this chapter. The first is a stepwise method, which starts from some initial model and successively adds or removes edges until some criterion is fulfilled. The second is a more global search technique proposed by Edwards and Havránek (1985, 1987), which seeks the simplest models consistent with the data. The third method is to select the model that optimizes one of the so-called information criteria (AIC or BIC). In Section 4 a brief comparison of the three approaches is made.

Section 5 describes a method to widen the scope of the CG-distribution by allowing power transformations of the continuous variables (Box and Cox, 1964). The last two sections describe techniques for checking whether the continuous variables satisfy the assumptions of multivariate normality.

We preface the chapter with some introductory remarks about model selection. Perhaps the first thing to be said is that all model selection methods should be used with caution, if not downright scepticism. Any method (or statistician) that takes a complex multivariate dataset and, from it, claims to identify one true model, is both naive and misleading. The techniques described below claim only to identify simple models consistent with the data, as judged by various criteria: this may be inadequate for various reasons.

For example, if important variables have been omitted or are unobservable, the models selected may be misleading (for some related issues, see Section 1.4). Some problems seem to require multiple models for an ade-
quate description (see the discussion of split models in Section 2.2.6), and for these, the adoption of one grand, all-embracing supramodel may be unhelpful. Finally, the purpose to which the models will be put and the scientific interpretation and relevance of the models ought to play decisive roles in the evaluation and comparison of different models.

The first two model selection approaches described here are based on significance tests; many tests may be performed in the selection process. This may be regarded as a misuse of significance testing, since the overall error properties are not related in any clear way to the error levels of the individual tests (see Section 6.4, however, for a qualification of this statement).

There is also a deeper problem. In statistical modelling we generally choose a model that fits the data well, and then proceed under the assumption that the model is true. The problem with this— the problem of model uncertainty — is that the validity of most model-based inference rests on the assumption that the model has not been chosen on the basis of the data. Typically, estimators that would be unbiased under a true, fixed model are biased when model choice is data-driven. Estimates of variance generally underestimate the true variance, for example. Similarly, hypothesis tests based on models chosen from the data often have supranominal type I error rates. Chatfield (1995) gives an accessible introduction to model uncertainty. With randomised studies the problem can be circumvented (Edwards, 1999).

To summarize: It is essential to regard model selection techniques as explorative tools rather than as truth-algorithms. In interplay with subject-matter considerations and the careful use of model control and diagnostic techniques, they may make a useful contribution to many analyses.

6.1 Stepwise Selection

This is an incremental search procedure. Starting from some initial model, edges are successively added or removed until some criterion is fulfilled. At each step, the inclusion or exclusion of eligible edges is decided using significance tests. Many variations are possible and are described in this section.

Stepwise selection is performed in MIM using the command Stepwise. The standard operation of this command is backward selection; that is to say, edges are successively removed from the initial model. At each step, the eligible edges are tested for removal using $\chi^2$-tests based on the deviance difference between successive models. The edge whose $\chi^2$-test has the largest (nonsignificant) p-value is removed. If all p-values are significant