Theoretical Aspects of Fitting Models

Ockham's razor and the Law of Parsimony

- Ockham's razor, sometimes known as the law of parsimony, is simply a maxim that states that simple explanations are usually better than complicated ones. **Ockham's razor** was originally proposed by a monk named William of Ockham. (He did not call it "Ockham's razor" or even "my razor." This is a name that has been given to it over time.)

- Another version of this principle is the Law of parsimony . This says that if you are choosing between two theories, choose the one with the fewest assumptions. Assumptions here means claims of fact that have no evidence.

- A theory that doesn't have many assumptions, and is very simple, is called a **parsimonious theory**.

- In the context of statistics, the law of parsimony can be interpreted as an adequate model which requires the fewest independent variables is the preferred model.

Parsimonious: The simplest plausible model with the fewest possible number of variables.

Akaike Information Criterion

- The Akaike information criterion is a measure of the relative goodness of fit of a statistical model. It was developed by Hirotsugu Akaike, under the name of "an information criterion" (AIC), and was first published by Akaike in 1974.

- The AIC methodology attempts to find the model that best explains the data with a minimum of parameters. (i.e. in keeping with the law of parsimony).

- The AIC is a **_model selection_** tool i.e. a method of comparing two or more candidate regression models. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best. (Although, a difference in AIC values of less than two is considered negligible).

- The AIC is calculated using the "likelihood function" and the number of parameters (*The likelihood function is not on the course*). The likelihood value is generally given in code output, as a complement to the AIC.

$$\text{AIC} = 2p - 2\ln(L)$$

- p is the number of free model parameters.

- L is the value of the Likelihood function for the model in question.

- For AIC to be optimal, n must be large compared to p.

Schwarz's Bayesian Information Criterion

An alternative to the AIC is the Schwarz BIC, which additionally takes into account the sample size n.

$$\text{BIC} = p \ln n - 2 \ln(L)$$

Model Building

- The traditional approach to statistical model building is to find the most parsimonious model that still explains the data. The more variables included in a model (overfitting), the more likely it becomes mathematically unstable, the greater the estimated standard errors become, and the more dependent the model becomes on the observed data.

- Choosing the most adequate and minimal number of explanatory variables helps to find out the main sources of influence on the response variable, and increases the predictive ability of the model.

- As a rule of thumb, there should be more than 10 observations for each variable in the model.

Overfitting

- Overfitting generally occurs when the model is excessively complex, such as having too many parameters (i.e. predictor variables) relative to the number of observations

- Equivalently, overfitting occurs when a statistical model does not adequately describe of the underlying relationship between variables in a regression model.

- A model which has been overfit will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data. The model predicts the fitted data very well, but predicts future observations poorly.

Variable-Selection Procedures

In regression analysis, variable-selection procedures are aimed at selecting a reduced set of the independent variables - the ones providing the best fit to the model, in keeping with the Law of Parsimony.