Akaike Information Criterion

AIC

- Akaike's information criterionis a measure of the goodness of fit of an estimated statistical model.
- The AIC was developed by Hirotsugu Akaike under the name of "an information criterion" in 1971.
- The AIC is a **model selection** tool i.e. a method of comparing two or more candidate regression models.
- 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 calculated using the "likelihood function" and the number of parameters (Likelihood function : not on course).
- The likelihood value is generally given in code output, as a complement to the AIC.
- 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).

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.

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

Akaike Information Criterion

- 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.

Information Criterions

We define two types of information criterion: the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). In AIC and BIC, we choose the model that has the minimum value of:

$$AIC = 2log(L) + 2m,$$

$$BIC = 2log(L) + mlogn$$

where

- L is the likelihood of the data with a certain model,
- n is the number of observations and
- m is the number of parameters in the model.

AIC

- The Akaike information criterion is a measure of the relative **goodness of fit** of a statistical model.
- When using the AIC for selecting the parametric model class, choose the model for which the AIC value is lowest.

Schwarz's Bayesian Information Criterion

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

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

AIC and BIC in Two-Step Cluster Analysis

(Removed from Last Week's Class due to Version Update)
Two-Step Cluster Analysis guides the decision of how many clusters to retain from the data by calculating measures-of-fit such as **Akaikes Information Criterion (AIC)** or **Bayes Information Criterion (BIC)**.

- These are relative measures of goodness-of-fit and are used to compare different solutions with different numbers of segments.(“Relative” means that these criteria are not scaled on a range of, for example, 0 to 1 but can generally take any value.)
- **Important**: Compared to an alternative solution with a different number of segments, smaller values in AIC or BIC indicate an increased fit.

SPSS computes solutions for different segment numbers (up to the maximum number of segments specified before) and chooses the appropriate solution by looking for the smallest value in the chosen criterion. However, which criterion should we choose?

- AIC is well-known for overestimating the correct number of segments
- BIC has a slight tendency to underestimate this number.

Thus, it is worthwhile comparing the clustering outcomes of both criteria and selecting a smaller number of segments than actually indicated by AIC. Nevertheless, when running two separate analyses, one based on AIC and the other based on BIC, SPSS usually renders the same results.

Once you make some choices or do nothing and go with the defaults, the clusters are formed. At this point, you can consider whether the number of clusters is "good". If automated cluster selection is used, SPSS prints a table of statistics for different numbers of clusters, an excerpt of which is shown in the figure below. You are interested in finding the number of clusters at which the Schwarz BIC becomes small , but also the change in BIC between adjacent number of clusters is small.
The decision of how much benefit accrued by another cluster is very subjective. In addition to the BIC, a high ratio of distance of measures is desirable. In the figure below, the number of clusters with this highest ratio is three.