1 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.

1.1 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.

1.2 Akaike Information Criterion

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).

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.

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

- 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.

1.2.1 Schwarz's Bayesian Information Criterion

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

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

1.3 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.

Information Criterions We define two types of information criterion: the Akaike Information Criterion (AIC) and the Schwarzs Bayesian Information Criterion (BIC). The Akaike information criterion is a measure of the relative goodness of fit of a statistical model. AIC = 2p 2 ln(L)

- p is the number of predictor variables in the model.
- 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.

Autoclustering statistics

		Schwarz's Bayesian Criterion (BIC)	BIC Change[1]	Ratio of BIC Changes[2]	Ratio of Distance Measures[3]
Number of Clusters	1	6827.387			
	2	5646.855	-1180.532	1.000	1.741
	3	5000.782	-646.073	.547	1.790
	4	4672.859	-327.923	.278	1.047
	5	4362.908	-309.951	.263	1.066
	6	4076.832	-286.076	.242	1.193
	7	3849.057	-227.775	.193	1.130
	8	3656.025	-193.032	.164	1.079
	9	3482.667	-173.358	.147	1.162
	10	3343.916	-138.751	.118	1.240
	11	3246.541	-97.376	.082	1.128
	12	3168.733	-77.808	.066	1.093
	13	3103.950	-64.783	.055	1.022
	14	3042.116	-61.835	.052	1.152
	15	2998.319	-43.796	.037	1.059

1. The changes are from the previous number of clusters in the table.
2. The ratios of changes are relative to the change for the two cluster solution.
3. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

Figure 1: Schwarz Bayesian Information Criterion

An alternative to the AIC is the Schwarz BIC, which additionally takes into account the sample size n. BIC = p ln n 2 ln(L) When using the AIC (or BIC) for selecting the optimal model, we choose the model for which the AIC (or BIC) value is lowest.

Akaike Information Criterion

- Akaikes 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 . The likelihood value is generally given in code output, as a complement to the AIC. (Likelihood function is not on our course)
- 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).

1.4 Model Metrics for Logistic Regression Models

- In order to understand how much variation in the dependent variable can be explained by a logistic regression model (the equivalent of R 2 in multiple regression), you should consult Model Summary statistics.

- Although there is no close analogous statistic in logistic regression to the coefficient of determination R 2 the Model Summary Table provides some approximations.
- Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many researchers have tried to come up with one.
- The SPSS output table below contains the Cox & Snell R Square and Nagelkerke R Square values, which are both methods of calculating the explained variation. These values are sometimes referred to as pseudo R 2 values (and will have lower values than in multiple regression).
- However, they are interpreted in the same manner, but with more caution. Therefore, the explained variation in the dependent variable based on our model ranges from 24.0to 33.0%, depending on whether you reference the Cox & Snell R 2 or Nagelkerke R 2 methods, respectively.
- Nagelkerke R 2 is a modification of Cox & Snell R 2 , the latter of which cannot achieve a value of 1. For this reason, it is preferable to report the Nagelkerke R 2 value.
- The Nagelkerke modification that does range from 0 to 1 is a more reliable measure of the relationship.
- Nagelkerkes R 2 will normally be higher than the Cox and Snell measure. Figure 1: SPSS output
- Cox and Snells R-Square attempts to imitate multiple R-Square based on likelihood, but its maximum can be (and usually is) less than 1.0, making it difficult to interpret. Here it is indicating that 55.2logistic model.

1.5 Pseudo R-squares

- Cox & Snell R Square and Nagelkerke R Square are two measures from the pseudo R-squares family of measures.
- There are a wide variety of pseudo-R-square statistics (these are only two of them). Because this statistic does not mean what R-squared means in OLS regression (the proportion of variance explained by the predictors), we suggest interpreting this statistic with great caution.

1.6 Cox & Snell R Square

Cox and Snells R-Square is an attempt to imitate the interpretation of multiple R-Square based on the likelihood, but its maximum can be (and usually is) less than 1.0, making it difficult to interpret. It is part of SPSS output.

1.7 Nagelkerkes R-Square

- Nagelkerkes R 2 is part of SPSS output in the Model Summary table and is the most-reported of the R-squared estimates.

- In our case it is 0.737, indicating a moderately strong relationship of 73.7predictors and the prediction.
- Nagelkerkes R-Square is a further modification of the Cox and Snell coefficient to assure that it can vary from 0 to 1. Nagelkerkes R-Square will normally be higher than the Cox and Snell measure. It is part of SPSS output and is the most-reported of the R-squared estimates.