

Model Selection Using Information Criteria (Made Easy in SAS®)

Wendy Christensen, University of California, Los Angeles

ABSTRACT

Today's statistical modeler has an unprecedented number of tools available to select the "best" model from a set of candidate models. Based on a focused search of SAS/STAT® procedure documentation, over 30 procedures include one or more information criteria as part of their output. It is, however, unusual for applied statistics courses to discuss information criteria beyond "smaller is better". The focus of this breakout session is twofold. First, I provide an accessible conceptual overview of what information criteria are and why they are useful for model selection. Second, I demonstrate how to use Base SAS® functions to compile information criteria from multiple models into a convenient and easy-to-read format.

INTRODUCTION

The scientific method is often presented as a cyclic process in which a researcher develops hypotheses based on theories and tests their veracity by their ability to predict or explain new observations. It is now common for researchers to express and test hypotheses and theories by building representative statistical models and determining how well the models fit collected data. To test the plausibility of a single theory, one statistical model representing that theory is fitted to the data, the predictive or explanatory power of that model assessed, and the parameters of that model interpreted. Researchers sometimes have competing theories about the mechanisms of a phenomenon and thus have two or more competing models. To determine which theory best explains the data actually observed, the models representing those theories, also referred to as candidate models, are compared using an objective method. The model considered the best by that method is selected as the "best" model, which becomes evidence in favor of the theory represented by that model.

Introductory applied statistics texts (e.g. Cohen, Cohen, West, & Aiken, 2003) often include instructions for model comparison methods that are grounded in the null hypothesis testing framework, such as R^2 change tests and likelihood ratio tests. Such tests are reasonably intuitive and simple to conduct, but there are situations for which they are unsuitable. Model selection using information criteria, which are grounded in the information-theoretic framework, offers greater flexibility than methods based on null hypothesis testing (Hamaker, van Huttum, Kuiper, & Hoijtink, 2010). One advantage is that information criteria do not require the candidate models to have any particular relationship with each other. Specifically, the candidate models do not have to be nested. Another advantage is that any number of models can be compared simultaneously using information criteria; that is, it is not necessary to test models incrementally (Model 1 vs. Model 2, Model 2 vs. Model 3, etc.), nor is it limited to comparing models two at a time. Finally, because information criteria are not based on the null hypothesis testing framework, there is no need to designate any model as a null model for comparison purposes.

This paper is divided into three parts. The first part is a brief conceptual overview of the logic and history of information criteria, how they are used for model selection, and which SAS/STAT procedures incorporate them by default or by user request. The second part is a series of examples demonstrating how to obtain information criteria from SAS/STAT procedure output using Base SAS functions. The third part is a demonstration of how to use Base SAS functions to compare models using information criteria in an easy and convenient way. Code excerpts and screen captures are incorporated into both demonstrations, and the full code with commentary is available on Github.

INFORMATION CRITERIA

Akaike (1974) introduced the first information criterion, which is now known as the Akaike Information Criterion (AIC). The logic of information criteria is grounded in information theory. More specifically, AIC was based on Kullback-Leibler (K-L) distance (Kullback & Leibler, 1951), which connected information theory to random variable distributions. A brief overview, which closely follows the conceptual summary provided in Burnham and Anderson's (2002) widely-cited text, of K-L distance is provided here.

K-L distance is a way of conceptualizing the distance, or discrepancy, between two models. One of these models, $f(x)$, is the “true” or “generating” model from which actual observations emerge. The other model, $g(x)$, is a model specified in accordance with the researcher’s theory about the “true” model. If $g(x)$ is exactly the same as $f(x)$, then the K-L distance between the two models is zero; otherwise, the K-L distance is greater than zero. Per information theory, K-L distance is understood as the information lost when $g(x)$ is used to approximate $f(x)$.

If the approximating model, $g(x)$, is a reasonable approximation of the “true” model, $f(x)$ – that is, if the researcher’s model includes what is most important to explain or predict the outcome of interest – then the information lost by using $g(x)$ to approximate $f(x)$ is minimized. In practice, the “true” model is unknown, so the absolute K-L distance between $f(x)$ and $g(x)$ is impossible to compute. If, however, two or more approximating models were being fit to the data and the “true” model were known, the K-L distances computed would all have the true model, $f(x)$, in common. In other words, $f(x)$ is a constant when computing these K-L distances. This means that, while *absolute* distances cannot be computed without the “true” model, *relative* distances can be computed, which makes the relative *rank* of the candidate models meaningful. Akaike’s (1974) key insight was that maximum likelihood estimation, with some correction based on the number of parameters in the model, is the expected K-L distance.

After the introduction of AIC, more information criteria were developed with differing mathematical properties and philosophies of model selection in mind. AIC and AICC (Hurvich & Tsai, 1989) are information criteria that are *efficient*, which means that they will select the best model when the generating model is of infinite dimension. A model’s dimensionality refers to the number of parameters estimated (e.g. predictor variables) as part of that model. This means that a model of infinite dimension is a model with an infinite number of parameters, which is impossible to fit in practice. Burnham and Anderson (2002) viewed this impossibility as matching the reality of research; because reality is too complex to be contained in any model, such a model will never be in the set of candidate models.

In contrast, BIC (Schwarz, 1978), CAIC (Bozdogan, 1987), and HQIC (Hannan & Quinn, 1979) are information criteria that are *consistent*, which means that as the size of the sample increases, these criteria will select a true model of finite dimension as long as it is included in the set of candidate models. AIC is an information criterion that is dimension inconsistent – that is, it has a non-zero probability of selecting an overly-complex model, even as the sample size approaches infinity (Bozdogan, 1987). While efficient and consistent information criteria represent different approaches to model selection, they are often presented alongside each other in software and used jointly in applied work. The formulae for these five information criteria are shown in Table 1, where LL is the model log likelihood estimate, K is the number of model parameters, and n is the sample size.

AIC: $-2LL + 2K$	AICC: $-2LL + 2K \left(\frac{n}{n-K-1} \right)$	
BIC: $-2LL + \ln(N)K$	CAIC: $-2LL + [\ln(n) + 1]K$	HQIC: $-2LL + 2K \ln(\ln(n))$

Table 1. Formulae for AIC, AICC, BIC, CAIC, and HQIC

COMPARING MODELS USING INFORMATION CRITERIA

Once the set of candidate models is defined, model selection using information criteria is a three-step process. The first step is to fit each candidate model to the same data, making sure that any transformation to the outcome variable is maintained across all candidate models (Burnham & Anderson, 2002). For example, if comparing four candidate models with regard to predicting outcome Y , one could do so as long as all four models had Y or $\log(Y)$ or any other transformation of Y as the outcome variable. If Y had been the outcome for two models, $\log(Y)$ for one model, and \sqrt{Y} for another model, the only sensible comparison within the set of candidate models would be between the two models with a shared outcome variable Y .

The second step is to obtain the desired information criteria for each model. Per the formulae in Table 1, each candidate model contributes uniquely to an information criterion’s value through its estimated log likelihood and the number of parameters used in the model. The sample size when computing sample

size-dependent information criteria is consistent across candidate models. The third step is to compare the candidate models by ranking them based on the information criteria being used. The model with the lowest value (i.e. closest to zero) is considered to be the “best” model. Burnham and Anderson (2002) likened this to a car race; the fastest car is awarded first place, the second-fastest car is awarded second place, and so on.

INFORMATION CRITERIA USED IN SAS/STAT PROCEDURES

Information criteria appear in the output of many SAS/STAT procedures. To better determine their prevalence, I conducted a focused search of the most recent documentation for each of the procedures listed in the SAS/STAT 14.3 User’s Guide (SAS Institute Inc., 2017). I used eight search terms (“aic”, “aicc”, “akaike”, “bic”, “caic”, “deviance”, “information criteria”, and “hqic”) to help guide my reading of each procedure’s documentation. To be included, a procedure needed to have the capability, by default or by specifying an option, of computing and displaying at least one information criterion.

I found 33 SAS/STAT procedures that provided at least one information criterion, all but three of which provided two or more. Thirty-one of these procedures and their associated information criteria are listed in Table 2. The most common information criterion was AIC (30), followed closely by BIC (27) and AICC (21). Nineteen procedures included all three. CAIC (4) and HQIC (3) appeared much less often, and were always provided along with AIC, AICC, and/or BIC. It is important to note that these information criteria are used for models estimated using maximum likelihood methods. There is an information criterion, DIC, that is used for models estimated using Bayesian methods. DIC is offered by six procedures total. Four of the six procedures are listed in Table 2 (indicated by an asterisk), which is possible because these procedures offer both maximum likelihood and Bayesian estimation methods. The two procedures that offer only DIC (the BCHOICE and MCMC procedures, not shown in Table 2) do so because the procedures exclusively use Bayesian estimation methods.

Procedure	AIC	BIC	AICC	CAIC	HQIC	Procedure	AIC	BIC	AICC	CAIC	HQIC
CALIS	Yes	Yes	No	Yes	No	IRT	Yes	Yes	No	No	No
FACTOR	Yes	Yes	No	No	No	LIFEREG*	Yes	Yes	Yes	No	No
FMM*	Yes	Yes	Yes	No	No	LOESS	No	No	Yes	No	No
GAMPL	Yes	Yes	Yes	No	No	LOGISTIC	Yes	Yes	Yes	No	No
GENMOD	Yes	Yes	Yes	No	No	MIXED	Yes	Yes	Yes	Yes	Yes
GLIMMIX	Yes	Yes	Yes	Yes	Yes	NLMIXED	Yes	Yes	Yes	No	No
GLMSELECT	Yes	Yes	Yes	No	No	PHREG*	Yes	Yes	No	No	No
HPFMM*	Yes	Yes	Yes	No	No	QUANTSELECT	Yes	Yes	Yes	No	No
HPGENSELECT	Yes	Yes	Yes	No	No	REG	Yes	Yes	No	No	No
HPLMIXED	Yes	Yes	Yes	No	No	ROBUSTREG	Yes	Yes	No	No	No
HPLOGISTIC	Yes	Yes	Yes	No	No	SPP	Yes	Yes	No	No	No
HPMIXED	Yes	Yes	Yes	Yes	Yes	SURVEYLOGISTIC	Yes	Yes	No	No	No
HPNLMOD	Yes	Yes	Yes	No	No	SURVEYPHREG	Yes	No	No	No	No
HPQUANTSELECT	Yes	Yes	Yes	No	No	TRANSREG	Yes	No	Yes	No	No
HPREG	Yes	Yes	Yes	No	No	VARIOGRAM	Yes	No	No	No	No
ICPHREG	Yes	Yes	Yes	No	No	<i>Criterion Total</i>	30	27	21	4	3

Table 2. Information Criteria Provided by SAS/STAT 14.3 Procedures

OBTAINING INFORMATION CRITERIA FROM PROCEDURE OUTPUT

This section demonstrates three methods to obtain information criteria for models fitted using the MIXED procedure. All three methods use the Output Delivery System (ODS) OUTPUT statement to capture relevant portions of the procedure's output, called tables, and turn them into data sets. These data sets are then processed and combined using a series of DATA steps to produce a tidy one-row data set, displaying all of the information criteria for a single model. The first two methods use one ODS OUTPUT statement to obtain information criteria that are provided by the procedure in a single table. The third method uses multiple ODS OUTPUT statements to obtain the tables that contain the component parts of the information criteria formulae (log likelihood, number of parameters, and sample size) to compute them manually.

Although all the code provided in this section is specific to PROC MIXED, it was written with adaptability to other SAS/STAT procedures in mind. To adapt the code to other procedures, the first step is to determine the names of the relevant tables produced by that procedure. This could be done either by looking for the table names in the procedure documentation or by running the model with ODS TRACE statements and looking for the table information in the log. The next step is to determine the best way to isolate the necessary table elements. Some trial and error may be needed to obtain elements of certain tables, but the processing and compilation code used in these examples is a good starting point.

The data and model used for this demonstration come from Singer (1998) and a UCLA Statistical Consulting Group seminar based on the same. The data set (HSB12) contains information from 7185 students in 160 schools (SCHOOL). The outcome of interest is the math achievement (MATHACH) for each of these students. The focus of this section is how to obtain information criteria from a single model: the unconditional means model. For more detail on the interpretation of this model, see Singer (1998).

EXAMPLE 1.1: PULLING INFORMATION CRITERIA PRODUCED BY DEFAULT

As of SAS/STAT 14.3, PROC MIXED computes and displays AIC, AICC, and BIC by default. As shown in Display 1, these information criteria are displayed in a table labeled Fit Statistics. If this model is run with ODS TRACE statements included, the log shows that the ODS table name for this table is FitStatistics.

The screenshot displays the SAS environment with three main windows:

- Example1 Codes.sas:** Contains the PROC MIXED code with ODS TRACE statements.
- Log - (Untitled):** Shows the output of the ODS TRACE statements, identifying the table name as FitStatistics.
- Results Viewer - SAS Output:** Displays the model results, including the Fit Statistics table.

The Fit Statistics table from the Results Viewer is as follows:

Criterion	Value
-2 Res Log Likelihood	47116.8
AIC (smaller is better)	47120.8
AICC (smaller is better)	47120.8
BIC (smaller is better)	47126.9

Display 1. ODS Table Name for Default Information Criteria Shown by ODS TRACE statements

Knowing this, the following code is used to create a new data set (called IC_default), which contains only the contents of the Fit Statistics table.

```

proc mixed data=SASGF.hsb12 covtest noclprint;
  class school;
  model mathach = /solution;
  random intercept / subject = school;
  ods output FitStatistics=IC_default;
run;

```

Next, the different information criteria are isolated into three separate data sets using a series of DATA steps. The code below shows how this is done for AIC. A new variable (AIC) is created by storing the value contained next to the description "AIC (smaller is better)". Note that the quoted portion, including the parentheses, exactly matches the row heading next to the AIC value on the Fit Statistics table.

```

data ModelAIC;
  set IC_default;
  if Descr="AIC (smaller is better)";
  AIC=Value;
  keep AIC;
run;

```

Once the three data sets containing isolated information criteria are created, they are merged into a single data set called Model_default. The contents of this data set are shown in Display 2. An additional variable is created (NAME) to give the model a descriptive name. The model name is superfluous when looking at information criteria for a single model, but it will become important when comparing information criteria from a set of models (as in Example 2). The values of the information criteria shown in Display 2 are slightly different than those shown in Display 1, but this is only because of rounding.

```

data Model_default;
  merge ModelAIC ModelAICC ModelBIC;
  NAME = "Unconditional Model";
run;

```

The screenshot shows a window titled "Results Viewer - SAS Output". Inside the window, there is a heading "Default information criteria produced for the unconditional means model" in blue. Below the heading is a table with the following data:

Obs	NAME	AIC	AICC	BIC
1	Unconditional Model	47120.79	47120.80	47126.94

Display 2. Data Set Containing Information Criteria Produced by Default

EXAMPLE 1.2: PULLING INFORMATION CRITERIA PRODUCED BY SPECIFICATION

In addition to the information criteria provided by default, PROC MIXED can also produce CAIC and HQIC by request. To do this, the IC option must be specified in the PROC MIXED statement. As shown in Display 3, the additional information criteria are shown in a table labeled Information Criteria. If this model is run with ODS TRACE statements included, the log shows that the ODS table name is InfoCrit. The Information Criteria table, however, does not replace the Fit Statistics table in the output. Display 3 also demonstrates that the values for AIC, AICC, and BIC in the Information Criteria table match those in the Fit Statistics table.

Example1 Code.sas *

```

/*****
/* Example 1.2 - Method 2: Pulling expanded information criteria in MIXED */
*****/

/* Unconditional means model - note the addition of the IC option */

ods trace on;
proc mixed data = SASGF.hsb12 covtest noclprint IC;
  class school;
  model mathach = / solution;
  random intercept / subject = school;
run;
ods trace off;

```

Log - (Untitled)

```

Label:      Fit Statistics
Template:   Stat.Mixed.FitStatistics
Path:       Mixed.FitStatistics
-----
Output Added:
Name:       InfoCrit
Label:      Information Criteria
Template:   Stat.Mixed.InfoCrit
Path:       Mixed.InfoCrit
-----
Output Added:
Name:       SolutionF
Label:      Solution for Fixed Effects
Template:   Stat.Mixed.SolutionF
Path:       Mixed.SolutionF

```

Results Viewer - SAS Output

Residual	39.1487	0.6607	59.26	<.0001
----------	---------	--------	-------	--------

Fit Statistics	
-2 Res Log Likelihood	47116.8
AIC (smaller is better)	47120.8
AICC (smaller is better)	47120.8
BIC (smaller is better)	47126.9

Information Criteria						
Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
47116.8	2	47120.8	47120.8	47123.3	47126.9	47128.9

Display 3. ODS Table Name for Expanded Information Criteria Shown by ODS TRACE statements

Knowing this, the following code is used to create a new data set (called IC_expanded), consisting of just the contents of the Information Criteria table.

```

proc mixed data=SASGF.hsb12 covtest noclprint IC;
  class school;
  model mathach = /solution;
  random intercept / subject = school;
  ods output InfoCrit=IC_expanded;
run;

```

Unlike the Fit Statistics table, which has row headers, the Information Criteria table has column headers. Because of this, it is not necessary to use multiple DATA steps to create a one-row data set (as was done in Example 1.1). The contents of Model_expanded are shown in Display 4. An additional variable is created (NAME) to give the model a descriptive name, which will be useful when comparing models.

```

data Model_expanded;
  set IC_expanded;
  NAME = "Unconditional Model";
  keep NAME AIC AICC BIC HQIC CAIC;
run;

```

Results Viewer - SAS Output

Expanded information criteria produced for the unconditional means model

Obs	NAME	AIC	AICC	BIC	CAIC	HQIC
1	Unconditional Model	47120.8	47120.8	47126.9	47128.9	47123.3

Display 4. Data Set Containing Information Criteria Produced by Specification

EXAMPLE 1.3: COMPUTING INFORMATION CRITERIA MANUALLY

There are two sorts of occasions where I have needed to compute information criteria manually. Some have been due to the procedure, such as computing CAIC for a model estimated in the NL MIXED procedure, which does not produce CAIC. Others have been due to considerations particular to the analysis method. For example, multilevel models require decisions to be made about each part of the information criteria formulae. In practice, these decisions are often made through software defaults. The details about these decisions and their consequences for multilevel model selection are outside the scope of this paper; for further discussion, see Gurka (2006).

For demonstration purposes, this example will focus on the implementation of one of these decisions. Multilevel models have two valid sample sizes to consider when computing sample size-dependent information criteria: the total number of observations (N) and the total number of clusters (m). Simulation studies have been conducted to explore how the choice of these influences the performance of different information criteria (e.g. Whittaker & Furlow, 2009). AIC is not sample size-dependent, so its computation does not change. AICC, BIC, CAIC, and HQIC are sample size-dependent, so their computation changes depending on whether N or m is chosen. In this example, nine information criteria are manually computed: AIC, and the N -based and m -based versions of AICC, BIC, CAIC, and HQIC.

As in the previous two examples, ODS OUTPUT statements are used to turn PROC MIXED tables containing the model's log likelihood, the number of model parameters, and the two sample sizes into separate data sets. Display 5 shows the Fit Statistics table, which will be used in this example to isolate the log likelihood. Display 5 shows the Dimensions table and the Number of Observations table. The Dimensions table contains the number of parameters PROC MIXED uses as part of its internal computation of the information criteria. It also contains the m -based sample size (160), which is next to the Subjects row heading. The Number of Observations table contains the N -based sample size (7185).

```
proc mixed data=SASGF.hsb12 covtest noclprint;
  class school;
  model mathach = /solution;
  random intercept / subject = school;
  ods output Dimensions=parms_m_sample;
  ods output Nobs=N_sample;
  ods output FitStatistics=IC_default;
run;
```

The screenshot shows the SAS Results Viewer interface. On the left, the 'Log - (Untitled)' window displays the following 'Output Added' messages:

- Dimensions:** Name: Dimensions, Label: Dimensions, Template: Stat.Mixed.Dimensions, Path: Mixed.Dimensions
- NObs:** Name: NObs, Label: Number of Observations, Template: Stat.Mixed.NObs, Path: Mixed.NObs
- IterHistory:** Name: IterHistory, Label: Iteration History, Template: Stat.Mixed.IterHistory, Path: Mixed.IterHistory

On the right, the 'Results Viewer - SAS Output' window displays three tables:

Dimensions	
Covariance Parameters	2
Columns in X	1
Columns in Z Per Subject	1
Subjects	160
Max Obs Per Subject	67

Number of Observations	
Number of Observations Read	7185
Number of Observations Used	7185
Number of Observations Not Used	0

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion

Display 5. ODS Table Names for Tables Containing the Number of Parameters, m , and N

Next, the relevant component parts of the tables are isolated into four separate data sets – Model_Likelihood, Parmns, N, and m – using a series of DATA steps. The code below shows how this is done for the number of parameters. Similar to Example 1.1, a new variable (Parmns) is created by storing the value contained next to the row header “Covariance Parameters”. Note that the quoted portion exactly matches the appropriate row heading in in the Dimensions table, as shown in Display 5.

```

data Model_Parms;
  set parms_m_sample;
  if Descr="Covariance Parameters";
  Parm=Value;
  keep Parm;
run;

```

The component parts are merged into one data set called Model_manual. After the MERGE statement, ten new variables are computed. The first one is a model name. The remaining ones are the formulae for the different information criteria, as shown in Table 1. The sample size-dependent information criteria have variable names that indicate which sample size was used in their computation. The values of the information criteria computed in this way are shown in Display 6.

```

data Model_manual;
  merge Model_Likelihood Model_Parms N m;
  NAME = "Unconditional Model";
  AIC = Likelihood + (2*Parm);
  AICC_N = Likelihood + (2*Parm)*(N/(N-Parm-1));
  AICC_m = Likelihood + (2*Parm)*(m/(m-Parm-1));
  BIC_N = Likelihood + log(N)*Parm;
  BIC_m = Likelihood + log(m)*Parm;
  CAIC_N = Likelihood + (log(N)+1)*Parm;
  CAIC_m = Likelihood + (log(m)+1)*Parm;
  HQIC_N = Likelihood + (2*Parm)*(log(log(N)));
  HQIC_m = Likelihood + (2*Parm)*(log(log(m)));
run;

```

The screenshot shows a window titled "Results Viewer - sashtml7.htm" with a table of information criteria. The table has 11 columns: Obs, NAME, AIC, AICC_N, AICC_m, BIC_N, BIC_m, CAIC_N, CAIC_m, HQIC_N, and HQIC_m. There is one row of data for the "Unconditional Model".

Obs	NAME	AIC	AICC_N	AICC_m	BIC_N	BIC_m	CAIC_N	CAIC_m	HQIC_N	HQIC_m
1	Unconditional Model	47120.79	47120.80	47120.87	47134.55	47126.94	47136.55	47128.94	47125.53	47123.29

Display 6. Data Set Containing Information Criteria Computed Manually

While discussion of multilevel modeling-specific concerns when computing information criteria is outside the scope of this paper, it is worth noting that the above code was written with another issue discussed by Gurka (2006) in mind. When PROC MIXED uses its default estimation method (REML), it only counts the number of covariance parameters (i.e. random effects). Gurka's (2006) simulation studies, however, counted fixed effects as well as random effects. The above code can be adapted to match this in one of two ways. The first way is to adapt the DATA step that creates the Model_Parms data set to make the Parm variable retain the appropriate number of parameters. The second way is to re-define the Parm variable in the code that creates Model_manual; that is, remove the Model_Parms data set from the MERGE statement and add a new line to set the value of the Parm variable to a constant (e.g. Parm=3, where 3 is the appropriate number of parameters).

The documentation for PROC MIXED indicates that default sample size for BIC, CAIC, and HQIC is *m*-based, while the default sample size for AICC is *N*-based. This can be demonstrated by merging Model_expanded (from Example 1.2) and Model_manual, then using the PRINT procedure to selectively print the joined information criteria. In the upper panel of Display 7, the information criteria computed by PROC MIXED are the same (within rounding) as the ones that were manually computed. In contrast, the lower panel shows that the information criteria computed with the non-default sample sizes are slightly different (even accounting for rounding) than those computed by PROC MIXED.

Comparing the manually-computed information criteria that match those produced by MIXED

Obs	NAME	AICC	AICC_N	BIC	BIC_m	CAIC	CAIC_m	HQIC	HQIC_m
1	Unconditional Model	47120.8	47120.80	47126.9	47126.94	47128.9	47128.94	47123.3	47123.29

Comparing the manually-computed information criteria that do not match those produced by MIXED

Obs	NAME	AICC	AICC_m	BIC	BIC_N	CAIC	CAIC_N	HQIC	HQIC_N
1	Unconditional Model	47120.8	47120.87	47126.9	47134.55	47128.9	47136.55	47123.3	47125.53

Display 7. Comparing Information Criteria Computed Manually with PROC MIXED Computations

COLLECTING INFORMATION CRITERIA FOR MODEL COMPARISON

Model selection using information criteria requires that the models be ranked relative to each other, but the display of the information criteria values is often inconvenient for this purpose. Manually recording the information criteria for each model and manually ranking the models is not overly difficult if there are only two or three candidate models. However, such methods quickly become unwieldy as more models are compared. If one of the three methods shown in the previous section is used, the RANK procedure offers an easy and convenient way to determine the “best” model based on any information criterion.

This demonstration is a natural continuation of the previous one. The data and model specifications used come from Singer (1998) and a UCLA Statistical Consulting Group seminar based on the same. The data set (HSB12) contains information from 7185 students in 160 schools (SCHOOL). The outcome of interest is the math achievement (MATHACH) for each of these students. There are two predictors at the school level (MEANSES and SECTOR) and one predictor at the student level (CSES). Four models were fit: an unconditional means model, a model with one school-level predictor, a model with the student-level predictor, and a model with all three predictors and cross-level interactions. The focus of this section is how to compare these four models using information criteria; for more detail on the interpretation and substantive meaning of these models, see Singer (1998).

EXAMPLE 2: COMPARING MODELS USING EXPANDED INFORMATION CRITERIA

In this example, the four candidate models are compared based on the expanded set of information criteria that can be obtained by specifying the IC option in the PROC MIXED statement. To start, some small changes are made to the code shown in Example 1.2. In order to conserve space, the adapted code is shown for the first candidate model only. The code shown below estimates the unconditional means model and outputs the contents of that model’s Information Criteria table (as shown in Display 4), creating a new data set called IC_expanded_1. The same is done for the other three models, creating new data sets called IC_expanded_2, IC_expanded_3, and IC_expanded_4, respectively.

```
proc mixed data = SASGF.hsb12 covtest noclprint IC;
  class school;
  model mathach = / solution;
  random intercept / subject = school;
  ods output InfoCrit=IC_expanded_1;
run;
```

As in Example 1.2, these new data sets are already properly shaped for compilation, but do not yet have names. A new data set, Model_expanded_1, includes a variable to give this model a name. The code shown below is for the model named “Unconditional model”, and there are 19 characters in this name. Similar code for the other candidate models creates data sets called Model_expanded_2 (“School-level

predictor”, 21 characters), Model_expanded_3 (“Student-level predictor”, 23 characters), and Model_expanded_4 (“Cross-level interaction”, 23 characters).

```
data Model_expanded_1;
  set IC_expanded_1;
  NAME = "Unconditional Model";
  keep NAME AIC AICC BIC HQIC CAIC;
run;
```

At this point, all of the Model_expanded data sets are compiled into a single data set called Model_compile. To incorporate the full name of each data set into Model_compile, the number of characters of the longest NAME variable needs to be included in the LENGTH statement. In this example, the longest model name was 23 characters, so the number 23 is next to the \$ symbol. The contents of this data set are shown in Display 8.

```
data Model_compile;
  length NAME $ 23;
  set Model_expanded_1 Model_expanded_2 Model_expanded_3 Model_expanded_4;
run;
```

Obs	NAME	AIC	AICC	HQIC	BIC	CAIC
1	Unconditional model	47120.8	47120.8	47123.3	47126.9	47128.9
2	School-level predictor	46965.3	46965.3	46967.8	46971.4	46973.4
3	Student-level predictor	46722.2	46722.2	46727.2	46734.5	46738.5
4	Cross-level interaction	46511.7	46511.7	46516.7	46524.0	46528.0

Display 8. Information Criteria for Four Candidate Models

Now that the candidate models and their respective information criteria are in the same data set, PROC RANK can be used to create ranks based on each information criterion. Rankings for all five of these information criteria are made using a single VAR statement, and new variables indicating the ranks for each are created in a single RANKS statement. A new data set called IC_ranks is created, which contains both the information criteria values and the ranks produced in PROC RANK. The contents of this data set are shown in Display 9.

```
proc rank data=Model_compile out=IC_ranks;
  var AIC AICC BIC CAIC HQIC;
  ranks Rank_AIC Rank_AICC Rank_BIC Rank_CAIC Rank_HQIC;
run;
```

Obs	NAME	AIC	AICC	HQIC	BIC	CAIC	Rank_AIC	Rank_AICC	Rank_BIC	Rank_CAIC	Rank_HQIC
1	Cross-level interaction	46511.7	46511.7	46516.7	46524.0	46528.0	1	1	1	1	1
2	Student-level predictor	46722.2	46722.2	46727.2	46734.5	46738.5	2	2	2	2	2
3	School-level predictor	46965.3	46965.3	46967.8	46971.4	46973.4	3	3	3	3	3
4	Unconditional model	47120.8	47120.8	47123.3	47126.9	47128.9	4	4	4	4	4

Display 9. Information Criteria and Rankings for Four Candidate Models, “Best” Model First

In this example, the model with a cross-level interaction has the lowest value for each information criterion. Given this, a researcher using information criteria to compare these four candidate models would conclude that the model with the cross-level interaction is the “best” model.

CONCLUSION

Information criteria offer a flexible framework for model comparison, and many SAS/STAT procedures include information criteria as part of their output. If a model can be estimated using a SAS/STAT procedure, thoughtful application of Base SAS functionality makes obtaining information criteria, collecting information criteria from many models, and ranking models on that basis easy and convenient.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.
- Burnham, K. P., & Anderson, D. R. (2003). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). New York, NY: Springer.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003) *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.
- Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician*, 60(1), 19-26.
- Hamaker, E. L., van Hattum, P., Kuiper, R. M., & Hoijtink, H. (2011). Model selection based on information criteria in multilevel modeling. In Hox, J. J. & Roberts, J. K (Eds.), *Handbook of advanced multilevel analysis* (pp. 231-255). New York, NY: Routledge.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 190-195.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 297-307.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.
- SAS Institute Inc. (2017). SAS/STAT® 14.3 user's guide. Retrieved from <http://documentation.sas.com/?docsetId=statug&docsetTarget=titlepage.htm&docsetVersion=14.3&locale=en>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23(4), 323-355.
- UCLA: Statistical Consulting Group. (n.d). Statistical computing seminars: Introduction to multilevel modeling using SAS. Retrieved from https://stats.idre.ucla.edu/sas/seminars/mlm_sas_seminar/
- Whittaker, T. A., & Furlow, C. F. (2009). The comparison of model selection criteria when selecting among competing hierarchical linear models. *Journal of Modern Applied Statistical Methods*, 8(1), 173-183.

ACKNOWLEDGMENTS

I would like to thank Dr. Han Du for her generous assistance with contextualizing DIC in relation to the other information criteria described.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Wendy Christensen
wchristensen@ucla.edu
<https://github.com/wendychristensen>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.