SAS® GLOBAL FORUM 2018

USERS PROGRAM

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF

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

Presenter

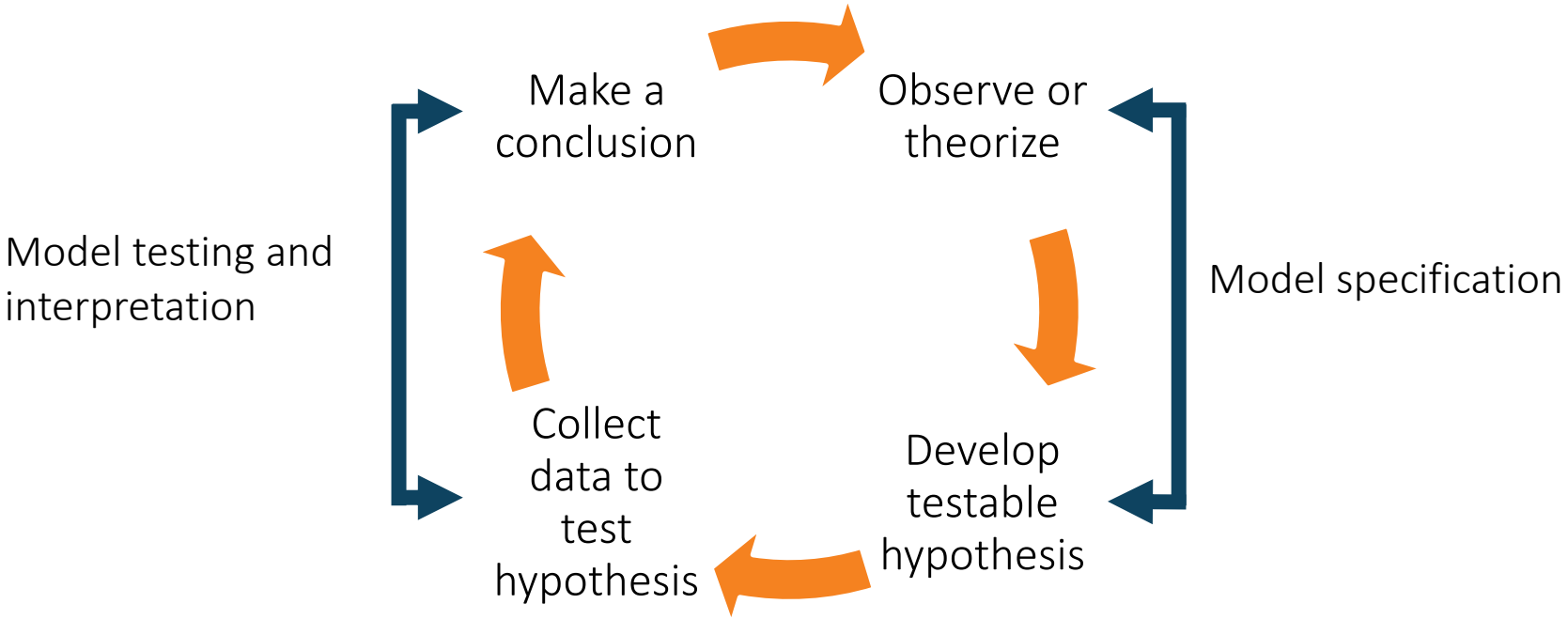
Wendy Christensen, Ph.D. candidate, University of California, Los Angeles

Wendy Christensen is a doctoral candidate in quantitative psychology at University of California, Los Angeles. Her research interests include statistical model selection, research design and data collection methods in behavioral research, and the role of psychological, behavioral, and cultural factors in health outcomes.

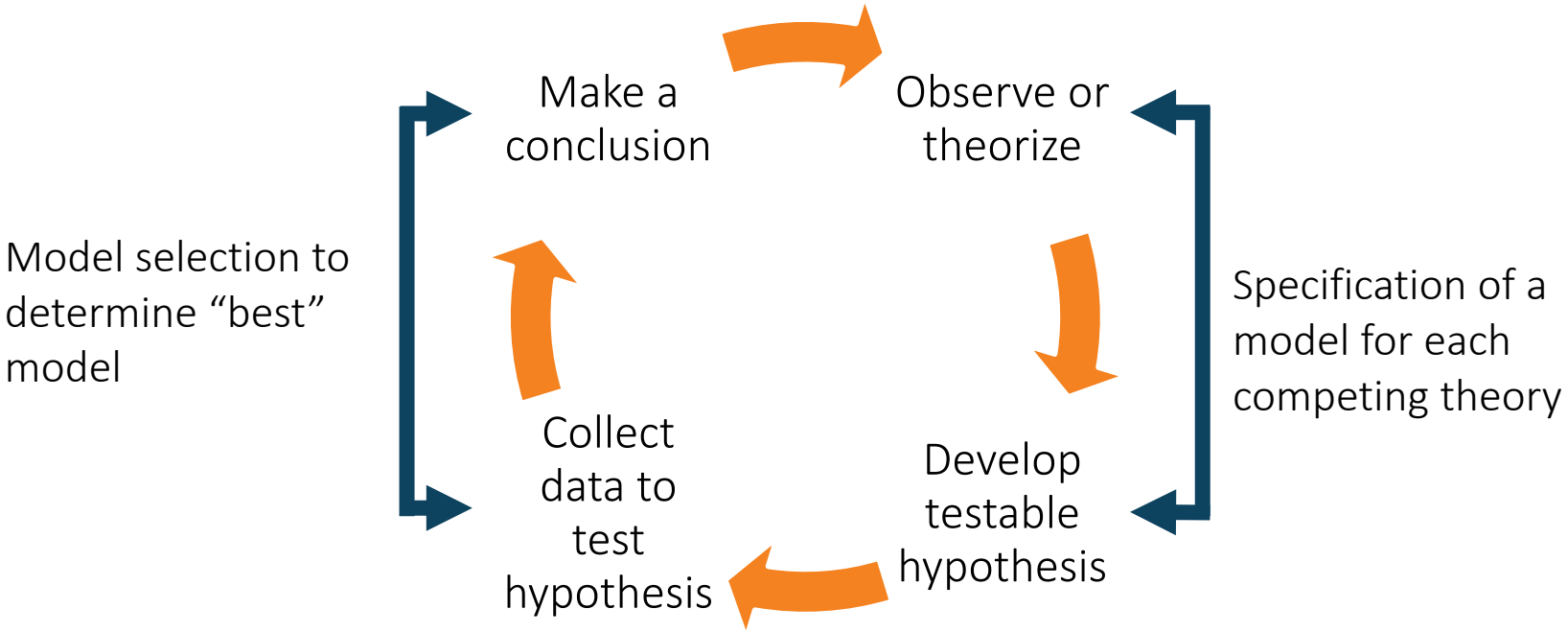
@wchristensen

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

The Scientific Method: Single Model



The Scientific Method: Multiple Models



Model Selection Using Null Hypothesis Testing

Models must be nested, such that the *reduced* model can be constructed by removing parameters from the *full* model. The reduced model acts as the null model, and rejecting it means that the added parameters in the full model significantly improve the model[1].

R^2 change test

F test of R^2 difference between two models

$$F_{test} = \frac{\frac{R^2_{full} - R^2_{red}}{df_{full} - df_{red}}}{\frac{1 - R^2_{full}}{df_{full}}}$$

Likelihood ratio test

χ^2 test of likelihood difference (or change in deviance) between two models

$$\chi^2_{test} = -2LL_{red} - (-2LL_{full})$$

Deviances

Model Selection Using Null Hypothesis Testing

Advantages

- Reasonably intuitive
- Simple to conduct
- Benefits of a parametric test

Disadvantages

- Models must be nested
- Only two models at a time
- Requires a null model be designated as the comparison model

There are situations where model selection using null hypothesis testing methods is inappropriate or undesirable

An Alternative: Information Criteria

Information criteria offer a flexible method of model selection, and can be used in situations for which null hypothesis testing methods are unsuitable[2].

- Models only need to have the outcome variable in common (i.e. models can be non-nested)
- Can simultaneously compare any number of models
- No need to designate null models for comparison
- Somewhat less intuitive, discussion about meaning is often limited to “smaller is better”

Model Selection Using Information Criteria

Model selection using information criteria is a four-step process[3]:

1) Define the set of models to be compared
2) Fit each model to the same data, using outcome *Y* for all models
3) Compute one or more information criteria for each model
4) Assign each model a rank by comparing the values of the information criteria. The model with the lowest value is considered the "best" model – but why?

Kullback-Liebler (K-L) Distance

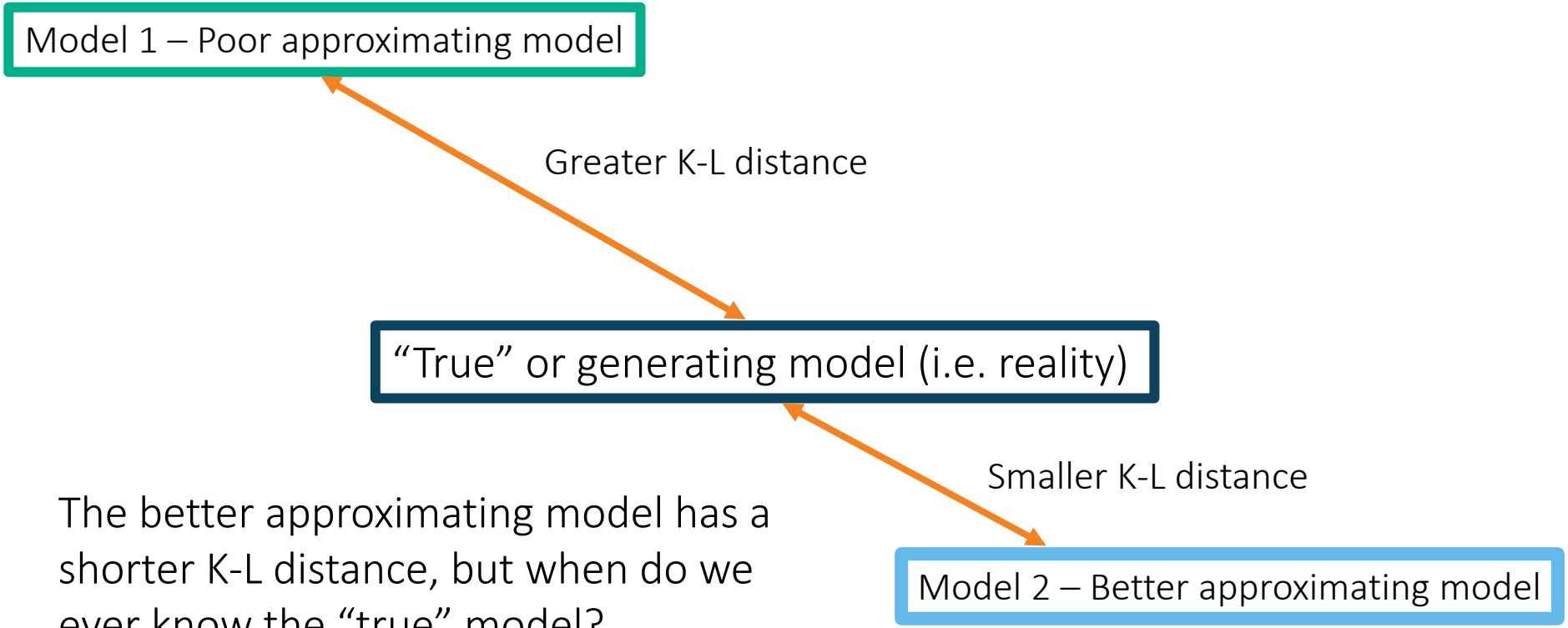
K-L distance refers to the distance/discrepancy between two probability distributions, and is understood as the information lost when using one distribution to approximate the other[3].

Distributions are identical → K-L distance = 0

Distributions are not identical → K-L distance > 0

Greater K-L distances imply greater divergence in the distributions.

Absolute K-L Distances



The better approximating model has a shorter K-L distance, but when do we ever know the "true" model?

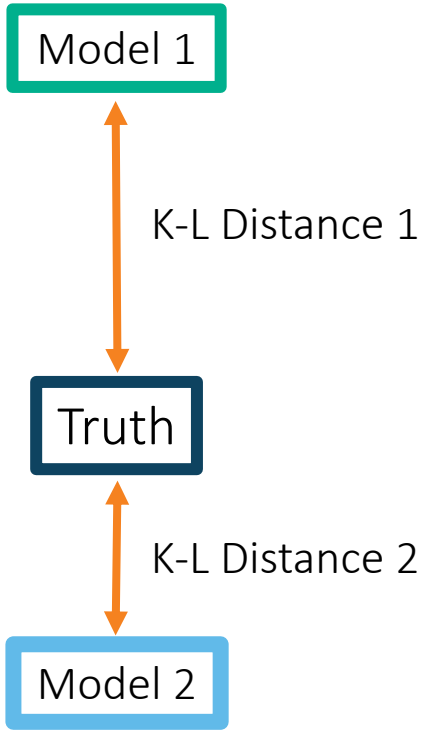
Relative K-L Distance

KL_1 = Truth – Model 1 → KL_1 – Truth = – Model 1

KL_2 = Truth – Model 2 → KL_2 – Truth = – Model 2

The “true” model is a constant across models!

Even though *absolute* K-L distances cannot be computed when the “true” model is unknown, *relative* K-L distances can be computed for each candidate model. This means that models can be ranked relative to each other[3].



The Akaike Information Criterion (AIC)

Akaike[4] developed the first information criterion by linking maximum likelihood estimation to K-L distance – specifically, a model's likelihood estimate (via maximum likelihood estimation) is related to the expected value of the relative K-L distance.

$$\text{AIC: } -2LL + 2K$$

Deviance (log-likelihood * -2)
"Complexity"

Number of model parameters
"Parsimony"

The model that best balances these will have the smallest AIC value, which is why "smaller is better".

Other Information Criteria

After the introduction of AIC, more information criteria were developed with differing mathematical properties and philosophies of model selection in mind (e.g. efficient vs. consistent criteria). In practice, they are often used in together for model selection purposes.

Efficient criteria	AIC: $-2LL + 2K$[4]	AICC: $-2LL + 2K\left(\frac{n}{n-K-1}\right)$[5]	
Consistent criteria	BIC: $-2LL + \ln(N)K$[6]	CAIC: $-2LL + [\ln(n) + 1]K$[7]	HQIC: $-2LL + 2K\ln(\ln(n))$[8]

Information Criteria in SAS/STAT® Procedures[9]

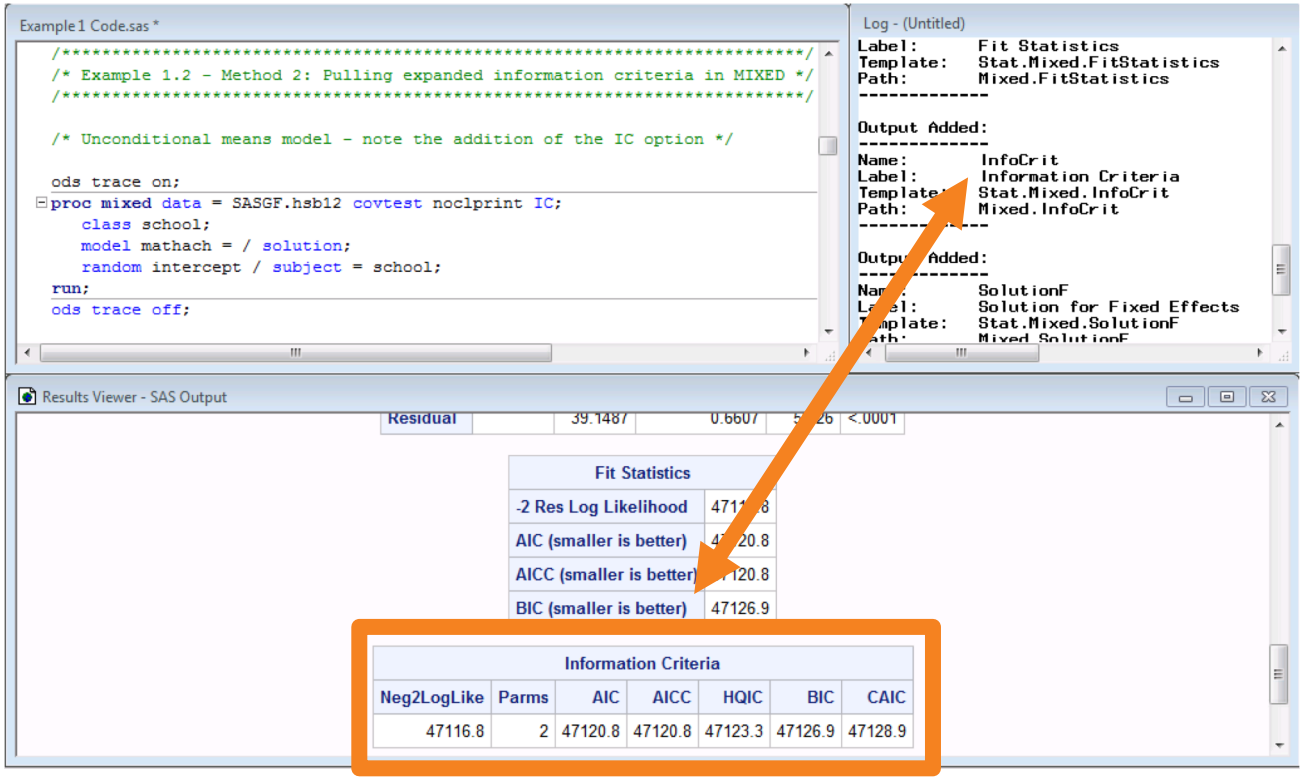
Procedure	AIC	BIC	AICC	CAIC	HQIC	Procedure	AIC	BIC	AICC	CAIC	HQIC
CALIS	Yes	Yes	No	Yes	No	IRT	Yes	Yes	No	No	No
FACTOR	Yes	Yes	No	No	No	LIFEREG	Yes	Yes	Yes	No	No
FMM	Yes	Yes	Yes	No	No	LOESS	No	No	Yes	No	No
GAMPL	Yes	Yes	Yes	No	No	LOGISTIC	Yes	Yes	Yes	No	No
GENMOD	Yes	Yes	Yes	No	No	MIXED	Yes	Yes	Yes	Yes	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	Criterion Total	30	27	21	4	3

Using Information Criteria: Made Easy in SAS

Four-step process of model selection:

1) Define the set of models to be compared.
2) Use the appropriate procedures to fit the models (with the same outcome *Y*) to the same data.
3) Use ODS statements and DATA steps to obtain information criteria from each model's output OR compute them manually for each model.
4) Use the RANK procedure to compute the relative rank of the models based on each information criteria of interest

Step 3: Information Criteria from Procedure Output



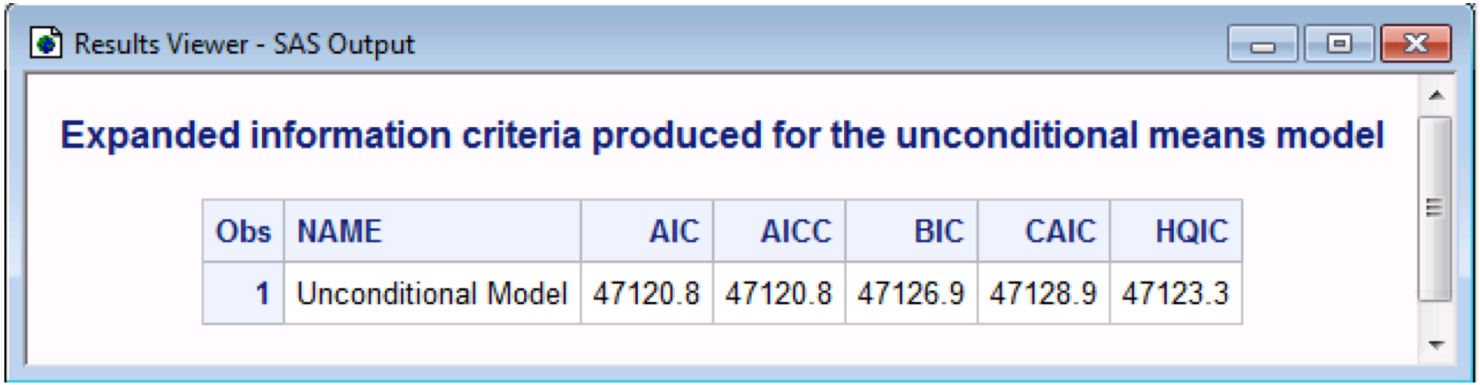
Step 3: Information Criteria from Procedure Output

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

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

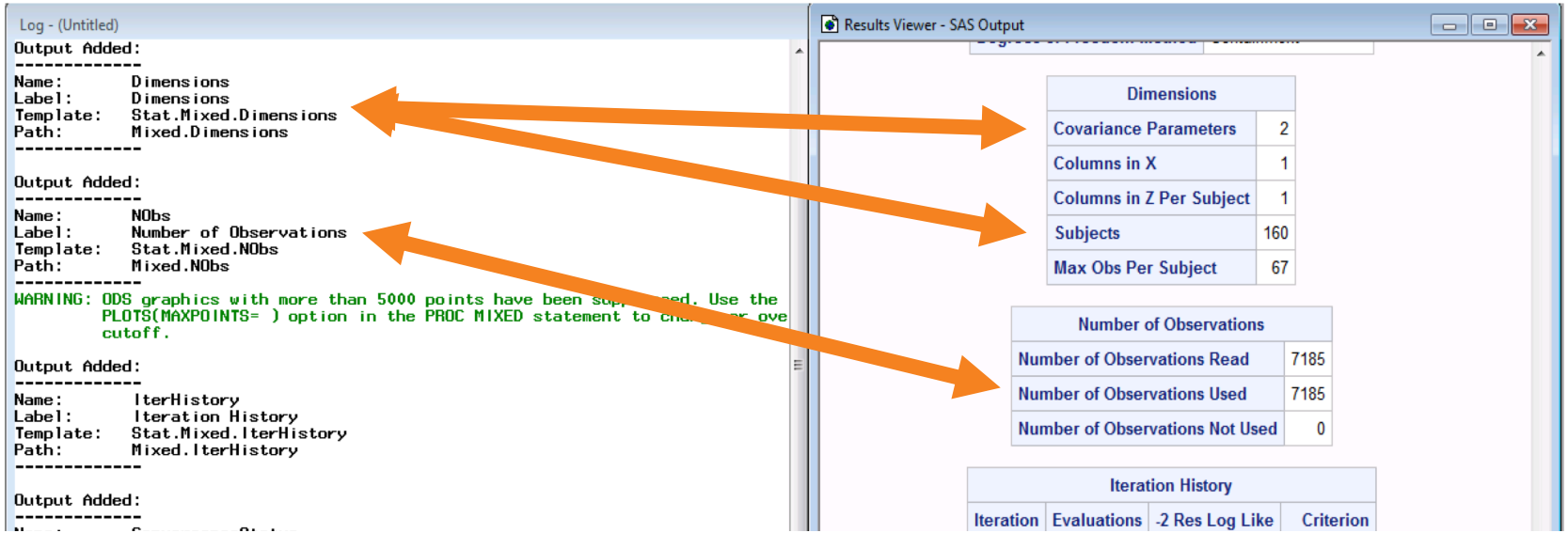
Step 3: Information Criteria from Procedure Output

If Model_expanded is printed, a tidy one-row data set containing the information criteria produced by the procedure and a model name is shown. This process is repeated for each candidate model.



Step 3: Computing Information Criteria Manually

Even if a procedure does not supply the desired information criteria, the output can still be used to pull the component parts of the formulae: deviance, number of parameters, and sample size



Step 3: Computing Information Criteria Manually

To compute the information criteria of interest, the different components need to be isolated. This is done with a series of DATA steps.

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

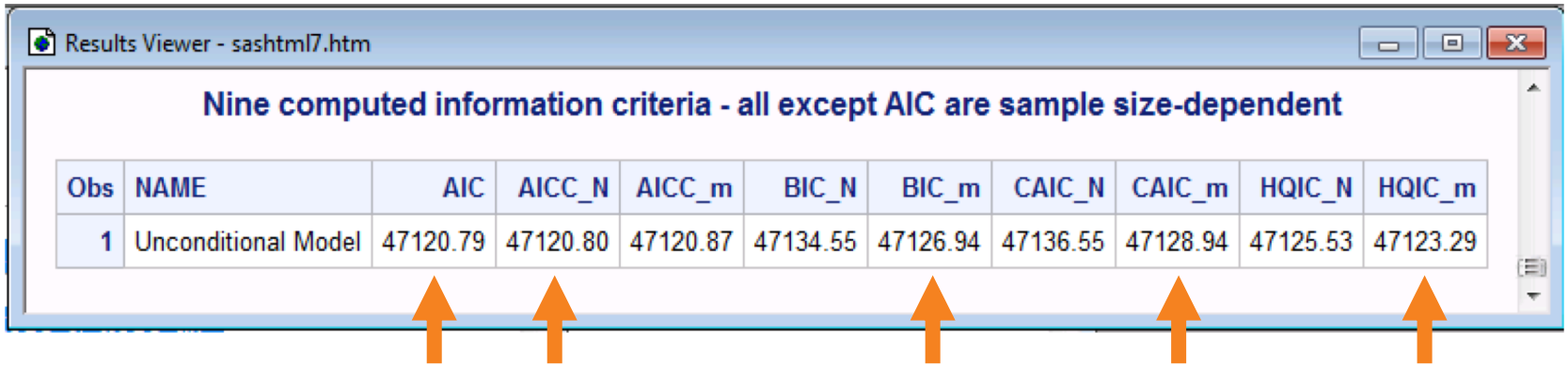
The parms_m_sample data set contains two component parts, but this DATA step isolates the number of parameters (*K*).

Step 3: Computing Information Criteria Manually

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

Step 3: Computing Information Criteria Manually

The final data set contains nine manually-computed information criteria.

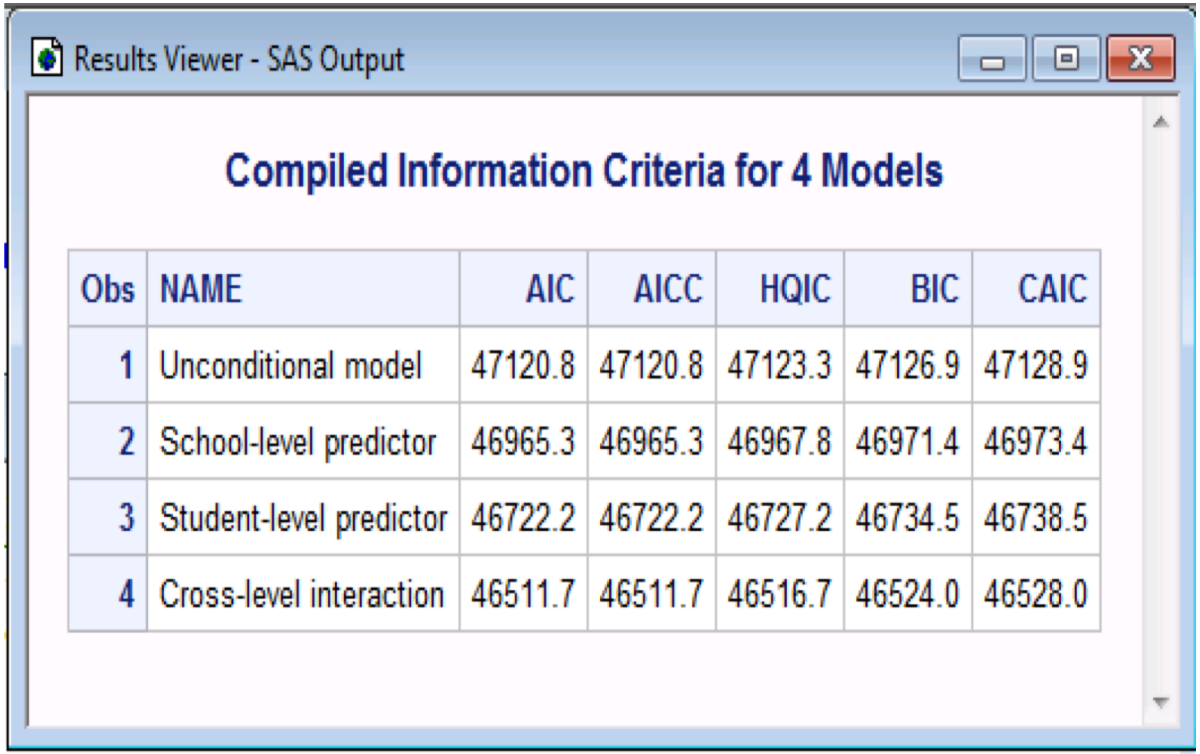


Obs	NAME	AIC	AICC_N	AICC_m	BIC_N	BIC_m	CAIC_N	CAIC_m	HQIC_N	HQIC_m
1	Unconditional Model	47120.79	47120.80	47120.87	47134.55	47126.94	47136.55	47128.94	47125.53	47123.29

PROC MIXED uses *N* for AICC and *m* for BIC, CAIC, and HQIC. The ones with arrows below them will match those produced by MIXED (apparent differences are due to rounding in the output).

Step 4: Finding the “best” model using PROC RANK

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



Compiled Information Criteria for 4 Models

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

Step 4: Finding the “best” model using PROC RANK

```
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;
```

Results Viewer - SAS Output

4 Candidate Models with Ranks

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

References

1. 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
2. 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.
3. Burnham, K. P., & Anderson, D. R. (2003). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). New York, NY: Springer.
4. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723.
5. Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 297-307.
6. Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.
7. Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345-370.
8. 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.
9. 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

Your Feedback Counts!

Don't forget to complete the session survey
in your conference mobile app.

1. Go to the Agenda icon in the conference app.
2. Find this session title and select it.
3. On the Sessions page, scroll down to Surveys and select the name of the survey.
4. Complete the survey and click Finish.