Title

estat ic — Display information criteria

Description

`estat ic` computes Akaike's (AIC), consistent Akaike's (CAIC), corrected Akaike's (AICc), and Schwarz's Bayesian (BIC) information criteria.

Quick start

Display AIC and BIC

```
estat ic
```

Display CAIC and BIC

```
estat ic, aicconsistent
```

Display AICc and BIC

```
estat ic, aiccorrected
```

Display AIC, BIC, AICc, and CAIC

```
estat ic, all
```

Specify N to be used in calculating BIC as 500

```
estat ic, n(500)
```

Specify N and degrees of freedom to be used in calculating all information criteria as 500 and 10, respectively

```
estat ic, n(500) df(10) all
```

Menu for estat

Statistics > Postestimation

Syntax

```
estat ic [ , options ]
```

options	Description
aiccorrected	report AICc instead of AIC
aiconsistent	report CAIC instead of AIC
all	report all four information criteria: AIC, BIC, AICc, and CAIC
n(#)	specify N to be used in calculating BIC, AICc, and CAIC; see [R] **IC note**
df(#)	specify degrees of freedom k to be used in calculating AIC, BIC, AICc, and CAIC

collect is allowed; see **[U] 11.1.10 Prefix commands**.

Options

aiccorrected specifies that AICc be computed instead of AIC. This information criterion is a second-order approximation and is recommended for small sample sizes.

Only one of aiccorrected, aiconsistent, or all is allowed.

aiconsistent specifies that CAIC be computed instead of AIC. This information criterion is a consistent version of AIC; that is, the probability of selecting the "true model" approaches 1 as sample size increases.

Only one of aiconsistent, aiccorrected, or all is allowed.

all produces a table showing all four information criteria: AIC, BIC, AICc, and CAIC.

Only one of all, aiccorrected, or aiconsistent is allowed.

n(#) specifies N to be used in calculating BIC, AICc, and CAIC; see [R] **IC note**.

df(#) specifies degrees of freedom k to be used in calculating AIC, BIC, AICc, and CAIC. By default, k is the number of estimated parameters.

Remarks and examples

stata.com

estat ic calculates four information criteria used to compare models fit to the same dataset. Unlike likelihood-ratio, Wald, and similar testing procedures, the models need not be nested to compare the information criteria. The information criteria are constructed as a function of the log likelihood $\ln L$, the number of estimated parameters (degrees of freedom) k, and, in some cases, the number of observations N. Because they are based on the log-likelihood function, information criteria are available only after commands that report the log likelihood.

The use of information criteria is subjective, and no formal inference can be drawn from the reported values. In a typical approach, a set of potential models is selected, and a superior model is selected from the values of information criteria. A superior model is the model with the lowest value of information criterion. For example, given two models, the model with the lowest AIC fits the data better than the model with the larger AIC. For details, see *Methods and formulas*.

▷ Example 1

In [R] **mlogit**, we fit a model explaining the type of insurance a person has on the basis of age, gender, race, and site of study. Here we refit the model with and without the site dummies and compare the models.

```
. use https://www.stata-press.com/data/r18/sysdsn1
(Health insurance data)
. mlogit insure age male nonwhite
  (output omitted)
. estat ic
Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	615	-555.8545	-545.5833	8	1107.167	1142.54

Note: BIC uses N = number of observations. See **[R] IC note**.

```
. mlogit insure age male nonwhite i.site
  (output omitted)
. estat ic
Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	615	-555.8545	-534.3616	12	1092.723	1145.783

Note: BIC uses N = number of observations. See **[R] IC note**.

The AIC indicates that the model including the site dummies fits the data better, whereas BIC indicates the opposite. As is often the case, different model-selection criteria have led to conflicting conclusions.

◁

Example 2

In example 1, we compared AIC and BIC. Here we focus on comparing AIC and AICc for small sample size. For simplicity, we are using the same health insurance dataset but running `mlogit` with the `age < 30` condition to reduce the sample size.

```
. mlogit insure age male nonwhite if age < 30
  (output omitted)
. estat ic, all
Information criteria
```

Model	N	ll(null)	ll(model)	df
.	87	-76.93025	-70.36684	8

```
Note: BIC, AICc, and CAIC use N = number of observations.
      See [R] IC note.
```

Model	AIC	BIC	AICc	CAIC
.	156.7337	176.4609	158.5798	184.4609

```
Legend: AIC is Akaike's information criterion.
        BIC is Bayesian information criterion.
        AICc is corrected Akaike's information criterion.
        CAIC is consistent Akaike's information criterion.
. mlogit insure age male nonwhite i.site if age < 30
  (output omitted)
. estat ic, all
Information criteria
```

Model	N	ll(null)	ll(model)	df
.	87	-76.93025	-66.03298	12

```
Note: BIC, AICc, and CAIC use N = number of observations.
      See [R] IC note.
```

Model	AIC	BIC	AICc	CAIC
.	156.066	185.6569	160.2822	197.6569

```
Legend: AIC is Akaike's information criterion.
        BIC is Bayesian information criterion.
        AICc is corrected Akaike's information criterion.
        CAIC is consistent Akaike's information criterion.
```

Burnham and Anderson (2002) recommend using AICc when the ratio $N/k < 40$. The AIC suggests that the model with the site dummies is preferred, whereas AICc reports the opposite result.

Example 3

As we discuss in the technical note below, for the linear mixed models fit using restricted maximum likelihood (REML), one needs to be careful when comparing models using the standard information criteria, especially when the fixed-effects specifications differ across models. In this example, we show how to use `n(#)` and `df(#)` to modify the the standard N and k used in the information criteria when we compare such models. As in [ME] **mixed**, we consider the dataset from Munnell (1990) and estimate a Cobb–Douglas production function, which examines the productivity of public capital in each state's private output (Baltagi, Song, and Jung 2001).

Suppose we want to compare two models:

```
. use https://www.stata-press.com/data/r18/productivity
(Public capital productivity)

. mixed gsp private emp hwy water other unemp || region: || state:, reml
  (output omitted )

. estimates store model1

. mixed gsp private emp hwy unemp || region: hwy || state: unemp, reml
  (output omitted )

. estimates store model2
```

The two models differ in both their fixed-effects and random-effects specifications. By default, the number of degrees of freedom in `estat ic` is calculated as $k = k_f + k_r$, where k_f and k_r are the number of estimated fixed-effects and random-effects parameters, respectively. For REML, Gurka (2006) evaluates the performance of various information criteria. He discusses using $k = k_r$ and different possible values for N. Here, we follow the Vonesh and Chinchilli (1997) approach and choose $N - k_f$. Finally, we run `estat ic` to compare the models:

```
. estimates restore model1
(results model1 are active now)

. estat ic, n(809) df(3)

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
model1	809	.	1404.71	3	-2803.42	-2789.333

```
. estimates restore model2
(results model2 are active now)

. estat ic, n(811) df(5)

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
model2	811	.	1413.557	5	-2817.114	-2793.623

Both AIC and BIC indicate that the second model is preferable.

◁

❑ Technical note

`glm` and `binreg, ml` report a slightly different version of AIC and BIC; see [R] **glm** for the formulas used. That version is commonly used within the generalized linear models literature; see, for example, Hardin and Hilbe (2018). The literature on information criteria is vast; see, among others, Akaike (1973), Sawa (1978), and Raftery (1995). Judge et al. (1985) discuss the use of information criteria in econometrics. Royston and Sauerbrei (2008, chap. 2) examine the use of information criteria as an alternative to stepwise procedures for selecting model variables.

For linear mixed models, when restricted maximum likelihood is used, the information criteria with default degrees of freedom and the number of observations cannot be used to compare models with varying sets of fixed effects, because the likelihood of restricted maximum likelihood is dependent on the fixed-effects design matrix (Harville 1974; Gurka 2006). By default, the degrees of freedom in `estat ic` is the sum of the dimension of fixed-effect parameters and the number of covariance parameters. Therefore, only models with the same sets of fixed effects can be compared. However, for each model, the `df(#)` option can be specified manually to allow comparison with different sets of fixed effects. There are also different views on which number should be used as N to calculate BIC, AICc, and CAIC. For example, see Vonesh and Chinchilli (1997) and Kass and Raftery (1995). Use the `n(#)` option to pass a desired number of observations to the `estat ic` command. For details, see [R] **IC note**.

❑

Stored results

`estat ic` stores the following in `r()`:

Matrices

`r(S)`	row vector with columns (`N`, `ll(null)`, `ll(model)`, `df`, and information criteria)

Methods and formulas

There are two main large-sample notions of information criteria: efficiency and consistency (Burnham and Anderson 2002). Efficient criteria target the best finite dimension model under the assumption that the unknown “true model” has infinite dimension. In contrast, assuming that the true data-generating model is finite and fixed, the consistent criterion selects the correct model with probability approaching 1 as $N \rightarrow \infty$. The AIC and AICc belong to the efficient class, while the BIC and CAIC to the consistent class.

Akaike's (1974) information criterion is defined as

$$\text{AIC} = -2\ln L + 2k$$

where $\ln L$ is the maximized log-likelihood of the model and k is the number of parameters estimated. Some authors define AIC as the expression above divided by the sample size.

AIC performs poorly when there are too many parameters in relation to the sample size. Hurvich and Tsai (1989) derived a second-order variant of AIC called AICc,

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{N-k-1}$$

where N is the sample size. See [R] **IC note** for additional information on calculating and interpreting N. Compared with AIC, AICc has an additional bias-correction term, and for large N and small k, this term is negligible. Burnham and Anderson (2002) recommend using AICc when the ratio $N/k < 40$.

Schwarz's (1978) Bayesian information criterion is another measure of fit defined as

$$\text{BIC} = -2\ln L + k\ln N$$

Bozdogan (1987) proposed a consistent version of AIC called CAIC,

$$\text{CAIC} = -2\ln L + k(\ln N + 1)$$

Burnham and Anderson (2002, chap. 6) argue that employing and comparing consistent and efficient information criteria in the same situation contrasts with the fact that they were designed to answer different questions. Thus, one needs to be careful when interpreting the results.

Hirotugu Akaike (1927–2009) was born in Fujinomiya City, Shizuoka Prefecture, Japan. He was the son of a silkworm farmer. He gained BA and DSc degrees from the University of Tokyo. Akaike's career from 1952 at the Institute of Statistical Mathematics in Japan culminated in service as Director General; after 1994, he was Professor Emeritus. His best-known work in a prolific career is on what is now known as the Akaike information criterion (AIC), which was formulated to help selection of the most appropriate model from a number of candidates.

Gideon E. Schwarz (1933–2007) was a professor of statistics at the Hebrew University, Jerusalem. He was born in Salzburg, Austria, and obtained an MSc in 1956 from the Hebrew University and a PhD in 1961 from Columbia University. His interests included stochastic processes, sequential analysis, probability, and geometry. He is best known for the Bayesian information criterion (BIC).

References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, ed. B. N. Petrov and F. Csaki, 267–281. Budapest: Akailseoniai–Kiudo.

———. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Baltagi, B. H., S. H. Song, and B. C. Jung. 2001. The unbalanced nested error component regression model. *Journal of Econometrics* 101: 357–381. https://doi.org/10.1016/S0304-4076(00)00089-0.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52: 345–370. https://doi.org/10.1007/BF02294361.

Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. New York: Springer.

Findley, D. F., and E. Parzen. 1995. A conversation with Hirotugu Akaike. *Statistical Science* 10: 104–117. https://doi.org/10.1214/ss/1177010133.

Gurka, M. J. 2006. Selecting the best linear mixed model under REML. *American Statistician* 60: 19–26. https://doi.org/10.1198/000313006X90396.

Hardin, J. W., and J. M. Hilbe. 2018. *Generalized Linear Models and Extensions*. 4th ed. College Station, TX: Stata Press.

Harville, D. A. 1974. Bayesian inference for variance components using only error contrasts. *Biometrika* 61: 383–385. https://doi.org/10.2307/2334370.

Hurvich, C. M., and C.-L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* 76: 297–307. https://doi.org/10.1093/biomet/76.2.297.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee. 1985. *The Theory and Practice of Econometrics*. 2nd ed. New York: Wiley.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90: 773–795. https://doi.org/10.1080/01621459.1995.10476572.

Munnell, A. H. 1990. Why has productivity growth declined? Productivity and public investment. *New England Economic Review* Jan./Feb.: 3–22.

Raftery, A. E. 1995. Bayesian model selection in social research. In Vol. 25 of *Sociological Methodology*, ed. P. V. Marsden, 111–163. Oxford: Blackwell.

Royston, P., and W. Sauerbrei. 2008. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Chichester, UK: Wiley.

Sawa, T. 1978. Information criteria for discriminating among alternative regression models. *Econometrica* 46: 1273–1291. https://doi.org/10.2307/1913828.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464. https://doi.org/10.1214/aos/1176344136.

Tong, H. 2010. Professor Hirotugu Akaike, 1927–2009. *Journal of the Royal Statistical Society, Series A* 173: 451–454. https://doi.org/10.1111/j.1467-985X.2009.00633.x.

Vonesh, E. F., and V. M. Chinchilli. 1997. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Dekker.

Also see

[R] **estat** — Postestimation statistics

[R] **estat summarize** — Summarize estimation sample

[R] **estat vce** — Display covariance matrix estimates

[R] **estimates stats** — Model-selection statistics