Statistics How To

Mallows' Cp: Definition, Example

[Regression Analysis](#) > Mallows' Cp

Mallows' Cp Criterion is a way to assess the fit of a [multiple regression](#) model. The metric is used to pick a "best" regression model from a set of smaller models.

The Mallows' Cp procedure involves comparing a full model to a subset model with p parameters ([explanatory variables](#)) and quantifying the remaining unexplained error. If the subset model is correct, then Cp should be approximately equal to p. The best model will be the one with the lowest Cp that is less than $p + 1$.

Mallows' Cp formula

Mallows' Cp calculates the standardized total mean square of estimation for the partial model using the formula [2]:

$$C_p = \frac{\mathrm{SS(Res)}_p}{s^2} + 2p' - n,$$

Where:

- n = [sample size](#),

- p = number of [predictor variables](#),

- $\mathrm{SS(Res)}_p$ = residual sum of squares (RSS) from a model with a set of $p - 1$ explanatory variables, along with a constant as an intercept.

- s^2 = the sample [variation](#) (estimate of the full model variance σ^2).

The full model variance sigma σ^2 can also be estimated as the [mean squared error (MSE)](#) of the full model. The ratio SS(Res)/MSE is an estimate of the [chi square statistic](#).

This means that the Mallows' CP formula compares the RSS from a model with $p - 1$ explanatory variables to the residual sum of squares from a model with p explanatory variables. The RSS is a measure of error in a model. The lower the RSS, the better the model fits the data. The Mallows' CP formula uses the residual sum of squares RSS to

penalize models with more explanatory variable because the more explanatory variables you have in a model, the more likely the model will overfit the data.

A frequently used approach is to conduct all possible regressions and then use Mallows' Cp to compare the outcomes.

How to Interpret Mallows' Cp

The general agreement is that smaller Cp values are preferred as they indicate less unexplained error. In other words, models with a Mallows' Cp value close to p +1 are considered to have low bias because they are not overfitting the data. Overfitting is a result of a model that is too complex and learns from noise in the data instead of the underlying relationships. This can lead to models that make have poor predictive value. However, you should consider the model's context and your field's knowledge of the data when using this statistic because the "best" model may not necessarily be the most "reasonable" one.

Comparing models using a plot of Cp vs p can be helpful. Another approach is to select the smallest model where Cp ≤ p holds true (meaning that you're only choosing from a set of unbiased models).

If all potential models have high Cp values, they are all overfitting the data which means that important predictor variables may be missing. Go back over your data and look for any missing variables that may be related to the target variable (the variable that the model is trying to predict). Then, rerun the Cp procedure with the missing variables.

Other ways to evaluate fit include the coefficient of determination (R^2) and adjusted R-squared. Akaike's Information Criterion (AIC) is another alternative, and it is equivalent to Mallows' Cp for Gaussian linear regression [3]. If several models have similar Cp values, take a look at both Mallows' Cp *and* adjusted R-squared.

Example of picking the best model with Mallows' Cp

Suppose a car salesman wants to use customer time in store, customer income, and group size (for example, single person, two spouses, family) as predictor variables in a regression model to predict the dollar amount of a sale. The salesman fits seven different regression models and calculates Mallows' Cp for each model:

Predictor variables	p + 1	Cp
Time in store	2	52.1
Customer income	2	33.4
Group size	2	27.7
Time in store, Customer income	3	4.2
Time in store, Group size	3	2.8
Customer income, Group size	3	2.5
Time in store, Customer income, Group size	4	4

If the value of Mallows' Cp is less than the number of coefficients in the model (p +1), the model is unbiased. There are two unbiased models in the table:

- Time in store and group size as predictor variables (Cp = 2.8, p + 1 = 3)

- Customer income and group size as predictor variables (Cp = 2.5, p + 1 = 3)

Between these two models, the with customer income and group size as predictor variables has the lowest Mallows' Cp value, which means it is the "best" model in terms of minimizing bias.

Can Mallows Cp be negative?

Mallows Cp should in theory be positive or zero, but it's possible to get negative values in practice. This may indicate one of the assumptions of linear models has been violated. Assumptions of linear models include:

- Normally distributed data.

- Independent data.

- The error terms are homoscedastic (have equal variance).

- The independent variables are not correlated with each other.

Be wary of any Cp value that is lower than the number of predictors as it may indicate underfitting. Underfitting happens when the model is too simple and does not capture the underlying relationships in the data. Like overfitting, this can also lead to models that make poor predictions on new data.

References

1. C. L. *Mallows*. Some Comments on C_P. Technometrics, Vol. 15, No. 4. (Nov., *1973*), pp. 661-675.
2. Hocking, R. The Analysis and Selection of Variables in Linear Regression. Biometrics, 32:1-49.
3. Boisbunon, A. et al. (2013). "AIC, Cp and estimators of loss for elliptically symmetric distributions". arXiv:1308.2766 [math.ST].

Comments? Need to post a correction? Please *Contact Us*.

Feel like "cheating" at Statistics? Check out our **"Practically Cheating Statistics Handbook**, which gives you hundreds of easy-to-follow answers in a convenient e-book.
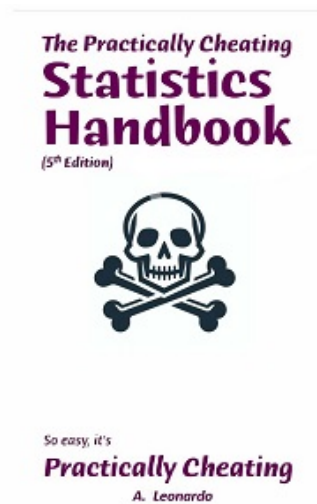
Looking for elementary **statistics help**? You've come to the right place. **Statistics How To** has more than 1,000 articles and **videos** for elementary statistics, probability, AP and advanced statistics topics. **Looking for a specific topic?** Type it into the search box at the top of the page.

Latest articles

Maximum Likelihood and Maximum Likelihood Estimation

What is a Parameter in Statistics?

Beta Level: Definition & Examples

Pairwise Independent, Mutually Independent: Definition, Example

Population Mean Definition, Example, Formula

Dispersion / Measures of Dispersion: Definition

Serial Correlation / Autocorrelation: Definition, Tests

Fisher Information / Expected Information: Definition

Mixed Derivative (Partial, Iterated)

Power Series Distributions

Coefficient, Leading Coefficient: Definition, Test

Reciprocal Distribution: Definition & Examples

© 2025 **Statistics How To** | About Us | Privacy Policy | Terms of Use