

# How can Mallows' Cp be calculated using R?

Authored by  
**stats writer**

April 30, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can Mallows' Cp be calculated using R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141373>

Mallows' Cp is a statistical technique used to evaluate the effectiveness of a regression model. It measures the trade-off between the simplicity and accuracy of the model. In R, Mallows' Cp can be calculated by using the "Cp()" function in the "leaps" package. This function takes the regression model as an input and returns a value that represents the goodness of fit of the model. The lower the Cp value, the better the model fits the data. By comparing the Cp values of different models, one can determine the most suitable model for the given data. Mallows' Cp calculation in R provides a quick and efficient way to assess the quality of regression models and make informed decisions about model selection.

## Calculate Mallows' Cp in R

**In regression analysis, is a metric that is used to pick the best regression model among several potential models.**

**We can identify the "best" regression model by identifying the model with the lowest Cp value that is close to  $p+1$ , where  $p$  is the number of predictor variables in the model.**

**The easiest way to calculate Mallows' Cp in R is to use the function from the olsrr package.**

**The following example shows how to use this function to calculate Mallows' Cp to pick the best regression model among several potential models in R.**

**Example: Calculating Mallows' Cp in R**

Suppose we would like to fit three different multiple linear regression models using variables from the mtcars dataset.

The following code shows how to fit the following regression models:

**Predictor variables in Full Model: All 10 variables**  
**Predictor variables in Model 1: disp, hp, wt, qsec**  
**Predictor variables in Model 2: disp, qsec**  
**Predictor variables in Model 3: disp, wt**

The following code shows how to fit each of these regression models and use the function to calculate the Mallows' Cp of each model:

```
library(olsrr)

#fit full model
full_model <- lm(mpg ~ ., data = mtcars)

#fit three smaller models
model1 <- lm(mpg ~ disp + hp + wt + qsec, data = mtcars)
model2 <- lm(mpg ~ disp + qsec, data = mtcars)
```

```
model3 <- lm(mpg ~ disp + wt, data = mtcars)
```

```
#calculate Mallows' Cp for each model
```

```
ols_mallows_cp(model1, full_model)
```

```
4.430434
```

```
ols_mallows_cp(model2, full_model)
```

```
18.64082
```

```
ols_mallows_cp(model3, full_model)
```

```
9.122225
```

Here's how to interpret the output:

Model 1:  $p + 1 = 5$ , Mallows' Cp = 4.43  
Model 2:  $p + 1 = 3$ , Mallows' Cp = 18.64  
Model 3:  $p + 1 = 30$ , Mallows' Cp = 9.12

We can see that model 1 has a value for Mallows' Cp that is closest to  $p + 1$ , which indicates that it's the best model that leads to the least amount of bias among the three potential models.

## Notes on Mallows' Cp

**Here are few things to keep in mind with regards to Mallows' Cp:**

**If every potential model has a high value for Mallows' Cp, this is an indication that some important predictor variables are likely missing from each model. If several potential models have low values for Mallows' Cp, choose the model with the lowest value as the best model to use.**

**Keep in mind that Mallows' Cp is only one way to identify the "best" regression model among several potential models.**

**When deciding which regression model is best among a list of several different models, it's recommended to look at both Mallows' Cp and adjusted R-squared.**