

How to Perform Forward Selection for Feature Selection

Authored by
stats writer

November 29, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Perform Forward Selection for Feature Selection*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=101537>

Forward selection is a feature selection technique which starts with an empty set of features and adds features one at a time until it reaches the desired level of accuracy. This technique helps to identify the most important features in a dataset that can lead to the best model performance. By adding features one at a time, the model is able to determine which features are the most important in terms of accuracy and can be used to make predictions.

In statistics, **stepwise selection** is a procedure we can use to build a regression model from a set of predictor variables by entering and removing predictors in a stepwise manner into the model until there is no statistically valid reason to enter or remove any more.

The goal of stepwise selection is to build a regression model that includes all of the predictor variables that are statistically significantly related to the response variable.

One of the most commonly used stepwise selection methods is known as **forward selection**, which works as follows:

Step 1: Fit an intercept-only regression model with no predictor variables. Calculate the AIC* value for the model.

Step 2: Fit every possible one-predictor regression model. Identify the model that produced the lowest AIC and also had a statistically significant reduction in AIC compared to the intercept-only model.

Step 3: Fit every possible two-predictor regression model. Identify the model that produced the lowest AIC and also had a statistically significant reduction in AIC compared to the one-predictor model.

Repeat the process until fitting a regression model with more predictor variables no longer leads to a statistically significant reduction in AIC.

*There are several metrics you could use to calculate the quality of fit of a regression model including cross-validation prediction error, Cp, BIC, AIC, or adjusted R². In the example below we choose to use AIC.

The following example shows how to perform forward selection in R.

Example: Forward Selection in R

For this example we'll use the built-in in R:

```
#view first six rows of mtcars  
head(mtcars)
```

```
mpg cyl disp hp drat wt qsec vs am gear carb
Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4
Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4
Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1
Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1
Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2
Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1
```

We will fit a multiple linear regression model using *mpg* (miles per gallon) as our response variable and all of the other 10 variables in the dataset as potential predictors variables.

The following code shows how to perform forward stepwise selection:

```
#define intercept-only model
```

```
intercept_only <- lm(mpg ~ 1, data=mtcars)
```

```
#define model with all predictors
```

```
all <- lm(mpg ~ ., data=mtcars)
```

```
#perform forward stepwise regression
```

```
forward <- step(intercept_only, direction='forward', scope=formula(all), trace=0)
```

```
#view results of forward stepwise regression
```

```
forward$anova
```

```
Step Df Deviance Resid. Df Resid. Dev AIC
1 NA NA 31 1126.0472 115.94345
2 + wt -1 847.72525 30 278.3219 73.21736
3 + cyl -1 87.14997 29 191.1720 63.19800
4 + hp -1 14.55145 28 176.6205 62.66456
```

```
#view final model
```

```
forward$coefficients
```

```
(Intercept) wt cyl hp
```

```
38.7517874 -3.1669731 -0.9416168 -0.0180381
```

Here is how to interpret the results:

First, we fit the intercept-only model. This model had an AIC of **115.94345**.

Next, we fit every possible two-predictor model. The model that produced the lowest AIC and also had a statistically significant reduction in AIC compared to the single-predictor model added the predictor *cyl*. This model had an AIC of **63.19800**.

Next, we fit every possible three-predictor model. The model that produced the lowest AIC and also had a statistically significant reduction in AIC compared to the two-predictor model added the predictor *hp*. This model had an AIC of **62.66456**.

Next, we fit every possible four-predictor model. It turned out that none of these models produced a significant reduction in AIC, thus we stopped the procedure.

Thus, the final model turns out to be:

$$\text{mpg} = 38.75 - 3.17 \cdot \text{wt} - 0.94 \cdot \text{cyl} - 0.02 \cdot \text{hyp}$$

It turns out that attempting to add more predictor variables to the model does not lead to a statistically significant reduction in AIC.

Thus, we conclude that the best model is the one with three predictor variables: *wt*, *cyl*, and *hp*.

A Note on Using AIC

In the previous example, we chose to use AIC as the metric for evaluating the fit of various regression models.

AIC stands for **Akaike information criterion** and is calculated as:

$$\text{AIC} = 2K - 2\ln(L)$$

where:

K: The number of model parameters.

$\ln(L)$: The log-likelihood of the model. This tells us how likely the model is, given the data.

However, there are other metrics you might choose to use to evaluate the fit of regression models including cross-validation prediction error, C_p , BIC, AIC, or adjusted R^2 .

Fortunately, most statistical software allows you to specify which metric you would like to use when performing forward selection.

The following tutorials provide additional information about regression models: