How to Perform Multiple Linear Regression in Stata: A Step-by-Step Guide

Authored by
stats writer

March 8, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform Multiple Linear Regression in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=134573

Multiple linear regression is a statistical technique used to analyze the relationship between multiple independent variables and a single dependent variable. In Stata, this can be done by using the "regress" command followed by the list of independent variables and the dependent variable. Stata also allows for the inclusion of interaction terms and categorical variables in the regression model. After running the regression, Stata provides output with the coefficients, standard errors, p-values, and other statistics to assess the significance and strength of the relationship between the variables. Additionally, Stata offers various diagnostic tools to check the assumptions of the regression model and evaluate its overall fit. Overall, performing multiple linear regression in Stata is a straightforward process that allows for in-depth analysis of the relationships between variables.

Perform Multiple Linear Regression in Stata

Multiple linear regression is a method you can use to understand the relationship between several explanatory variables and a response variable.

This tutorial explains how to perform multiple linear regression in Stata.

Example: Multiple Linear Regression in Stata

Suppose we want to know if miles per gallon and weight impact the price of a car. To test this, we can perform a multiple linear regression using miles per gallon and weight as the two explanatory variables and price as the response variable.

Perform the following steps in Stata to conduct a multiple linear regression using the dataset called *auto*,

which contains data on 74 different cars.

Step 1: Load the data.

Load the data by typing the following into the Command box:

use http://www.stata-press.com/data/r13/auto

Step 2: Get a summary of the data.

Gain a quick understanding of the data you're working with by typing the following into the Command box:

summarize

```
. use http://www.stata-press.com/data/r13/auto
(1978 Automobile Data)

. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

We can see that there are 12 different variables in the dataset, but the only ones we care about are *mpg*, *weight*, and *price*.

We can see the following basic summary statistics about these three variables:

price | mean = $6,165, min = $3,291, max $15,906

mpg | mean = 21.29, min = 12, max = 41

weight | mean = 3,019 pounds, min = 1,760 pounds, max = 4,840 pounds

Type the following into the Command box to perform a multiple linear regression using mpg and weight as explanatory variables and price as a response variable.

regress price mpg weight

. **regress price mpg weight**

Source	SS	df	MS			
Model	186321280	2	93160639.9	Number of obs	=	74
Residual	448744116	71	6320339.67	F(2, 71)	=	14.74
				Prob > F	=	0.0000
				R-squared	=	0.2934
				Adj R-squared	=	0.2735
Total	635065396	73	8699525.97	Root MSE	=	2514

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
mpg	-49.51222	86.15604	-0.57	0.567	-221.3025	122.278
weight	1.746559	.6413538	2.72	0.008	.467736	3.025382
_cons	1946.069	3597.05	0.54	0.590	-5226.245	9118.382

Here is how to interpret the most interesting numbers in the output:

Prob > F: 0.000. This is the p-value for the overall regression. Since this value is less than 0.05, this indicates that the combined explanatory variables of *mpg* and *weight* have a statistically significant relationship with the response variable *price*.

R-squared: 0.2934. This is the proportion of the variance in the response variable that can be explained

by the explanatory variables. In this example, 29.34% of the variation in price can be explained by mpg and weight.

Coef (mpg): -49.512. This tells us the average change in price that is associated with a one unit increase in the mpg, *assuming weight is held constant*. In this example, each one unit increase in mpg is associated with an average decrease of about $49.51 in price, assuming weight is held constant.

For example, suppose cars A and B both weigh 2,000 pounds. If car A gets 20 mpg and car B only gets 19 mpg, we would expect the price of car A to be $49.51 less than the price of car B.

P>|t| (mpg): 0.567. This is the p-value associated with the test statistic for mpg. Since this value is not less than 0.05, we don't have evidence to say that mpg has a statistically significant relationship with price.

Coef (weight): 1.746. This tells us the average change in price that is associated with a one unit increase in weight, *assuming mpg is held constant*. In this example, each one unit increase in weight is associated with an

average increase of about $1.74 in price, assuming mpg is held constant.

For example, suppose cars A and B both get 20 mpg. If car A weighs one pound more than car B, then car A is expected to cost $1.74 more.

P>|t| (weight): 0.008. This is the p-value associated with the test statistic for weight. Since this value is less than 0.05, we have sufficient evidence to say that weight has a statistically significant relationship with price.

Coef (_cons): 1946.069. This tells us the average price of a car when both mpg and weight are zero. In this example, the average price is $1,946 when both weight and mpg are zero. This doesn't actually make much sense to interpret since the weight and mpg of a car can't be zero, but the number 1946.069 is needed to form a regression equation.

Step 4: Report the results.

Lastly, we want to report the results of our multiple linear regression. Here is an example of how to do so:

Multiple linear regression was performed to quantify the

relationship between the weight and mpg of a car and its price. A sample of 74 cars was used in the analysis.

Results showed that there was a statistically significant relationship between weight and price (t = 2.72, p = .008), but there was not a statistically significant relationship between mpg and price (and mpg (t = -.57, p = 0.567).