

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

Authored by
stats writer

March 8, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform Simple Linear Regression in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134564>

Simple linear regression is a statistical method used to analyze the relationship between two continuous variables, where one variable (called the dependent variable) is believed to be influenced by the other variable (called the independent variable). In Stata, simple linear regression can be performed by using the "regress" command. This command allows the user to specify the dependent and independent variables, and provides output that includes the regression coefficients, standard errors, and other relevant statistics. Stata also offers various options for diagnostic tests and visualizations, making it a comprehensive tool for conducting and interpreting simple linear regression analysis.

Perform Simple Linear Regression in Stata

is a method you can use to understand the relationship between an explanatory variable, x , and a response variable, y .

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

Example: Simple Linear Regression in Stata

Suppose we are interested in understanding the relationship between the weight of a car and its miles per gallon. To explore this relationship, we can perform simple linear regression using weight as an explanatory variable and miles per gallon as a response variable.

Perform the following steps in Stata to conduct a simple 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

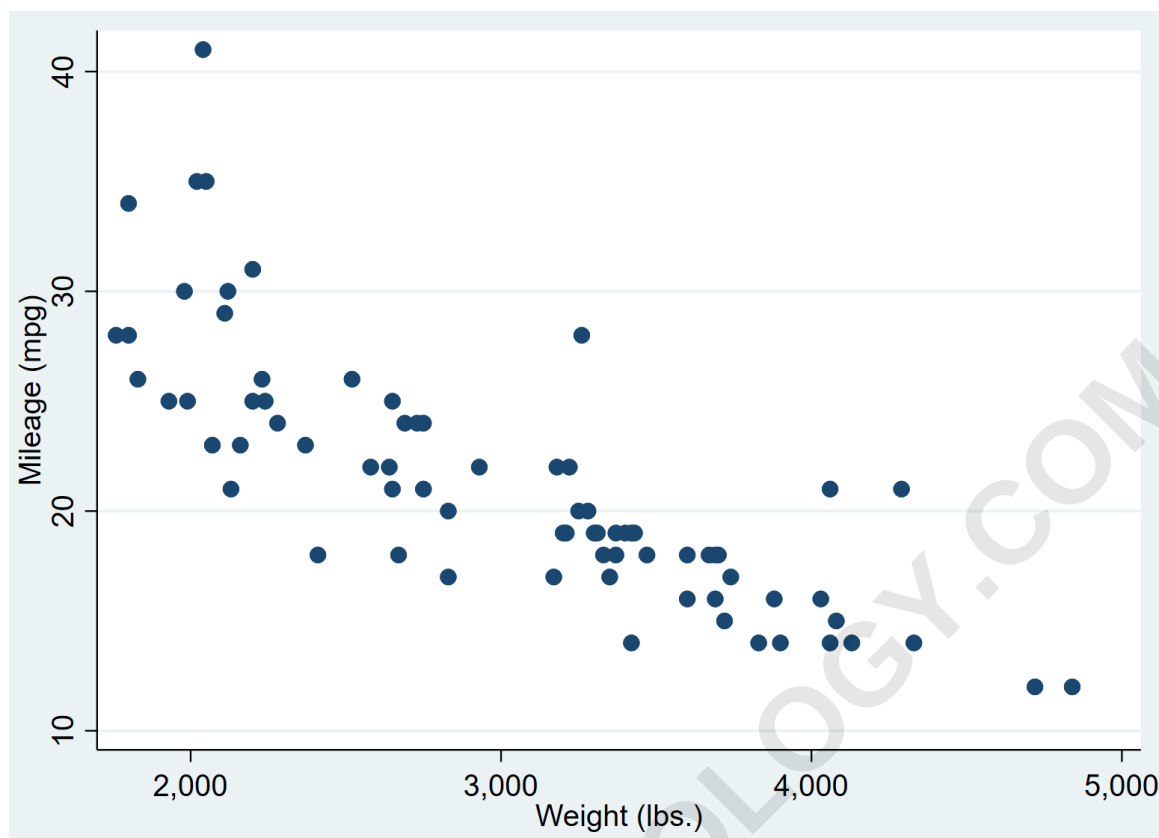
We can see that there are 12 different variables in the dataset, but the only two that we care about are *mpg* and *weight*.

Step 3: Visualize the data.

Before we perform simple linear regression, let's first create a of weight vs. mpg so we can visualize the relationship between these two variables and check for any obvious outliers. Type the following into the Command box to create a scatterplot:

```
scatter mpg weight
```

This produces the following scatterplot:



We can see that cars with higher weights tend to have lower miles per gallon. To quantify this relationship, we will now perform a simple linear regression.

Step 4: Perform simple linear regression.

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

regress mpg weight

```
. regress mpg weight
```

Source	SS	df	MS	Number of obs	=	74
Model	1591.9902	1	1591.9902	F(1, 72)	=	134.62
Residual	851.469256	72	11.8259619	Prob > F	=	0.0000
				R-squared	=	0.6515
				Adj R-squared	=	0.6467
Total	2443.45946	73	33.4720474	Root MSE	=	3.4389

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
weight	-.0060087	.0005179	-11.60	0.000	-.0070411	-.0049763
_cons	39.44028	1.614003	24.44	0.000	36.22283	42.65774

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

R-squared: 0.6515. This is the proportion of the variance in the response variable that can be explained by the explanatory variable. In this example, 65.15% of the variation in mpg can be explained by weight.

Coef (weight): -0.006. This tells us the average change in the response variable associated with a one unit increase in the explanatory variable. In this example, each one pound increase in weight is associated with a decrease of 0.006 in mpg, on average.

Coef (_cons): 39.44028. This tells us the average value of the response variable when the explanatory variable

is zero. In this example, the average mpg is 39.44028 when the weight of a car is zero. This doesn't actually make much sense to interpret since the weight of a car can't be zero, but the number 39.44028 is needed to form a regression equation.

$P>|t|$ (weight): 0.000. This is the p-value associated with the test statistic for weight. In this case, since this value is less than 0.05, we can conclude that there is a statistically significant relationship between weight and mpg.

Regression Equation: Lastly, we can form a regression equation using the two coefficient values. In this case, the equation would be:

$$\text{predicted mpg} = 39.44028 - 0.0060087 * (\text{weight})$$

We can use this equation to find the predicted mpg for a car, given its weight. For example, a car that weighs 4,000 pounds is predicted to have mpg of 15.405:

$$\text{predicted mpg} = 39.44028 - 0.0060087 * (4000) = 15.405$$

Step 5: Report the results.

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

A linear regression was performed to quantify the relationship between the weight of a car and its miles per gallon. A sample of 74 cars was used in the analysis.

Results showed that there was a statistically significant relationship between weight and mpg ($t = -11.60$, $p < 0.0001$) and weight accounted for 65.15% of explained variability in mpg.

The regression equation was found to be:

$$\text{predicted mpg} = 39.44 - 0.006(\text{weight})$$

Each additional pound was associated with a decrease, on average, of -.006 miles per gallon.