

How to Create a Correlation Matrix in Stata: A Step-by-Step Guide

Authored by
stats writer

March 8, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Create a Correlation Matrix in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134556>

Correlation matrix is a useful statistical tool that displays the correlation coefficients between all possible pairs of variables in a dataset. In Stata, creating a correlation matrix can be done by using the "corr" command followed by the list of variables. This command will generate a table with the correlation coefficients and their corresponding p-values. Users can also customize the output by specifying options such as the type of correlation coefficient (e.g. Pearson, Spearman), the significance level, and the variables to be included in the matrix. Overall, creating a correlation matrix in Stata is a simple and effective way to explore the relationships between variables in a dataset.

Create a Correlation Matrix in Stata

In statistics, we're often interested in understanding the relationship between two variables. For example, we might want to understand the relationship between the number of hours a student studies and the exam score they receive.

One way to quantify this relationship is to use the , which is a measure of the linear association between two variables. It has a value between -1 and 1 where:

**-1 indicates a perfectly negative linear correlation between two variables
0 indicates no linear correlation between two variables
1 indicates a perfectly positive linear correlation between two variables**

The further away the correlation coefficient is from zero, the stronger the relationship between the two variables.

But in some cases we want to understand the correlation between more than just one pair of variables. In these cases, we can create a , which is a square table that shows the the correlation coefficients between several pairwise combination of variables.

In this tutorial we explain how to create a correlation matrix in Stata.

How to Create a Correlation Matrix in Stata

The command `corr` can be used to produce a correlation matrix for a particular dataset in Stata.

To illustrate this, let's load the 1980 census data into Stata by typing the following into the command box:

use <http://www.stata-press.com/data/r13/census13>

We can then get a quick summary of the dataset by typing the following into the command box:

summarize

This produces the following table:

```
. use http://www.stata-press.com/data/r13/census13
(1980 Census data by state)
```

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
state	50	25.5	14.57738	1	50
brate	50	167.94	29.34552	125	286
pop	50	4518149	4715038	401851	2.37e+07
medage	50	29.54	1.693445	24.2	34.7
division	50	5.12	2.560612	1	9
region	50	2.66	1.061574	1	4
mrgrate	50	.0186789	.0257327	.0103731	.1955433
dvcrate	50	.0079769	.0031552	.0039954	.0236739
medagesq	50	875.422	99.87562	585.64	1204.09

We see that the dataset contains nine different variables. To create a correlation matrix for every pairwise combination of variables in the dataset, we can type the following into the command box:

```
corr
```

This produces the following correlation matrix:

```
. corr
(obs=50)
```

	state	brate	pop	medage	division	region	mrgrate	dvcrate
state	1.0000							
brate	0.0208	1.0000						
pop	-0.0540	-0.2830	1.0000					
medage	-0.0624	-0.8800	0.3294	1.0000				
division	-0.1345	0.6356	-0.1081	-0.5207	1.0000			
region	-0.1339	0.6086	-0.1515	-0.5292	0.9688	1.0000		
mrgrate	0.0509	0.0677	-0.1502	-0.0177	0.2280	0.2490	1.0000	
dvcrate	-0.0655	0.3508	-0.2064	-0.2229	0.5522	0.5682	0.7700	1.0000
medagesq	-0.0621	-0.8609	0.3324	0.9984	-0.5162	-0.5239	-0.0202	-0.2192

The numbers shown in the table represent the Pearson Correlation Coefficients for each pairwise combination of variables. For example, the correlation between *pop* and *state* is -0.0540. This indicates that these two variables are slightly negatively correlated.

Notice that the correlation along the diagonals of the table are each 1.0000, since each variable is perfectly correlated with itself.

You can also create a correlation matrix for only a certain subset of variables in a dataset by specifying the variables after the `corr` command. For example, here is how to create a correlation matrix for just the variables *pop*, *medage*, and *region*:

```
corr pop medage region
```

This produces the following correlation matrix for just these three variables:

```
. corr pop medage region  
(obs=50)
```

	pop	medage	region
pop	1.0000		
medage	0.3294	1.0000	
region	-0.1515	-0.5292	1.0000

It's also possible to put a star next to the correlation coefficients that are statistically significant at a certain significance level by using the `pwcorr` command (which produces the same result as `corr`) along with the `star()` command.

For example, the following code produces a correlation matrix for every variable in the census dataset and places a star next to the correlation coefficients that are statistically significant at $\alpha = 0.05$:

```
pwcorr, star(.05)
```



```
. pwcorr, star(.01)
```

	state	brate	pop	medage	division	region	mrgrate
state	1.0000						
brate	0.0208	1.0000					
pop	-0.0540	-0.2830	1.0000				
medage	-0.0624	-0.8800*	0.3294	1.0000			
division	-0.1345	0.6356*	-0.1081	-0.5207*	1.0000		
region	-0.1339	0.6086*	-0.1515	-0.5292*	0.9688*	1.0000	
mrgrate	0.0509	0.0677	-0.1502	-0.0177	0.2280	0.2490	1.0000
dvcrate	-0.0655	0.3508	-0.2064	-0.2229	0.5522*	0.5682*	0.7700*
medagesq	-0.0621	-0.8609*	0.3324	0.9984*	-0.5162*	-0.5239*	-0.0202
		dvcrate	medagesq				
dvcrate	1.0000						
medagesq	-0.2192	1.0000					

Notice how fewer correlation coefficients have a star next to them.