

# What are some examples of data analysis in Stata that involve correlation?

Authored by  
**stats writer**

June 29, 2024

## RECOMMENDED CITATION

stats writer (2024). *What are some examples of data analysis in Stata that involve correlation?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=158972>

Data analysis in Stata often involves identifying relationships between variables, and one commonly used technique for this is correlation. Correlation measures the strength and direction of the linear relationship between two variables. Some examples of data analysis in Stata that involve correlation include examining the relationship between income and education level, analyzing the correlation between customer satisfaction and sales, and investigating the correlation between age and job performance. Other applications of correlation in Stata may include studying the relationship between variables in a time series dataset or exploring the correlation between different economic indicators. Overall, correlation in Stata is a powerful tool for understanding and identifying patterns and relationships in data.

## Stata Data Analysis Examples | Correlation

### Example

**A researcher is trying to develop a new, less expensive, test to detect a particular chemical in soil samples. The old test correlates to the criterion (i.e., gold standard measurement) at  $r = 0.89$ . The new test correlates to the criterion at  $r = 0.76$ . Two research assistants are asked to collect soil samples to compare the old test with the new test. The first research assistant collects 42 independent soil samples; the second research assistant collects 47 samples. Each runs a power analysis to determine the observed power.**

**Next, the research assistants are asked to calculate the number of independent samples necessary to detect a difference between 0.89 and 0.76 for power values of .7, .8 and .9.**

## Prelude to the Power Analysis

There are two different aspects of power analysis. One is to calculate the observed power for a specified sample size as in the first part of the example. The other aspect is to calculate the necessary sample size when given a specific power as in the second part of the example. Technically, power is the probability of rejecting the null hypothesis when the specific alternative hypothesis is true. For all examples, we will assume  $\alpha = 0.05$ .

## Power Analysis

In Stata, it is fairly straightforward to perform power analysis for comparing correlations. For example, we can use Stata's power command for our calculation as shown below. We first specify the onecorrelation keyword, indicating that we have a criterion correlation value (0.89 in our example) and an alternative correlation value (0.76 in our example). After the comma, we specify the number of samples obtained. The first research assistant collected 42 soil samples, so we specify 42 in the n option. We will run the analysis a second time with the n value of 47, the

**number of samples the second research assistant collected.**

**power onecorrelation .89 .76, n(42)**

**Estimated power for a one-sample correlation test**

**Fisher's z test**

**H0:  $r = r_0$  versus Ha:  $r \neq r_0$**

**Study parameters:**

**alpha = 0.0500**

**N = 42**

**delta = -0.1300**

**$r_0 = 0.8900$**

**ra = 0.7600**

**Estimated power:**

**power = 0.7576**

**power onecorrelation .89 .76, n(47)**

**Estimated power for a one-sample correlation test**

**Fisher's z test**

**H0:  $r = r_0$  versus Ha:  $r \neq r_0$**

## Study parameters:

**alpha = 0.0500**

**N = 47**

**delta = -0.1300**

**r0 = 0.8900**

**ra = 0.7600**

## Estimated power:

**power = 0.8062**

We can see that the first research assistant observed a power of approximately 0.76, while the second research assistant observed a power of approximately 0.81.

Now the research assistants need to calculate the necessary sample sizes for power values of .7, .8 and .9. In Stata, we could run three separate power analyses, or we could run one analysis and get the results for all three levels of power in a single table. Let's do that.

```
power onecorrelation .89 .76, power(0.7(.1).9)
```

## Performing iteration ...

## Estimated sample size for a one-sample correlation test Fisher's z test

$H_0: r = r_0$  versus  $H_a: r \neq r_0$

```

+-----+
| alpha power N delta r0 ra |
+-----+
| .05 .7 38 -.13 .89 .76 |
| .05 .8 47 -.13 .89 .76 |
| .05 .9 61 -.13 .89 .76 |
+-----+

```

The required sample size for a power of .7 is 38. The required sample size for a power of .8 is 47, and the required sample size for a power of .9 is 61. This makes sense, because as power increases, the sample size must increase, assuming that alpha and the effect size (e.g., delta), are held constant.