

## 12.12: Spearman Rank Correlation

### Learning Objectives

- To use Spearman rank correlation to test the association between two ranked variables, or one ranked variable and one measurement variable. You can also use Spearman rank correlation instead of linear regression/correlation for two measurement variables if you're worried about non-normality, but this is not usually necessary.

### When to use it

Use Spearman rank correlation when you have two ranked variables, and you want to see whether the two variables covary; whether, as one variable increases, the other variable tends to increase or decrease. You also use Spearman rank correlation if you have one measurement variable and one ranked variable; in this case, you convert the measurement variable to ranks and use Spearman rank correlation on the two sets of ranks.

For example, Melfi and Poyser (2007) observed the behavior of 6 male colobus monkeys (*Colobus guereza*) in a zoo. By seeing which monkeys pushed other monkeys out of their way, they were able to rank the monkeys in a dominance hierarchy, from most dominant to least dominant. This is a ranked variable; while the researchers know that Erroll is dominant over Milo because Erroll pushes Milo out of his way, and Milo is dominant over Fraiser, they don't know whether the difference in dominance between Erroll and Milo is larger or smaller than the difference in dominance between Milo and Fraiser. After determining the dominance rankings, Melfi and Poyser (2007) counted eggs of *Trichuris* nematodes per gram of monkey feces, a measurement variable. They wanted to know whether social dominance was associated with the number of nematode eggs, so they converted eggs per gram of feces to ranks and used Spearman rank correlation.

Monkey name	Dominance rank	Eggs per gram	Eggs per gram (rank)
Erroll	1	5777	1
Milo	2	4225	2
Fraiser	3	2674	3
Fergus	4	1249	4
Kabul	5	749	6
Hope	6	870	5

Some people use Spearman rank correlation as a non-parametric alternative to linear regression and correlation when they have two measurement variables and one or both of them may not be normally distributed; this requires converting both measurements to ranks. Linear regression and correlation that the data are normally distributed, while Spearman rank correlation does not make this assumption, so people think that Spearman correlation is better. In fact, numerous simulation studies have shown that linear regression and correlation are not sensitive to non-normality; one or both measurement variables can be very non-normal, and the probability of a false positive ( $P < 0.05$ , when the null hypothesis is true) is still about 0.05 (Edgell and Noon 1984, and references therein). It's not incorrect to use Spearman rank correlation for two measurement variables, but linear regression and correlation are much more commonly used and are familiar to more people, so I recommend using linear regression and correlation any time you have two measurement variables, even if they look non-normal.

### Null hypothesis

The null hypothesis is that the Spearman correlation coefficient,  $\rho$  ("rho"), is 0. A  $\rho$  of 0 means that the ranks of one variable do not covary with the ranks of the other variable; in other words, as the ranks of one variable increase, the ranks of the other variable do not increase (or decrease).

### Assumption

When you use Spearman rank correlation on one or two measurement variables converted to ranks, it does not assume that the measurements are normal or homoscedastic. It also doesn't assume the relationship is linear; you can use Spearman rank correlation

even if the association between the variables is curved, as long as the underlying relationship is monotonic (as  $X$  gets larger,  $Y$  keeps getting larger, or keeps getting smaller). If you have a non-monotonic relationship (as  $X$  gets larger,  $Y$  gets larger and then gets smaller, or  $Y$  gets smaller and then gets larger, or something more complicated), you shouldn't use Spearman rank correlation.

Like linear regression and correlation, Spearman rank correlation assumes that the observations are independent.

### How the test works

Spearman rank correlation calculates the  $P$  value the same way as linear regression and correlation, except that you do it on ranks, not measurements. To convert a measurement variable to ranks, make the largest value 1, second largest 2, etc. Use the average ranks for ties; for example, if two observations are tied for the second-highest rank, give them a rank of 2.5 (the average of 2 and 3).

When you use linear regression and correlation on the ranks, the Pearson correlation coefficient ( $r$ ) is now the Spearman correlation coefficient,  $\rho$ , and you can use it as a measure of the strength of the association. For 11 or more observations, you calculate the test statistic using the same equation as for linear regression and correlation, substituting  $\rho$  for  $r$ :  $t_s = \frac{\sqrt{d.f.} \times \rho}{\sqrt{1-\rho^2}}$ . If the null hypothesis (that  $\rho = 0$ ) is true,  $t_s$  is  $t$ -distributed with  $n - 2$  degrees of freedom.

If you have 10 or fewer observations, the  $P$  value calculated from the  $t$ -distribution is somewhat inaccurate. In that case, you should look up the  $P$  value in a table of Spearman  $t$ -statistics for your sample size. My Spearman spreadsheet does this for you.

You will almost never use a regression line for either description or prediction when you do Spearman rank correlation, so don't calculate the equivalent of a regression line.

For the Colobus monkey example, Spearman's  $\rho$  is 0.943, and the  $P$  value from the table is less than 0.025, so the association between social dominance and nematode eggs is significant.

### Example



Fig. 5.2.1 Magnificent frigatebird, *Fregata magnificens*.

Volume (cm <sup>3</sup> )	Frequency (Hz)
1760	529
2040	566
2440	473
2550	461
2730	465
2740	532
3010	484

3080	527
3370	488
3740	485
4910	478
5090	434
5090	468
5380	449
5850	425
6730	389
6990	421
7960	416

Males of the magnificent frigatebird (*Fregata magnificens*) have a large red throat pouch. They visually display this pouch and use it to make a drumming sound when seeking mates. Madsen et al. (2004) wanted to know whether females, who presumably choose mates based on their pouch size, could use the pitch of the drumming sound as an indicator of pouch size. The authors estimated the volume of the pouch and the fundamental frequency of the drumming sound in 18 males.

There are two measurement variables, pouch size and pitch. The authors analyzed the data using Spearman rank correlation, which converts the measurement variables to ranks, and the relationship between the variables is significant (Spearman's  $\rho = -0.76$ , 16 d.f.,  $P = 0.0002$ ). The authors do not explain why they used Spearman rank correlation; if they had used regular correlation, they would have obtained  $r = -0.82$ ,  $P = 0.00003$ .

## Graphing the results

You can graph Spearman rank correlation data the same way you would for a linear regression or correlation. Don't put a regression line on the graph, however; it would be misleading to put a linear regression line on a graph when you've analyzed it with rank correlation.

## How to do the test

### Spreadsheet

I've put together a spreadsheet that will perform a Spearman rank correlation spearman.xls on up to 1000 observations. With small numbers of observations (10 or fewer), the spreadsheet looks up the  $P$  value in a table of critical values.

### Web page

This web page will do Spearman rank correlation.

### R

Salvatore Mangiafico's *R Companion* has a sample R program for Spearman rank correlation.

### SAS

Use PROC CORR with the SPEARMAN option to do Spearman rank correlation. Here is an example using the bird data from the correlation and regression web page:

```
PROC CORR DATA=birds SPEARMAN;  
VAR species latitude;  
RUN;
```

The results include the Spearman correlation coefficient  $\rho$ , analogous to the  $r$  value of a regular correlation, and the  $P$  value:

Spearman Correlation Coefficients,  $N = 17$

Prob > |r| under  $H_0$ :  $Rho=0$

species latitude

species 1.00000 -0.36263 **Spearman correlation coefficient**

0.1526 **P value**

latitude -0.36263 1.00000

0.1526

## References

1. Picture of magnificent frigatebird from CalPhoto, by Lloyd Glenn Ingles, © California Academy of Sciences.
2. Edgell, S.E., and S.M. Noon. 1984. Effect of violation of normality on the  $t$ -test of the correlation coefficient. *Psychological Bulletin* 95: 576-583.
3. Madsen, V., T.J.S. Balsby, T. Dabelsteen, and J.L. Osorno. 2004. Bimodal signaling of a sexually selected trait: gular pouch drumming in the magnificent frigatebird. *Condor* 106: 156-160.
4. Melfi, V., and F. Poyser. 2007. *Trichuris* burdens in zoo-housed *Colobus guereza*. *International Journal of Primatology* 28: 1449-1456.

---

This page titled [12.12: Spearman Rank Correlation](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [John H. McDonald](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [5.2: Spearman Rank Correlation](#) by [John H. McDonald](#) has no license indicated. Original source: <http://www.biostat handbook.com>.