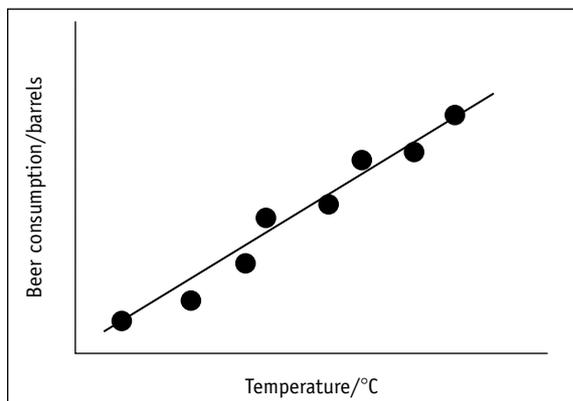


# Maths/stats support 12 Spearman's rank correlation

## Using Spearman's rank correlation

Use a Spearman's rank correlation test when you've got two variables and you want to see if they are correlated. Your calculated Spearman's rank correlation coefficient ( $r_s$ ) lets you test to see if you've got a correlation between two variables, i.e. if a change in one variable is accompanied by a change in the other variable. It also measures the strength of any correlation. You need at least eight pairs of data in matched pairs. Being in matched pairs means that one piece of data is associated with only one other piece of data. For example, if you were measuring speed of hair growth at different ages, each hair growth measurement would belong with only one age measurement. If you mixed the matched pairs up the data would be meaningless.

## Types of correlation



▲ **Figure 1** A positive correlation.

### Positive correlations

Figure 1 shows the (fictional) relationship between amount of beer drunk and temperature. It seems that the hotter it gets, the more beer is drunk. As one variable goes up so does the other. This is an example of a perfect **positive correlation**. If you calculated  $r_s$  for these data you would get a value of +1 (plus one). Rarely will you get exactly +1 but strongly positively correlated variables will usually give you a value approaching +1.



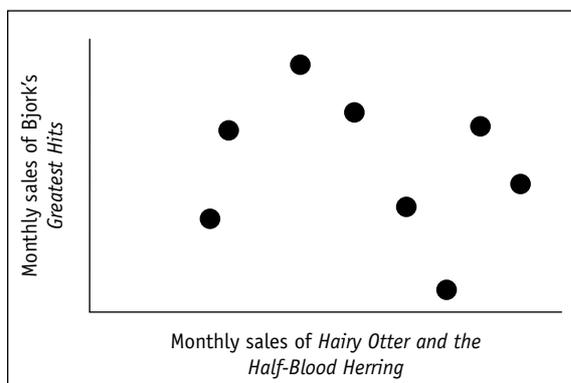
▲ **Figure 2** A negative correlation.

### Negative correlations

Figure 2 shows students' happiness quotient plotted against the number of hours spent doing statistics. Extraordinary as it may seem, it would appear that the longer you spend doing statistics the more unhappy you become. As one variable (time spent doing stats) goes up, the other (happiness quotient) goes down. This is a perfect **negative correlation**.

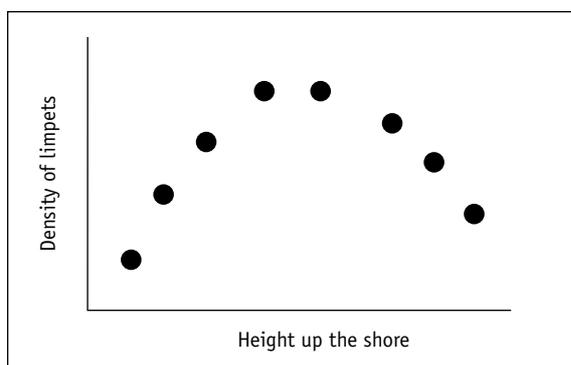
If you calculated  $r_s$  for these data you would get a value of -1 (minus one). Rarely will you get exactly -1. Strongly negatively correlated variables will usually give you a value approaching -1.

## No correlation



▲ **Figure 3** No correlation.

This graph purports to show the monthly sales of *Hairy Otter and the Half-Blood Herring* in Reykjavik plotted against monthly Icelandic sales of Bjork's *Greatest Hits*. There does not seem to be any kind of relationship here. Changing one variable has no predictable effect on the other. There is **no correlation**. If you calculated  $r_s$  for these data you would get a value of close to 0 (zero). Rarely will you get exactly 0. Uncorrelated variables will usually give you a value approaching 0.



▲ **Figure 4** Spearman's rank correlation cannot be used on this type of data.

You should be aware that occasionally you might plot sales of *Hairy Otter and the Half-Blood Herring* against sales of Bjork's *Greatest Hits* and come up with a straight line. This does not necessarily imply a *causal* relationship. (This is also true of the other two examples.) As with all statistics the interpretation of the meaning of the data is up to you. That's the bit that most people find interesting. Remember also that you might have a relationship between two variables that does not take the form of

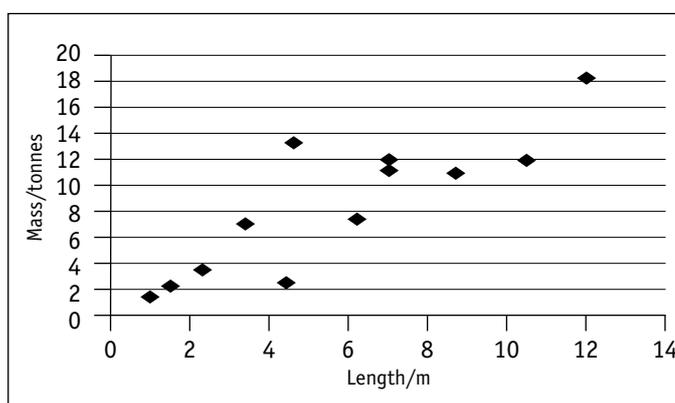
a straight line. The example in Figure 4 is one such relationship. This type of correlation is beyond the scope of Spearman's rank correlation.

### Calculating Spearman's rank correlation coefficient: a worked example

Using a fishing net, a spring balance and a metre rule, a student of above average strength collected the data in Table 1.

▼ **Table 1** Length and mass of blue whales caught at Sunny Bay, Pembrokeshire.

Length/m	Mass/tonnes
1	1.5
1.5	2.3
2.3	3.6
3.4	7.1
4.4	2.6
4.6	13.3
6.2	7.5
7	11.2
7	12.1
8.7	11
10.5	12
12	18.2



▲ **Figure 5** Mass of different length whales.

**Plot the data**

Do blue whales get heavier as they get longer? It looks as if they do. The first step is to plot out the data to see if you have a straight-line graph. This has been done in Figure 5.

There does appear to be a positive correlation between the length and mass of whales so you can proceed to the next stage, which is to calculate  $r_s$ .

**Calculating  $r_s$** **1: Rank the data**

Each set of data is ranked (put in order from lowest to highest). The lowest value is given the lowest rank value, 1. The next lowest value is given the next rank, 2, and so on. When two values are tied, for example the two lengths with a value of 7, they each get the average of the next two available ranks, 8 and 9 in this case, i.e. the same rank, 8.5. If you have more than two tied items of data, do the same thing: add up the appropriate ranks and share them equally between the tied items of data. Continue until you reach the end of the list (in our example the final rank is 12). Set out the data as in Table 2.

Table 2 Ranking the data and calculating differences.

Length/m	Rank length	Mass/tonnes	Rank mass	Difference/ $D$	$D^2$
1.0	1	1.5	1	0	0
1.5	2	2.3	2	0	0
2.3	3	3.6	4	-1	1
3.4	4	7.1	5	-1	1
4.4	5	2.6	3	2	4
4.6	6	13.3	11	-5	25
6.2	7	7.5	6	1	1
7.0	8.5	11.2	8	0.5	0.25
7.0	8.5	12.1	10	-1.5	2.25
8.7	10	11.0	7	3	9
10.5	11	12.0	9	2	4
12.0	12	18.2	12	0	0
			$\Sigma$	0	47.5

From now on we work with the ranks, not the original data. This means you can use this test on interval or ordinal data (see Maths/stats support 9 What test should I use? (M9.09S)).

**2: Work out the differences and square them**

Subtract one rank from the other, giving the difference in ranks. This has been done in the fifth column ( $D$ ). Then square every difference ( $D^2$ ). You can see this has been done in the sixth column of the table above.

**3: Calculate  $r_s$** 

Add up the  $D^2$  column ( $\Sigma D^2$ ).  $\Sigma D^2 = 47.5$

Calculate  $r_s$  from this formula:  $r_s = 1 - \frac{6(\Sigma D^2)}{n(n^2 - 1)}$

where  $r_s$  = Spearman's rank correlation coefficient

$\Sigma$  = sum of

$D$  = Difference

$n$  = number of pairs of data

For the example above, we get:

$$\begin{aligned} r_s &= 1 - \frac{6(47.5)}{12(12^2 - 1)} \\ &= 1 - 0.166 \\ &= 0.834. \end{aligned}$$

This is an answer close to +1; hence we suspect that length and weight are positively correlated.

Even more exciting, we can use  $r_s$  as a hypothesis-testing statistic. Every time you use this test your null hypothesis will be:

*There is no correlation between the two sets of data.*

To know whether to accept or reject this statement you must compare your calculated value of  $r_s$  with the critical value obtained from the appropriate table of critical values (Table 3 on page 5). Reject your null hypothesis if your calculated value is bigger than or equal to the critical value at the chosen significance level for your number of pairs of data ( $n$ ).

We have 12 pairs of length/mass data so our critical value lies along the  $n = 12$  row. Next we decide at what significance level we want to accept or reject the null hypothesis. Usually with biological data a significance level of 5% ( $p < 0.05$ ) is deemed acceptable (although there is no law about this so you can pick a different one if you have a good reason to; however you must know what it means).

In our case the critical value at 5% significance is 0.591. Our value is bigger than this so we reject the null hypothesis and say that there is a positive correlation ( $p < 0.05$ ). The critical value for a 1% significance level is 0.777; our calculated value is also greater than this so we can reject the null hypothesis at this significance level too and say that there is positive correlation at the 1% significance level. In other words, the chance that there isn't a correlation between blue whale length and mass is less than one in a hundred.

▼ **Table 3** Critical values of Spearman's rank correlation coefficient.

Number of pairs/ <i>n</i>	Significance level			
	10%	5%	2%	1%
5	0.900	1.000	1.000	–
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.746	0.794
12	0.506	0.591	0.712	0.777
14	0.456	0.544	0.645	0.715
16	0.425	0.506	0.601	0.665
18	0.399	0.475	0.564	0.625
20	0.377	0.450	0.534	0.591
22	0.359	0.428	0.508	0.562
24	0.343	0.409	0.485	0.537
26	0.329	0.392	0.465	0.515
28	0.317	0.377	0.448	0.496
30	0.306	0.364	0.432	0.478