

## Correlation

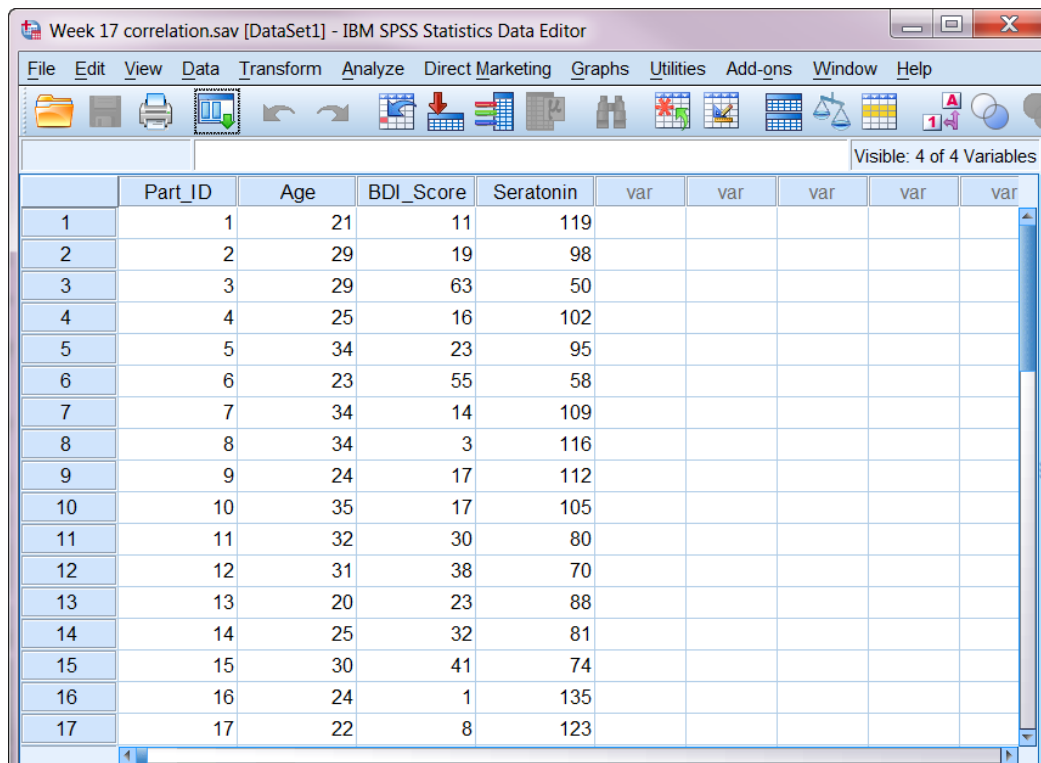
In the following tutorial you will be shown how to carry out a simple correlation analysis. Correlations tell us about the relationship between pairs of variables (for example height and weight; or age and memory performance).

The example here is based on a fictional study investigating the relationship between mood and serotonin levels. As some drugs that are given to people to treat depression work by stimulating serotonin pathways, we might expect to see a relationship between depression scores and serotonin levels in the blood. Specifically, we might predict that people with *lower* levels of serotonin have *higher* levels of depression.

To test the relationship between mood and serotonin levels, we first need a way of measuring these two things.

- First, we could give participants a standardised test to reliably measure their depression levels. In this example, we could use the Beck Depression Inventory, which involves filling out a short questionnaire about their feelings and depressive symptoms. This is then numerically scored.
- Serotonin level could be measured by taking blood from each participant and assessing the level of the neurotransmitter detected in the samples.

If we collected this data, we could then input it into SPSS as follows:



The screenshot shows the IBM SPSS Statistics Data Editor window for a file named 'Week 17 correlation.sav'. The data is organized into a table with 17 rows and 5 columns. The columns are labeled 'Part\_ID', 'Age', 'BDI\_Score', 'Serotonin', and 'var'. The 'var' column contains five empty cells, indicating that the variable is not yet defined. The data points are as follows:

	Part_ID	Age	BDI_Score	Serotonin	var	var	var	var	var
1	1	21	11	119					
2	2	29	19	98					
3	3	29	63	50					
4	4	25	16	102					
5	5	34	23	95					
6	6	23	55	58					
7	7	34	14	109					
8	8	34	3	116					
9	9	24	17	112					
10	10	35	17	105					
11	11	32	30	80					
12	12	31	38	70					
13	13	20	23	88					
14	14	25	32	81					
15	15	30	41	74					
16	16	24	1	135					
17	17	22	8	123					

The data above can be found in the SPSS file: 'Week 17 correlation.sav'.

The different columns display the following data:

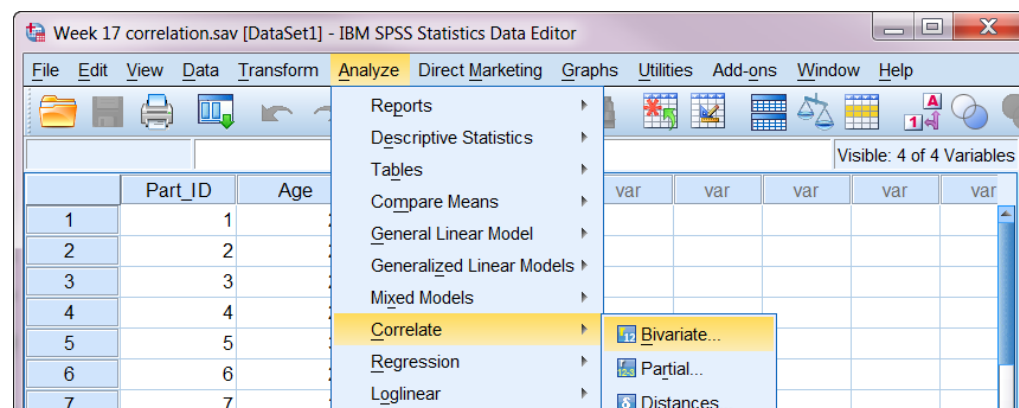
**Part\_ID** refers to the ID number assigned to the participants. We use these numbers as identifiers instead of participant names, as this allows us to collect data while keeping the participants anonymous. This is good practice in psychology, especially when collecting potentially sensitive data, such as that about mental health.

**Age** is usually recorded to allow the researcher to rule out age as a possible confounding variable.

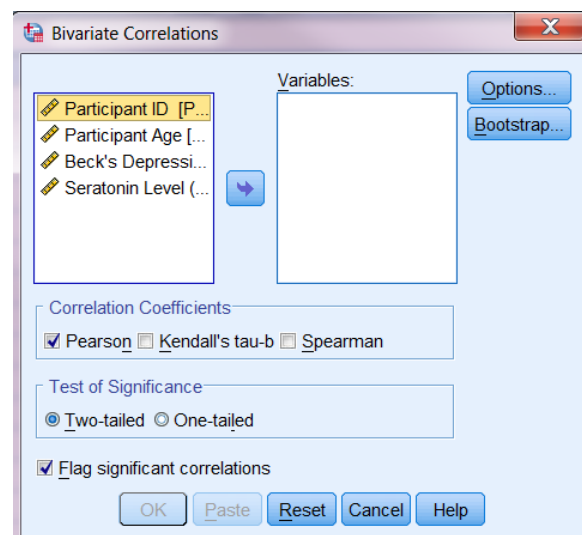
**BDI\_Score** is our first variable of interest: Depression Score. This is measured by the Beck Depression Inventory, and is scored between 0-63. Higher scores indicate higher levels of depressed mood.

**Serotonin** is our second variable of interest. In this case, levels of the neurotransmitter in participants' blood samples were measured in ng/ml.

To start the correlation analysis, begin by **CLICKING** on the **Analyze** menu, select the **Correlate** option, and then the **Bivariate** sub-option.



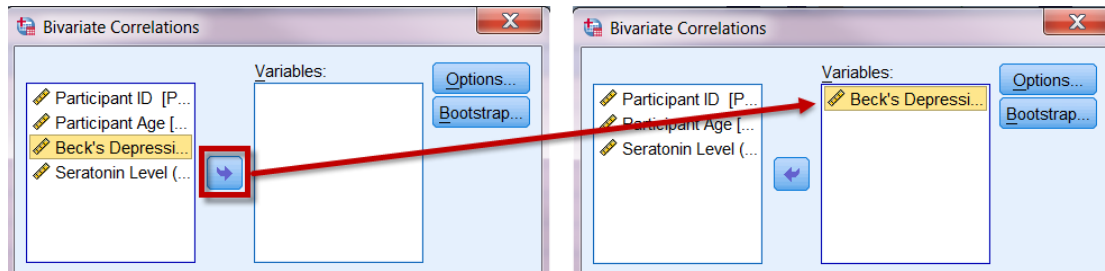
This brings up the Bivariate Correlations dialog box.



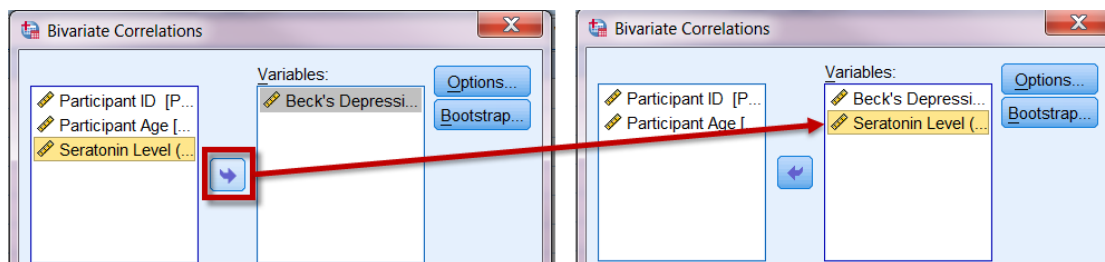
In this dialogue box you will see all of the variables recorded in the data file displayed on the box in the left.

To tell SPSS what we want to analyse we need to move our variables of interest from the left hand list to *Variables* box on the right.

First, **SELECT** the **Depression** variable and add it to the analysis by **CLICKING** on the blue arrow.

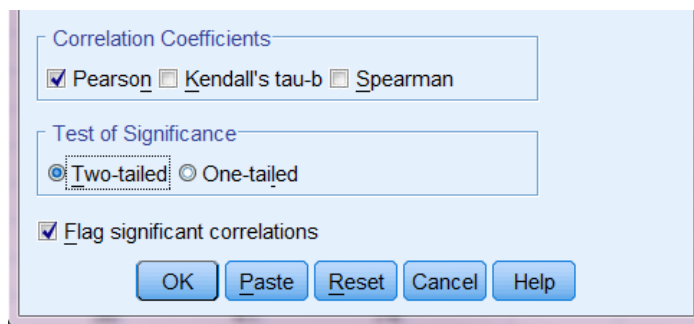


Now do the same with **Serotonin Level**:



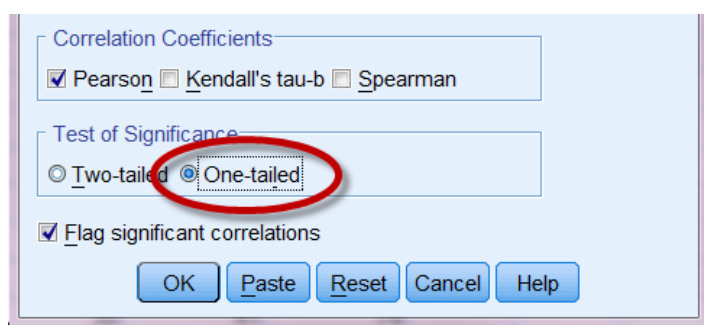
You have now told SPSS which variables you want to add to the analysis.

Now look at the bottom half of the Bivariate Correlations dialog box.



This section allows you to select the type of correlation and significance level that you want. By default this is set to a 2-tailed Pearson correlation (Pearson's  $r$ ). You can change these settings by clicking on the relevant radio buttons or check boxes.

Our hypothesis (or prediction) is that there will be a *negative correlation* between depression scores and serotonin level.

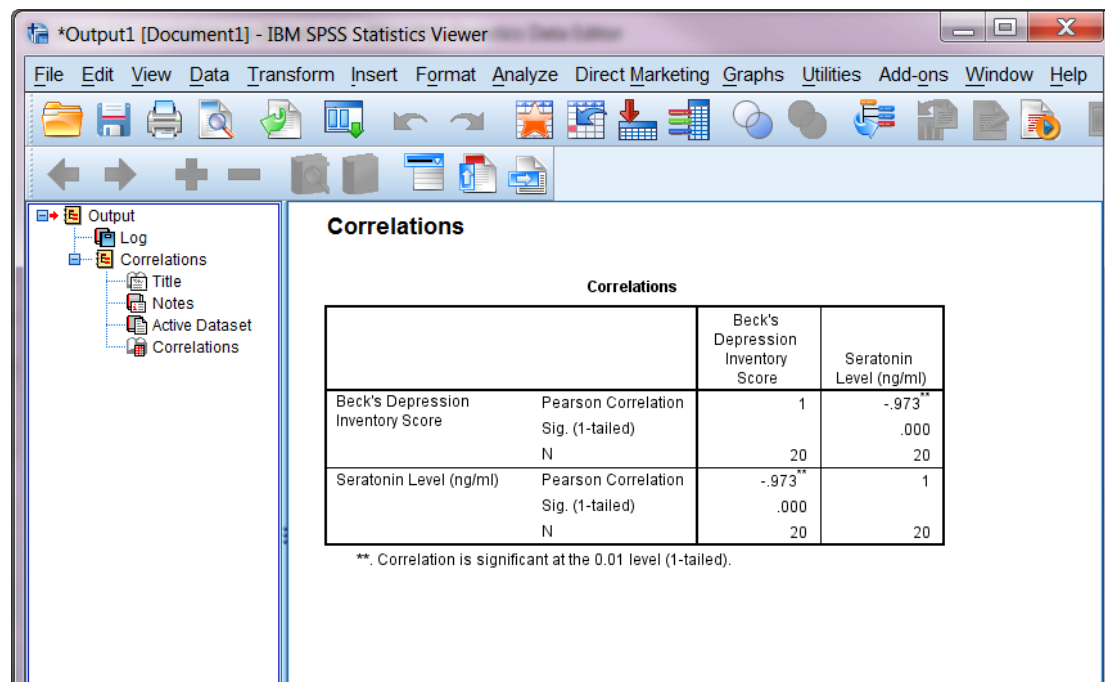


Because we are predicting the *direction* of the correlation, this is a **one-tailed hypothesis**. We therefore have to tell SPSS to calculate a one-tailed correlation. Do this by **CLICKING** on the **one-tailed** radio button.

Now that you have told SPSS what analysis you want to do, you can now run the correlation analysis. To do this, **CLICK** on **OK** to proceed.



You can now view the results in the output window:



But what does this show you?

The Correlations table only has two variables in it, so it's not too hard to read in this example... but sometimes you might be investigating the relationship between several variables all at once. If that were the case, you would have multiple variables in your table. Regardless of the number of variables you have in this table, the way you read it is always the same.

To find the correlation between two variables (in this case BDI Score and Serotonin Level), you start by moving across the **row** that represents your first variable (BDI Score).

		Beck's Depression Inventory Score	Serotonin Level (ng/ml)
Beck's Depression Inventory Score	Pearson Correlation	1	-.973
	Sig. (1-tailed)		.000
	N	20	20
Serotonin Level (ng/ml)	Pearson Correlation	-.973**	1
	Sig. (1-tailed)	.000	
	N	20	20

Next, you need to move down the **column** that represents your second variable (Serotonin level).

\*\* . Correlation is significant at the 0.01 level (1-tailed).

The cell where they **intersect** contains the relevant correlation statistics.

Alternatively you can do this the other way around, and move across the **row** for Serotonin Level and then down the **column** for BDI Score. Again, the cell where they **intersect** contains the relevant correlation statistics.

**Correlations**

		Beck's Depression Inventory Score	Serotonin Level (ng/ml)
Beck's Depression Inventory Score	Pearson Correlation	1	-.973**
	Sig. (1-tailed)		.000
	N	20	20
Serotonin Level (ng/ml)	Pearson Correlation	-.973**	1
	Sig. (1-tailed)	.000	
	N	20	20

\*\* . Correlation is significant at the 0.01 level (1-tailed).

These cells are identical, so it doesn't matter which one you use. These correlation statistics tell us three things:

1. The **direction** of the correlation between the variables
2. The **strength** of the correlation between the variables
3. Whether this correlation is **significant** or not

We can tell the **direction** of the relationship between the variables from the *Pearson Correlation* line.

**Correlations**

		Beck's Depression Inventory Score	Serotonin Level (ng/ml)
Beck's Depression Inventory Score	Pearson Correlation	1	-.973**
	Sig. (1-tailed)		.000
	N	20	20
Serotonin Level (ng/ml)	Pearson Correlation	-.973**	1
	Sig. (1-tailed)	.000	
	N	20	20

\*\* . Correlation is significant at the 0.01 level (1-tailed).

- If the Pearson's coefficient ( $r$ ) is *positive*, this means that as the value of one variable goes up, the value of the other variable also increases.
- In contrast, if the relationship is *negative* this means that as the value of one variable goes up, the value of the other variable goes down.

The **strength** can also be read from the *Pearson Correlation* line. Ignoring the direction of this value, the Pearson's coefficient (*r*) tells you the strength of the relationship.

- 0.8 or above = very strong
- 0.5 or above = strong
- 0.3 or above = medium
- Less than 0.3 = weak

The *Sig* row tells us whether or not this relationship is **significant**. In psychology we tend to accept values of less than  $p=0.05$  as significant.

**Correlations**

		Beck's Depression Inventory Score	Serotonin Level (ng/ml)
Beck's Depression Inventory Score	Pearson Correlation	1	-.973**
	Sig. (1-tailed)		.000
	N	20	20
Serotonin Level (ng/ml)	Pearson Correlation	-.973**	1
	Sig. (1-tailed)	.000	
	N	20	20

\*\* . Correlation is significant at the 0.01 level (1-tailed).

This row also tells you whether the test you are doing is 1-tailed or 2-tailed (as this has implications for significance). A 1-tailed test is done when the hypothesis (or prediction) is directional. You use a 2-tailed test when you don't know what kind of relationship you are looking for.

### Quick Quiz

**Q1. What does the correlation table show?**

- A. There is no significant correlation between **depression score** and **serotonin level**
- B. That **depression score** and **serotonin level** are highly positively correlated
- C. That **depression score** and **serotonin level** are highly negatively correlated
- D. The correlation between **depression score** and **serotonin level** is an example of a perfect correlation?

**Answer:**

The answer is C. There *is* a correlation between depression score and serotonin level, which you can tell by looking at the Pearson Correlation coefficient (so A is incorrect). Looking at the value of the coefficient, it is neither positive (so B is incorrect) nor a perfect correlation, which is either 1.0 or -1.0 (so D is incorrect). The two variables are highly negatively correlated.

**Q2. Do the results in the table support the hypothesis that there will be a significant negative correlation between depression score and serotonin level?**

- A. Yes
- B. No
- C. They neither support nor refute this claim

**Answer:**

Significance testing uses a rule to decide whether we should accept or reject the null hypothesis ( $H_0$ ) in favour of the research hypothesis ( $H_1$ ). To determine this, the p-value needs to be less than 0.05. As this is the case here ( $p < .001$ ), the answer is A... The negative correlation between depression score and serotonin level is significant, so we can reject the null hypothesis (that there is no relationship between the variables) in favour of our research hypothesis.

**So... what do our findings tell us?**

When writing up your results you need to report the strength, direction and significance of the correlation along with the correct statistics *and* give a meaningful interpretation of your findings. A Pearson's correlation is reported using the small letter *r*. So, in this case we might say something like:

*The results showed a very strong, significant negative relationship between Depression Score and Serotonin Level ( $r = -0.973$ ,  $p < 0.001$ ). As serotonin scores increased, the participants' depression scores decreased.*

**This brings us to the end of the correlation tutorial!**

**Just a few take home messages from the work you have done so far:**

1. Correlation analyses are selected via the **Analyze** menu
2. To find the relevant statistics, look along the row for your first variable and down the column for your second: the cell where they intersect has the information you want
3. When writing up your correlation you need to report the *direction*, *strength* and *significance* of your findings

Now you have learned how to carry out a correlation in SPSS, try to get more familiar with SPSS and correlations by practicing the skills you have learned in this tutorial on your own. Download the data file used in this tutorial (using the link below) and see if you can produce the same output yourself. You could also try adding age into the correlational analysis to see what you find.