

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 (such as height and weight, or age and memory performance).

Worked example

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.

- 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.

This is what the data looks like in Jamovi.

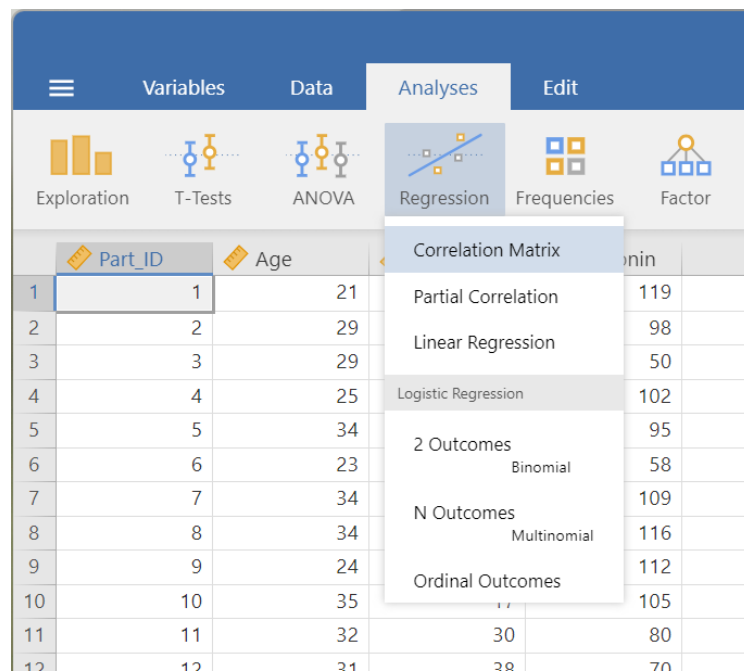
☰ Variables Data Analyses Edit						
Exploration T-Tests ANOVA Regression Frequencies Factor						
	Part_ID	Age	BDI_Score	Serotonin		
1	1	21	11	119		
2	2	29	19	98		
3	3	29	63	50		
4	4	25	16	102		
5	5	34	23	95		
6	6	23	55	58		
7	7	34	14	109		
8	8	34	3	116		
9	9	24	17	112		
10	10	35	17	105		
11	11	32	30	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		
18	18	32	49	65		
19	19	28	4	126		
20	20	34	40	67		

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 and 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.

Running the correlation

To start the analysis, click on Analyses, select Regression and click "Correlation Matrix".



This brings up the Correlation Matrix dialog box.

Here we can see all our variables from the data file displayed in the box on the left.

Correlation Matrix

Part_ID
Age
BDI_Score
Serotonin

→

Correlation Coefficients

☒ Pearson
☐ Spearman
☐ Kendall's tau-b

Additional Options

☒ Report significance
☐ Flag significant correlations
☐ N
☐ Confidence intervals
Interval 95 %

Hypothesis

☒ Correlated
☐ Correlated positively
☐ Correlated negatively

Plot

☐ Correlation matrix
☐ Densities for variables
☐ Statistics

To tell Jamovi what we want to analyse we need to move our variables to the box on the right.

First, **SELECT** BDI_Score in the left-hand box and **CLICK** on the arrow to move it to the right-hand box.

Then, **SELECT** Serotonin in the left-hand box and **CLICK** on the arrow to move it to the right-hand box (the same box that BDI_Score has moved to).

Now you will see both variables in the right-hand box, like below.

Correlation Matrix

Variables: Part_ID, Age

Target Variables: BDI_Score, Serotonin

Correlation Coefficients

- ☒ Pearson
- ☐ Spearman
- ☐ Kendall's tau-b

Hypothesis

- ☒ Correlated
- ☐ Correlated positively
- ☐ Correlated negatively

Additional Options

- ☒ Report significance
- ☐ Flag significant correlations
- ☐ N
- ☐ Confidence intervals
- Interval: 95 %

Plot

- ☒ Correlation matrix
- ☐ Densities for variables
- ☐ Statistics

We next want to make sure that we have ticked certain things in the boxes below to make sure they are included in our output.

Under **Correlation Coefficients**, we need to make sure Pearson is selected. Normally, Jamovi selects Pearson as default. As you can see, we have the option to select Spearman or Kendall's tau-b, but for the purpose of this example, we want to use Pearson.

Under **Additional Options**, we want to select:

- Report significance
- Flag significant correlations
- N (sample size)

What we select under **Hypothesis** depends on our predictions about the outcome.

- If we were unsure whether the relationship between our variables is likely to be positive (as one variable increases, so does the other) or negative (as one variable increases, the other decreases), we would choose **Correlated**.

- If we had good scientific reason to expect the relationship to be positive (as one variable increases, so does the other), we would select **Correlated positively**
- If we had good scientific reason to expect the relationship to be negative (as one variable increases, the other decreases), we would select **Correlated negatively**

We stated above that we expect serotonin levels to fall as depression increases (i.e., a negative correlation), so we should select:

Correlated negatively.

Once you've done this, your Correlation Matrix dialog box should mirror the image below:

Correlation Matrix

Part_ID
Age

BDI_Score
Serotonin

Correlation Coefficients

☒ Pearson
☐ Spearman
☐ Kendall's tau-b

Hypothesis

☐ Correlated
☐ Correlated positively
☒ Correlated negatively

Additional Options

☒ Report significance
☒ Flag significant correlations
☒ N
☐ Confidence intervals
Interval 95 %

Plot

☐ Correlation matrix
☐ Densities for variables
☐ Statistics

The output will then update on the right-hand side based on our selection.

The Output

So, what does the output 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.

Correlation Matrix

Correlation Matrix			
		BDI_Score	Serotonin
BDI_Score	Pearson's r	—	
	df	⊗	—○
	p-value	—	
	N	—	
Serotonin	Pearson's r	-0.97 ***	—
	df	18	—
	p-value	< .001	—
	N	20	—

Note. H_a is negative correlation
 Note. * p < .05, ** p < .01, *** p < .001, one-tailed

In our example, you can see we have the variables displayed on the left hand-side of the table, and across the top. The part of the table that contains numbers, is the section that we want to focus on (this is highlighted in the red box in the image above).

These correlation statistics tell us three things:

1. The **direction** of the correlation between the variables (Pearson's *r*)
2. The **strength** of the correlation between the variables (Pearson's *r*)
3. Whether this correlation is **significant** or not (*p*-value)

Pearson's *r*:

The Pearson's *r* is used to determine the **direction** and **strength** of the correlation.

Correlation Matrix

Correlation Matrix			
		BDI_Score	Serotonin
BDI_Score	Pearson's r	—	
	df	⊗	—○
	p-value	—	
	N	—	
Serotonin	Pearson's r	-0.97 ***	—
	df	18	—
	p-value	< .001	—
	N	20	—

Note. H_a is negative correlation
 Note. * p < .05, ** p < .01, *** p < .001, one-tailed

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

- 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.

As you can see, our Pearson's r is **-0.97**. This indicates a negative relationship between our two variables.

The **strength** can also be read from the *Pearson Correlation* line. Ignoring the direction of this value (i.e. whether it's positive or negative), 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

Looking at our Pearson's r (**-0.97**), we can see there is a very strong, negative relationship between our two variables, as Pearson's r is larger than 0.8.

p-value:

The *p*-value 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.

Correlation Matrix

Correlation Matrix		BDI_Score	Serotonin
BDI_Score	Pearson's r	—	
	df	⊗	—○
	p-value	—	
	N	—	
Serotonin	Pearson's r	−0.97 ***	—
	df	18	—
	p-value	< .001	—
	N	20	—

Note. H_a is negative correlation
 Note. * p < .05, ** p < .01, *** p < .001, one-tailed

As you can see, our p -value is **<.001**, indicating that the negative correlation between depression score and serotonin level *is significant*.

As such, we can confirm that our hypothesis, which predicted that there will be a negative correlation between depression scores and serotonin level, is correct.

You will also notice at the bottom of the table, Jamovi has noted that our hypothesis is one-tailed. This means that we have tested a directional prediction (i.e., that the relationship between BDI_Score and Serotonin will be negative), and the p -value reflects that.

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's r (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:

Null hypothesis significance testing uses a rule to decide whether you 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 you can reject the null hypothesis (that there is no relationship between the variables) in favour of your research hypothesis.

Writing up the results

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 , followed by the degrees of freedom (df) in parentheses. So, in this case we might say something like:

Correlation Matrix

Correlation Matrix			
		BDI_Score	Serotonin
BDI_Score	Pearson's r	—	
	df	—	
	p-value	—	
	N	—	
Serotonin	Pearson's r	-0.97 ***	—
	df	18	—
	p-value	< .001	—
	N	20	—

Note. H_a is negative correlation
Note. * $p < .05$, ** $p < .01$, *** $p < .001$, one-tailed

The results showed a very strong, significant negative relationship between Depression Score and Serotonin Level, $r(18) = -0.97$, $p < .001$, one-tailed test. As serotonin scores increased, the participants' depression scores decreased.

What next?

This brings us to the end of the tutorial. Why not download the dataset used in this tutorial and see if you can produce the same output on your own. You could also try adding age into the correlational analysis to see what you find.