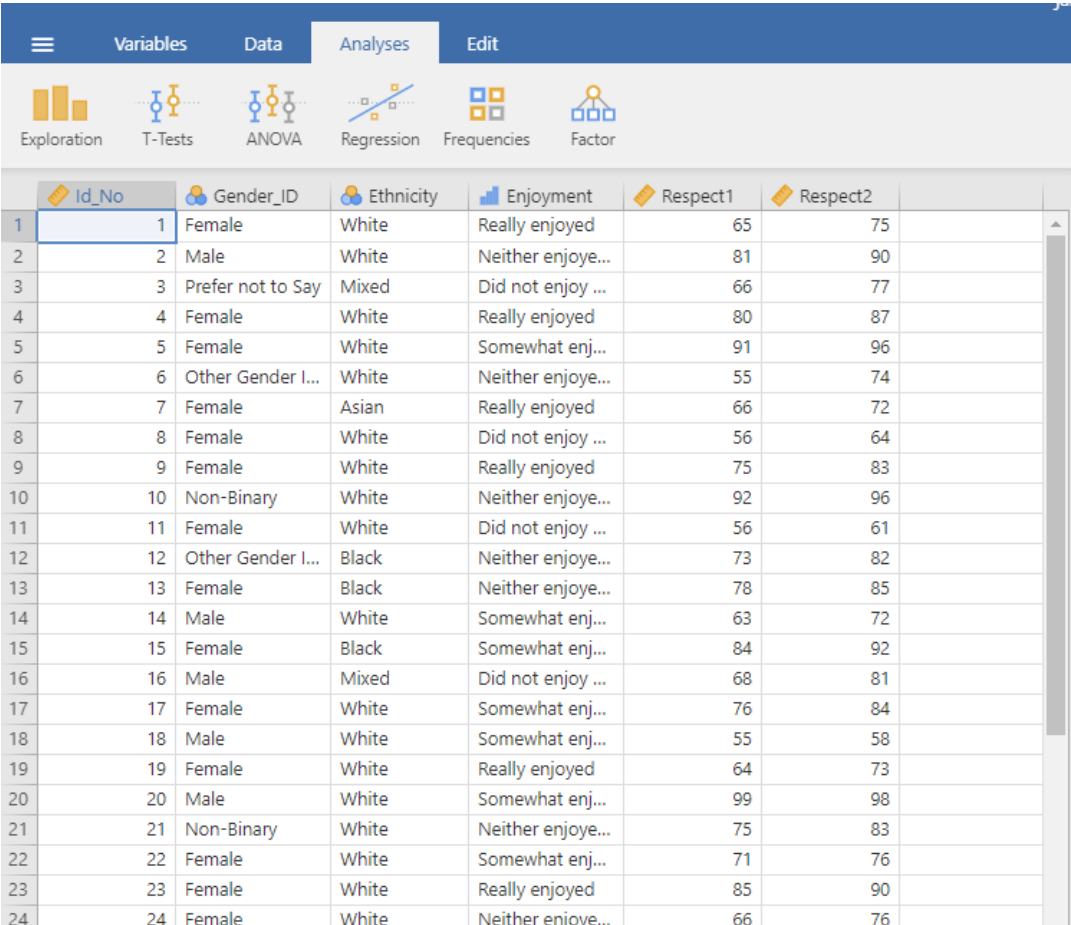


Descriptive Statistics

Using a simple dataset, this tutorial will show you how to produce descriptive statistics using Jamovi.

The data in this example is loosely based on the evaluation of the Schools Linking Network. As the name suggests, this project links schools in different communities to put 'the contact hypothesis' into practice. The contact hypothesis suggests that by increasing contact between people from diverse backgrounds, prejudice can be reduced and positive attitudes towards 'out-groups' can be fostered.

This is what the data looks like in Jamovi.



	Id_No	Gender_ID	Ethnicity	Enjoyment	Respect1	Respect2
1	1	Female	White	Really enjoyed	65	75
2	2	Male	White	Neither enjoye...	81	90
3	3	Prefer not to Say	Mixed	Did not enjoy ...	66	77
4	4	Female	White	Really enjoyed	80	87
5	5	Female	White	Somewhat enj...	91	96
6	6	Other Gender I...	White	Neither enjoye...	55	74
7	7	Female	Asian	Really enjoyed	66	72
8	8	Female	White	Did not enjoy ...	56	64
9	9	Female	White	Really enjoyed	75	83
10	10	Non-Binary	White	Neither enjoye...	92	96
11	11	Female	White	Did not enjoy ...	56	61
12	12	Other Gender I...	Black	Neither enjoye...	73	82
13	13	Female	Black	Neither enjoye...	78	85
14	14	Male	White	Somewhat enj...	63	72
15	15	Female	Black	Somewhat enj...	84	92
16	16	Male	Mixed	Did not enjoy ...	68	81
17	17	Female	White	Somewhat enj...	76	84
18	18	Male	White	Somewhat enj...	55	58
19	19	Female	White	Really enjoyed	64	73
20	20	Male	White	Somewhat enj...	99	98
21	21	Non-Binary	White	Neither enjoye...	75	83
22	22	Female	White	Somewhat enj...	71	76
23	23	Female	White	Really enjoyed	85	90
24	24	Female	White	Neither eniove...	66	76

Each column represents a different variable, and each row contains the data from one participant. The different columns display the following data:

ID_No: This 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. As anonymity is

ethically important in psychology research, this is generally considered good practice.

Gender_ID: This column contains categorical (nominal) information about participants' gender identity.

Ethnicity: This column contains information about participants' ethnicity. Like gender identity, this is a categorical (or nominal) variable, as participants belong to different groups or categories.

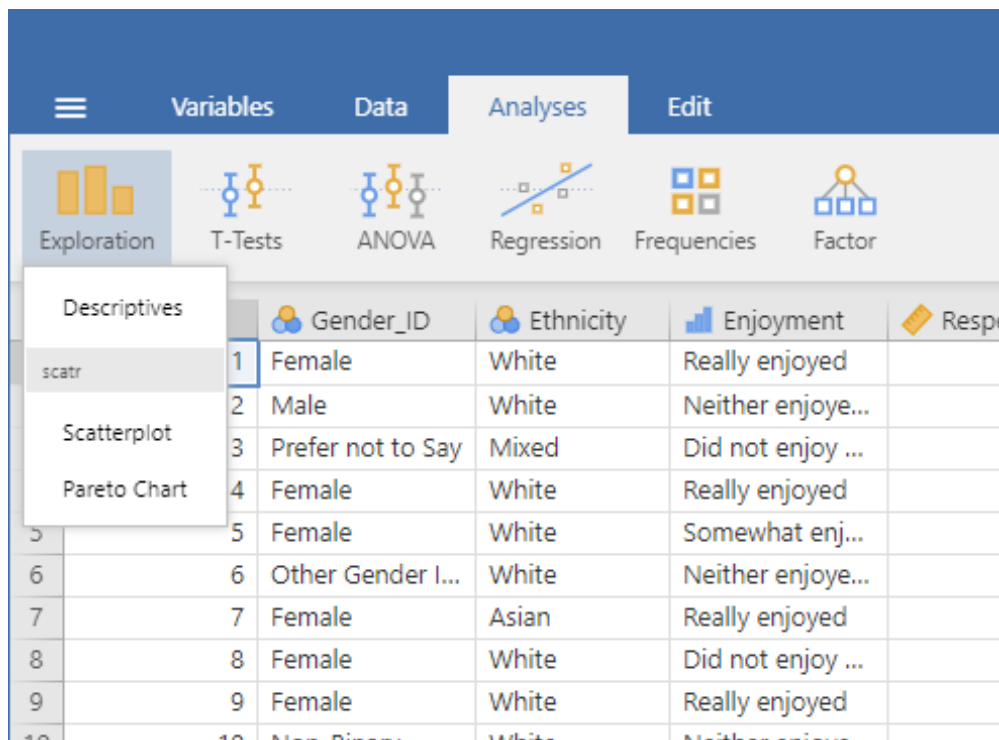
Enjoyment: This variable measures pupils' enjoyment of meeting new people through the Linking School Network on a scale of 1-5 (where 1 = did not enjoy at all, and 5 = really enjoyed). As this is a single scale with 5 ordered categories to choose from, it's an *ordinal* variable.

Respect1: The Linking School Network aims to reduce prejudice between different groups. To assess this, a questionnaire that measures participants' respect for the rights of others was given to students both before and after their participation in the program. This variable represents their respect for others *before* the intervention. As this is measured using a standardised questionnaire, this is a continuous variable.

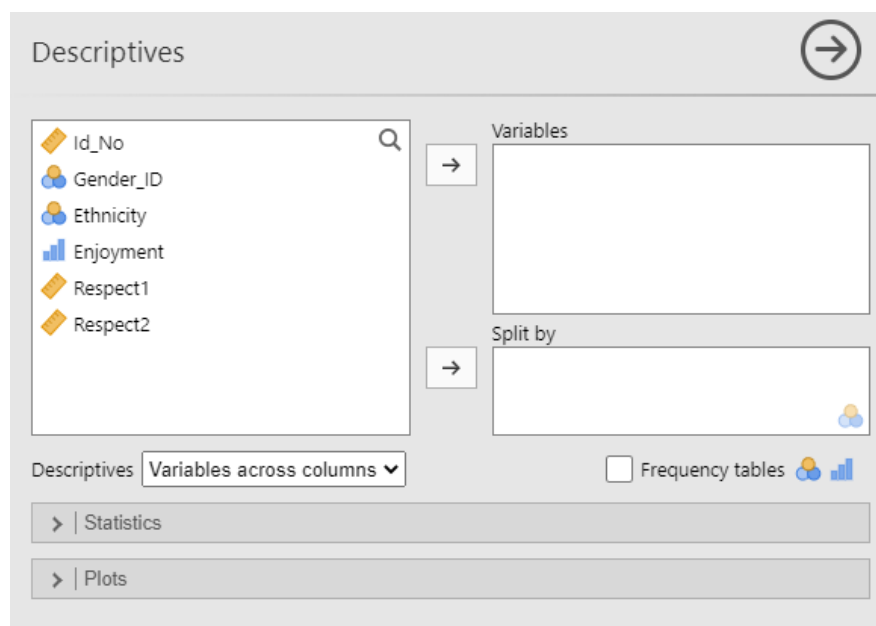
Respect2: Following Respect1, this variable represents participants' respect for others *after* the intervention. It uses the same questionnaire and is calculated in the same way as 'Respect1'. Again, this is a continuous variable.

Now you have had a chance to explore what the different variables are, you need a way to inspect and summarise the data, so you can get a better idea of any patterns that may exist within it. You can do this by looking at descriptive statistics for the data.

To obtain the descriptive statistics, you will need to click on Analyses, select Exploration and click "Descriptives".



This will bring up the Descriptives dialog box.

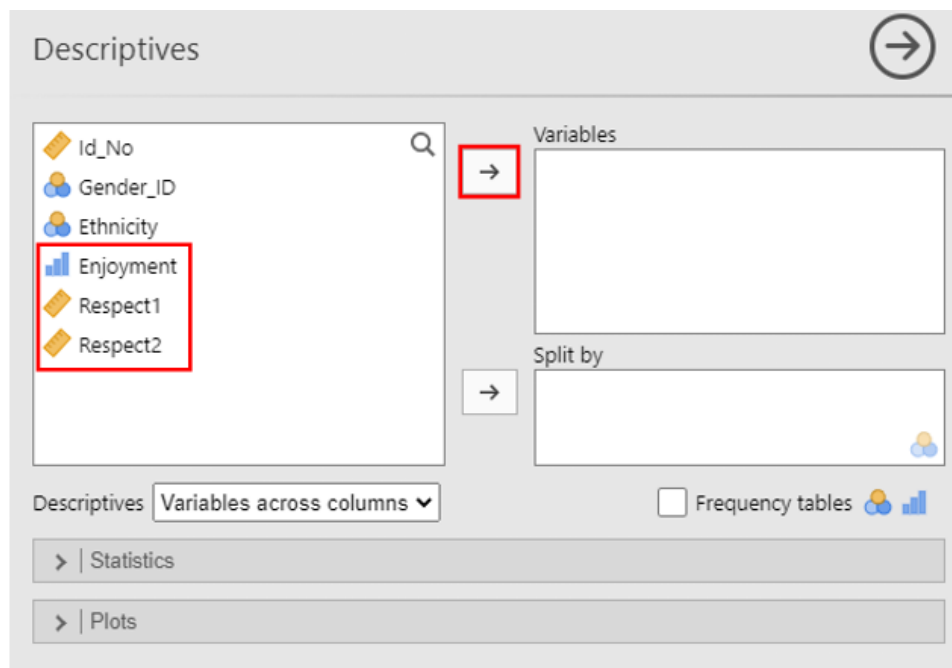


It has the names of your variables in the pane on the left-hand side, and empty panes on the right-hand side labelled **Variables** and **Split by**.

When producing descriptive statistics, you only want to do so for continuous and ordinal variables. It doesn't make sense to ask Jamovi to produce means and standard deviations for nominal data.

Continuous variables are represented by a ruler symbol 📏; ordinal variables by a bar chart 📊; and categorical by the three circles 🍌. Always remember to check that these are set correctly before beginning any analyses (see previous tutorials for details).

To tell Jamovi which variables you want descriptives statistics for, you need to move the variables from the box on the left-hand side to the boxes on the right-hand side. This process is quite common for many of the analyses you will undertake using Jamovi.

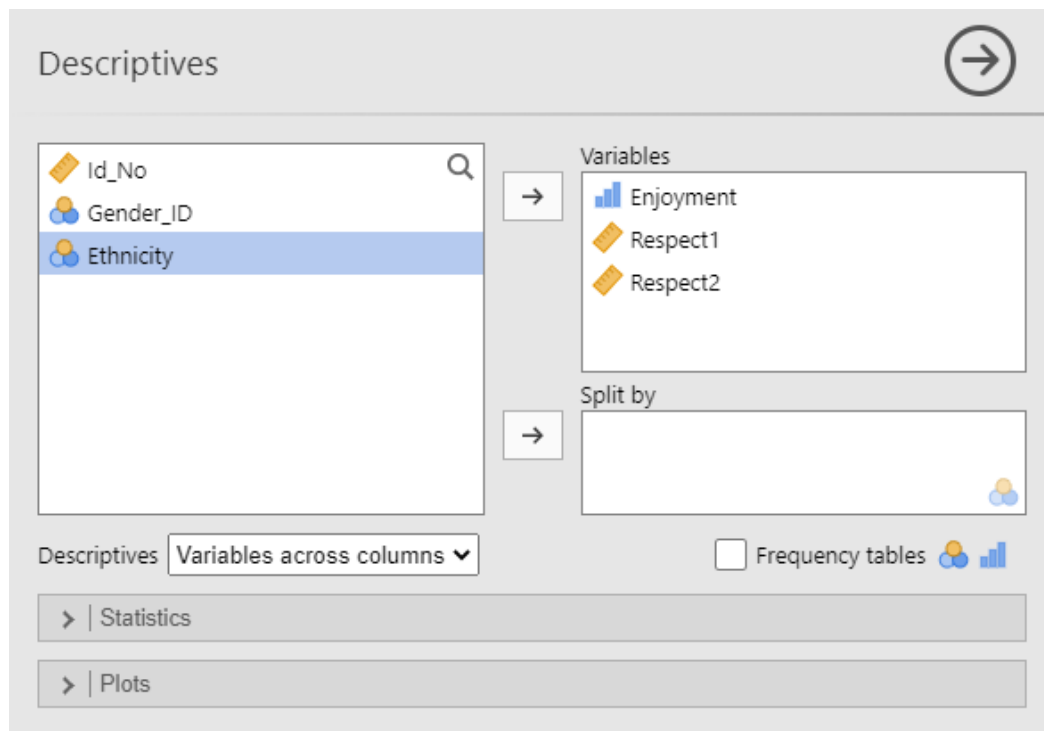


First, **SELECT** the Enjoyment variable in the left-hand box and **CLICK** on the arrow to move it to the right-hand box titled Variables.

Then, **SELECT** the Respect1 variable in the left-hand box and **CLICK** on the arrow to move it to the right-hand box titled Variables.

Next, **SELECT** the Respect2 variable in the left-hand box and **CLICK** on the arrow to move it to the right-hand box titled Variables.

Once you've done this, it should look like this:



(If you wanted to produce separate statistics for different groups of participants, you would add the appropriate grouping variable to the **Split by** box. For example, if you moved Gender_ID into the **Split by** box, Jamovi would calculate separate enjoyment and respect scores for each gender identity. For the purposes of this tutorial, though, just leave this box empty.)

Next, click on the arrow next to Statistics to expand the options available. Once you've expanded this box, you will see the descriptive statistics that Jamovi selects by default:

Statistics

Sample Size

☒ N
☒ Missing

Percentile Values

☐ Cut points for equal groups
☐ Percentiles

Dispersion

☒ Std. deviation
☒ Minimum
☐ Variance
☒ Maximum
☐ Range
☐ IQR

Mean Dispersion

☐ Std. error of Mean
☐ Confidence interval for Mean %

Central Tendency

☒ Mean
☒ Median
☐ Mode
☐ Sum

Distribution

☐ Skewness
☐ Kurtosis

Normality

☐ Shapiro-Wilk

Outliers

☐ Most extreme values

For the purpose of this tutorial, let's add the **Range** to our descriptive statistics. To do this, **select** the box to the left of this option, so the box becomes ticked. Once this has all been done, the output will update on the right-hand side of the window.

Descriptives

Descriptives

	Enjoyment	Respect1	Respect2
N	50	50	50
Missing	0	0	0
Mean	3.74	71.10	79.44
Median	4.00	72.00	81.00
Standard deviation	1.14	13.60	11.70
Range	4	62.00	48.00
Minimum	1	38.00	52.00
Maximum	5	100.00	100.00

The **Descriptive Statistics** table displays all the information you have requested. Variable names are listed at the top (Enjoyment, Respect1, Respect2), with the Descriptives listed in the left-hand column. These Descriptive Statistics represent the following:

- **N:** 'N' stands for 'number of participants'. This column simply tells you how many participants you have data for, for each variable. We have 50 participants in our dataset, so our $N = 50$ for each variable.
- **Missing:** This tells us if we have any data missing from our dataset. In this case, Missing = 0, so we do not have any data missing from this dataset.
- **Mean:** The Mean column shows the mathematical average for each of the variables. This is the sum of the scores divided by the number of scores (the N).
- **Median:** The Median represents the middle number, or data point, in the dataset, if the data was arranged from the lowest value (datapoint) to highest value (datapoint).
- **Standard deviation:** This column displays the Standard Deviation for each variable. This refers to the spread of the scores around the mean and represents how much variation you have in the data. The larger the values relative to the means, the more dispersed the scores are.
- **Range:** This column measures the spread of the scores obtained. Essentially the range is the difference between the lowest and the highest value.
- **Minimum:** This refers to the lower end of the range. In the minimum column is the lowest score of the variable in each row.
- **Maximum:** This refers to the upper end of the range. In the maximum column is the highest score of the variable in each row.

When carrying out research that has continuous data, you should always report the means and standard deviations of your variables, as each statistic provides a useful summary to help make sense of your data. They are usually reported to **2 decimal places**.

	Enjoyment	Respect1	Respect2
N	50	50	50
Missing	0	0	0
Mean	3.74	71.10	79.44
Median	4.00	72.00	81.00
Standard deviation	1.14	13.60	11.70
Range	4	62.00	48.00
Minimum	1	38.00	52.00
Maximum	5	100.00	100.00

In this case, looking at the table you could say:

- On average pupils seemed to enjoy the Linking Schools Project, with a mean Enjoyment score of **3.74** (SD = **1.14**).
- In addition, participants' Respect scores before taking part in the project appeared lower (mean = **71.10**; SD = **13.60**) than afterwards (mean = **79.44**; SD = **11.70**).

Now you have been shown how to enter data and produce descriptive statistics, why don't you try creating your own dataset and calculating the means, standard deviations and other descriptive statistics? Or download the data file used in this tutorial and see if you can produce the same output yourself.

Why not explore these options and see what output you can produce? Practicing using Jamovi yourself will help to increase your confidence with using the program and analysing statistics, so we really do recommend that you take advantage of any opportunities for practice that are available!