JASP

BAYESIAN INFERENCE IN JASP:

A GUIDE FOR STUDENTS

Mark A Goss-Sampson

Johnny van Doorn & EJ Wagenmakers

JASP v0.12.2

DOI 10.17605/OSF.IO/CKNXM

CONTENTS

PREFACE

JASP stands for **Jeffrey's Amazing Statistics Program** in recognition of the pioneer of Bayesian inference Sir Harold Jeffreys. This is a **free** multi-platform open-source statistics package, developed and continually updated by a group of researchers at the University of Amsterdam. Their aim was to develop a free, open-source programme that includes both frequentist, Bayesian and more advanced statistical techniques with a major emphasis on providing a simple intuitive user interface.

In contrast to many statistical packages, JASP provides a simple drag and drop interface, easy access menus, intuitive analysis with real-time computation and display of all results. All tables and graphs are presented in APA format and can be copied directly and/or saved independently. Tables can also be exported from JASP in LaTeX format

JASP can be downloaded free from the website https://jasp-stats.org/ and is available for Windows, Mac OS X and Linux. You can also download a pre-installed Windows version that will run directly from a USB or external hard drive without the need to install it locally. The WIX installer for Windows enables you to choose a path for the installation of JASP – however, this may be blocked in some institutions by local Administrative rights.

The programme also includes a data library with an initial collection of over 50 datasets from Andy Fields book, Discovering Statistics using IBM SPSS statistics[1] and The Introduction to the Practice of Statistics[2] by Moore, McCabe and Craig.

Since May 2018 JASP can also be run directly in your browser via rollApp™ without having to install it on your computer (https://www.rollapp.com/app/jasp). However, this may not be the latest version of JASP.

Keep an eye on the JASP site since there are regular updates as well as helpful videos and blog posts!!

Please note that the underlying concepts of Bayesian analyses are not covered in this book since there many other books and reviews that cover these in much more depth. Some easy reading papers with reference to JASP are listed on the next page. This is a collection of standalone handouts covering the most common Bayesian statistical analyses available in JASP for students studying Biological Sciences. Datasets used in this document are available for download from https://osf.io/8qtu2/

I would also like to acknowledge both EJ Wagenmakers and Johnny van Doorn at the University of Amsterdam for their support, in-depth advice and help in compiling this guide.

Dr Mark Goss-Sampson
Centre for Science and Medicine in Sport & Exercise
University of Greenwich

[1] A Field. (2017) Discovering Statistics Using IBM SPSS Statistics (5th Ed.) SAGE Publications.
[2] D Moore, G McCabe, B Craig. (2011) Introduction to the Practice of Statistics (7th Ed.) W H Freeman.

RECCOMENDED EASY READING

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, A. J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., Morey, R. D. (2018). **Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications.** *Psychonomic Bulletin & Review, 25*, 35-57.

Wagenmakers, E.-J et al (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*, 58-76. (**Preprint**).

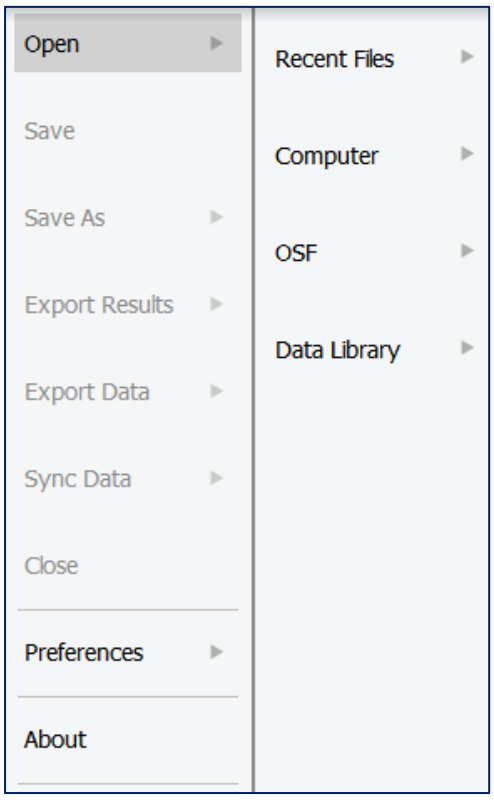
Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry, 2018 (18)*, 178. DOI: **10.1186/s12888-018-1761-4**. (**Open Access**).

Van Doorn J et al (2019) The JASP Guidelines for Conducting and Reporting a Bayesian Analysis. https://psyarxiv.com/yqxfr

USING THE JASP ENVIRONMENT

Open JASP.

The main menu can be accessed by clicking on the top-left icon. 

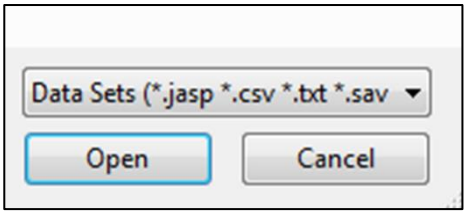


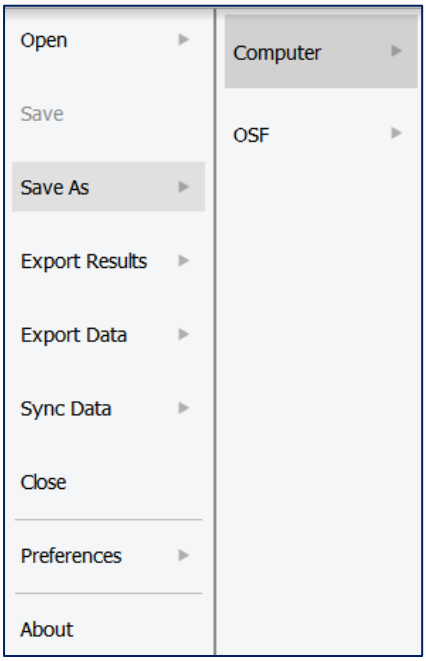
Open:

JASP has its own .**jasp** format but can open a variety of different dataset formats such as:

- **.csv** (comma separated values) can be saved in Excel
- **.txt** (plain text) also can be saved in Excel
- **.tsv** (tab-separated values) also can be saved in Excel
- .**sav** (IBM SPSS data file)
- .**ods** (Open Document spreadsheet)

You can open recent files, browse your computer files, access the Open Science Framework (OSF) or open the wide range of examples that are packaged with the Data Library in JASP.





Save/Save as:

Using these options the data file, any annotations and the analysis can be saved in the .jasp format

Export:

Results can be exported to an HTML file

Data can be exported to either a .csv or .txt file

Sync data:

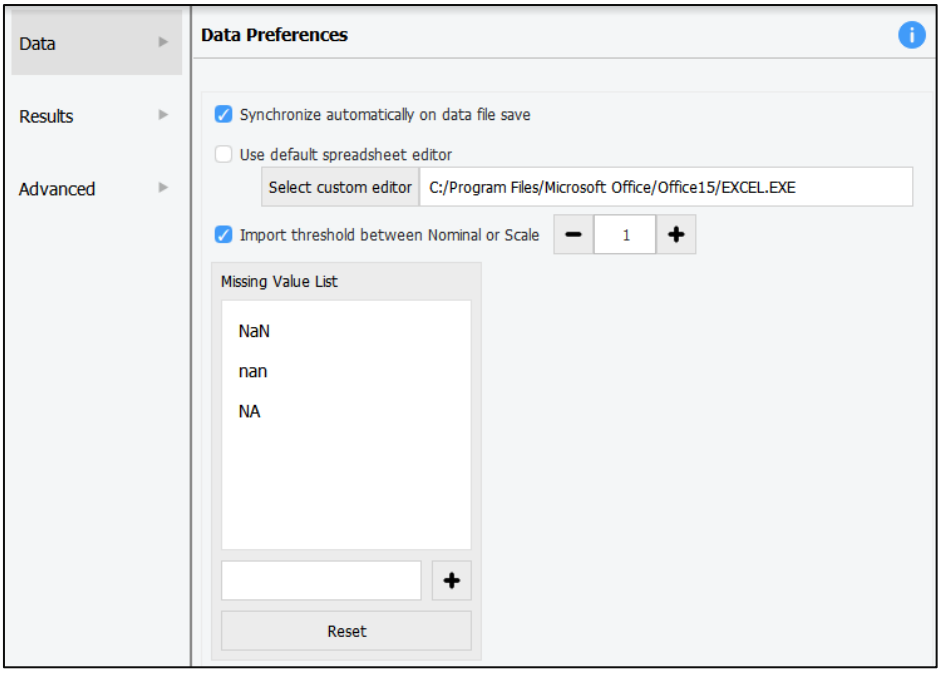
Used to synchronize with any updates in the current data file (also can use Ctrl-Y)

Close:

As it states - it closes the current file but not JASP

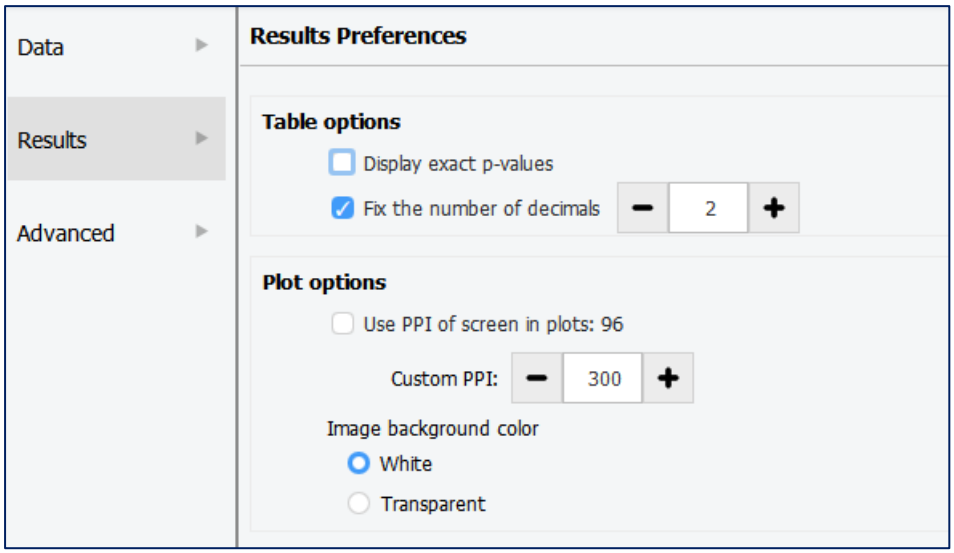
Preferences:

There are three sections that users can use to tweak JASP to suit their needs



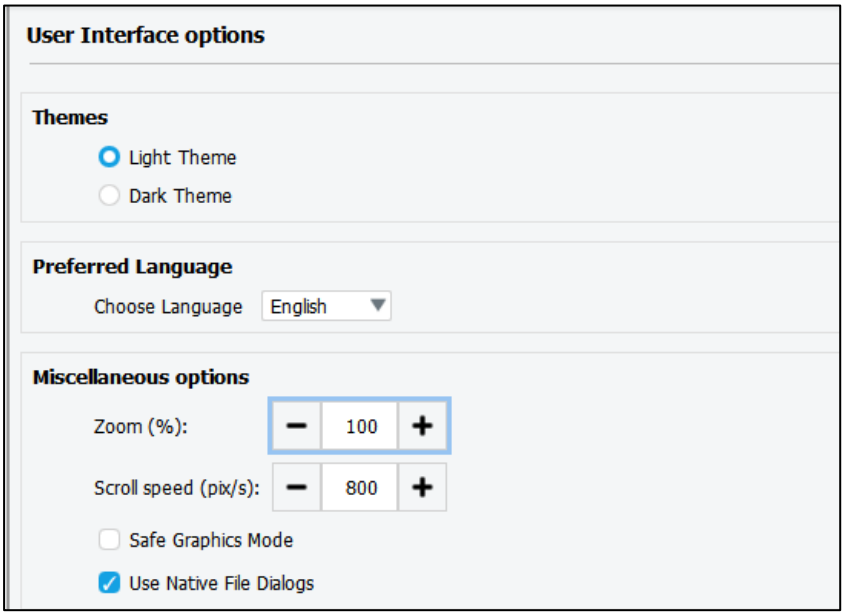
In the ***Data Preferences*** section users can:

- Synchronize/update the data automatically when the data file is saved (default)
- Set the default spreadsheet editor (i.e. Excel, SPSS etc)
- Change the threshold so that JASP more readily distinguishes between nominal and scale data
- Add a custom missing value code



In the ***Results Preferences*** section users can:

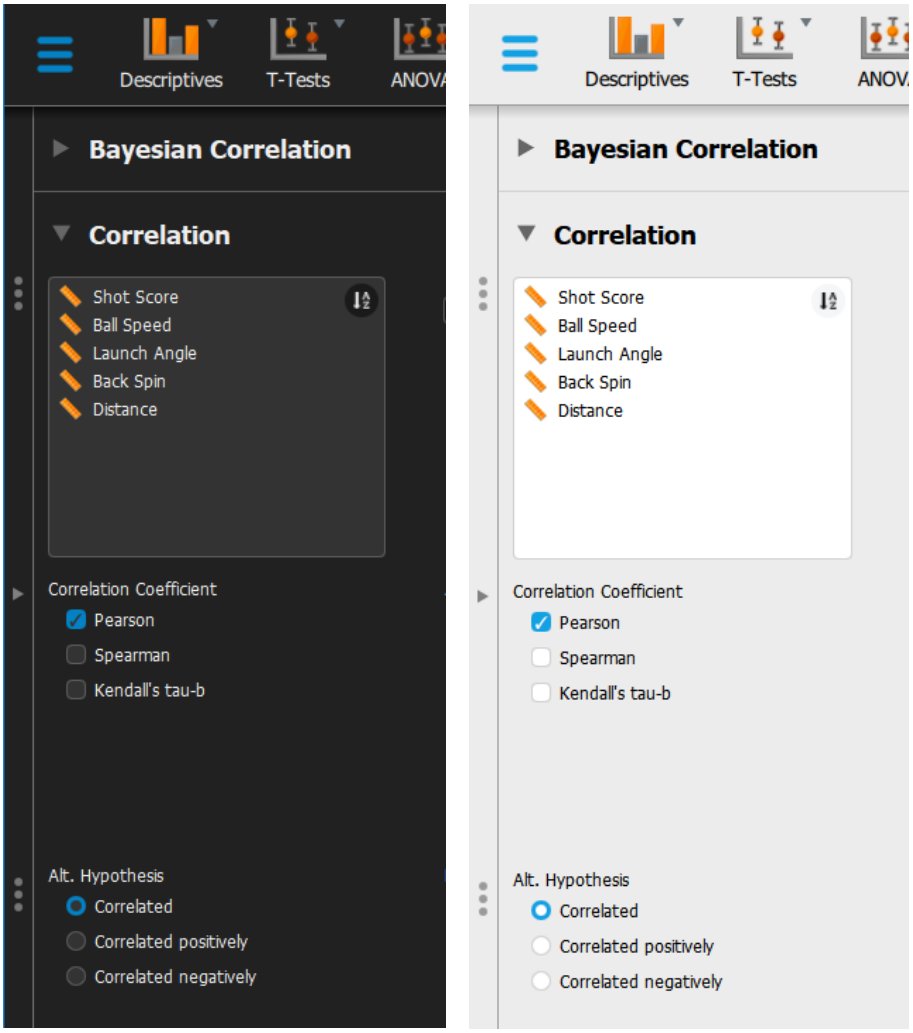
- Fix the number of decimals for data in tables – makes tables easier to read/publish
- Change the pixel resolution of the graph plots
- Select when copying graphs whether they have a white or transparent background.



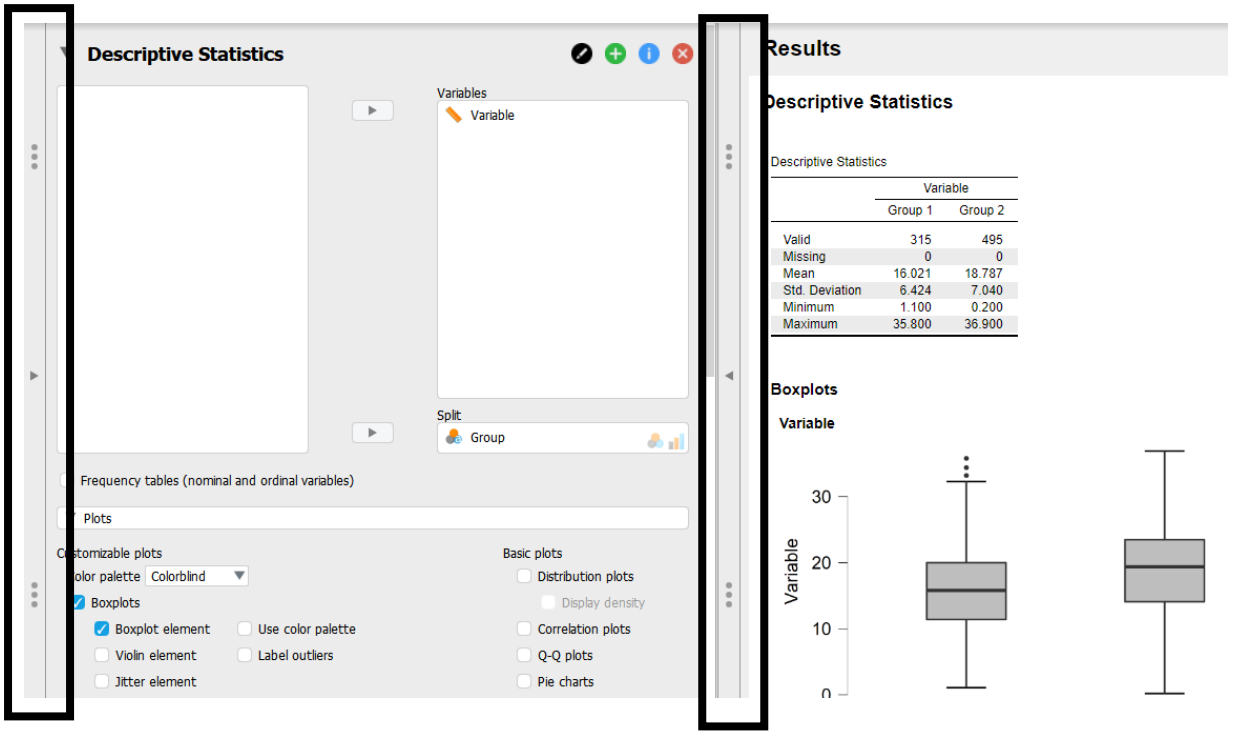
In the ***Interface Preferences*** section users can now pick between two different themes; a light theme (default) and a dark theme. The preferred language currently supports English and Dutch only. In this section, there is also the ability to change the system font size for accessibility and the scroll speeds.

In the ***Advanced Preferences*** section, most users will probably never have to change any of the default settings.

Comparison of the dark and light themes in JASP



JASP has a streamlined interface to switch between the spreadsheet, analysis and results views.



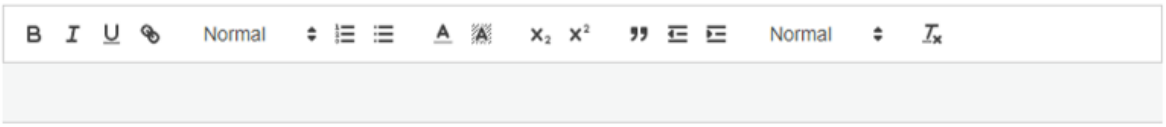
The vertical bars highlighted above allows for the windows to be dragged right or left by clicking and dragging the three vertical dots 

The individual windows can also be completely collapsed using the right or left arrow icons 

If you hover the cursor over the Results a  icon appears, clicking on this provides a range of options including:

- **Remove all** analyses from the output window
- **Remove** selected analysis
- **Collapse** the output
- **Add notes** to each output
- **Copy**
- **Copy special (LaTeX code)**
- **Save image as**

The 'add notes' option allows the results output to be easily annotated and then exported to an HTML file by going to File > Export Results.



The Add notes menu provides many options to change text font, colour size etc.

Bayesian ANOVA

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Group	0.500	1.000	1.425e +28	1.000	
Null model	0.500	7.016e −29	7.016e −29	7.016e −29	6.507e −6

One-way Bayesian independent ANOVA

You can change the size of all the tables and graphs using ctrl+ (increase) ctrl- (decrease) ctrl= (back to default size). Graphs can also be resized by dragging the bottom right corner of the graph.

As previously mentioned, all tables and figures are APA standard and can just be copied into any other document. Since all images can be copied/saved with either a white or transparent background. This can be selected in Preferences > Advanced as described earlier.

There are many further resources on using JASP on the website https://jasp-stats.org/

DATA HANDLING IN JASP

For this section open England injuries.csv

All files must have a header label in the first row. Once loaded, the dataset appears in the window:

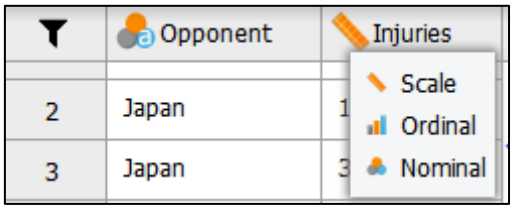
	Opponent	Injuries
1	Japan	4
2	Japan	1
3	Japan	3
4	Japan	6
5	Japan	2
6	Japan	3
7	Japan	4
8	Japan	0
9	Japan	5
10	Japan	2
11	Japan	2
12	New Zealand	2
13	New Zealand	4

For large datasets, there is a hand icon which allows easy scrolling through the data. 

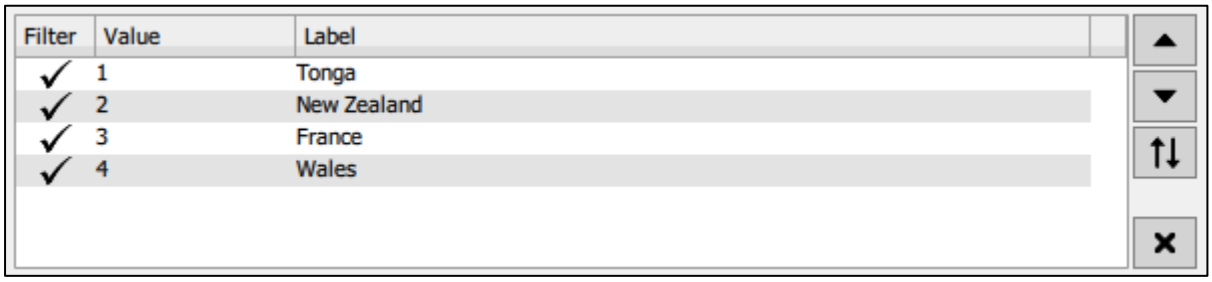
On import JASP makes a best guess at assigning data to the different variable types:

Nominal  **Ordinal**  **Continuous** 

If JASP has incorrectly identified the data type just click on the appropriate variable data icon in the column title to change it to the correct format.

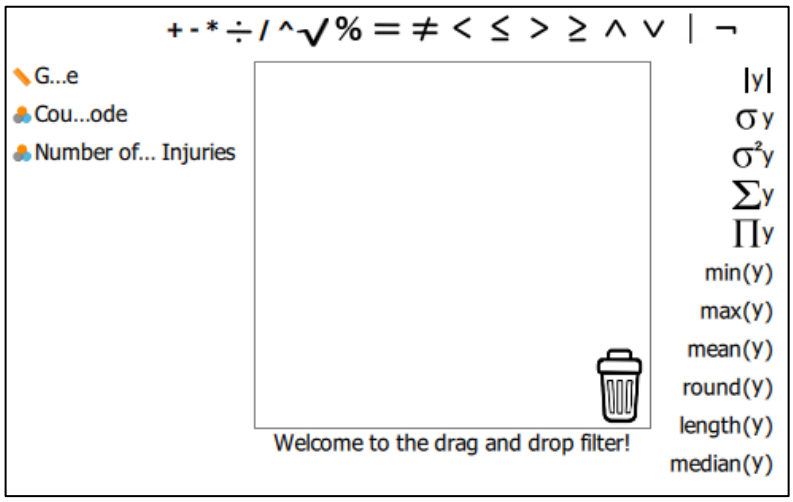


If you have coded the data, you can click on the variable name to open up the following window in which you can label each code. These labels now replace the codes in the spreadsheet view. If you save this as a .**jasp** file these codes, as well as all analyses and notes, will be saved automatically. This makes the data analysis fully reproducible.



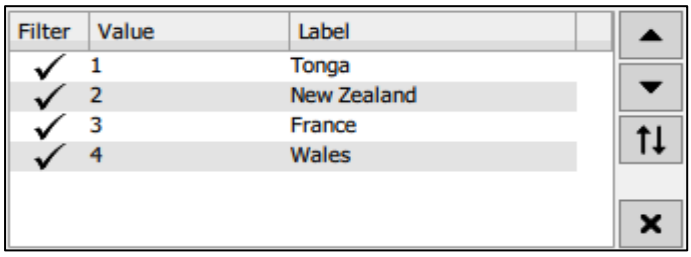
In this window, you can also carry out simple filtering of data, for example, if you untick the Wales label it will not be used in subsequent analyses.

 Clicking this icon in the spreadsheet window opens a much more comprehensive set of data filtering options:



Using this option will not be covered in this document. For detailed information on using more complex filters refer to the following link: https://jasp-stats.org/2018/06/27/how-to-filter-your-data-in-jasp/

By default, JASP plots data in the Value order (i.e. 1-4). The order can be changed by highlighting the label and moving it up or down using the appropriate arrows:

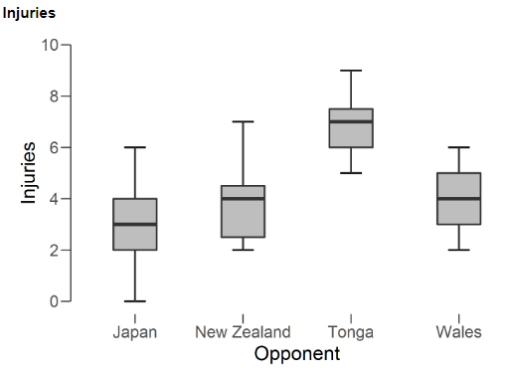


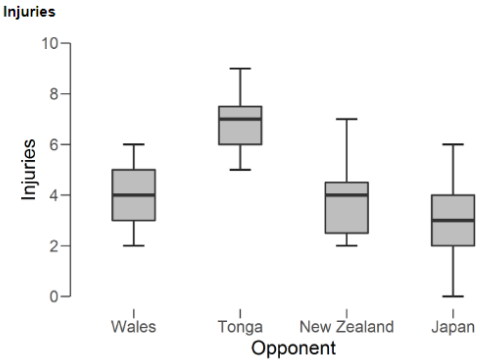
Move up

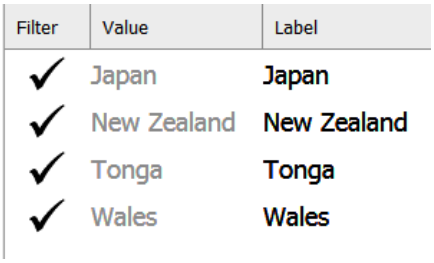
Move down

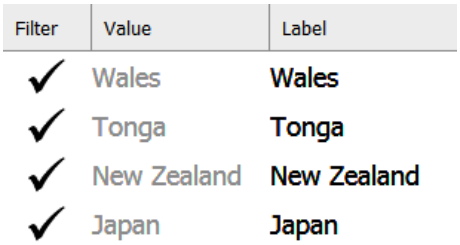
Reverse order

Close









If you need to edit the data in the spreadsheet just double click on a cell and the data should open up in the original spreadsheet i.e. Excel. Once you have edited your data and saved the original spreadsheet JASP will automatically update to reflect the changes that were made, provided that you have not changed the file name.

JASP ANALYSIS MENU

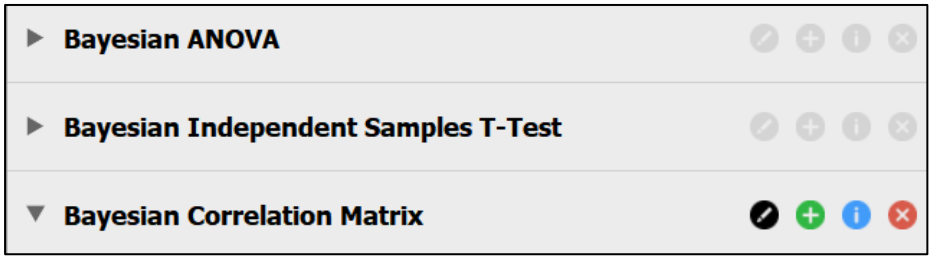
The **main** analysis options can be accessed from the main toolbar. JASP offers a range of frequentist and Bayesian (parametric and non-parametric) statistics and for the purpose of this guide the following alternative Bayesian tests are described:

Descriptives • Descriptive stats	**Bayesian Correlation & Regression** • Correlation • Linear regression
Bayesian T-Tests • Independent • Paired • One sample	**Bayesian Frequencies** • Binomial test • Multinomial test • Contingency tables
Bayesian ANOVA • Independent • Repeated measures • Mixed factor	**BAIN** • Bayesian informative hypotheses evaluation

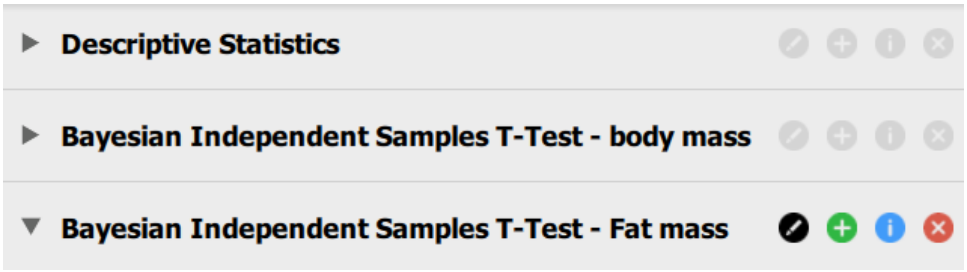
BY clicking on the + icon on the top-right menu bar you can also access advanced options including; Network analysis, Meta-Analysis, Structural Equation Modelling and Bayesian Summary stats.

Once you have selected your required analysis all the possible statistical options appear in the left window and output in the right window.

JASP has the ability to rename and 'stack' the results output thereby organising multiple analyses.



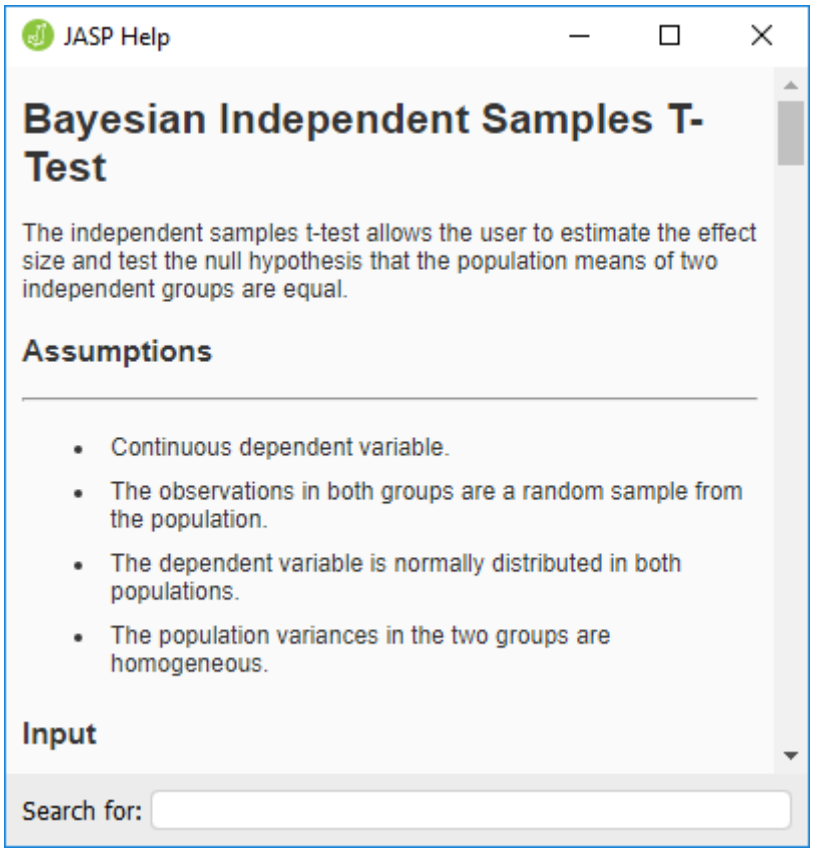
The individual analyses can be renamed using the pen icon or deleted using the red cross.



By clicking on the analysis in this list will then take you to the appropriate part of the results output window. They can also be rearranged by dragging and dropping each of the analyses.

The green + icon produces a copy of the chosen analysis

The blue information icon provides detailed information on each of the statistical procedures used and a search option.



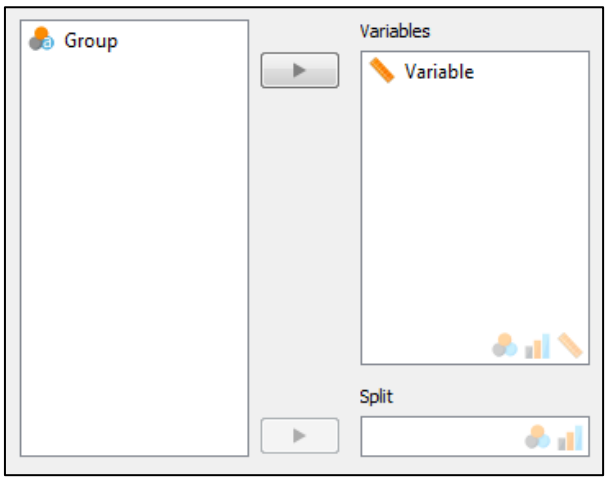
DESCRIPTIVE STATISTICS

Presentation of all the raw data is very difficult for a reader to visualise or to draw any inference on. Descriptive statistics and related plots are a succinct way of describing and summarising data but do not test any hypotheses. There are various types of statistics that are used to describe data:

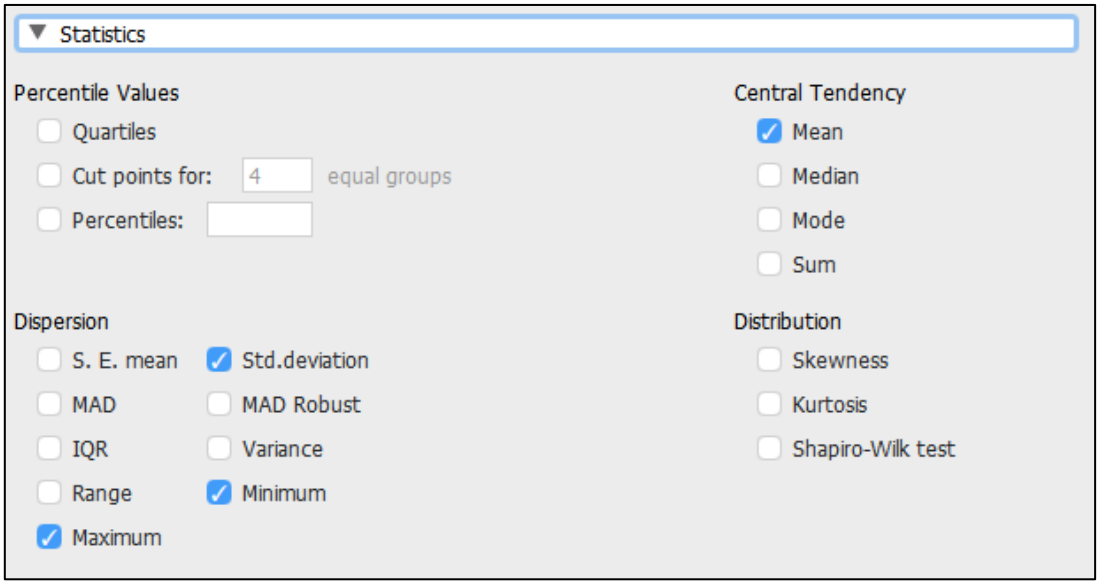
- Measures of central tendency
- Measures of dispersion
- Percentile values
- Measures of distribution
- Descriptive plots

In order to explore these measures, load **Descriptive data.csv** into JASP. Go to Descriptives > Descriptive statistics and move the Variable data to the Variables box on the right.

	Group	Variable
1	Group 1	26.4
2	Group 1	8.4
3	Group 1	8.5
4	Group 1	22.9
5	Group 1	21.7
6	Group 1	14.1
7	Group 1	13.8
8	Group 1	15
9	Group 1	20.5
10	Group 1	21.7
11	Group 1	32.3
12	Group 1	9.7



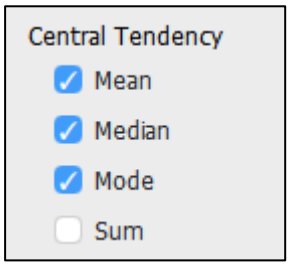
The Statistics menu can now be opened to see the various options available.



CENTRAL TENDENCY.

This can be defined as the tendency for variable values to cluster around a central value. The three ways of describing this central value are mean, median or mode. If the whole population is considered, we the term population mean / median/mode is used. If a sample/subset of the population is being analysed the term sample mean/ median/mode is used. The measures of central tendency move toward a constant value when the sample size is sufficient to be representative of the population.

In the Statistics options make sure that everything is unticked apart from mean, median and mode.



Descriptive Statistics

	Variable
Valid	810
Missing	0
Mean	17.71
Median	17.90
Mode	20.00

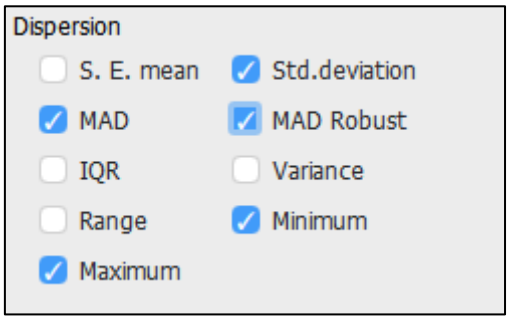
The ***mean***, **M or $\bar{x}$** (17.71) is equal to the sum of all the values divided by the number of values in the dataset i.e. the average of the values. It is used for describing continuous data. It provides a simple statistical model of the centre of distribution of the values and is a theoretical estimate of the 'typical value'. However, it can be influenced heavily by 'extreme' scores.

The ***median***, **Mdn** (17.9) is the middle value in a dataset that has been ordered from the smallest to largest value and is the normal measure used for ordinal or non-parametric continuous data. Less sensitive to outliers and skewed data

The ***mode*** (20.0) is the most frequent value in the dataset and is usually the highest bar in a distribution histogram

DISPERSION

In the Statistics options make sure that the following options are ticked



Standard deviation, S or SD (6.94) is used to quantify the amount of dispersion of data values around the mean. A low standard deviation indicates that the values are close to the mean, while a high standard deviation indicates that the values are dispersed over a wider range.

Variance (S^2 = 48.1) is another estimate of how far the data is spread from the mean. It is also the square of the standard deviation.

The standard error of the mean, SE (0.24) is a measure of how far the sample mean of the data is expected to be from the true population mean. As the size of the sample data grows larger the SE decreases compared to S and the true mean of the population is known with greater specificity.

MAD, median absolute deviation, a robust measure of the spread of data. It is relatively unaffected by data that is not normally distributed. Reporting median +/- MAD for data that is not normally distributed is equivalent to mean +/- SD for normally distributed data.

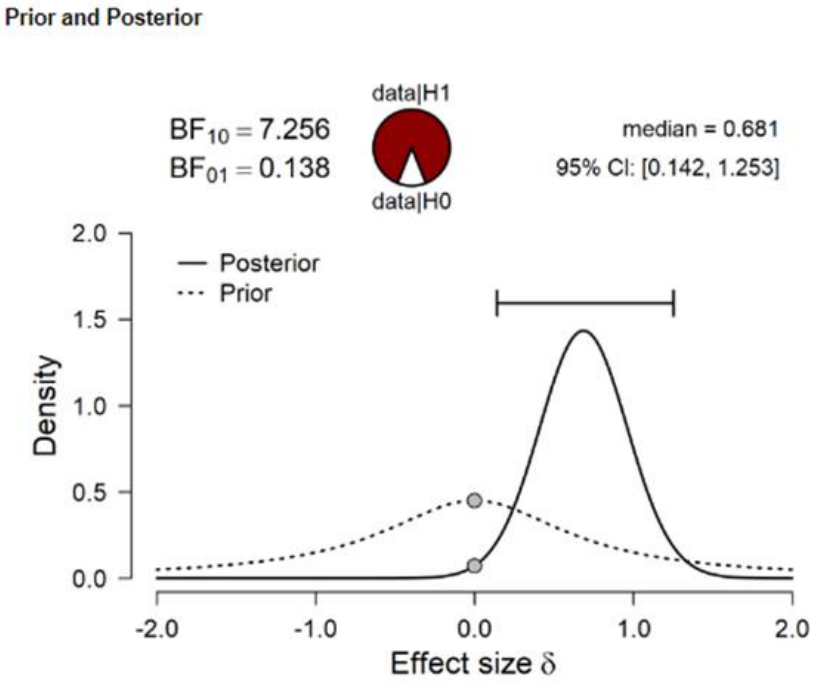
MAD Robust: Median absolute deviation of the data points, adjusted by a factor for asymptotically normal consistency.

IQR - Interquartile Range is similar to the MAD but is less robust (see Boxplots).

Variance: Variance of the data points

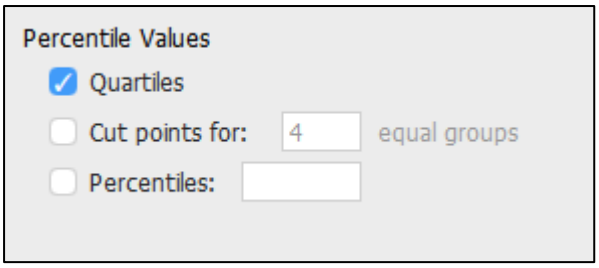
Credible intervals (CI), although not shown in the general Descriptive statistics output, these are used in many other statistical tests. They are an important concept when looking at Bayesian inference and are somewhat similar to confidence intervals used in frequentist statistics although their meaning is very different.

Bayesian analyses produce a posterior distribution of the possible effect values. A 95% credible interval is simply the central portion of the posterior distribution that contains 95% of the values i.e. given the observed data, the effect has a 95% probability of falling within this range.



QUARTILES

In the Statistics options make sure that everything is unticked apart from Quartiles.



Descriptive Statistics

	Variable
Valid	810
Missing	0
25th percentile	13.05
50th percentile	17.90
75th percentile	22.30

Quartiles are where datasets are split into 4 equal quarters, normally based on rank ordering of median values. For example, in this dataset

1	1	2	2	3	3	4	4	4	4	5	5	5	6	7	8	8	9	10	10	10
				25%						50%					75%						

The median value that splits data by 50% = 50th percentile = 5

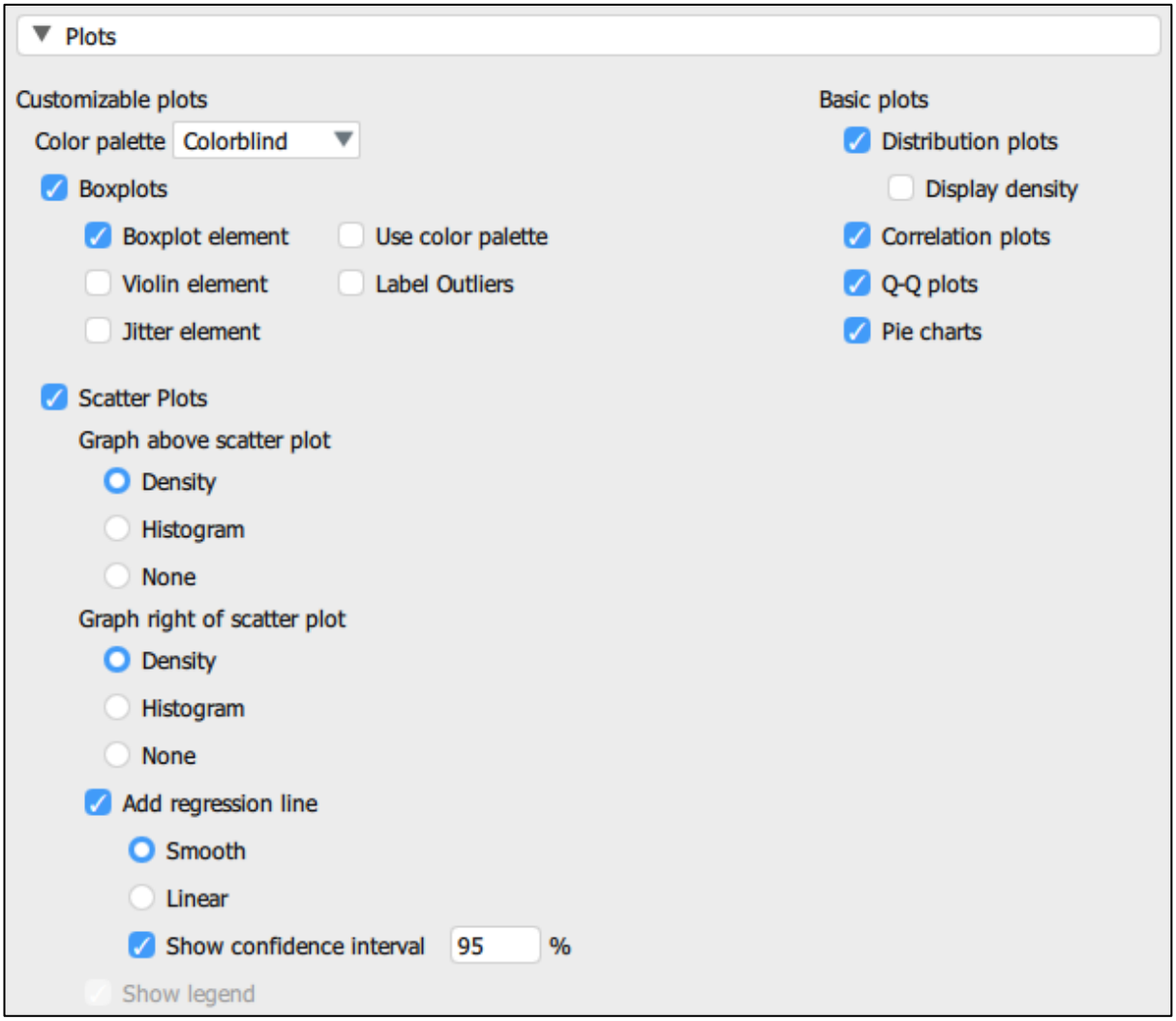
The median value of left side = 25th percentile = 3

The median value of right side = 75th percentile = 8

From this the Interquartile range (IQR) range can be calculated, this is the difference between the 75th and 25th percentiles i.e. 5. These values are used to construct the descriptive boxplots later. The IQR can also be shown by ticking this option in the Dispersion menu.

DESCRIPTIVE PLOTS IN JASP

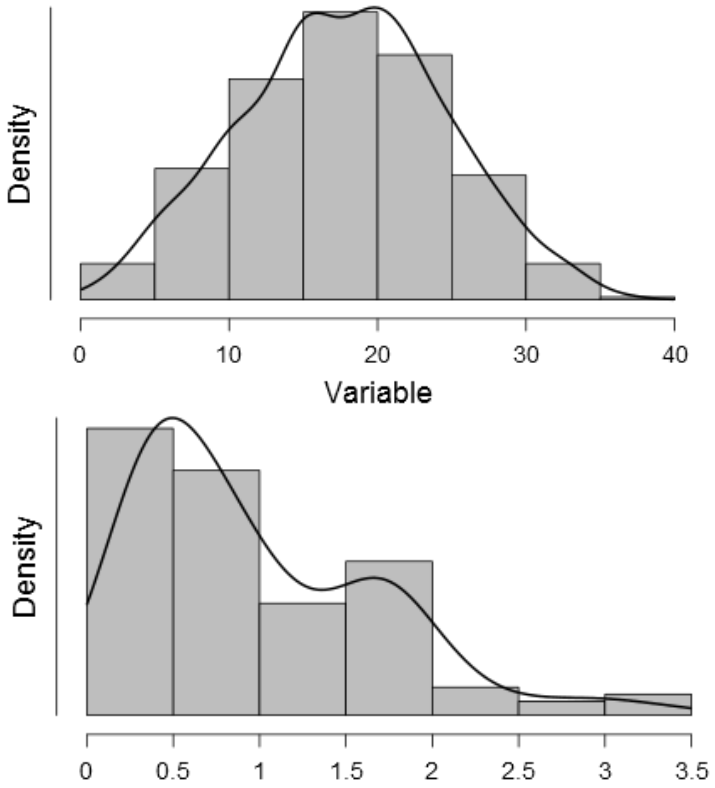
JASP can produce a range of descriptive plots:



Again, using **Descriptive data.csv** with the variable data in the Variables box, go to the statistics options and under Plots tick Distribution plots, Boxplots – Boxplot Element and Q-Q plots.

Distribution plots

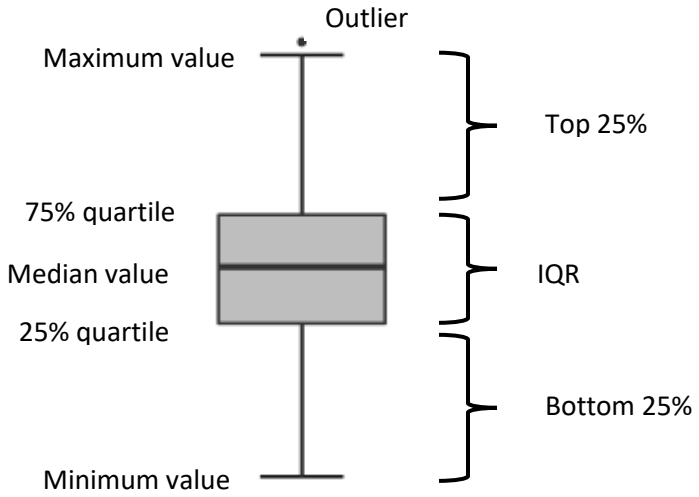
The Distribution plot is based on splitting the data into frequency bins, this is then overlaid with the distribution curve. As mentioned before, the highest bar is the mode (most frequent value of the dataset. In this case, the curve looks approximately symmetrical suggesting that the data is approximately normally distributed. The second distribution plot is from another dataset which shows that the data is positively skewed.



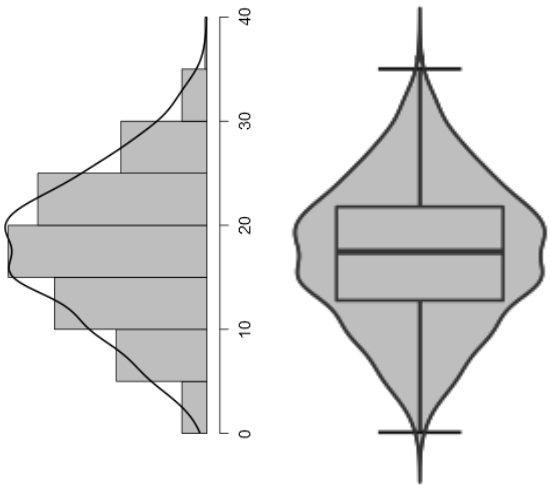
Boxplots

The boxplots visualise several statistics described above in one plot:

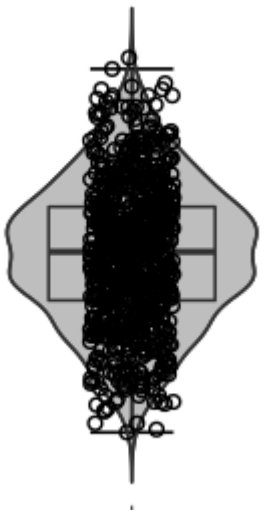
- Median value
- 25 and 75% quartiles
- Interquartile range (IQR) i.e. 75% - 25% quartile values
- Maximum and minimum values plotted with outliers excluded
- Outliers are shown if requested



Go back to the statistics options, in Descriptive plots tick both Boxplot and Violin Element, look at how the plot has changed. Next tick Boxplot, Violin and Jitter Elements. The Violin plot has taken the smoothed distribution curve from the Distribution plot, rotated it 90° and superimposed it on the boxplot. The jitter plot has further added all the data points.

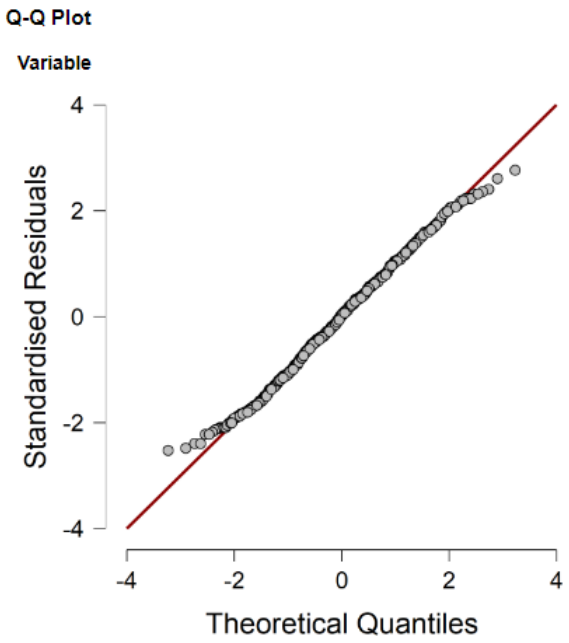


Boxplot + Violin plot



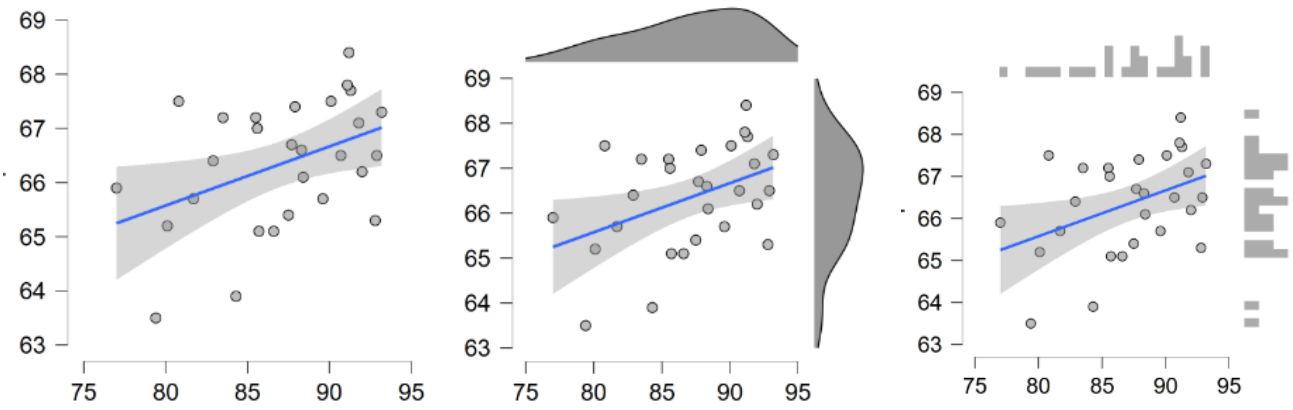
Boxplot + Violin + Jitter plot

A Q-Q plot (quantile-quantile plot) can be used to visually assess if a set of data comes from a normal distribution. Q-Q plots take the sample data, sort it in ascending order, and then plot them against quantiles (percentiles) calculated from a theoretical distribution. If the data is normally distributed, the points will fall on or close to the 45-degree reference line. If the data is not normally distributed, the points will deviate from the reference line.



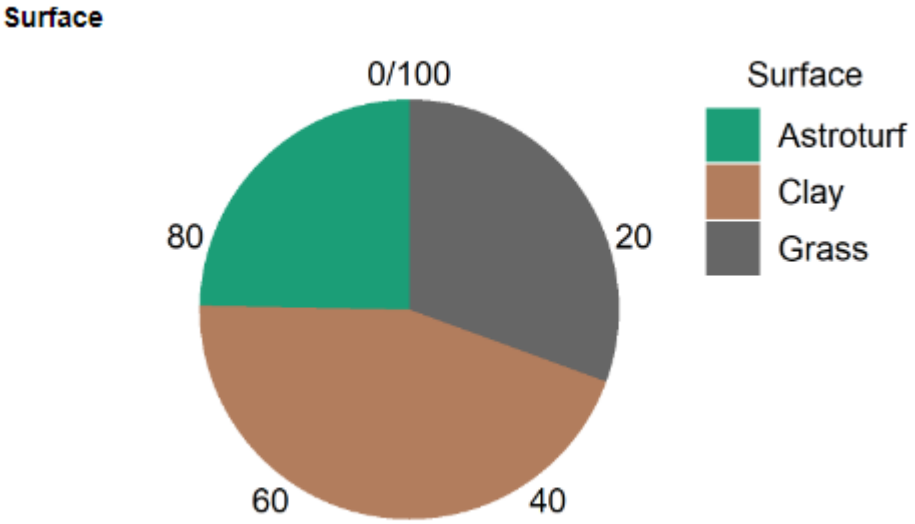
Scatter plots

JASP v0.12 introduced the ability to produce scatterplots of various types and to be able to include smooth or linear regression lines. There are also options to add distributions to these either in the form of density plots or histograms.



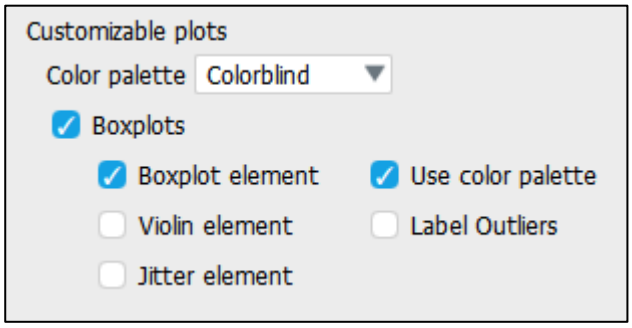
Pie charts

Also introduced was the ability to plot piecharts when working with categorical or other frequency data.



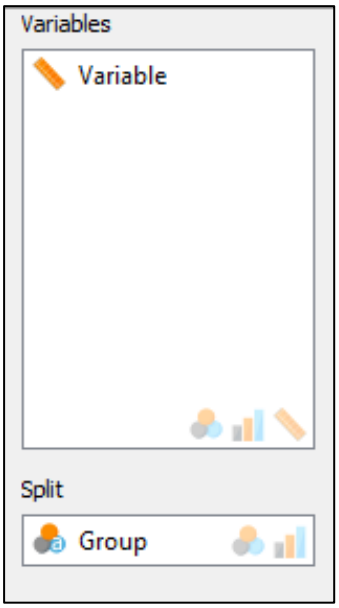
Plot colour palettes

Users can choose from between 5 different colour palettes using the drop-down menu



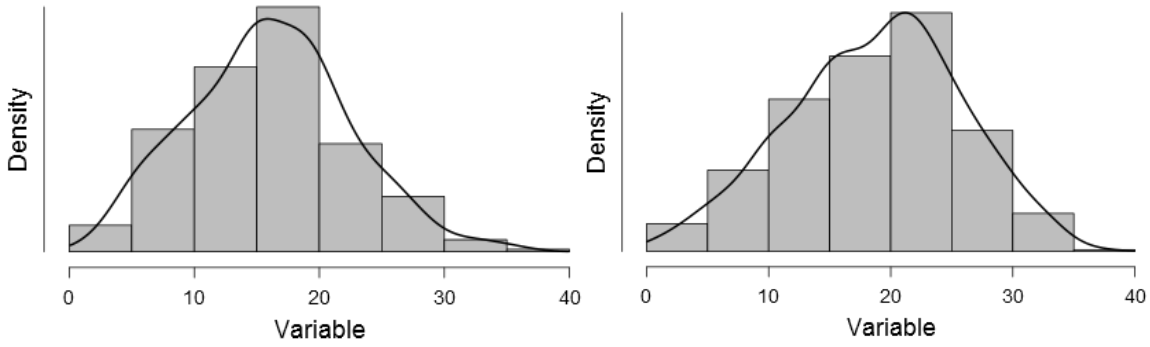
SPLITTING DATA FILES

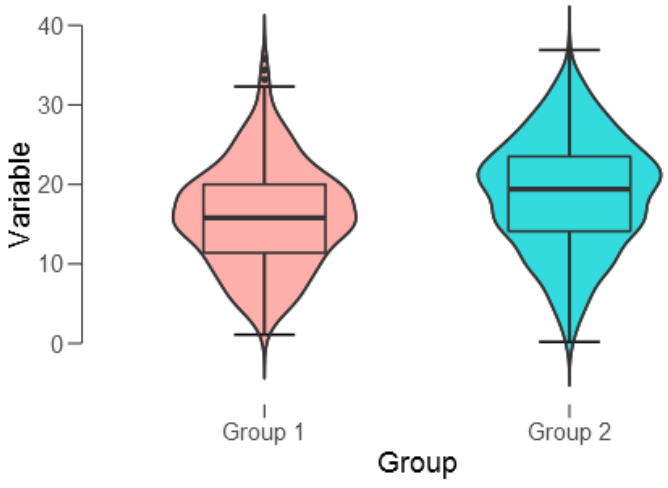
If there is a grouping variable (categorical or ordinal) descriptive statistics and plots can be produced for each group. Using **Descriptive data.csv** with the variable data in the Variables box now add Group to the Split box.



Descriptive Statistics

	Variable	
	Group 1	Group 2
Valid	315	495
Missing	0	0
Mean	16.021	18.787
Median	15.800	19.400
MAD	4.200	5.000
Minimum	1.100	0.200
Maximum	35.800	36.900





EXPLORING DATA INTEGRITY

Sample data is used to estimate parameters of the population whereby a parameter is a measurable characteristic of a population, such as a mean, standard deviation, standard error or confidence intervals etc.

What is the difference between a statistic and a parameter? If you randomly polled a selection of students about the quality of their student bar and you find that 75% of them were happy with it. That is a sample **statistic** since only a sample of the population were asked. You calculated what the population was likely to do based on the sample. If you asked **all** the students in the university and 90% were happy you have a **parameter** since you asked the whole university population.

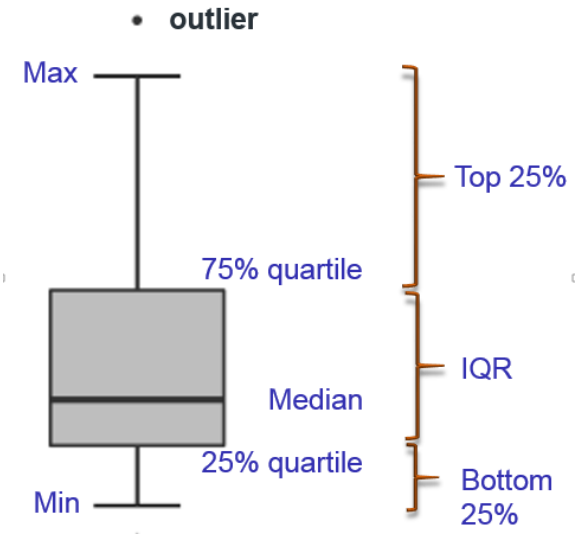
Bias can be defined as the tendency of a measurement to over- or under-estimate the value of a population parameter. There are many types of bias that can appear in research design and data collection including:

- Participant selection bias – some being more likely to be selected for study than others
- Participant exclusion bias - due to the systematic exclusion of certain individuals from the study
- Analytical bias - due to the way that the results are evaluated

However statistical bias can affect a) parameter estimates, b) standard errors and confidence intervals or c) test statistics and *p* values. So how can we check for bias?

IS YOUR DATA CORRECT?

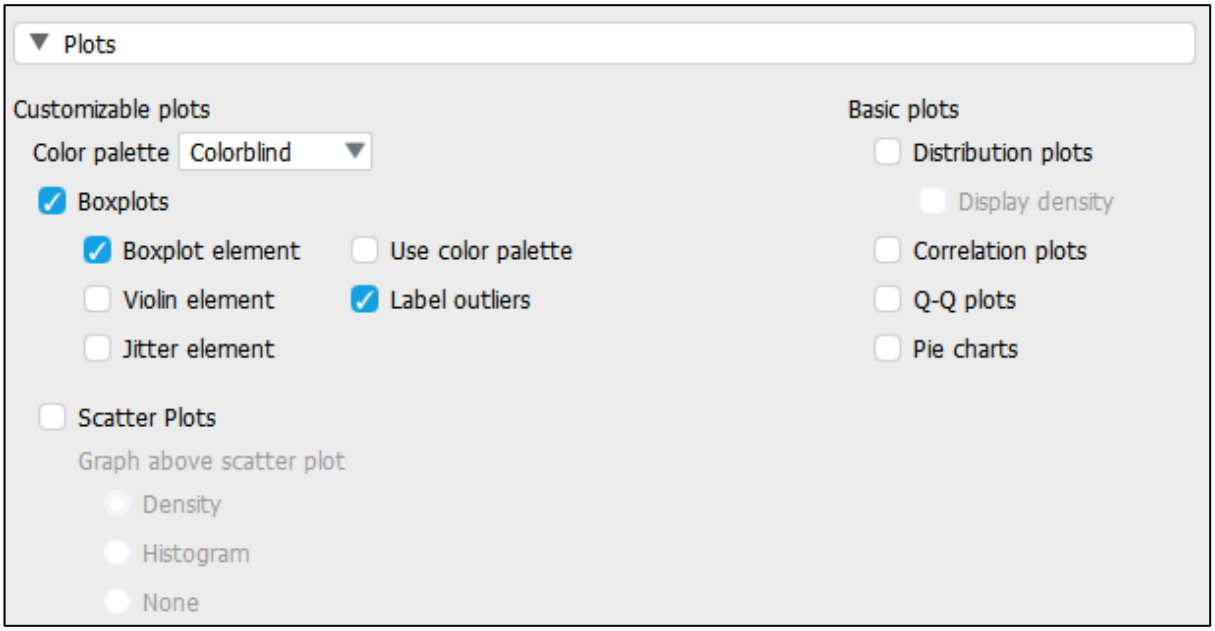
Outliers are data points that are abnormally outside all other data points. Outliers can be due to a variety of things such as errors in data input or analytical errors at the point of data collection Boxplots are an easy way to visualise such data points where outliers are outside the upper (75% + 1.5 * IQR) or lower (25% - 1.5 * IQR) quartiles



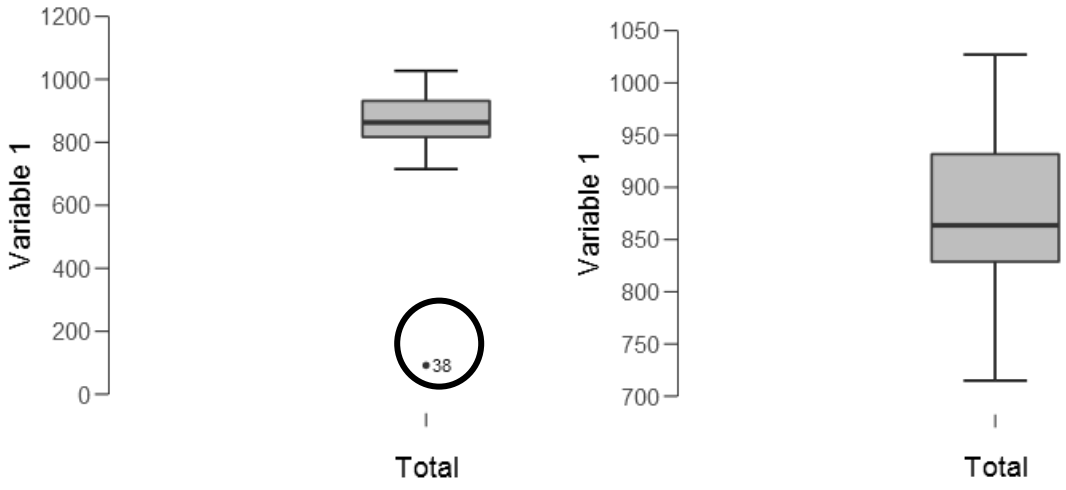
Boxplots show:

- Median value
- 25 & 75% quartiles
- IQR – Inter quartile range
- Max & min values plotted with outliers excluded
- Outliers shown if requested

Load **Exploring Data.csv** into JASP. Under Descriptives > Descriptive Statistics, add Variable 1 to the Variables box. In Plots tick the following Boxplots, Label Outliers, and BoxPlot Element.



The resulting Boxplot on the left looks very compressed and an obvious outlier is labelled as being in row 38 of the dataset. This can be traced back to a data input error in which 91.7 was input instead of 917. The graph on the right shows the BoxPlot for the 'clean' data.



How you deal with an outlier depends on the cause. Most parametric tests are highly sensitive to outliers while non-parametric tests are generally not.

Correct it? – Check the original data to make sure that it isn't an input error, if it is, correct it, and rerun the analysis.

Keep it? - Even in datasets of normally distributed, data outliers may be expected for large sample sizes and should not automatically be discarded if that is the case.

Delete it? – This is a controversial practice in small datasets where a normal distribution cannot be assumed. Outliers resulting from an instrument reading error may be excluded but it should be verified first.

Replace it? – Also known as 'winsorizing'. This technique replaces the outlier values with the relevant maximum and/or minimum values found after excluding the outlier.

Whatever method you use must be justified in your statistical methodology and subsequent analysis.

WE MAKE MANY ASSUMPTIONS ABOUT OUR DATA.

When using parametric tests, we make a series of assumptions about our data and bias will occur if these assumptions are violated, in particular:

- Normality
- Homogeneity of variance or homoscedasticity

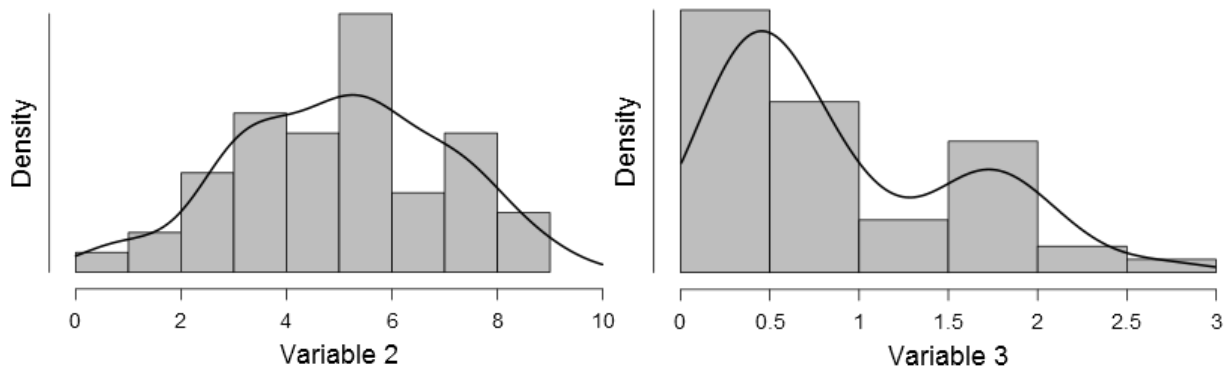
Many statistical tests are an omnibus of tests of which some will check these assumptions.

ASSESSING THE ASSUMPTION OF NORMALITY

Normality does not mean necessarily that the data is normally distributed per se but it is whether or not the dataset can be well modelled by a normal distribution. Normality can be explored in a variety of ways:

- Numerically
- Visually / graphically
- Statistically

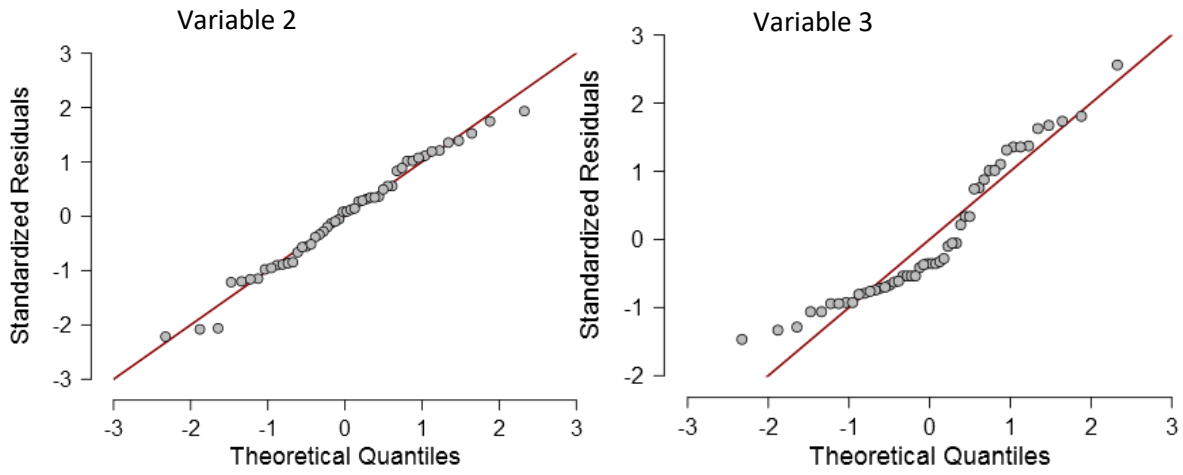
Using **Exploring data.csv**, go to Descriptives>Descriptive Statistics move Variables 2 and 3 to the Variables box and in Plots tick Distribution plot. This will show the following two graphs:



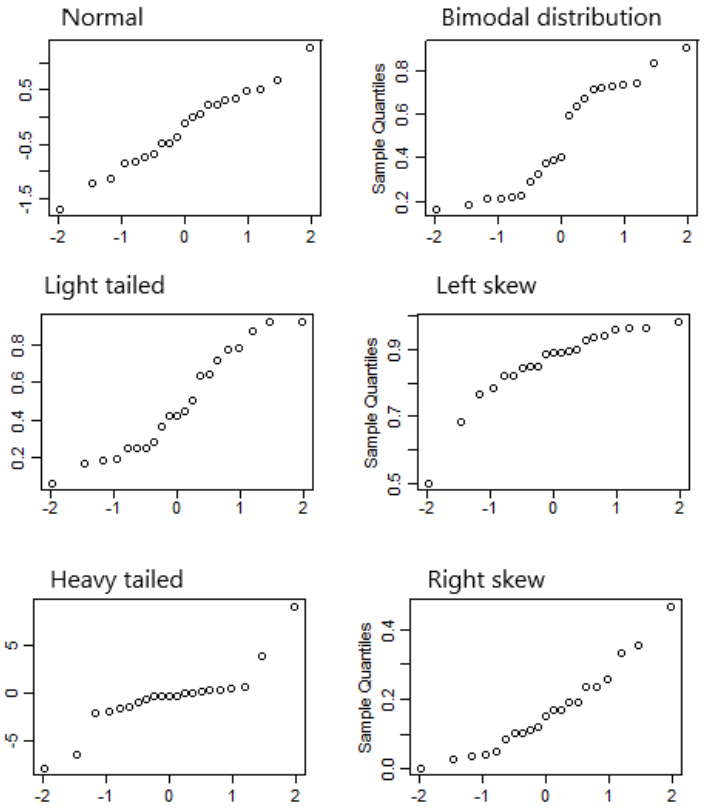
It is quite easy to visualise that Variable 2 has a symmetrical distribution. Variable 3 is skewed to the left.

Another graphical check for normality is a Q-Q plot. These show the quantiles of the actual data against those expected for a normal distribution.

If data are normally distributed all the points will be close to the diagonal reference line. If the points 'sag' above or below the line, there is a problem with kurtosis. If the points snake around the line, then the problem is skewness. Below are Q-Q plots for Variables 2 and 3. Compare these to the previous distribution plots.



The following Q-Q plot scenarios are possible:



Currently, there is no Bayesian equivalent of the Shapiro-Wilk test in JASP to check the assumption of normality.

Testing the assumption of normality – A cautionary note!

For most parametric tests to be reliable, one of the assumptions is that the data is **approximately** normally distributed. A normal distribution peaks in the middle and is symmetrical about the mean. However, data does not need to be perfectly normally distributed for the tests to be reliable.

So, having gone on about testing for normality – is it necessary?

The Central Limit Theorem states that as the sample size gets larger i.e. >30 data points the distribution of the sampling means approaches a normal distribution. So, the more data points you have the more normal the distribution will look and the closer your sample mean approximates the population mean.

However, data that does not meet the assumption of normality is going to result in poor results for certain types of test (i.e. ones that state that the assumption must be met!). How closely does your data need to be normally distributed? This is a judgment call best made by eyeballing the data.

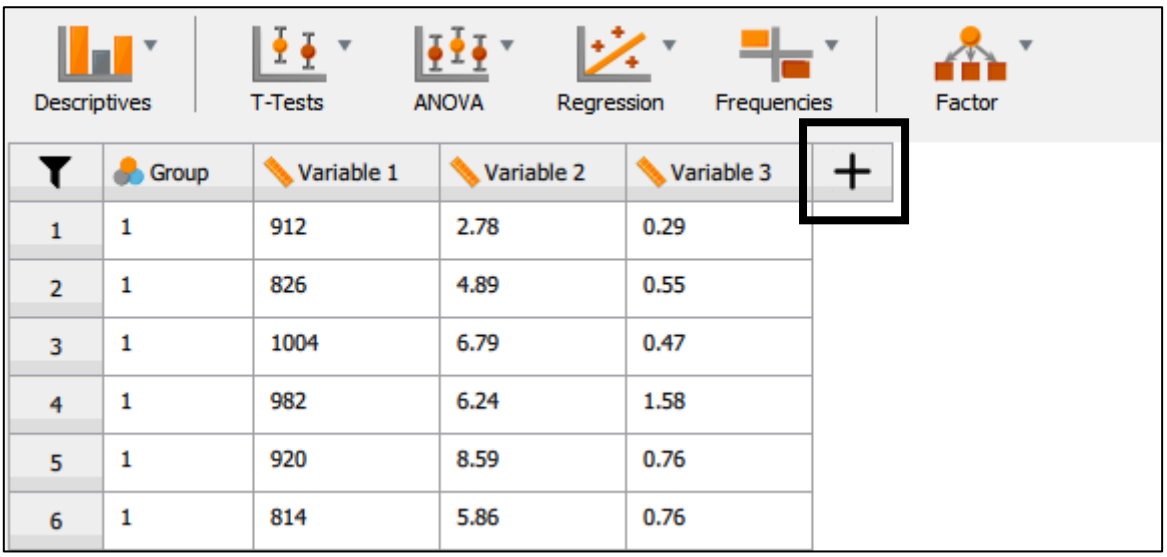
WHAT DO I DO IF MY DATA IS REALLY NOT NORMALLY DISTRIBUTED?

Transform the data and redo the normality checks on the transformed data. Common transformations include taking the log or square root of the data.

Use non-parametric Bayesian tests since these are distribution-free tests and can be used instead of their parametric equivalent.

DATA TRANSFORMATION

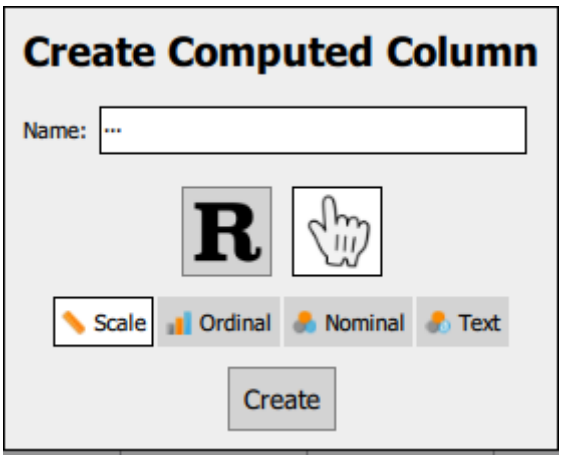
In some cases, it may be useful to compute the differences between repeated measures or, to make a dataset more normally distributed, you can apply a log transform for example. When a dataset is opened there will be a plus sign (**+**) at the end of the columns.



	Group	Variable 1	Variable 2	Variable 3	+
1	1	912	2.78	0.29	
2	1	826	4.89	0.55	
3	1	1004	6.79	0.47	
4	1	982	6.24	1.58	
5	1	920	8.59	0.76	
6	1	814	5.86	0.76	

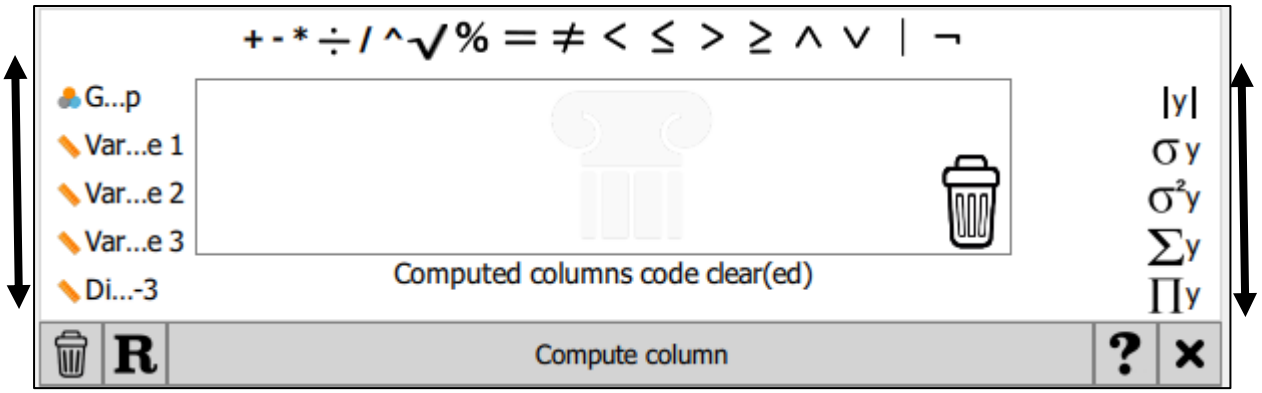
Clicking on the **+** opens a small dialogue window where you can:

- Enter the name of a new variable or the transformed variable
- Select whether you enter the R code directly or use the commands built into JASP
- Select what data type is required

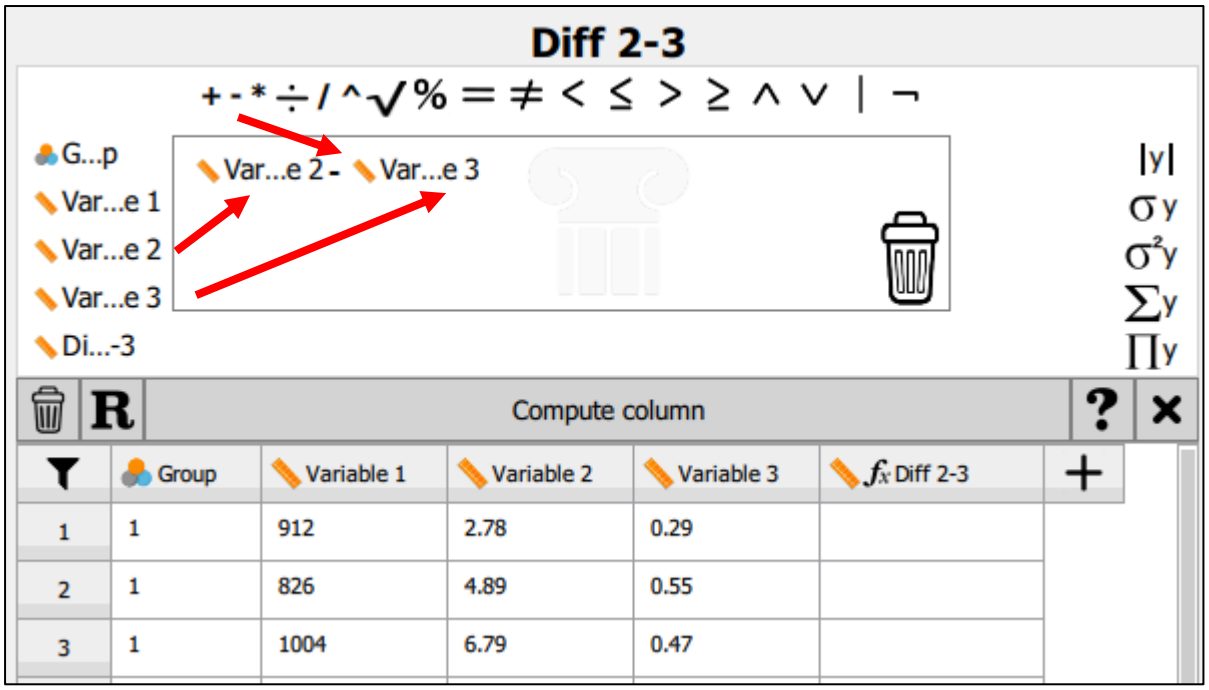


Once you have named the new variable and chose the other options – click create.

If you choose the manual option rather than the R code, this opens all the built-in create and transform options. Although not obvious, you can scroll the left and right-hand options to see more variables or more operators respectively.

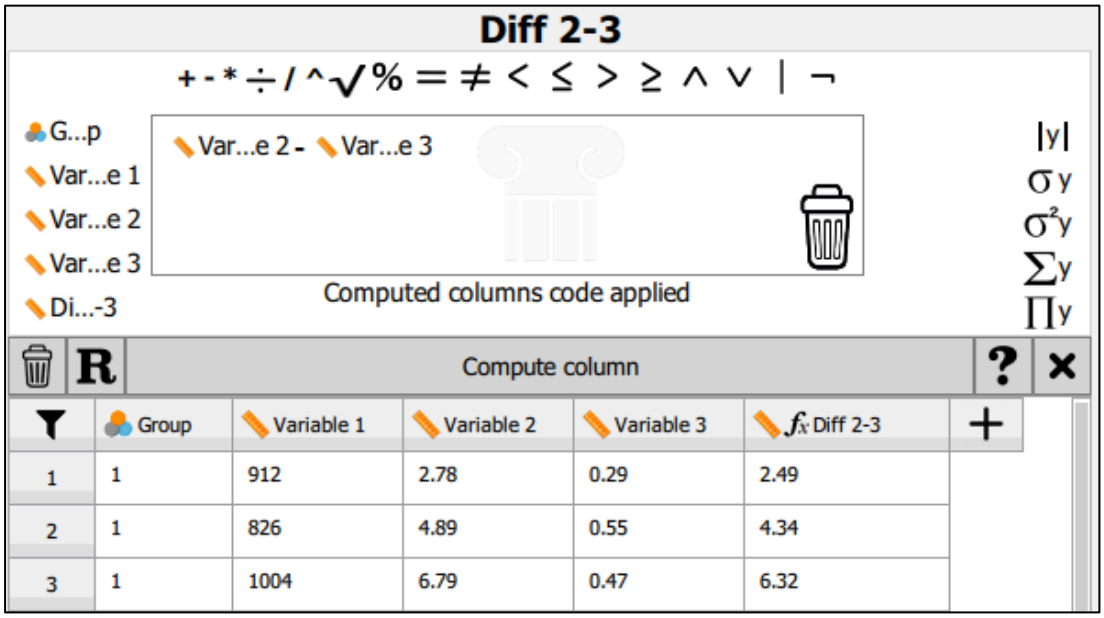


For example, we want to create a column of data showing the difference between variable 2 and variable 3. Once you have entered the column name in the Create Computed Column dialogue window, its name will appear in the spreadsheet window. The mathematical operation now needs to be defined. In this case drag variable 2 into the equation box, drag the 'minus' sign down and then drag in variable 3.



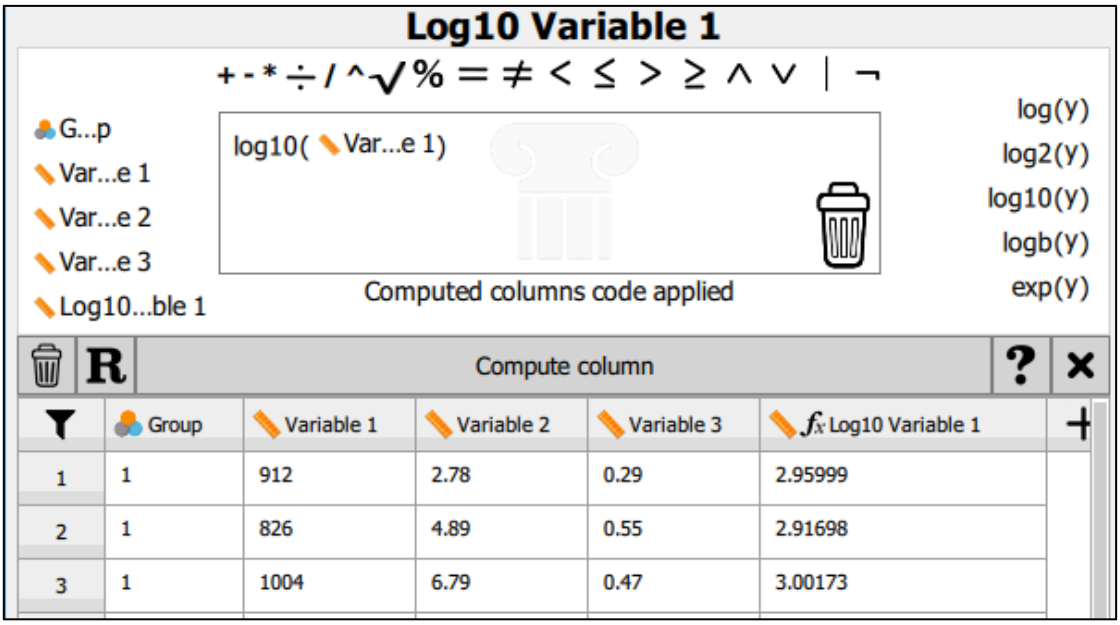
If you have made a mistake, i.e. used the wrong variable or operator, remove it by dragging the item into the dustbin in the bottom right corner.

When you are happy with the equation/operation, click compute column and the data will be entered.

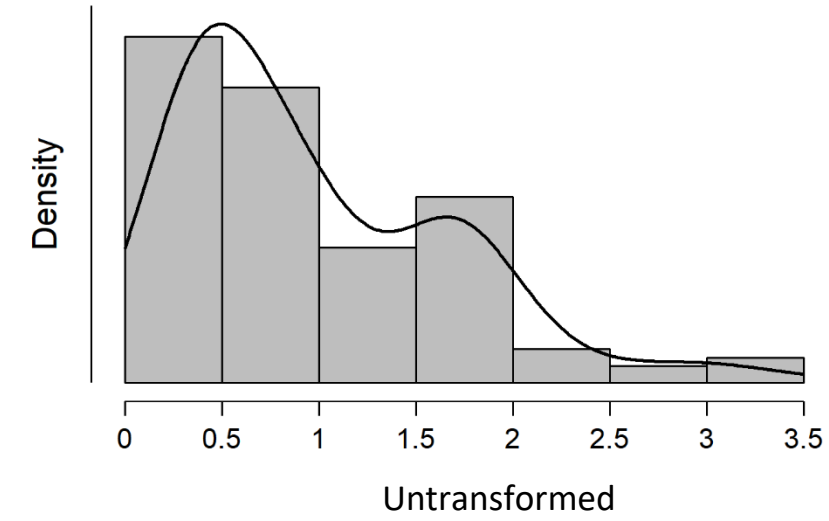


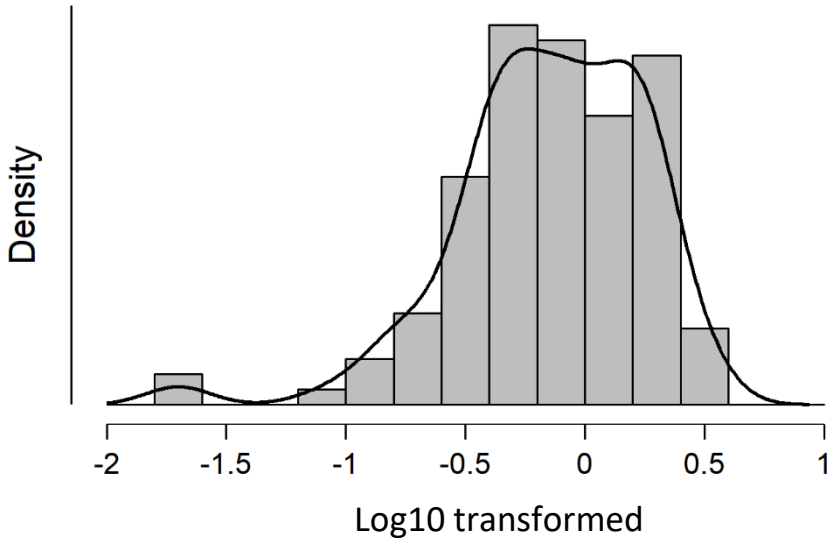
If you decide that you do not want to keep the derived data, you can remove the column by clicking the other dustbin icon next to the **R**.

Another example is to do a log transformation of the data. In the following case variable 1 has been transformed by scrolling the operators on the left and selecting the log10(y) option. Replace the "y" with the variable that you want to transform and then click Compute column. When finished, click the **X** to close the dialogue.



The two graphs below show the untransformed and the log10 transformed data. The obviously skewed data has been transformed into a profile with a more normal distribution





The Export function will also export any new data variables that have been created.

BAYESIAN INFERENCE METHODS

SOME BAYESIAN TERMINOLOGY[3]

Bayesian Statistics

A statistical tool that can be used to combine background knowledge of population parameters with current data to obtain estimates via the resulting posterior distribution.

Bayes Factor

Evaluates the conditional probability between two competing hypotheses. The aim is to quantify support levels for each hypothesis, which can be updated as new information becomes available, instead of generating definitive accept or reject hypothesis decisions.

Credibility Interval

The Bayesian version of the traditional confidence interval. Can be interpreted as the (e.g. 95%) probability that the population parameter is between the particular upper and lower bounds determined by the Bayesian credibility interval

Likelihood Function

Represents the observed data likelihood. This weights the prior distribution in Bayesian statistics to obtain the posterior distribution from which we draw inferences.

Markov Chain Monte Carlo (MCMC)

A simulation-based estimation method that is used to make simulated draws from a distribution and form a Markov chain that represents the posterior distribution.

Prior distribution

A statistical distribution that can be used to capture the amount of (un)certainty in a population parameter. This distribution is then weighted by the sample data to obtain the posterior, which is used to make an inference.

Prior odds

The odds of the outcome before the evidence is considered. These can be uninformative (assigning equal probabilities to all possibilities) or informative based on previous findings/knowledge.

Posterior distribution

The distribution that is obtained once combining the prior and the likelihood in the Bayesian estimation process.

Posterior odds

Posterior odds = Bayes factor × prior odds. From this formula, we see that the Bayes' factor (BF) tells us whether the data provides evidence for or against the hypothesis
assigns equal probabilities to all possibilities

[3] Adapted from Schoot, Rens & Depaoli, Sarah. (2014). Bayesian analyses: Where to start and what to report. European Health Psychologist. 16. 75-84.

GUIDELINES TO UNDERSTANDING PLOTS IN BAYESIAN ANALYSES

This section sets out to explain the meaning of the common plots seen in some of the Bayesian analyses. More complex analyses such as Regression and two-way ANOVAs have their own specific plots and will be dealt with in their individual sections.

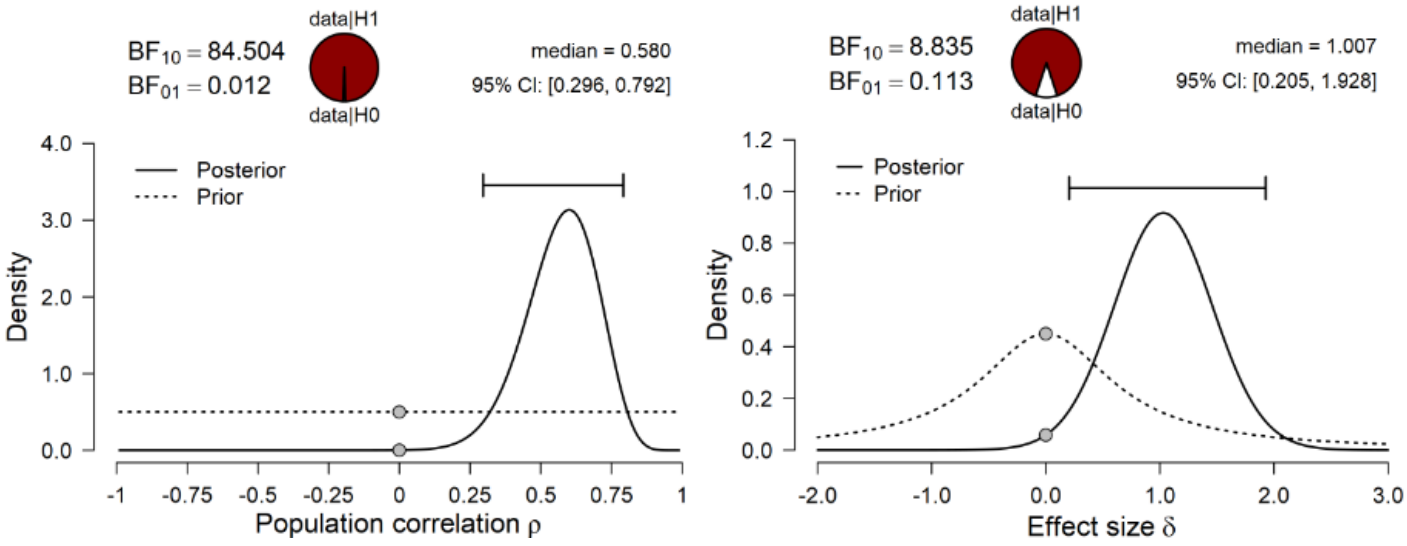
Posterior and Prior Plots – parameter estimation.

Example plots for a two-sided ($H_1 \neq H_0$) correlation and independent t-test are shown below (left and right respectively). These plots provide information for parameter estimation, as well as hypothesis testing.

In each case, the dotted line represents the prior, the probability distribution of the parameter under the alternative hypothesis *before* actually seeing the data.

For a 2-sided correlation, the default stretched beta distribution states that any correlation coefficient (rho: ρ) between -1, and 1 is possible, and is equally likely *a priori*, hence the uniform distribution. In the case of hypothesis testing, the two rival hypotheses tested are H_0: ρ = 0 and H1: ρ ≠ 0 (more specifically: H: ρ ~ Uniform (-1, 1)).

For the 2-sided independent t-test, the prior is defined by a Cauchy distribution centred on a zero effect size (δ) and a width/scale of .707 (default in JASP). This distribution reflects our beliefs about likely values of the population parameter, before seeing the data. The prior distribution depicted below reflects the belief that values of the effect size close to 0 are relatively plausible, whereas values greater than 1 are less plausible.



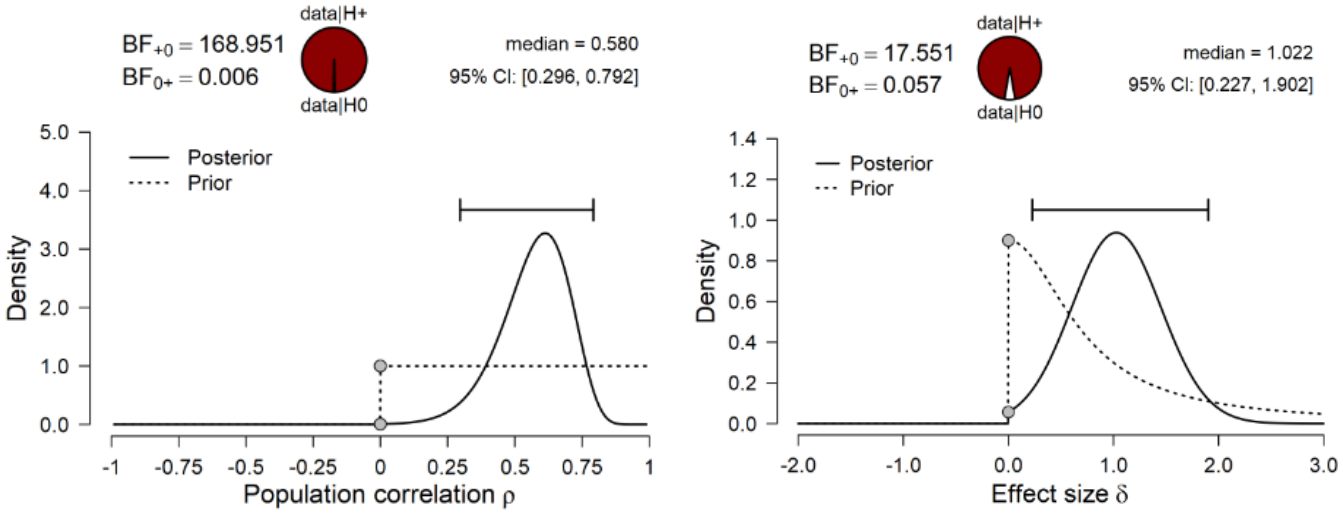
In the case of hypothesis testing, the two rival hypotheses tested are H_0: δ = 0 and H1: δ ≠ 0 (more specifically, H1: δ ~ Cauchy (0.707)).

The solid lines show the posterior distribution (which is conditional on H1 being true), i.e. the updated probability distribution of the parameter of interest *after* seeing the data. The horizontal bar represents the 95% credible intervals around the median correlation or effect size.

The two grey dots indicate the density values of the two distributions where the parameter of interest is equal to the point of testing (e.g., 0 for the correlation or effect size). The ratio of these two values is called the Savage-Dickey density ratio, which gives the BF_{10} and BF_{01} values (also shown above), depending on which way around the division is done. Visually, if the grey dot of the posterior distribution is lower than that of the prior distribution the evidence supports the alternative hypothesis and v.v.

Posterior and Prior Plots – hypothesis testing.

The initial results above are based on two-sided non-directional alternative hypotheses. Bayesian tests also enable one-sided directional hypotheses to be tested. To test the associated one-sided hypothesis, you can specify "Correlated positively", and "Group 1 > Group 2". The new prior– posterior plots are shown below (left and right respectively).



Now the prior distribution densities are concentrated to the right of 0 in each case, reflecting the directionality of the alternative hypothesis. Both Bayes factors have increased in magnitude, compared to the two-sided tests, thus favouring the alternative directional hypotheses (H+).

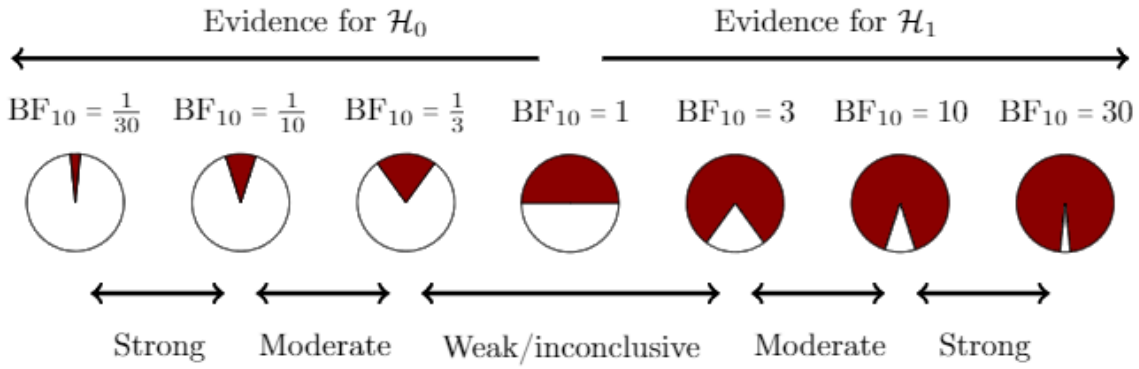
How strong is the evidence?

Different descriptive classifications have been used to interpret Bayes factors. The one adopted by JASP is an adaption of Jeffery's scheme that proposes a series of labels for which specific Bayes factor values can be considered either "anecdotal", "moderate", "strong", "very strong", or "decisive" relative evidence for a hypothesis.

BF_{10}	$Log_e BF_{10}$	Evidence	In favour of
>100	>4.6	Decisive	Alternative hypothesis
30 to 100	3.4 to 4.6	Very strong	Alternative hypothesis
10 to 30	2.3 to 3.4	Strong	Alternative hypothesis
3 to 10	1.1 to 2.3	Moderate	Alternative hypothesis
1 to 3	0 to 1.1	Anecdotal	Alternative hypothesis
1	0	No evidence	Neither
1 to 0.33	0 to -1.1	Anecdotal	Null Hypothesis
0.33 to 0.1	-1.1 to -2.3	Moderate	Null Hypothesis
0.1 to 0.033	-2.3 to -3.4	Strong	Null Hypothesis
0.033 to 0.01	-3.4 to -4.6	Very strong	Null Hypothesis
<0.01	< -4.6	Decisive	Null Hypothesis

However, these are merely a simplified heuristic for interpreting Bayes factors, but that the Bayes factor really is a continuous metric of evidence.

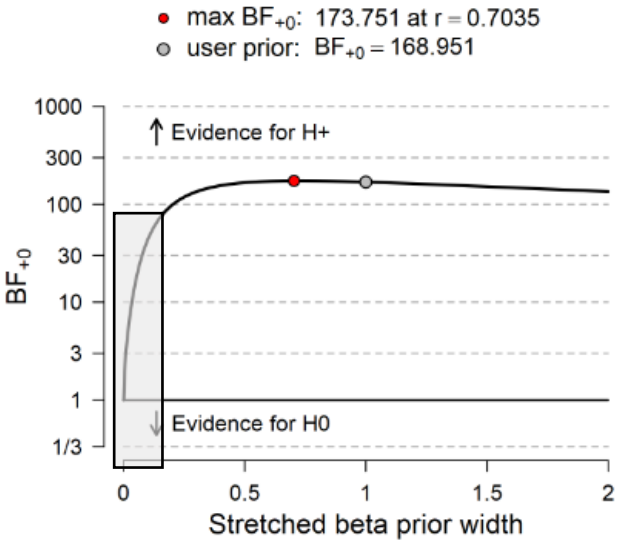
The pizza plots show the transformed odds of two Bayes factors (between 0 and 1). This allows the strength of evidence for each Bayes factor to be easily visualised[4].



[4] Van Doorn J et al (2019) The JASP Guidelines for Conducting and Reporting a Bayesian Analysis. https://psyarxiv.com/yqxfr

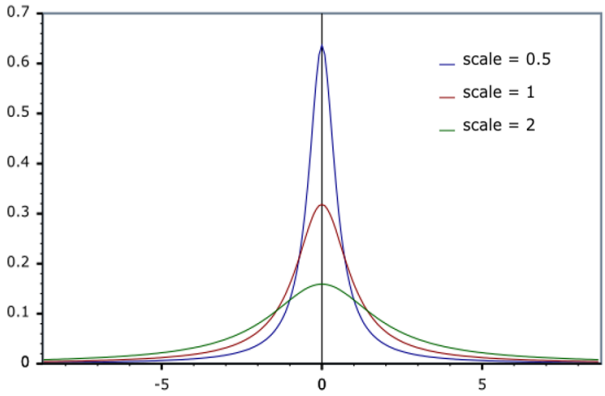
Bayes factor robustness checks

Bayes factors are known to be sensitive to how the prior distribution is specified. For the analysis to be "robust", Bayes factors should be relatively consistent over a range of different prior specifications. The robustness analysis for the one-sided correlation analysis is shown below:

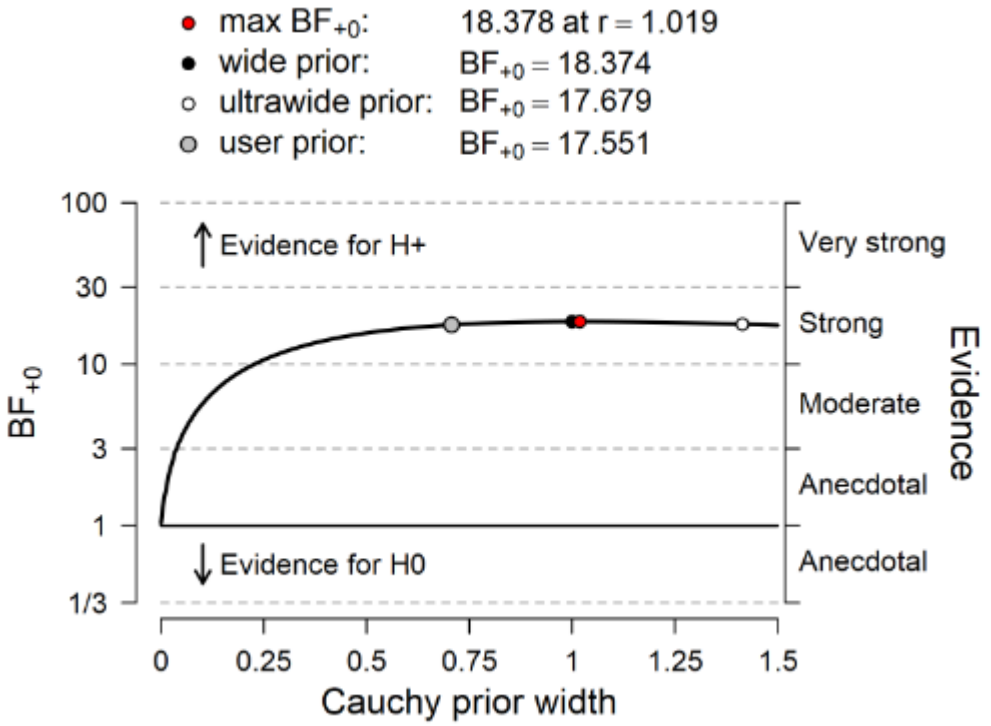


For the "positively correlated" alternative hypothesis (BF_{+0}), the robustness analysis computes BF_{+0} values for all prior shape parameters between 0 and 2. This shows to what extent the Bayes factor fluctuates based on the prior specification. Except for very small prior widths (i.e., very extreme/informative prior specifications), there is very little change in BF+0 which consistently supports "extreme" evidence for the alternative hypothesis over the null.

In terms of the Cauchy distribution, if the location is maintained as being centred on 0, changing the prior width (scale) changes the shape of the distribution. An example of this is shown below. Note that the default Cauchy prior is set to 0.707. This scale parameter for the Cauchy distribution works as follows: 50% of the probability mass is situated between -(scale) and +(scale). For instance, a Cauchy distribution with scale = 1.5 will have 50% of its probability mass between -1.5 and 1.5.

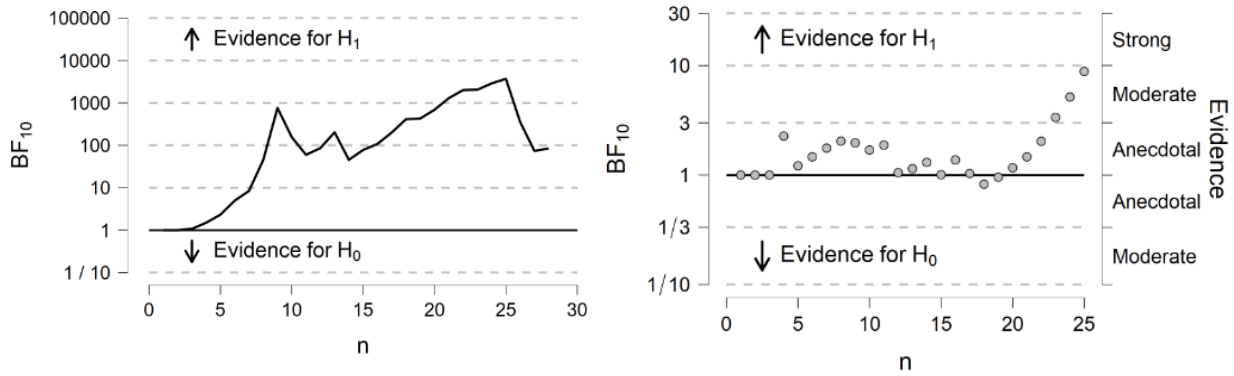


The robustness test for the one-sided independent t-test is shown below. As can be seen, the Bayes factors are calculated over a range of prior width values from 0 to 1.5. The analysis also provides BF_{+0} values over a selection of 4 prior widths (max: maximum attainable Bayes factor, user: user-specified prior, wide: width of 1, and ultrawide: 1.4). As with the correlation example, except for very small prior widths, the BF_{+0} values consistently show strong evidence in support of the alternative hypothesis.



Sequential Analyses

The sequential analyses for the correlation and independent t-tests are shown below (left and right respectively). This shows the sequential development of the evidence as the data accumulate.



Sequential analysis is generally only of interest in monitoring the sampling plan in the original research design. For example, to either stop collecting data after a set number of trials or when a pre-defined Bayes factor is achieved.

BAYESIAN ONE SAMPLE T-TEST

Research is normally carried out in sample populations, but how close does the sample reflect the whole population? The Bayesian one-sample t-test determines whether a sample mean is the same or different from a known or hypothesized population mean.

The 2-sided null hypothesis (H_o) tested is that the effect size (δ) = 0 while the alternative hypothesis is that the effect size ≠ 0

ASSUMPTIONS

Three assumptions are required for a one-sample t-test to provide a valid result:

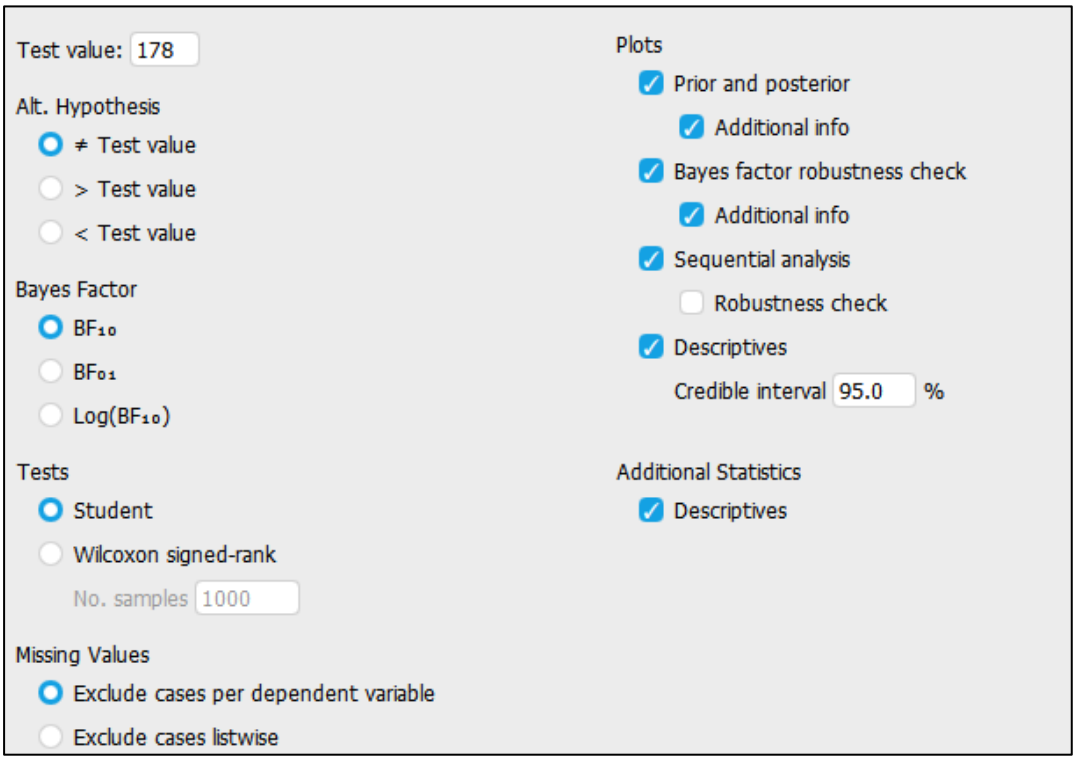
- The test variable should be measured on a **continuous** scale.
- The test variable data should be **independent** i.e. no relationship between any of the data points.
- The data should be approximately **normally distributed**
- There should be no **outliers** in the differences between the 2 groups.

The last two assumptions should be checked by doing a descriptives analysis.

RUNNING THE BAYESIAN ONE SAMPLE T-TEST

Open **One sample t-test.csv**, this contains two columns of data representing the height (cm) and body masses (kg) of a sample population of males used in a study. In 2017 the average adult male in the UK population was **178 cm** tall and has a body mass of **83.6 kg**.

Go to T-Tests > Bayesian One-Sample t-test and in the first instance add height to the analysis box on the right. Then tick the following options and add **178** as the test value:



UNDERSTANDING THE OUTPUT

The output should contain two tables and four graphs.

The results show that the BF_{10} favouring the alternative hypothesis, that data is not equal to the test value, is less than one. Switch the Bayes factor from BF_{10} to BF_{01} which will report in favour of the null hypothesis.

Bayesian One Sample T-Test

	BF_{10}	error %
height	0.234	0.034

Note. For all tests, the alternative hypothesis specifies that the population mean is different from 178.

Bayesian One Sample T-Test

	BF_{01}	error %
height	4.279	0.034

Note. For all tests, the alternative hypothesis specifies that the population mean is different from 178.

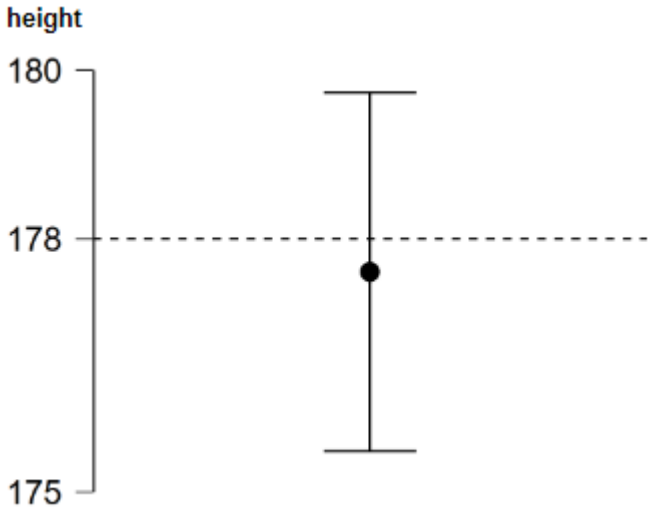
As BF_{01} = 4.28, this indicates the null model is 4.28 more favoured than the alternative model, given the data. Not only does this provide moderate evidence for H_0 relative to H_1 — something not possible with p-values — but it also describes the magnitude of this evidence.

If the data is not normally distributed, JASP provides the option to run the Wilcoxon signed-rank test instead of the default Student test.

Descriptives

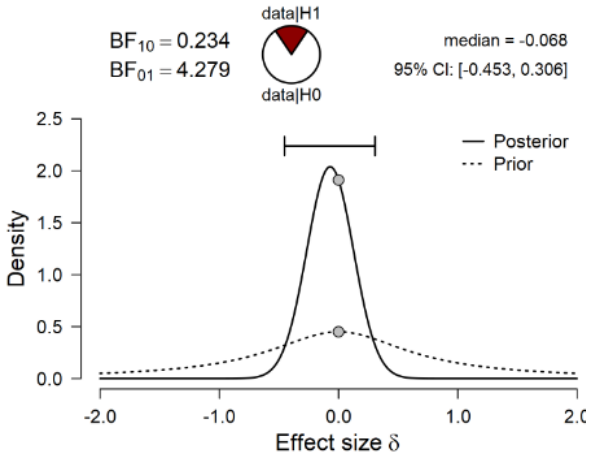
	N	Mean	SD	SE	95% Credible Interval Lower	95% Credible Interval Upper
height	23	177.609	4.915	1.025	175.483	179.734

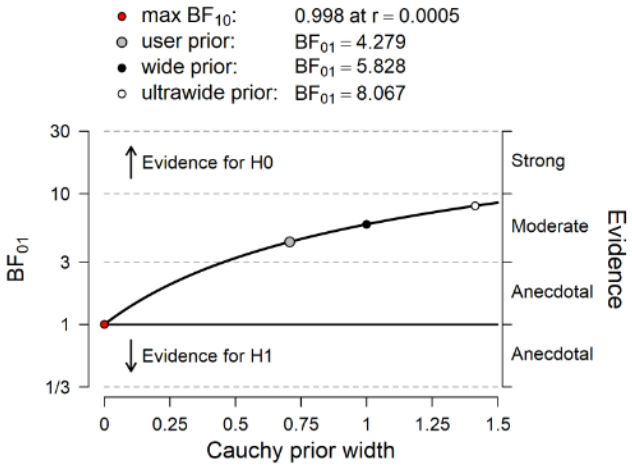
The descriptive data shows that the mean height of the sample population was 177.6 cm compared to the average 178 cm UK male. This is shown graphically with the mean ± 95% credible intervals below.



The posterior – prior plot shows that the posterior distribution is centred very close to an effect size of 0 (median = -0.006) with the pizza plot favouring the data under the null rather than the alternative hypothesis.

The robustness test also shows evidence in favour of the null hypothesis with a range of prior widths.





Repeat the procedure by replacing height with mass and change the test value to 83.6 and test for the alternative hypothesis ≠ test value.

Bayesian One Sample T-Test

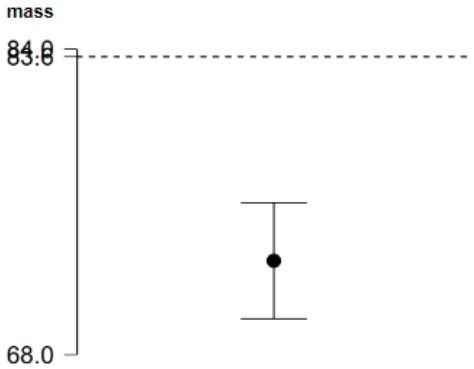
	BF_{10}	error %
mass	61312.288	7.279e −8

Note. For all tests, the alternative hypothesis specifies that the population mean is different from 83.6.

Descriptives

	N	Mean	SD	SE	95% Credible Interval Lower	95% Credible Interval Upper
mass	23	72.913	7.025	1.465	69.875	75.951

The Bayes factor is reported as 61312, i.e. the data is 61312 times more likely under the alternative hypothesis than the null. The mean weight of the participants (72.9 kg) is less than the test value defined (83.5 kg).



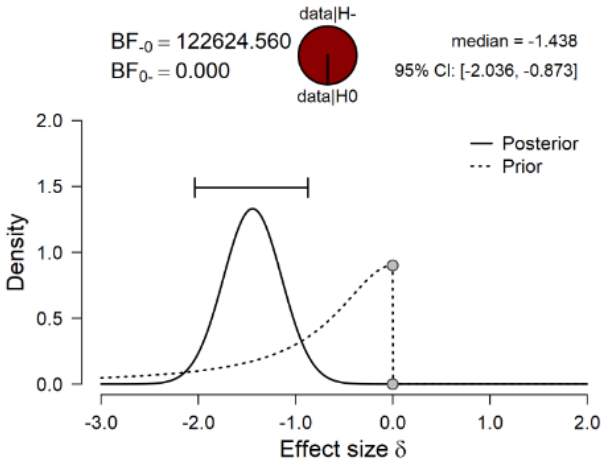
Re-run the test changing the alternative hypothesis < test value.

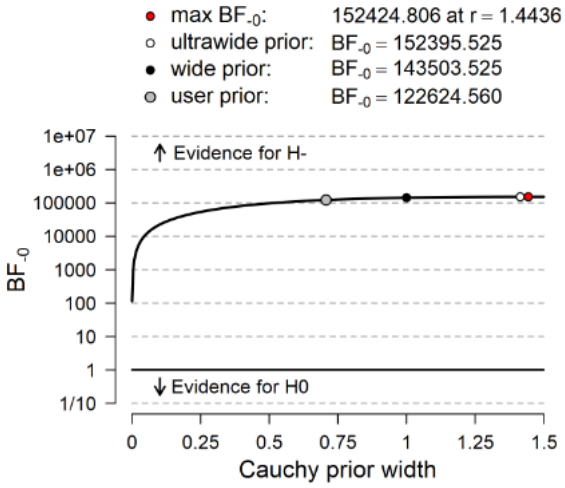
Bayesian One Sample T-Test

	BF_{-0}	error %
mass	122624.560	NaN

Note. For all tests, the alternative hypothesis specifies that the mean is less than 83.6.

It can now be seen that the Bayes factor extreme evidence in favour of the alternative hypothesis with the data being 122,625 times more likely than under the null hypothesis.





REPORTING THE RESULTS

A 2-sided Bayesian one-sample t-test comparing the sample population height (177.6 cm) to the UK adult norm (178 cm) returns a BF_{01} of 4.3 indicating moderate evidence in favour of the null hypothesis. This means that the data is 4.3 times more likely to have occurred under the null than under the alternative hypothesis.

A one-sided Bayesian one-sample t-test, where H_1 is less than the test value, comparing the sample population mass (72.9 kg) to the UK adult norm (83.6 kg) returns a BF_{10} of 122,625 indicating decisive evidence in favour of the alternative hypothesis. This means that the data 122,625 times more likely to have occurred under the alternative than under the null hypothesis.

BAYESIAN BINOMIAL TEST

The binomial test is effectively a non-parametric version of the one-sample t-test for use with dichotomous (i.e. yes/no) categorical datasets. This tests whether the sample differs from a known or hypothesized population proportion (test value).

The null hypothesis (H_o) postulates that the population proportion is equal to the test value.

The alternative hypotheses that can be tested are:

- ≠ *Test value*: Two-sided alternative hypothesis that the population proportion is not equal to test value.
- > *Test value*: One-sided alternative hypothesis that the population proportion is larger than the test value.
- < *Test value*: One-sided alternative hypothesis that the population proportion is smaller than the test value.

If a one-sided test is requested, the BF_{10} (or BF_{01}): Bayes factor is denoted as:

- BF+0: Bayes factor that quantifies evidence for the one-sided alternative hypothesis that the population proportion is **larger** than the test value, relative to the null hypothesis.
- BF-0: Bayes factor that quantifies evidence for the one-sided alternative hypothesis that the population proportion is **smaller** than the test value, relative to the null hypothesis.
- BF0+: Bayes factor that quantifies evidence for the null hypothesis, relative to the one-sided alternative hypothesis that the population proportion is **larger** than the test value.
- BF0-: Bayes factor that quantifies evidence for the null hypothesis, relative to the one-sided alternative hypothesis that the population proportion is **smaller** than the test value.

ASSUMPTIONS

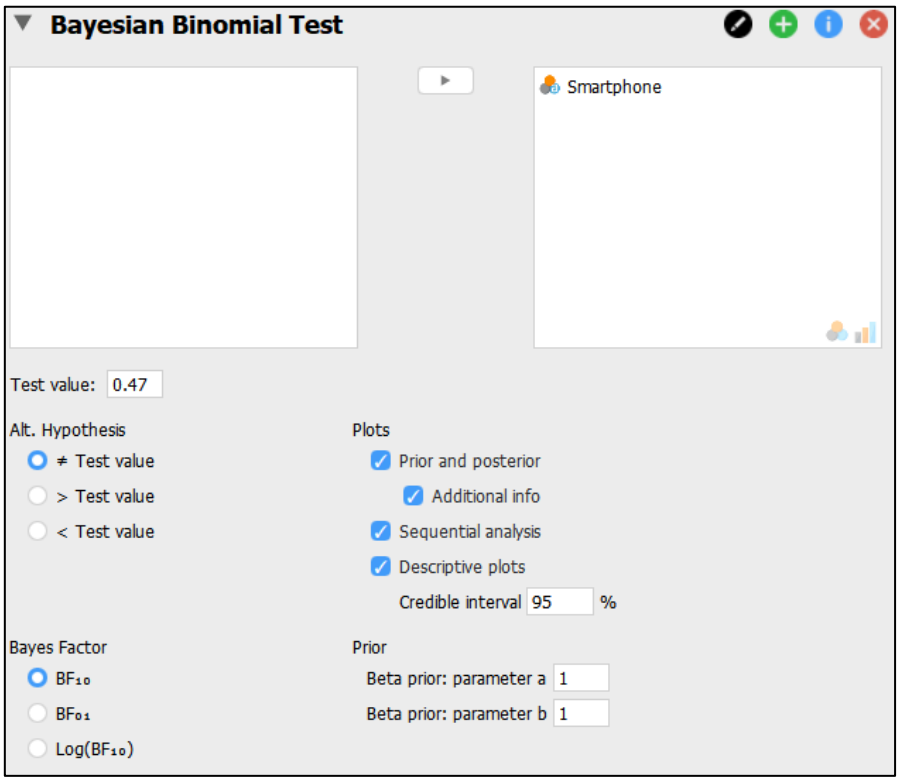
Three assumptions are required for a binomial test to provide a valid result:

- The test variable should be on a dichotomous scale (such as yes/no, male/female etc.).
- The sample responses should be independent

RUNNING THE BINOMIAL TEST

Open **Bayesian binomial.csv**, this contains one column of data showing the number of students in a first-year class using either an iPhone or another smartphone. In August 2019, when comparing smartphone ownership in the UK, the market share of the iPhones was 47%.[5]
Go to Frequencies >Bayesian Binomial test. Move the Smartphone variable to the data window and set the Test value to 0.47 (47%). Also, tick all plot options.

5 https://www.statista.com/statistics/271195/apple-ios-market-share-in-the-united-kingdom-uk/

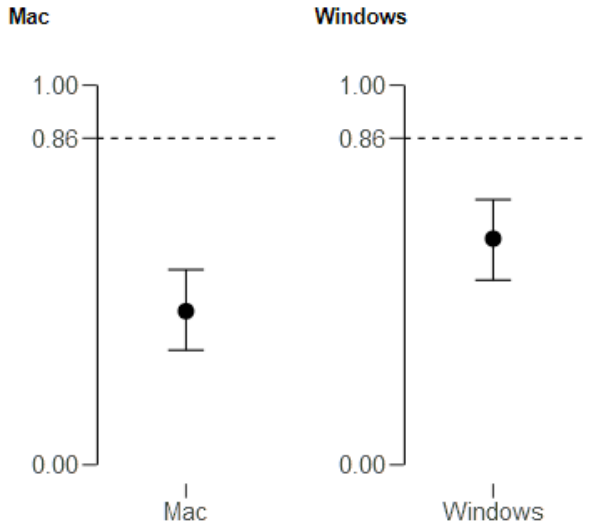


The following table and graph show that the proportions of both smartphones were 59% and 41% for iPhones and other makes, respectively, in the student cohort compared to the market proportions being 47% and 53%.

Bayesian Binomial Test

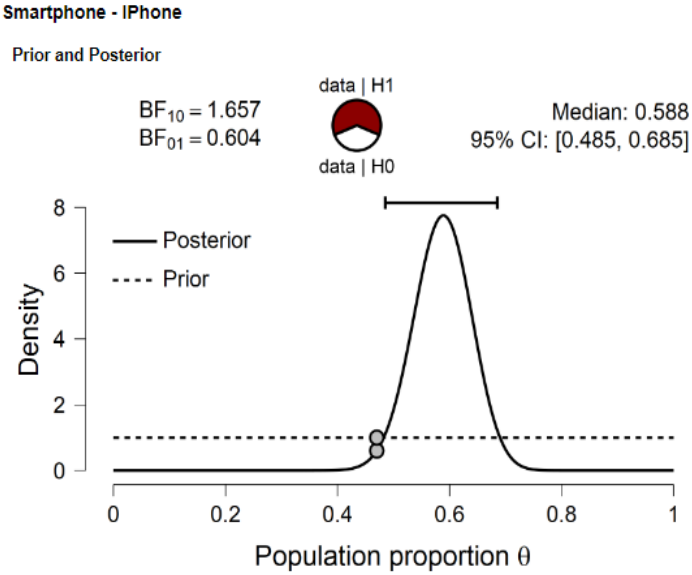
	Level	Counts	Total	Proportion	BF_{10}
Smartphone	IPhone	53	90	0.589	1.657
	Other	37	90	0.411	0.242

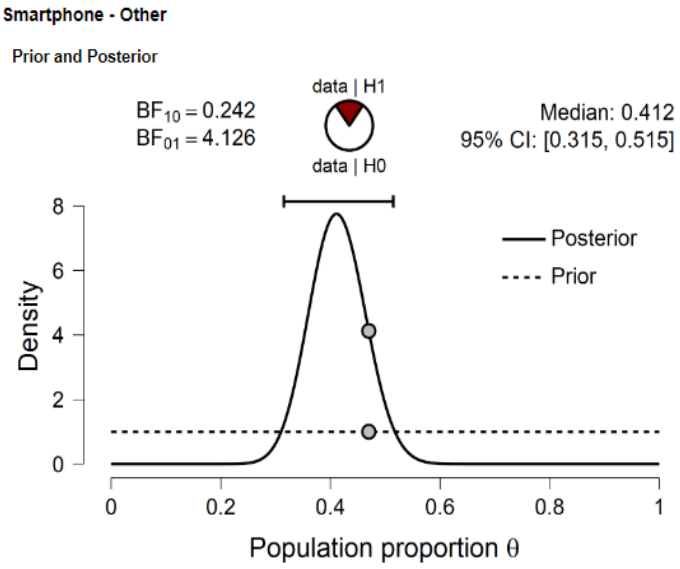
Note. Proportions tested against value: 0.47.

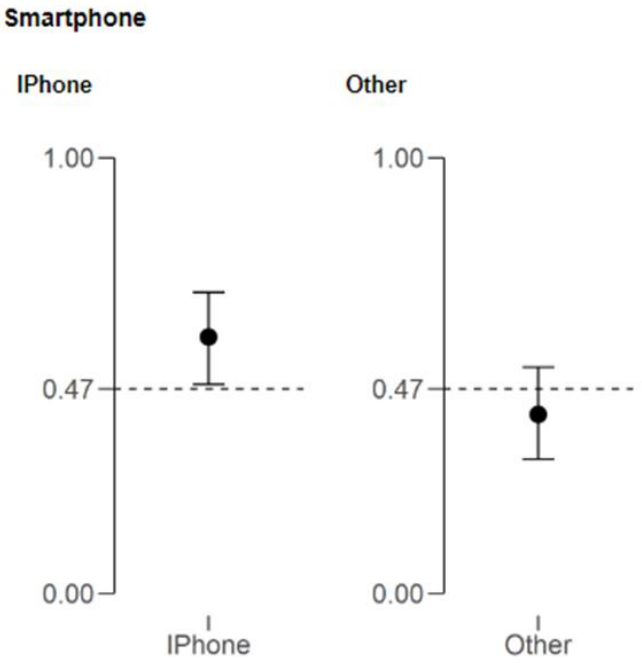


For a 2-sided test, the Bayes factors show that for the iPhone proportion there was insufficient evidence to accept or reject the null hypothesis (BF_{10} = 1.657).

This can be further visualised in the pizza plots presented with the Prior and Posterior plots. The two grey dots indicate the density values of the two distributions at the test value. The ratio of these two values is called the Savage-Dickey density ratio, which gives the BF_{10} and BF_{01} values (also shown above), depending on which way around the division is done. Visually, if the grey dot on the posterior distribution is higher than that on the prior distribution the evidence supports the null hypothesis and vice versa.







ONE-SIDED TESTING.

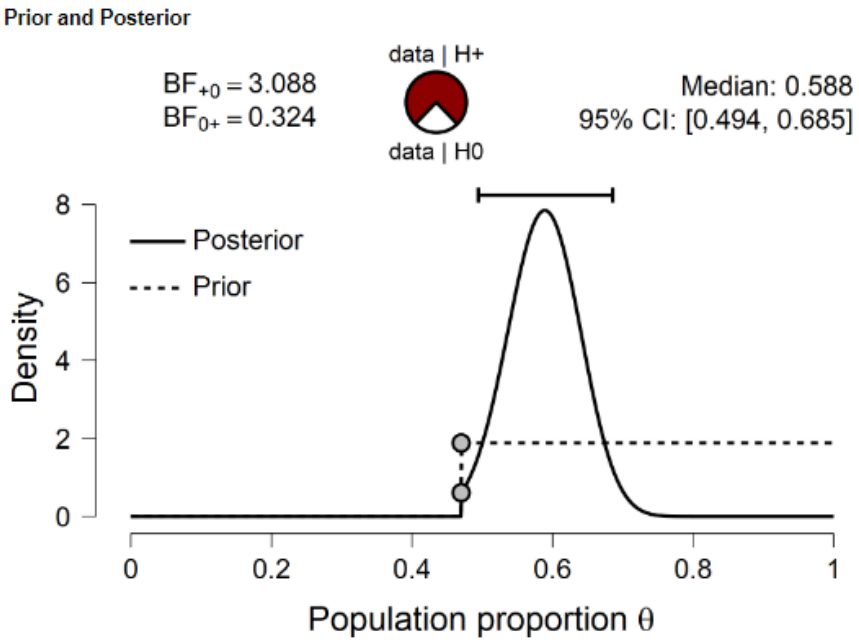
This can be followed up with one-sided hypothesis testing i.e. is the proportion of iPhones used in class (58.9%) greater than the expected UK market proportion of 47%? To do so, change the alternative hypothesis to > Test value.

Bayesian Binomial Test

	Level	Counts	Total	Proportion	BF_{+0}
Smartphone	IPhone	53	90	0.589	3.088
	Other	37	90	0.411	0.061

Note. For all tests, the alternative hypothesis specifies that the proportion is greater than 0.47.

The results show that there is only moderate evidence in support of the one-sided alternative hypothesis that the proportion of student iPhone users is higher than the UK market sales proportion.



REPORTING THE RESULTS

The UK market proportion of iPhone and other smartphone users was reported to be 47% and 53% respectively. In a cohort of University students (N=90) this proportion was found to be 58.9% and 41.1%.

Are these young students more susceptible to the glossy Apple marketing machine than the normal population? A one-sided Bayesian Binomial test based on the alternative hypothesis that the proportion of student iPhone users was higher than in the general population when the market proportion was carried out. The resulting BF was 3.09 which only provides anecdotal/moderate evidence favouring the alternative hypothesis.

BAYESIAN MULTINOMIAL TEST

The multinomial test is effectively an extended version of the Binomial test for use with categorical datasets containing three or more factors. This tests whether the sample distribution is different from a hypothesized population distribution (multinomial test) or a known distribution (Chi-square 'goodness-of-fit' test).

The null hypothesis (H_0) is that the sample counts are generated by a specified set of population proportions. The alternative hypothesis (H_1) is that the sample counts are not generated by those population proportions.

ASSUMPTIONS

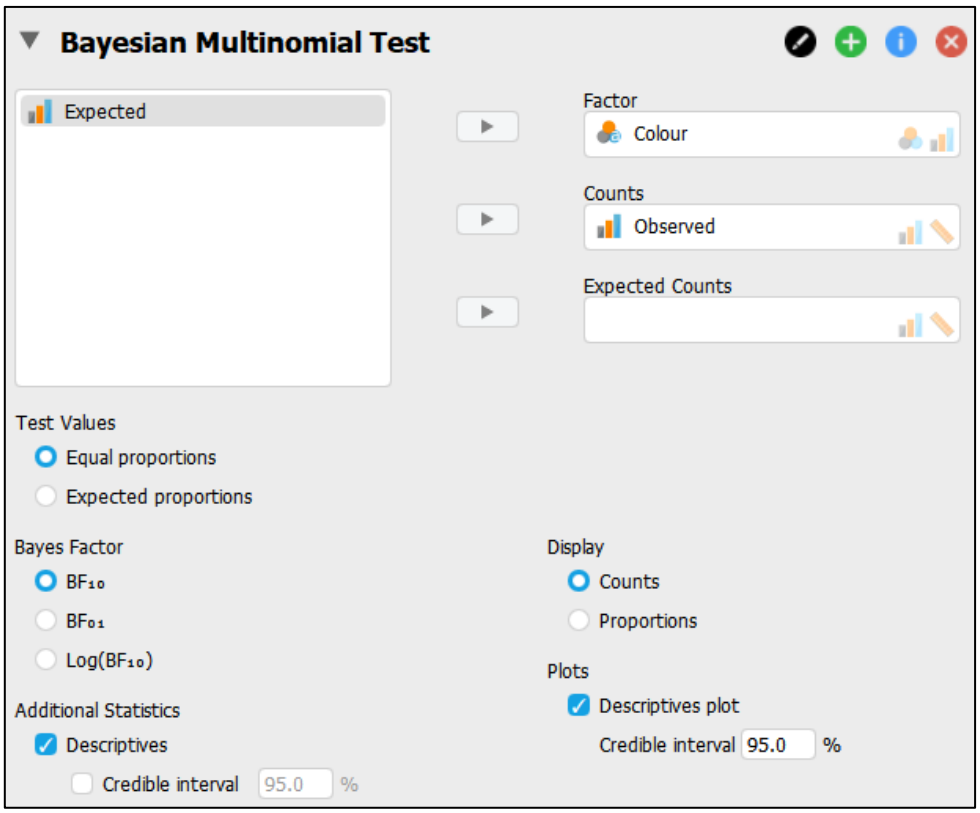
Three assumptions are required for a multinomial test to provide a valid result:

- The test variable should be a categorical scale containing 3 or more factors
- The sample responses should be independent

RUNNING THE MULTINOMIAL TEST

Open **Bayesian Multinomial.csv**. This contains three columns of data including the number of different coloured M&Ms counted in a total of five bags (observed). Without any prior knowledge, it could be assumed that the different coloured M&Ms are equally distributed. Therefore, the priors are all set to be equal i.e. 1.

Go to Frequencies > Bayesian Multinomial test. Move colour of the M&Ms to Factor and the observed number of M&Ms to counts. Tick Descriptives and Descriptives Plots.



As can be seen in the Descriptive table, the test assumes an equal expectation for the proportions of coloured M&Ms (36 of each colour). The Multinomial test results show a BF_{10} of 2512 suggesting that the data are 2512 times more likely under the alternative hypothesis than the null hypothesis.

Bayesian Multinomial Test

	Levels	BF_{10}
Multinomial	6	2512.077

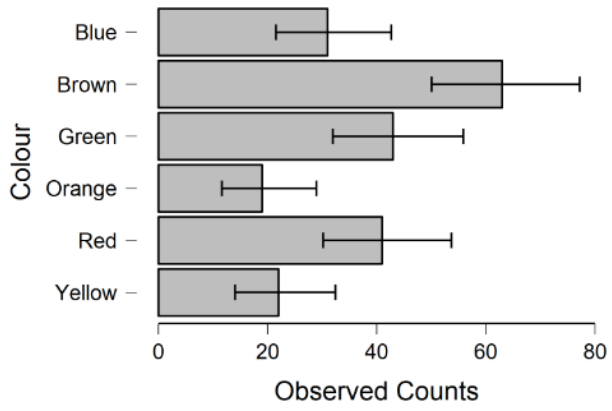
Use the Display options to switch between either counts or proportions

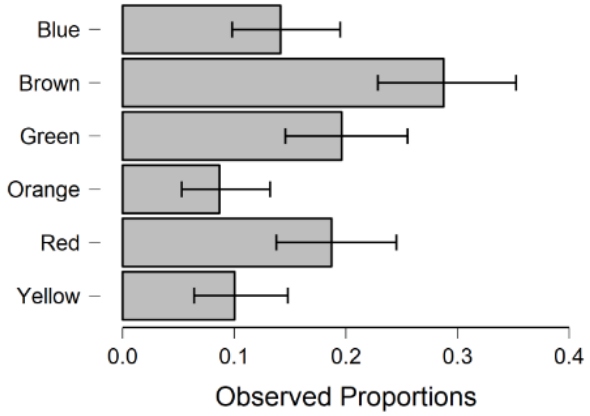
Descriptives

Colour	Observed	Expected: Multinomial
Blue	31	36
Brown	63	36
Green	43	36
Orange	19	36
Red	41	36
Yellow	22	36

Descriptives

Colour	Observed	Expected: Multinomial
Blue	0.142	0.167
Brown	0.288	0.167
Green	0.196	0.167
Orange	0.087	0.167
Red	0.187	0.167
Yellow	0.100	0.167





In 2008, Mars, the manufacturers of M&Ms changed the colour distribution to the following.

Colour	Blue	Brown	Green	Orange	Red	Yellow
Proportion	24	13	16	20	13	14

Sometime later, the proportions were removed from the manufactures web site and have not been restored since. These last published values will now be used as the expected counts, so move the Expected variable to the Expected Counts box. As can be seen in the Descriptives table, JASP has calculated the expected numbers of the different coloured M&Ms based on the manufacturers reported production ratio. The results of the test result in a BF_{10} of $4.3 * 10^{10}$ and provide decisive evidence in favour of the alternative hypothesis where the observed counts of M&Ms are not generated by the last proportions stated by the manufacturer.

Bayesian Multinomial Test

	Levels	BF_{10}
Expected	6	4.332e +10

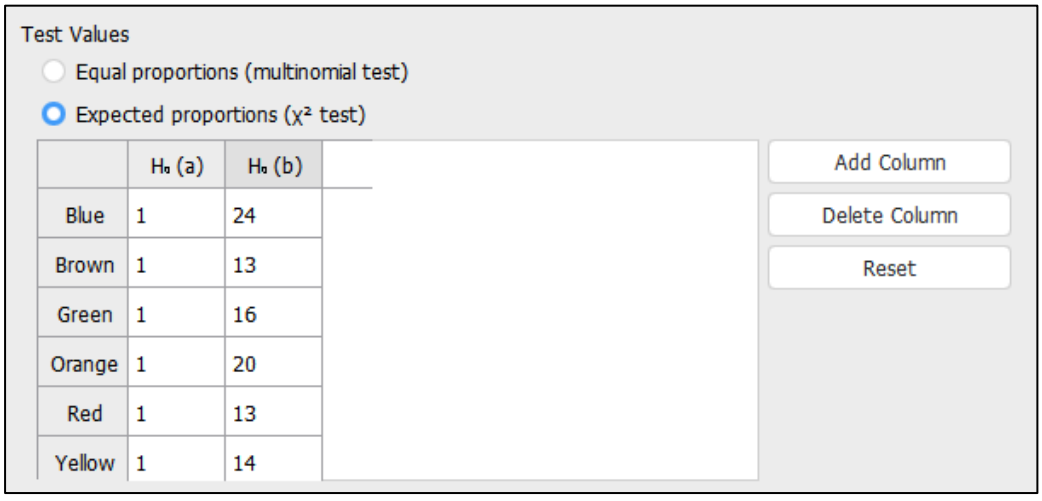
Descriptives

Colour	Observed	Expected: Expected
Blue	0.142	0.240
Brown	0.288	0.130
Green	0.196	0.160
Orange	0.087	0.200
Red	0.187	0.130
Yellow	0.100	0.140

MULTIPLE HYPOTHESES

JASP also provides another option whereby different hypotheses can be run at the same time. Go back to the Options window and only add Colour to the Factor and Observed to the Counts boxes, remove the expected counts if the variable is still there. In Test values, tick Expected proportions. This will open a small spreadsheet window showing the colour and H_o (a) with each cell have 1 in it. This is assuming that the proportions of each colour are equal (multinomial test).

In this window, add another column which will automatically be labelled H_o (b). The expected proportions of each colour can now be typed in.



Now when the analysis is run, the results of the tests for the two hypotheses are shown. H_o (a) is the null hypothesis that the population counts are equal, while H_o (b) is the null hypothesis that the population counts are the same as those specified by the manufacturer. As can be seen, the Bayes factors reject both null hypotheses decisively.

Bayesian Multinomial Test

	Levels	BF_{10}
H_o (a)	6	2512.077
H_o (b)	6	4.332e +10

Descriptives

		Expected	
Colour	Observed	H_o (a)	H_o (b)
Blue	0.142	0.167	0.240
Brown	0.288	0.167	0.130
Green	0.196	0.167	0.160
Orange	0.087	0.167	0.200
Red	0.187	0.167	0.130
Yellow	0.100	0.167	0.140

BAYESIAN INDEPENDENT SAMPLES T-TEST

Like its frequentist counterpart, the Bayesian Independent t-test test is used to determine if there is a difference between two independent groups. The test requires a continuous dependent variable (i.e. weight loss following a 4-week diet) and an independent variable comprising 2 groups (i.e. males and females). The two hypotheses tested are:

H_0**: males and females have similar weight loss (i.e. effect sizes for each group are equal (δ = 0))**

H_1**: males and females have different weight loss (i.e. effect sizes for each group are not equal (δ ≠ 0))**

ASSUMPTIONS

Group independence:

Both groups must be independent of each other. Each participant will provide one data point for one group only. For example, participant 1 can only be in either a male or female group – not both.

Normality of the dependent variable:

The dependent variable should also be measured on a continuous scale and be approximately normally distributed with no outliers. This can be checked visually using the Q-Q plots.

If normality is violated you can try transforming the data (for example log values, square root values) or, and if the group sizes are very different, use the Mann-Whitney U test which is a non-parametric equivalent that does not require the assumption of normality (see the end of this chapter).

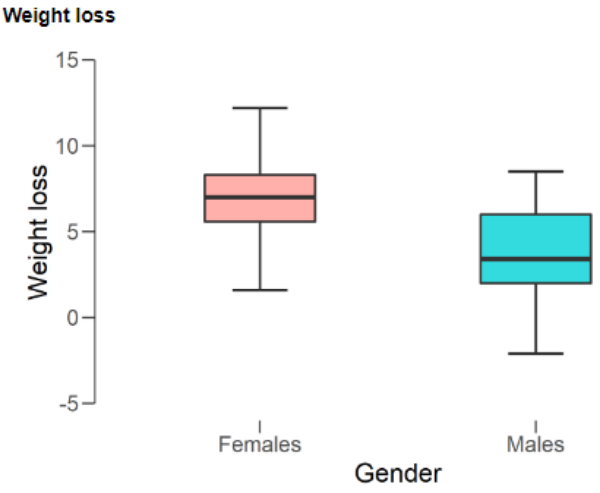
Homogeneity of variance:

The variances of the dependent variable should be equal in each group. This can be tested using Levene's Test of Equality of Variances.

Open **Bayesian Independent t-test.csv** into JASP. Go to Descriptives and look at Weight loss split by gender. Check for outliers and normal distribution (Shapiro-Wilk). In this case, the data looks like the assumptions have been met.

Descriptive Statistics

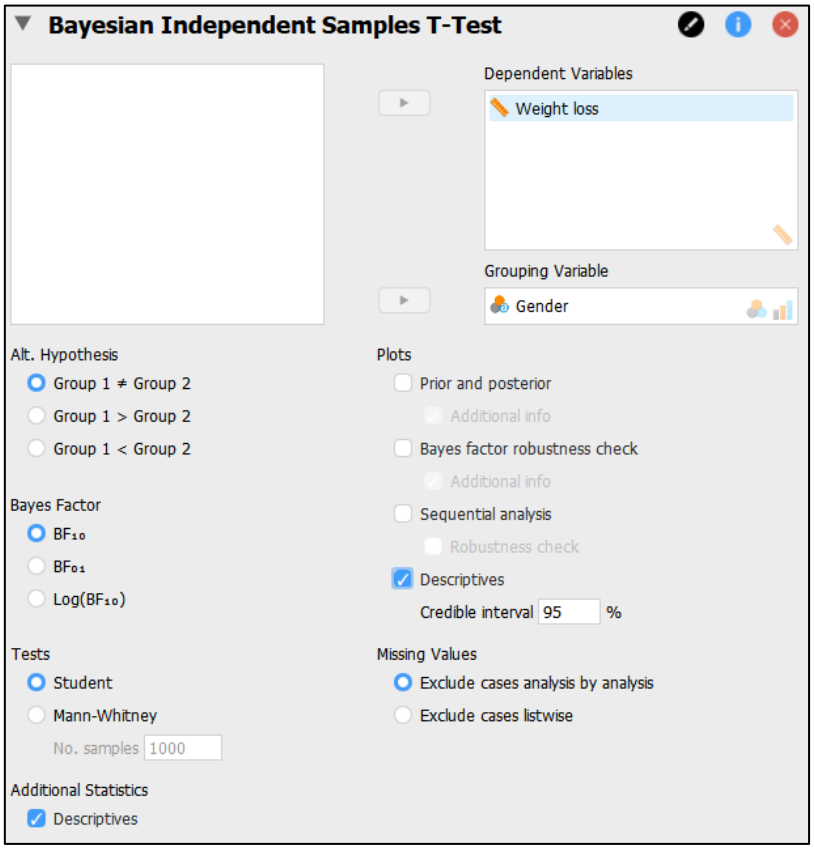
	Weight loss	
	Females	Males
Valid	42	45
Missing	0	0
Mean	6.929	3.720
Std. Deviation	2.242	2.588
Shapiro-Wilk	0.968	0.971
P-value of Shapiro-Wilk	0.282	0.310
Minimum	1.600	−2.100
Maximum	12.200	8.500



RUNNING THE ANALYSIS

Go to T-Tests > Bayesian Independent Samples t-test. Move the weight loss variable into the dependent variable and Gender into the Grouping variable on the right. In the first instance, tick

- ✓ the hypothesis to be the alternative hypothesis (Group 1 ≠ Group 2)
- ✓ BF_{10}
- ✓ Descriptives
- ✓ Descriptive plots



UNDERSTANDING THE OUTPUT

The tables below show the Bayes factors for both BF_{10} and Log BF_{10}.

Bayesian Independent Samples T-Test

	BF_{10}	error %
Weight loss	442346.504	6.421e −9

Bayesian Independent Samples T-Test

	$Log(BF_{10})$	error %
Weight loss	13.000	6.421e −9

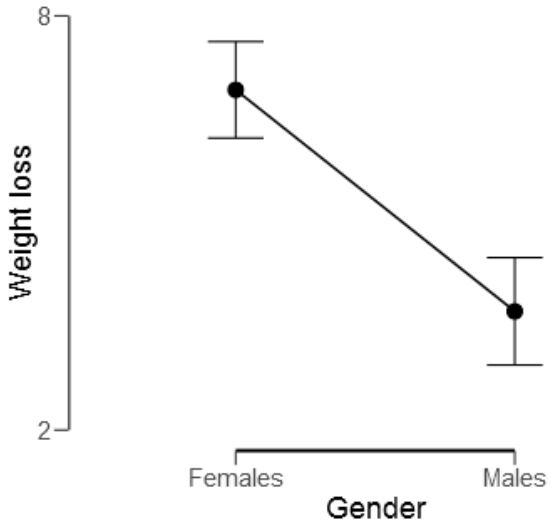
The BF_{10} value is the ratio of the $\frac{\text{Likelihood of data given the alternative hypothesis ()}}{\text{Likelihood of data given the null hypothesis ()}}$

In this case, the data are 442346 times more likely under the alternative hypothesis (H_1) than the null hypothesis. The error % is based on the accuracy of the Bayes factor calculations, if this is less than 10% this can be ignored. Using Jeffrey's criterion, this is decisive evidence in favour of the alternative hypothesis although the Bayes factor is non-directional (unlike the t statistic) in that it does not show how they differ.

This can be seen in the Descriptives table where weight loss is higher in females compared to males.

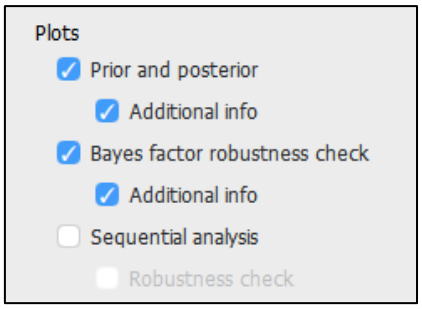
Group Descriptives

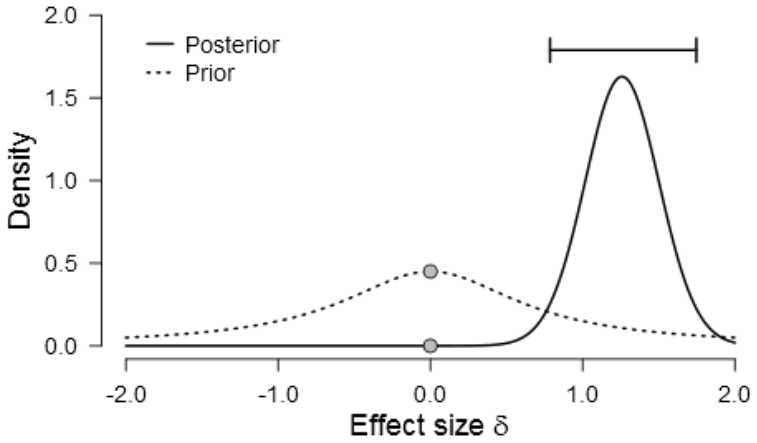
	Group	N	Mean	SD	SE	95% Credible Interval	
						Lower	Upper
Weight loss	Females	42	6.929	2.242	0.346	6.230	7.627
	Males	45	3.720	2.588	0.386	2.942	4.498



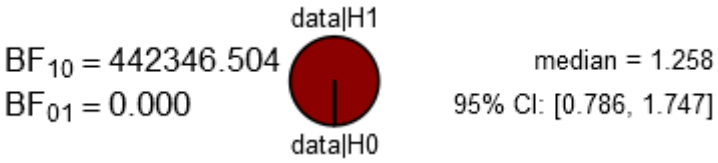
FURTHER CHECKS

Go back to the statistical options and tick all the Prior and Posterior, as well as the Bayes factor robustness check options:

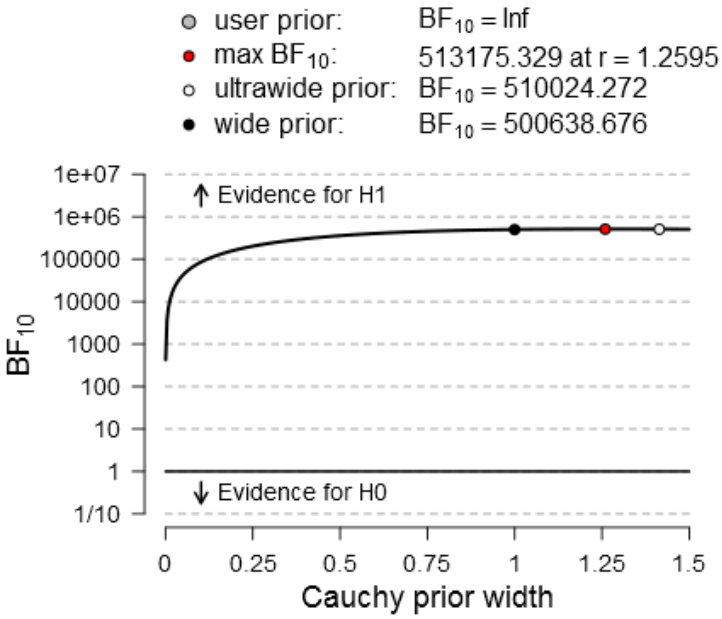




The dashed line shows the prior distribution and the solid line the posterior distribution (based on the dataset). The posterior distribution is shifted to the right over large positive effect sizes. Each of the distributions has a grey dot at the 0.0 effect size. If the dot on the prior distribution is higher than the one on the posterior distribution, then the Bayes factor supports the alternative hypothesis.



Bayes factors supporting the alternative / null hypothesis (BF_{10}) and null/alternative hypothesis (BF_{01}). The pizza plot distribution shows the proportion of evidence for the H_1 (red) and H_0 (white) hypothesis. In this example, the pizza plot is completely red. The median effect size of 1.258 and 95% credible intervals are also shown.



The width (uncertainty) of the prior distribution is set as 0.707 by default in JASP. This graph shows a range of prior widths, which in this case are relatively consistent and do not greatly change the BF_{10} value with all values being over 100. Therefore, it can be concluded that this test is robust to changes in the prior width.

REFINING THE HYPOTHESIS TESTING

The descriptive data show that females exhibit greater weight loss than males, the analysis can, therefore, be re-run but now selecting the alternative hypothesis Group 1 (females) > group 2 (males).

Group Descriptives

	Group	N	Mean	SD
Weight loss	Females	42	6.929	2.242
	Males	45	3.720	2.588

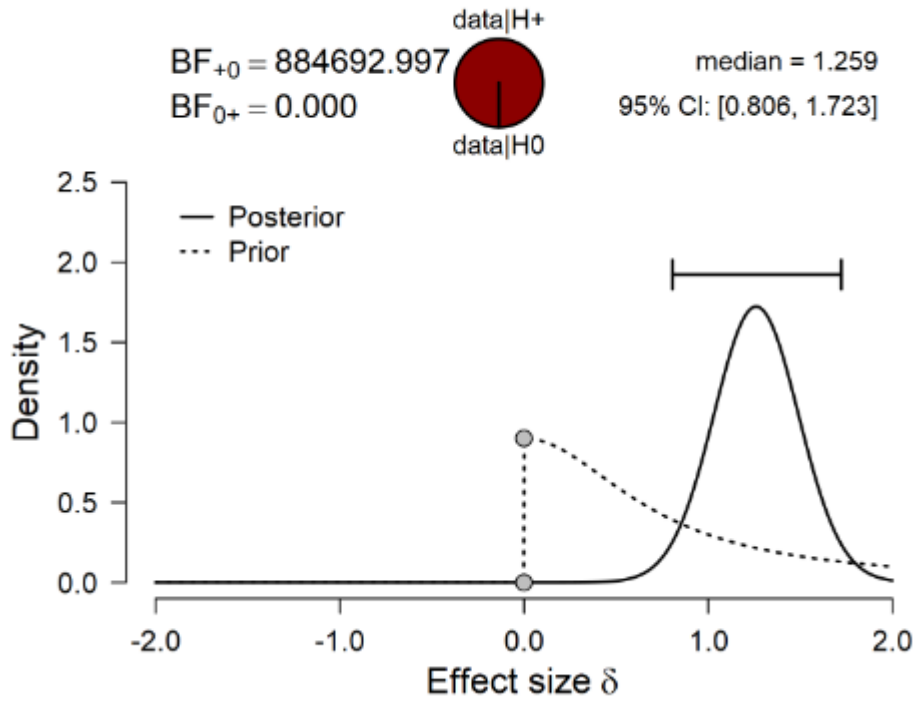
Bayesian Independent Samples T-Test

	BF_{+0}	error %
Weight loss	884692.997	NaN

Note. For all tests, the alternative hypothesis specifies that group *Females* is greater than group *Males*.

This shows that the evidence for this one-sided alternative hypothesis (BF+0) is now 884,693 times more likely than under the null hypothesis. The error% is reported a NaN since the error is incredibly small.

Prior and Posterior



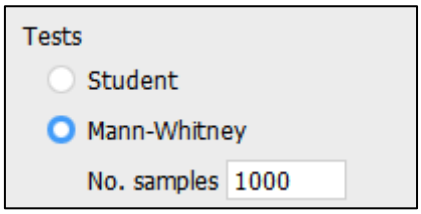
The prior and posterior graph now shows the one-sided prior with all its weight on the positive effect size side.

REPORTING THE RESULTS

Following a 4-week diet, females lost on average 6.93 kg compared to males who lost 3.72 kg. A two-sided analysis revealed a Bayes factor (BF_{10}) that the data were 442346 times more likely under the alternative than the null hypothesis. A subsequent one-sided test based on the alternative positive directional hypothesis that females lost more weight than males (BF+0) resulted in a Bayes factor indicating that the data were 884,692 times more likely under this directional alternative hypothesis than the null with a median effect size of 1.26.

BAYESIAN MANN-WHITNEY TEST

JASP has an option to run a Mann-Whitney test for nominal or non-normally distributed data as an alternative to the Student T-test.



The first thing to notice is that the analysis takes longer than when running the Student T-test. Secondly, if the analysis is repeated on the same data, although the W statistic is the same, the BF is usually quite different. Below is the output for 3 analyses of the same data:

Bayesian Mann-Whitney U Test

	BF_{10}	W	R^
Pain score	30.582	207.000	1.003

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

	BF_{10}	W	R^
Pain score	46.327	207.000	1.001

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

	BF_{10}	W	R^
Pain score	33.614	207.000	1.000

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

The following explanation has been paraphrased from the JASP forum

"The underlying algorithm introduces some degree of variation …… it runs multiple chains and bases the Bayes factor off that.

This variation is especially prevalent when there is either a low sample size or a low number of MCMC-samples. For now, maybe it helps to increase that number to the maximum."

I have found a more stable repeated BF by increasing the number of samples/ iterations from 1000 to 10,000.

THE BAYESIAN PAIRED SAMPLES T-TEST

This test, like, the classical parametric paired-samples t-test compares the means between two related groups on the same continuous, dependent variable. For example, looking at weight loss pre and post 2 weeks jogging programme. The two-sided version of this test compares two hypotheses for effect size δ:

H_0**: the null hypothesis - that the effect size is absent (i.e., δ = 0)**

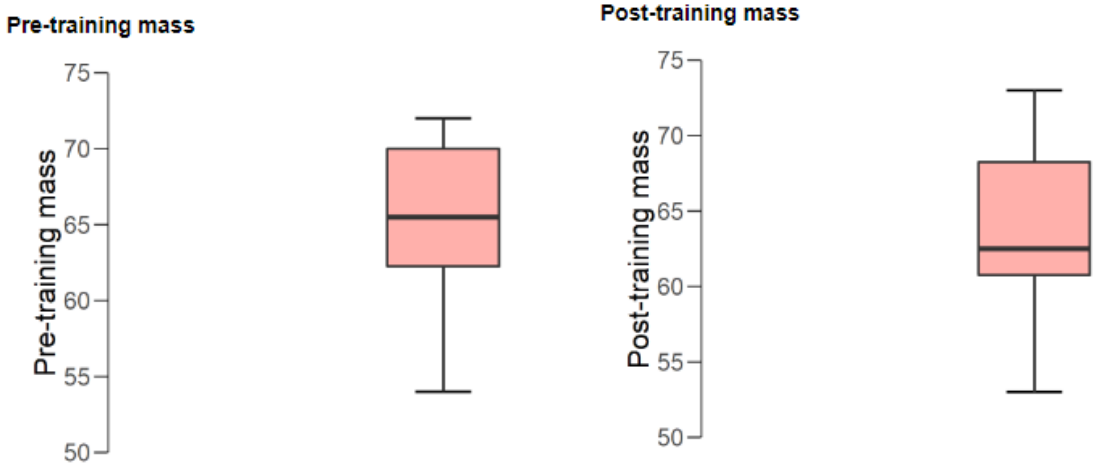
H_1**: the alternative hypothesis - that the effect size δ ≠ 0**

ASSUMPTIONS

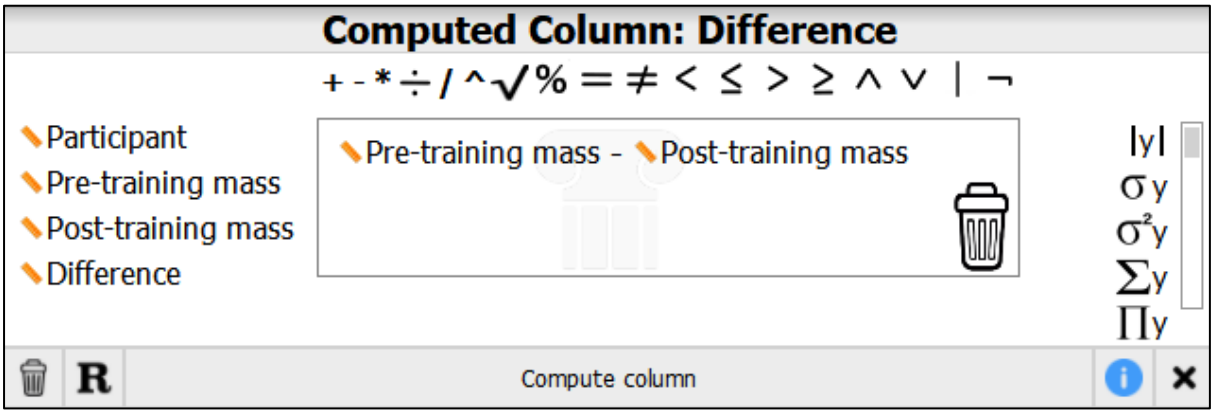
Four assumptions are required for a paired t-test to provide a valid result:

1. The dependent variable should be measured on a continuous scale.
2. The independent variable should consist of 2 categorical related/matched groups, i.e. each participant is matched in both groups
3. The differences between the matched pairs should be approximately normally distributed
4. There should be no outliers in the differences between the 2 groups.

Open **Bayesian paired t-test.csv** into JASP. As a matter of good practice check the data using the Descriptives analysis. As can be seen, there are no outliers so assumption 4 is fine.



To check the normality of the paired differences, go to the spreadsheet view and click on the black cross in the column header row to add a computed column. Name the new column "difference" and make sure that it is a Scale variable. In the dialogue box drag pre-training mass to the main box, click on the minus sign and drag over the post-training mass then click Compute column.



Now in Descriptives, the difference column can be used to check for data normality using Shapiro-Wilk. In this case, assumption 3 has not been violated since Shapiro-Wilk is not significant.

Descriptive Statistics

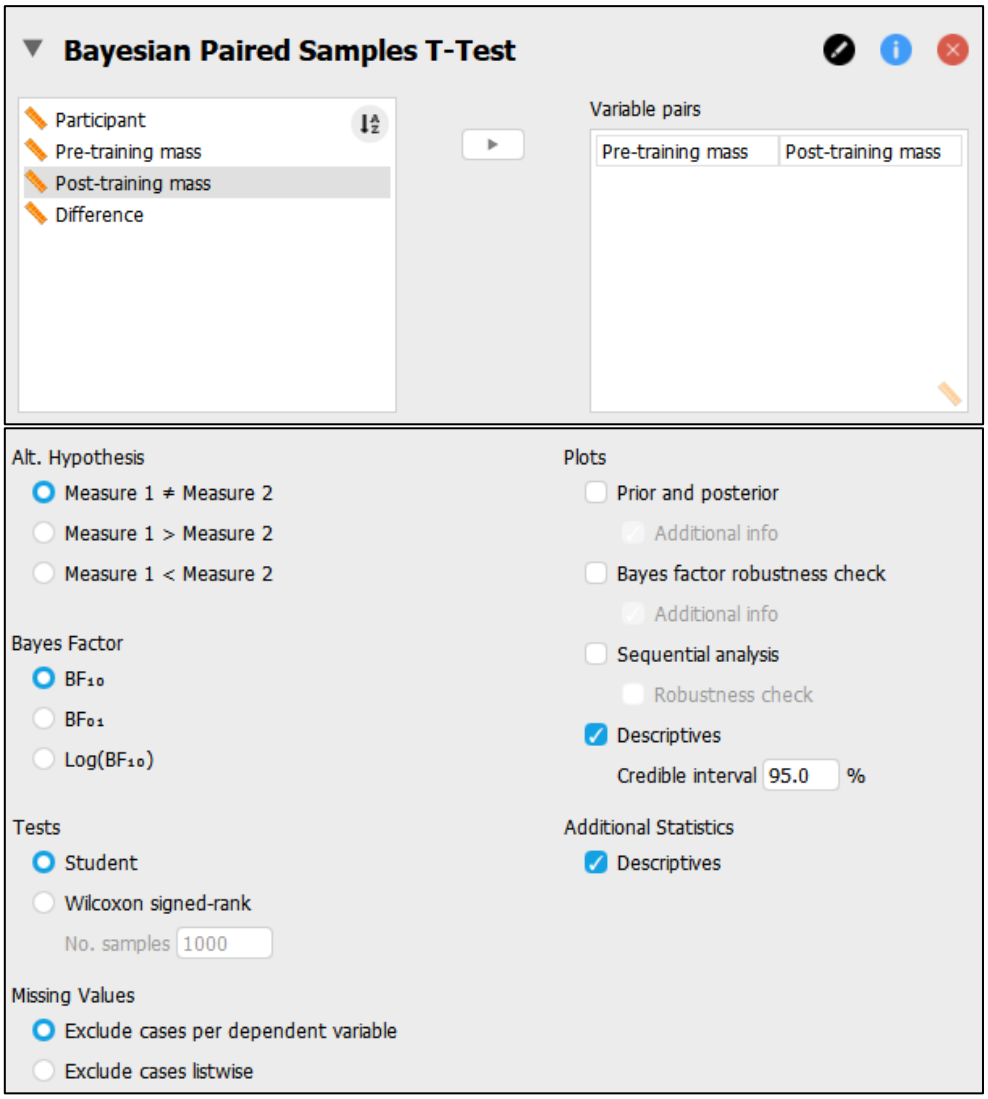
	Difference
Valid	16
Missing	0
Mean	2.13
Std. Deviation	2.73
Shapiro-Wilk	0.97
P-value of Shapiro-Wilk	0.78
Minimum	−3.00
Maximum	8.00

NOTE: To date, a non-parametric version (i.e. Wilcoxon's test) of a Bayesian paired samples t-test has not been implemented in JASP 0.10.2 but will be added soon.

RUNNING THE ANALYSIS

Go to T-Tests > Bayesian Paired Samples t-test. Move the paired variables into the analysis box on the right. In the first instance, tick

- ✓ the alternative hypothesis to be Measure 1 ≠ Measure 2
- ✓ BF_{10}
- ✓ Descriptives
- ✓ Plots - Descriptive



UNDERSTANDING THE OUTPUT

The output should consist of two tables and one graph.

Bayesian Paired Samples T-Test

			BF_{10}	error %
Pre-training mass	-	Post-training mass	7.26	1.96e −4

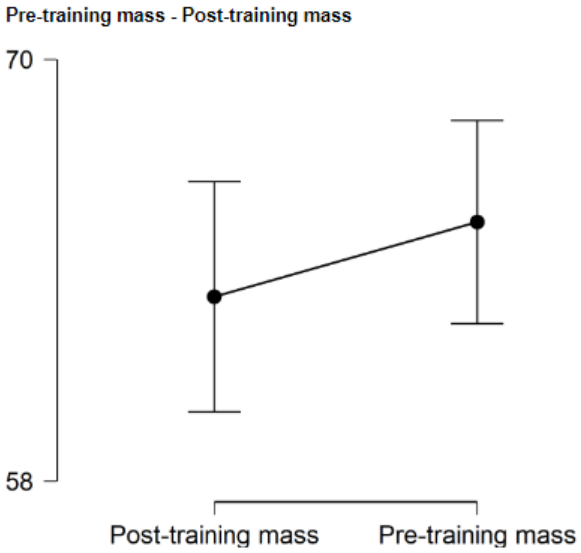
The BF_{10} value is the ratio of the $\frac{\text{Likelihood of data given the alternative hypothesis ()}}{\text{Likelihood of data given the null hypothesis ()}}$

in this case, the alternative hypothesis (H_1) is 7.26 times more likely than the null hypothesis. Using Jeffrey's criterion, this is moderate evidence in favour of the alternative hypothesis.

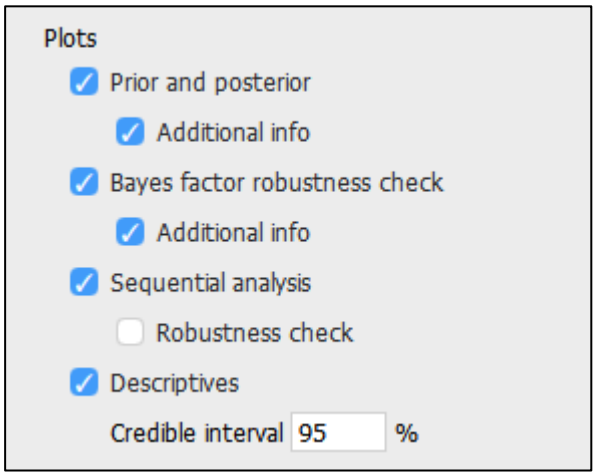
The error % is based on the accuracy of the Bayes factor calculations, if this is less than 10% this can be ignored. Although there is moderate evidence for a difference between the two groups the Bayes factor does not show in which direction they differ.
This can be seen in the Descriptives table where body mass is lower 2 weeks post-training with a mean difference of 2.13 kg. The descriptives plot shows the mean values and their 'credible intervals'.

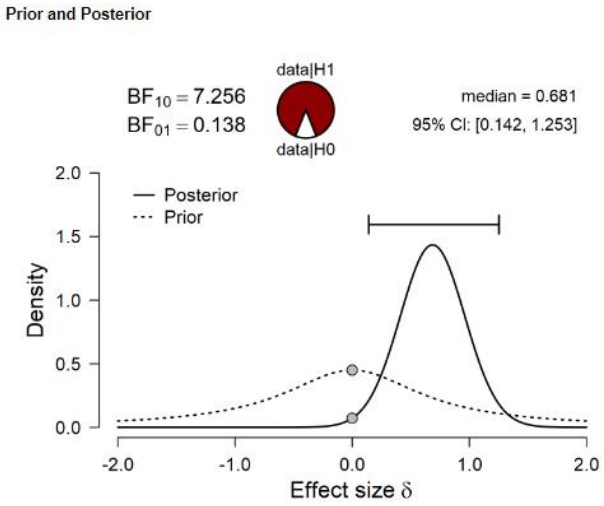
Descriptives

					95% Credible Interval	
	N	Mean	SD	SE	Lower	Upper
Pre-training mass	16	65.38	5.43	1.36	62.48	68.27
Post-training mass	16	63.25	6.15	1.54	59.97	66.53



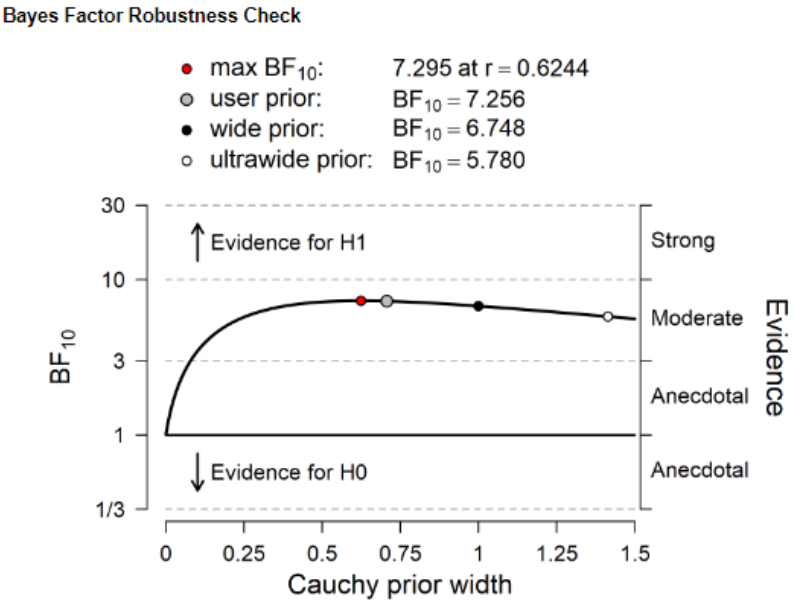
FURTHER CHECKS
Go back to the statistical options and tick all the other available Plots options which will result in 3 more graphs:



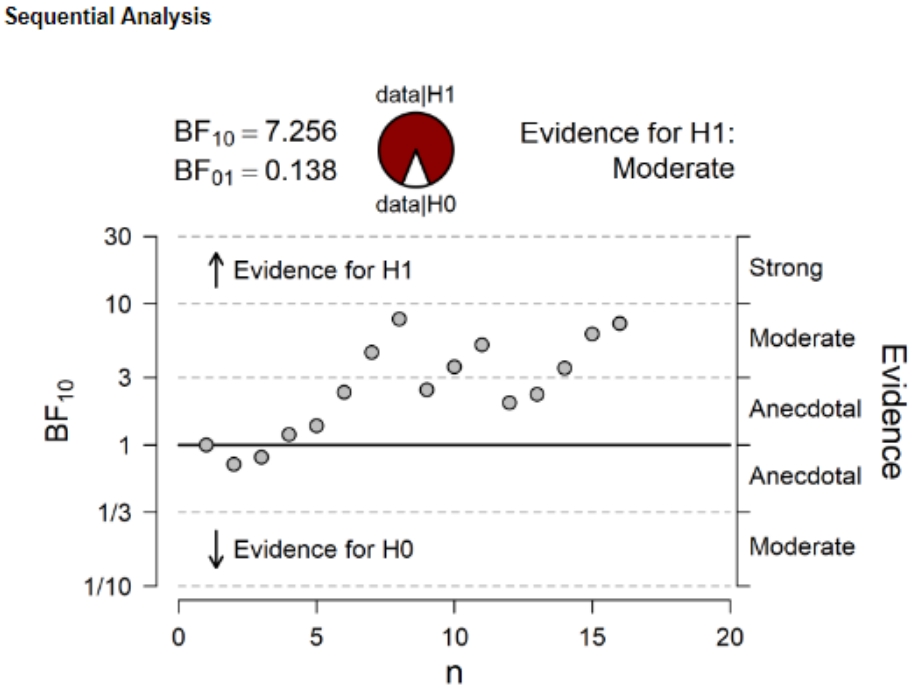


The dashed line shows the prior distribution and the solid line the posterior distribution (based on the dataset). The posterior distribution is shifted to the right over positive effect sizes. Each of the distributions has a grey dot at the 0.0 effect size. If the dot on the prior distribution is higher than the one on the posterior distribution, then the Bayes factor supports the alternative hypothesis.
Bayes factors supporting the alternative/null hypothesis (BF_{10}) and null/alternative hypothesis (BF_{01}). The pizza plot distribution shows the proportion of evidence for the H_1 (red) and H_0 (white) hypothesis. In this example, the pizza plot is predominantly red. The median effect size and 95% credible intervals are also shown.

Robustness relates to the strength of the model and is used when the data are collected from a wide range of probability distributions that are largely unaffected by outliers or small violations of model assumptions.
The width (uncertainty) of the prior distribution is set as 0.707 by default in JASP. This graph shows a range of prior widths, which in this case are relatively consistent and do not greatly change the BF_{10} value. Therefore, it can be concluded that this test is robust to changes in the prior width.



The sequential plot shows how the Bayes factor changes after every data point are added with the BF_{10} fluctuating between anecdotal and moderate evidence in support of the alternative hypothesis.



REFINING THE HYPOTHESIS TESTING

The descriptive data show that the group had lower body mass after 2 weeks jogging exercise, the analysis can, therefore, be re-run but now selecting the directional alternative hypothesis Measure 1 (pre-training) > Measure 2 (post-training).

Descriptives

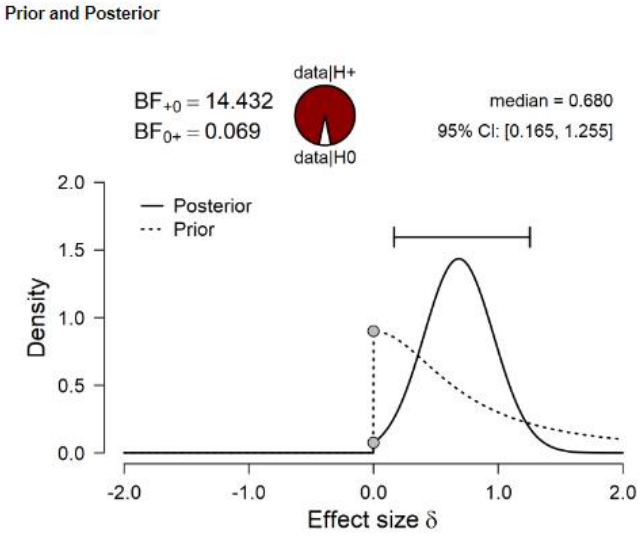
	N	Mean	SD
Pre-training mass	16	65.38	5.43
Post-training mass	16	63.25	6.15

Bayesian Paired Samples T-Test

			BF_{+0}	error %
Pre-training mass	-	Post-training mass	14.43	~ 3.08e −4

Based on the interpretation of the posterior probability having seen the data, the one-sided alternative hypothesis (BF+0) is now 14.43 times more likely than under the null hypothesis.

The prior and posterior graph now shows the one-sided prior with all its weight on the positive effect size side.



REPORTING THE RESULTS

Following 2 weeks of jogging training, the study group lost on average 2.13 kg (pre-training: 65.4 ± 5,4 kg, Post Training: 63.25 ± 6.16 kg). A two-sided analysis revealed a Bayes factor (BF_{10}) suggesting that the data were 7.2 times more likely under the alternative than the null hypothesis. A subsequent one-sided test based on the alternative positive directional hypothesis that body mass post-training was less than pre-training (BF+0) resulted in a Bayes factor indicating that the data were 14.43 times more likely under this directional alternative hypothesis than the null with a median effect size of 0.63.

BAYESIAN CORRELATION

Correlation is a statistical technique that can be used to determine if, and how strongly, pairs of variables are associated. Correlation is only appropriate for quantifiable data in which numbers are meaningful such as continuous or ordinal data. As in frequentist correlation both parametric (Pearson's) and non-parametric (Kendall's tau) correlation coefficients are reported. However, p values and confidence intervals are replaced by Bayes factors (BF) and credible intervals.

The test assesses whether the data are more likely to occur under the null hypothesis (H_0)

i.e. that there is no linear association between the two variables), or under the alternative hypothesis (H_1 **i.e. there is an association between the two variables). Then, after observing the data, Bayes' theorem is applied to obtain the posterior probability of both hypotheses.**

ASSUMPTIONS

Four assumptions are required for a correlation to provide a valid result:

1. The two variables should be measured on a continuous scale.
2. There is a linear relationship between the two variables
3. The data should be approximately normally distributed (can use Kendall's tau-b option if this assumption not met)
4. There should be no outliers in the 2 variables.

RUNNING THE BAYESIAN CORRELATION

Open **Bayesian correlation.csv** in JASP. This contains real data comprising a series of variables that can be measured during a golf drive:

Variables:

- Shot score (best value = 100, lowest = 0)
- Ball speed (m/s)
- Launch angle (degrees)
- Backspin (rpm)
- Distance (m)

Run a descriptive analysis to check for data normality and the presence of any outliers. In this case, none of the variables shows a deviation from normality (see Q-Q plots)

Descriptive Statistics

	Shot Score	Ball Speed	Launch Angle	Back Spin	Distance
Valid	28	28	28	28	28
Missing	0	0	0	0	0
Mean	87.1	66.4	15.2	2744.5	244.5
Std. Deviation	4.5	1.2	2.2	632.2	8.0
Minimum	77.0	63.5	11.1	1809.0	226.5
Maximum	93.2	68.4	20.6	4378.0	255.8

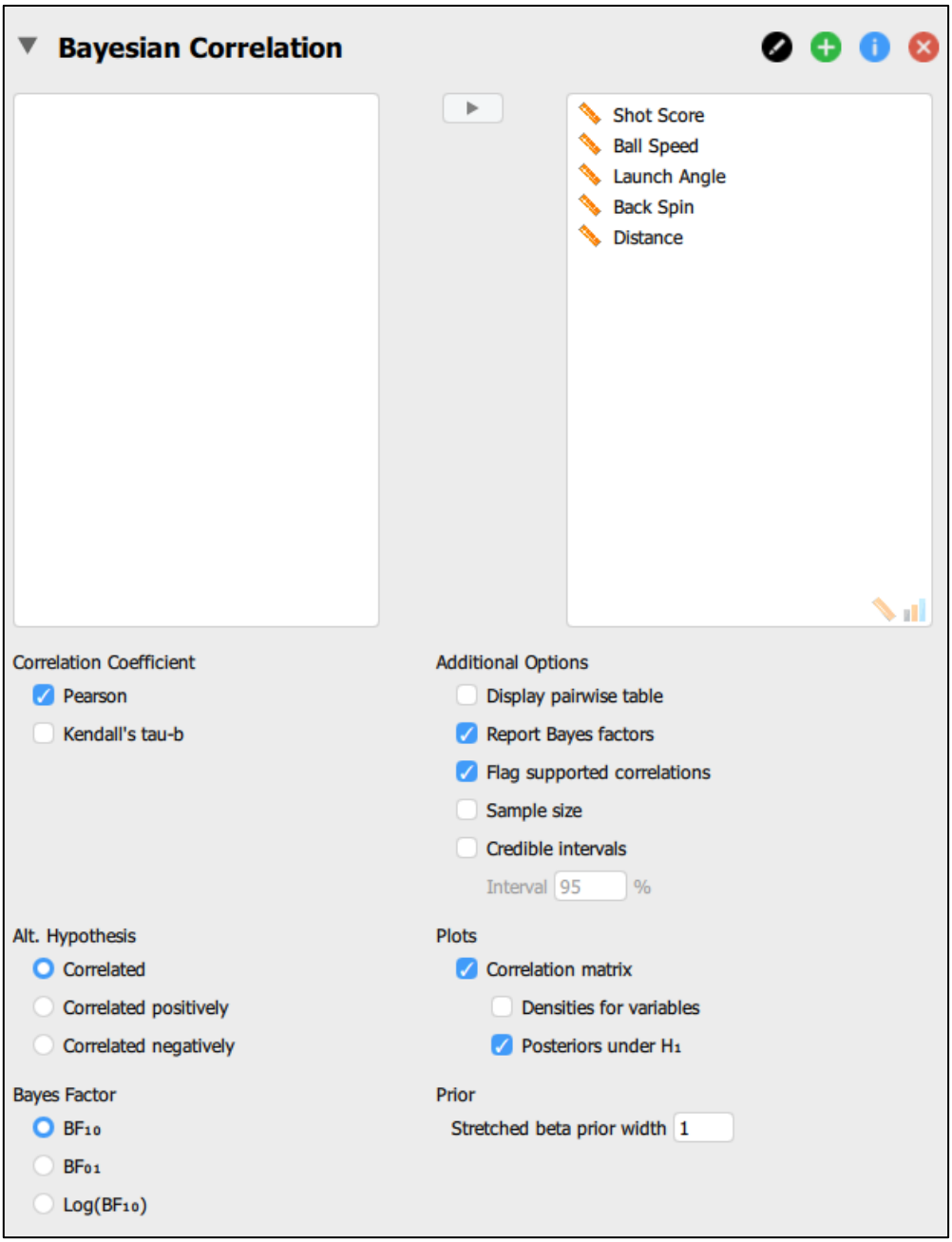
Go to Regression > Bayesian correlation. Move all variables into the analysis box on the right.

In the statistics options, tick

- ✓ Pearson's rho (or Kendall's tau if data is not normally distributed)
- ✓ Alternative hypothesis = correlated
- ✓ BF_{10}
- ✓ (will present the Bayes factor in favour of the alternative hypothesis)
- ✓ Report Bayes factors
- ✓ Flag supported correlations

Plots – Correlation matrix and posteriors under H_1

A stretched beta prior width of 1 is set by default i.e. all correlations between -1 and +1 are given an equal prior probability.



UNDERSTANDING THE OUTPUT

The Bayesian Pearson correlation matrix is shown below as is the normal frequentist Pearson correlation matrix for comparison.

Pearson Correlations

		Shot Score	Ball Speed	Launch Angle	Back Spin	Distance
Shot Score	Pearson's r	—				
	p-value	—				
Ball Speed	Pearson's r	0.42*	—			
	p-value	0.03	—			
Launch Angle	Pearson's r	0.40*	−0.11	—		
	p-value	0.04	0.57	—		
Back Spin	Pearson's r	−0.14	−0.36	0.01	—	
	p-value	0.49	0.06	0.95	—	
Distance	Pearson's r	0.67***	0.62***	0.50**	−0.58**	—
	p-value	< .001	< .001	0.01	0.00	—

* p < .05, ** p < .01, *** p < .001

As can be seen, both the Bayesian and frequentist analysis report the same Pearson's r-values. Nonetheless, with p-values, it cannot be certain if non-significance is due to data insensitivity or evidence supporting a lack of relationship between these two variables.

Whereas Pearson's correlation flags significant correlations for ball speed, launch angle, backspin with the distance the BF_{10} value for distance and Launch angle is only 7.124 suggesting that there is only moderate evidence for a correlation between the two. Bayesian correlation between shot score and ball speed/launch angle report low BF values in the anecdotal evidence range whereas they are flagged as significant in the conventional correlation test. This suggests that the Bayesian approach is more conservative and only flags significance when the evidence is strong i.e. BF>10.

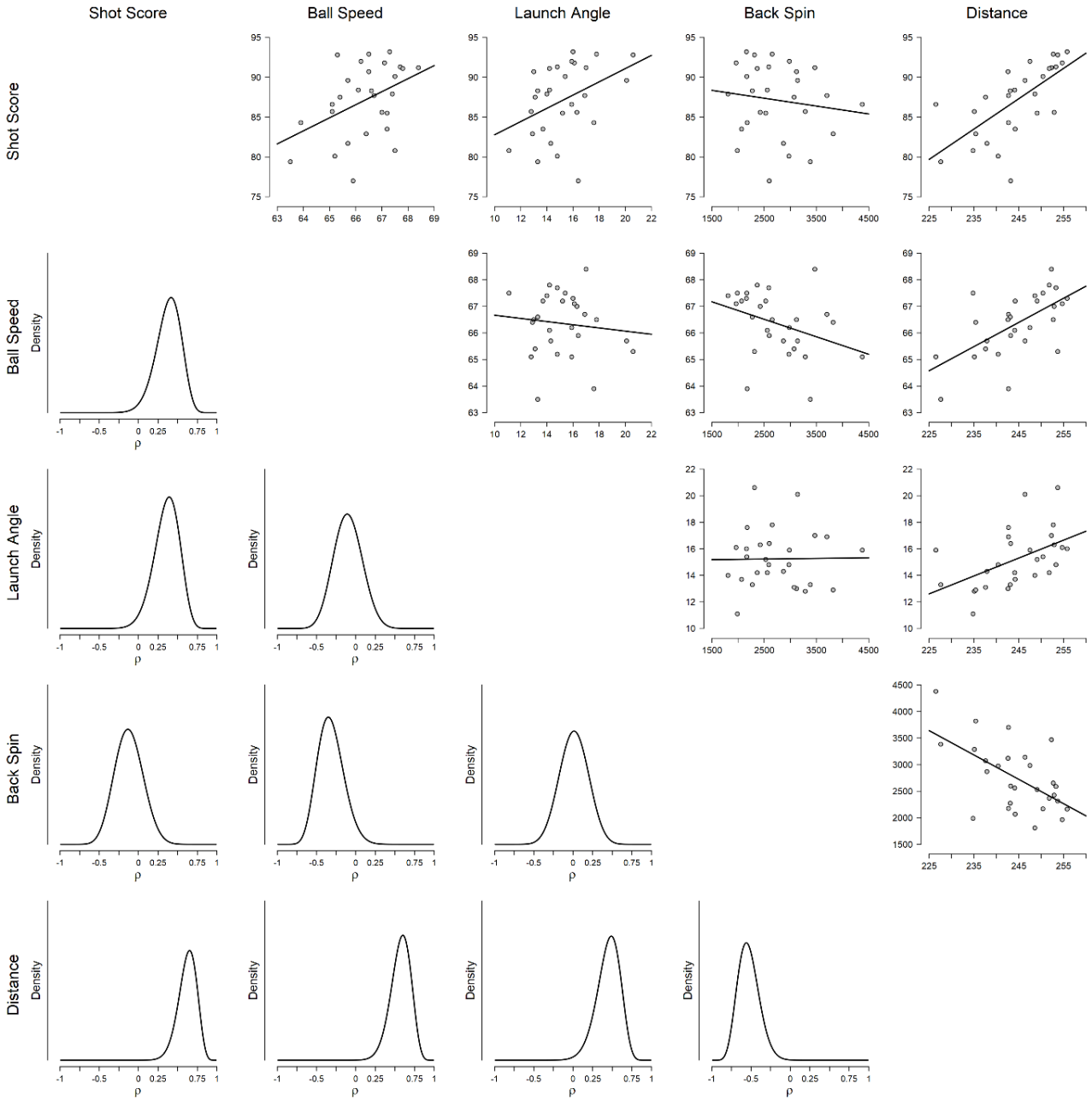
Bayesian Pearson Correlations

		Shot Score	Ball Speed	Launch Angle	Back Spin	Distance
Shot Score	Pearson's r	—				
	BF_{10}	—				
Ball Speed	Pearson's r	0.42	—			
	BF_{10}	2.56	—			
Launch Angle	Pearson's r	0.40	−0.11	—		
	BF_{10}	1.91	0.27	—		
Back Spin	Pearson's r	−0.14	−0.36	0.01	—	
	BF_{10}	0.30	1.21	0.24	—	
Distance	Pearson's r	0.67***	0.62**	0.50	−0.58**	—
	BF_{10}	326.00	84.50	7.12	33.95	—

* BF_{10} > 10, ** BF_{10} > 30, *** BF_{10} > 100

The Bayes factors report very strong evidence in favour of the alternative hypothesis (i.e. a relationship between variables) for distance with shot score (BF_{10}= 326), ball speed BF_{10} = 84.5 and Backspin (BF_{10}=33.95). So, for example, it is 326 times more likely that distance and ball speed are related than not. Based on the posterior probability the data are 326 times more likely under H_0 than under H_1. There was only moderate evidence for distance and launch angle (BF_{10} =7.12)

The correlations and posteriors under H_1 are plotted together. The posterior distributions are plotted on a horizontal scale centred on a correlation coefficient of ρ= 0. It can be seen that posteriors relating to negative correlations are weighted to the left of 0 and positive ones to the right of 0.



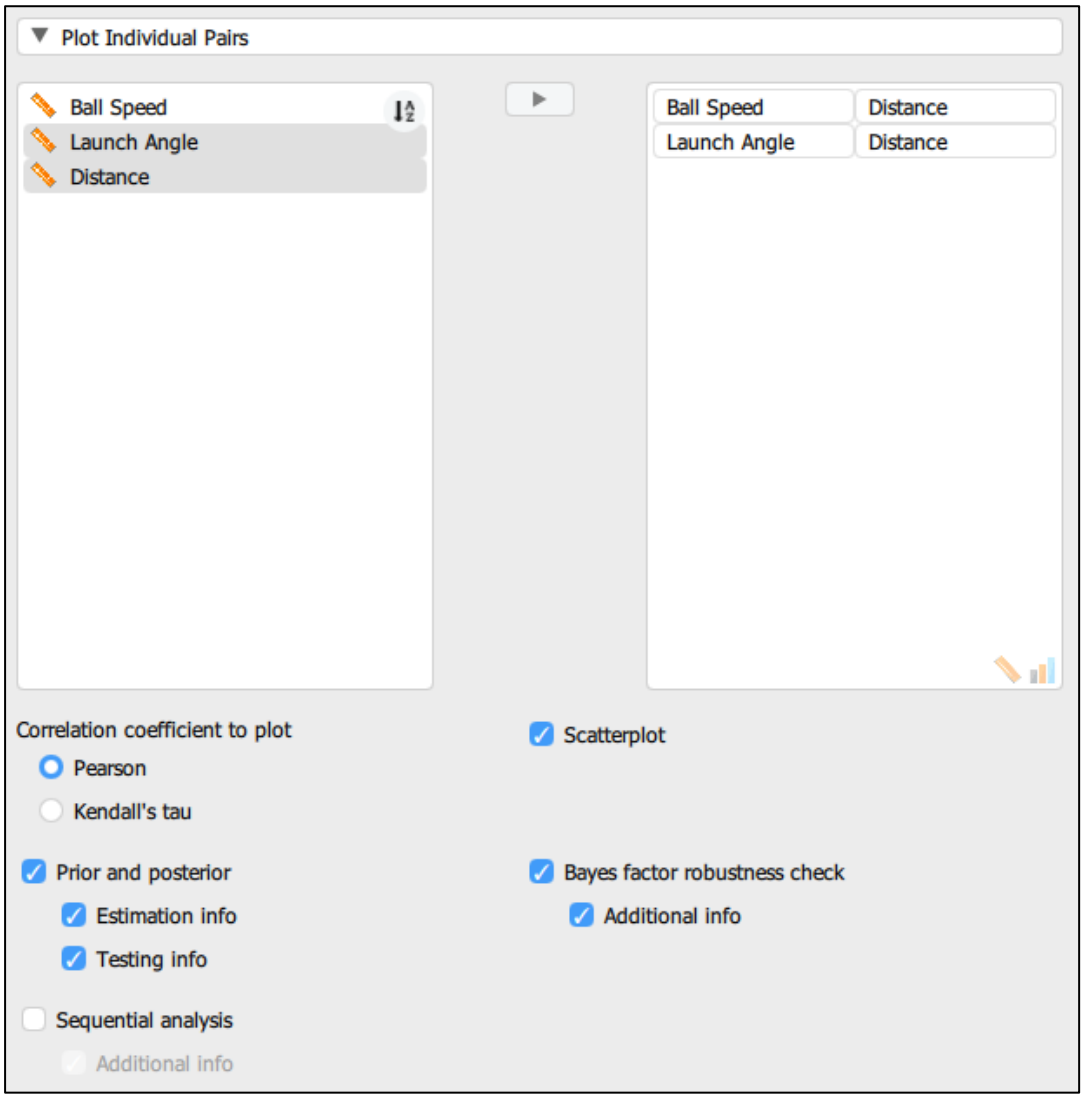
BAYESIAN CORRELATION PAIRS OPTION

The correlation between distance with ball speed and launch angle were both reported as having positive r-values in the correlation matrix with only distance and ball speed being marked as supported correlations. Therefore, correlations with a directional alternative hypothesis (correlated positively) can be run.

Return to the analysis options and remove the variables just keeping distance, launch angle and ball speed.

- ✓ Alt hypothesis: Correlated positively
- ✓ Display pairwise table
- ✓ Report Bayes factors

Open up the Plot Individual Pairs tab, Add the values pairwise to the right box. Then tick all the options as shown below:



UNDERSTANDING THE OUTPUT

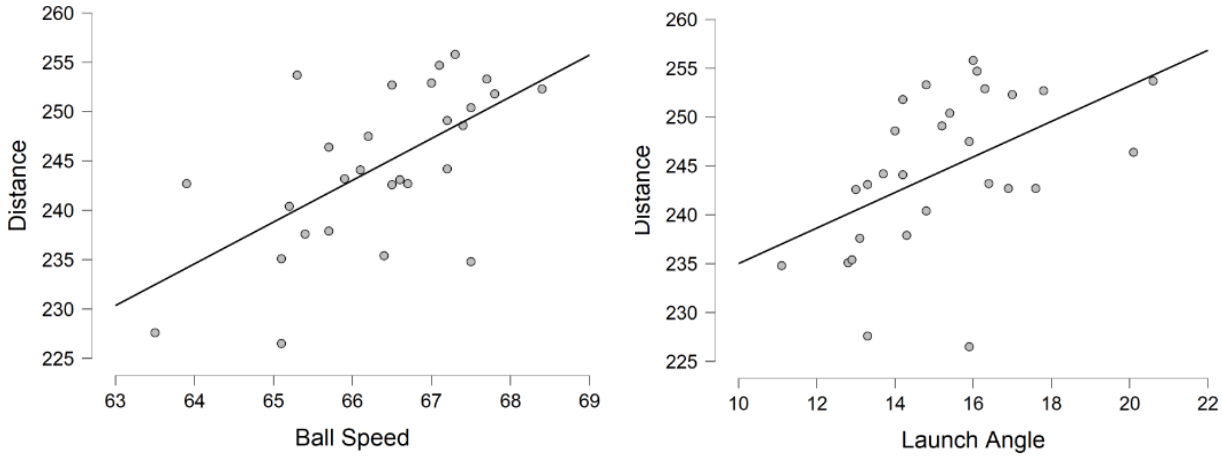
This method now produces a table of pairwise correlations instead of the correlation matrix. The reported r-values are the same, however, the Bayes factors are different. There is very strong evidence (168 times more likely) supporting a positive correlation between ball speed and distance i.e. the alternative hypothesis.

Whereas in the 2-sided correlation matrix there was only anecdotal /moderate evidence in support of a correlation between launch angle and distance, now there is strong evidence in support of the alternative positive correlation (BF_{10} = 14).

Bayesian Pearson Correlations

			Pearson's r	BF_{-0}
Ball Speed	-	Launch Angle	−0.111	0.160
Ball Speed	-	Distance	0.620	168.951
Launch Angle	-	Distance	0.495	14.180

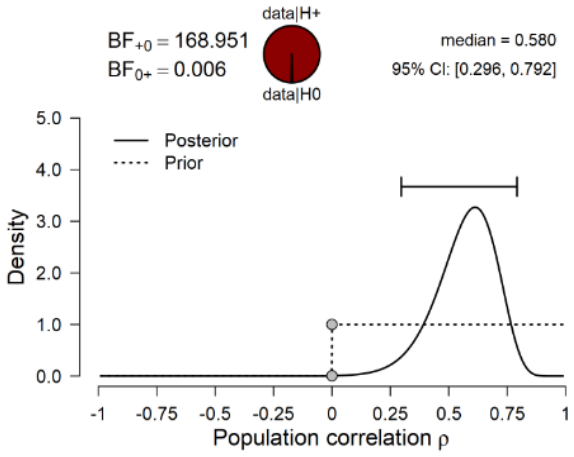
Note. For all tests, the alternative hypothesis specifies that the correlation is positive.

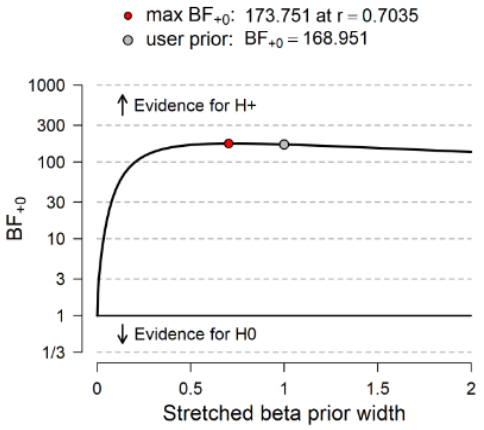


The posterior-prior plot for distance and ball speed, assuming a positive correlation, show the data fully distributed to the right of rho = 0, with a median value of 0.58 as was indicated by the large Bayes factor.

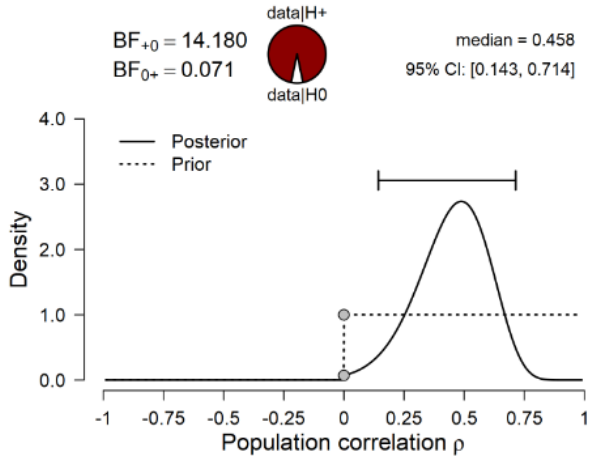
The dashed line shows the uniform prior distribution and the solid line the posterior distribution (based on the dataset). Each of the distributions has a grey dot at the 0.0 effect size. If the dot on the prior distribution is higher than the one on the posterior distribution, then the Bayes factor is more supportive of the alternative hypothesis.

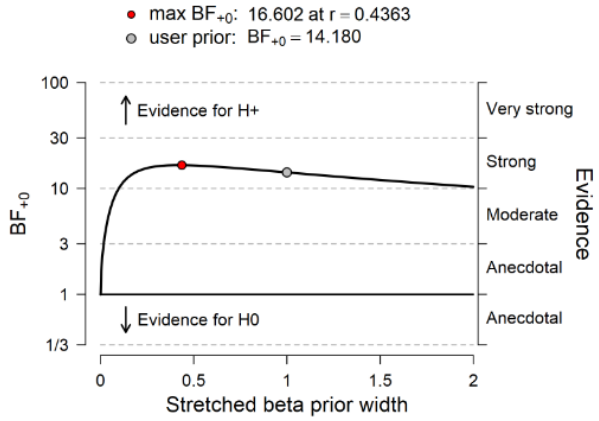
The robustness analysis allows one to inspect what BF would be obtained if the alternative model were specified differently. The analysis shows the outcomes of specifying a range of different prior values from 0 to 2.





The prior width is set as 1.0 by default in JASP. If the results are insensitive to changes in the prior width the Bayes factor should be stable. Except for very small prior widths, the Bayes factors are relatively stable therefore confirming the robustness of the analysis.
A similar picture is shown below when correlating launch angle with distance.





REPORTING THE RESULTS

Using a one-sided alternative hypothesis there was a positive correlation for distance with respect to ball speed (r = 0.620) this was accompanied by a Bayes factor BF_{10} = 169 indicating a decisive likelihood ("evidence") of this occurring under the H_1 than H_0.

Using a one-sided alternative hypothesis there was a positive correlation for distance with respect to the launch angle (r = 0.495) this was accompanied by a Bayes factor BF_{10} = 14.2 indicating a strong likelihood ("evidence") of this occurring under the H_1 than H_0.

BAYESIAN REGRESSION ANALYSIS

Both linear regression and Bayesian regression can be used for predictive analysis, i.e. to predict a dependent outcome variable from one (simple regression) or more (multiple regression) independent predictor variables.

Simple regression results in a hypothetical model of the relationship between the outcome and predictor variable(s). The model used is a linear one defined by the formula:

Simple regression

$$y = b_0 + b_1 * x_1 \quad + b_2 * x_2 + b_3 * x_3 + \ldots b_n * x_n$$

Multiple regression

- y = estimated dependent outcome variable score,
- b_0 = constant (intercept),
- b_1 = regression coefficient(s) (slope)and
- x = score on the independent predictor variable (s)

NOTE: Linear regression provides both the constant and regression coefficient(s). Bayesian regression also provides these but in a slightly different way in that, the constant is centred on the mean value of the outcome variable.

Regression tests the following hypotheses

H_o: that there will be no prediction of the dependent (outcome) variable by the predictor variable(s).

H_1: $H_1 \neq H_o$

ASSUMPTIONS

1. Linear relationship: important to check for outliers since linear regression is sensitive to their effects.
2. Independence of variables
3. Multivariate normality: requires all variables to be normally distributed
4. Homoscedasticity: homogeneity of variance of the residuals
5. Minimal multicollinearity /autocorrelation: when the independent variables/residuals are too highly correlated with each other.

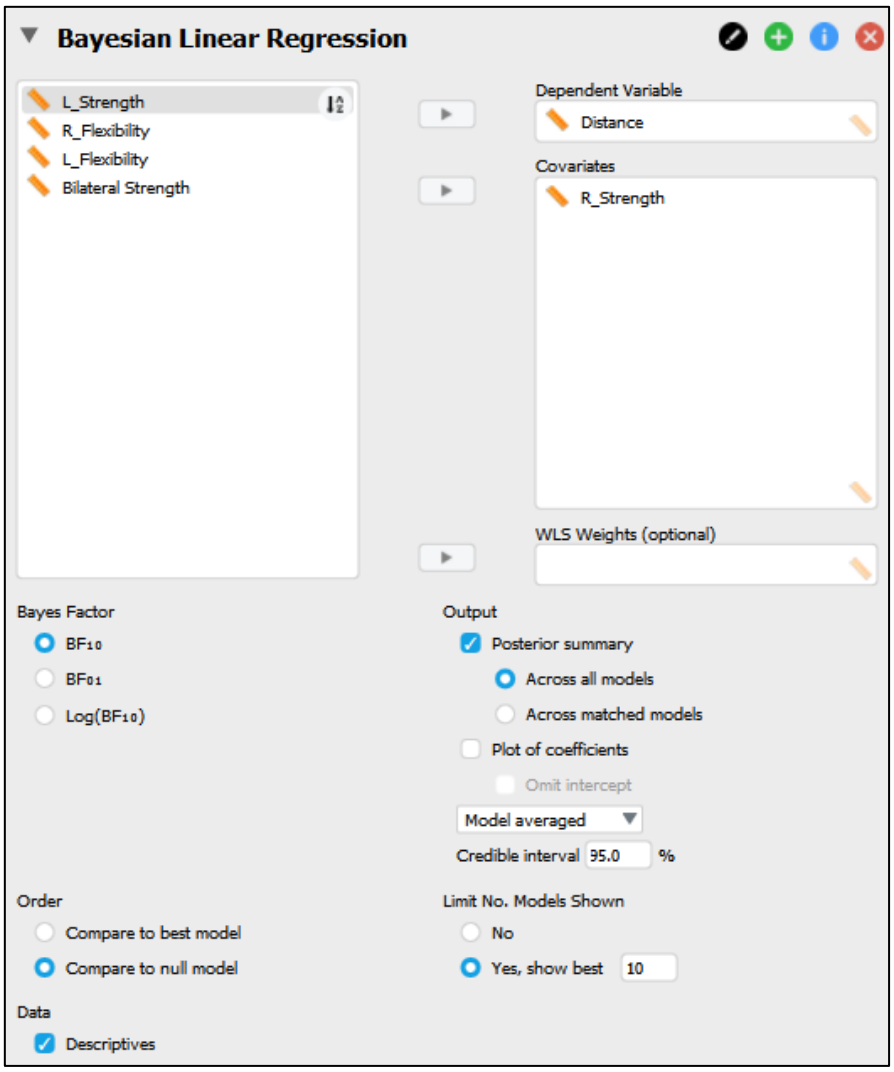
SIMPLE BAYESIAN REGRESSION

Regression compares the data to two hypotheses, the null hypothesis (H_o) that there will be no prediction of the dependent (outcome) variable by the predictor variable(s) against an alternative hypothesis (H_1) that does include predictor(s).

Open **Bayesian regression.csv**. This data set contains rugby kick data including distance (feet) kicked, right/left leg strength and flexibility and bilateral leg strength. Firstly, go to Descriptives > Descriptive statistics and check the boxplots for any outliers. In this case, there should be none, though it is good practice to check.

For this simple regression go to Regression > Bayesian Linear regression and put distance into the Dependent Variable (outcome) and R_Strength into the Covariates (Predictor) box. Tick the following options in the Statistics options:

- ✓ BF_{10} (Bayes factor favouring the alternative hypothesis over the null hypothesis)
- ✓ Compare to the null model
- ✓ Posterior summary – across all models
- ✓ Descriptives



UNDERSTANDING THE OUTPUT

You will now get the following outputs:

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2
Null model	0.500	0.046	0.048	1.000	0.000
R_Strength	0.500	0.954	20.728	20.728	0.614

P(M) = prior model probability. Since there are only two models the prior probability of each model is assigned an uninformed prior where both models have equal probabilities P(M) = 0.5.

P(M I data) is the probability of the posterior distribution having taken into account the data which can be seen as having gone from 50 to 95.4% probability in the model containing right leg strength.

BF_M shows how much the model has improved after seeing the data.

The BF_{10} value (20.728) suggests that there is strong evidence for the alternative model containing right leg strength compared to the null model. However, the R^2 value suggests that right leg strength alone only accounts for 61.4% variance in the model.

Posterior Summaries of Coefficients

	Coefficient	Mean	SD	P(incl)	P(incl|data)	$BF_{inclusion}$	95% Credible Interval	
							Lower	Upper
b_0	Intercept	486.077	15.812	1.000	1.000	1.000	452.354	515.995
b_1	R_Strength	5.227	1.798	0.500	0.954	20.728	0.000	7.928

This table gives the coefficients that can be put into the linear equation.
y = b_0 + b_1*x_1

y = estimated dependent outcome variable score,
b_0 = constant (intercept),
b_1 = regression coefficient (**R_strength)**
x_1= score difference for the independent predictor variable (= x – mean x)

Descriptives

	N	Mean	SD
Distance	13	486.077	85.250
R_Strength	13	66.769	10.402

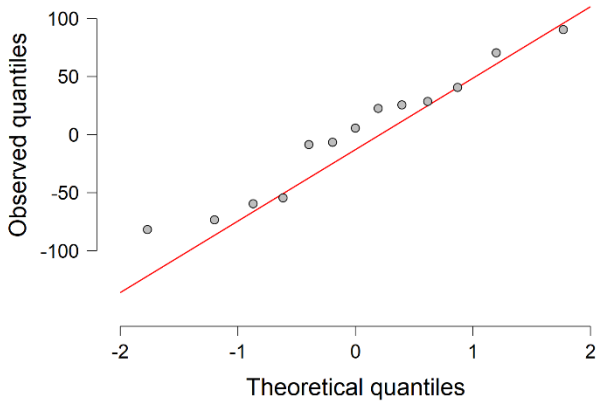
The prediction equation is therefore slightly different from the one used in normal linear regression. For example, for a leg strength of 77 kg, the distance kicked can be predicted by the following - remember x_1 the score difference for the independent predictor variable (= x – mean x):

(x – mean x)

Distance = 486.077 + (5.479 * [77 – 66.769]) = 543.7 feet

FURTHER CHECKS

In the analysis menu, under plots, now tick Q-Q plot of model-averaged residuals. The Q-Q plot shows that the standardized residuals fit fairly well along the diagonal suggesting that both assumptions or normality and linearity have also not been violated.



REPORTING THE RESULTS

A simple Bayesian regression was carried out using right leg strength as a predictor of rugby kicking distance. An uninformed uniform prior [P(M)] of 0.5 was set for each possible model. There was strong evidence for a regression model including right leg strength (BF_{10} = 20.73) compared to the null model.

It is suggested that the Model Comparison and Posterior summaries of coefficients tables are also shown along with the regression equation.

MULTIPLE REGRESSION

The model used is still a linear one defined by the formula:

$y = b_0 + b_1 * x$

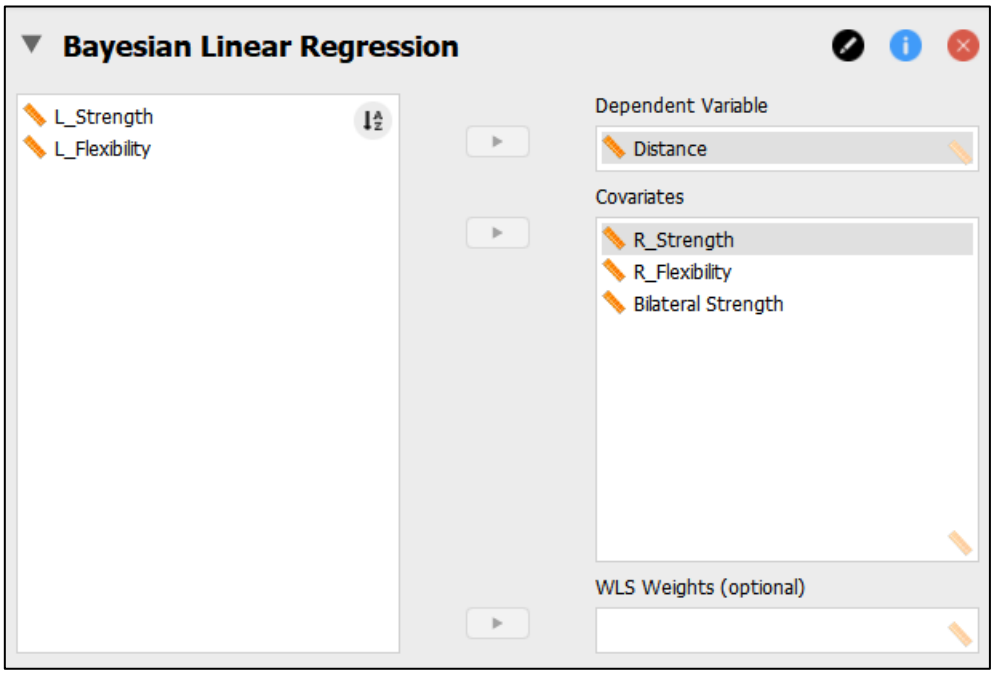
- y = estimated dependent outcome variable score,
- c = constant,
- b = regression coefficient and
- x = score on the independent predictor variable

However, we now have more than 1 regression coefficient and predictor score i.e.

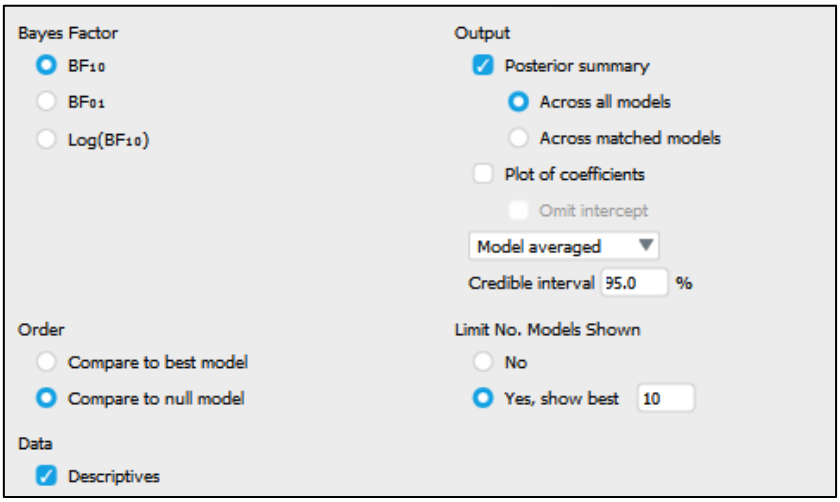
$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 \ldots\ldots b_n * x_n$

RUNNING MULTIPLE BAYESIAN REGRESSION

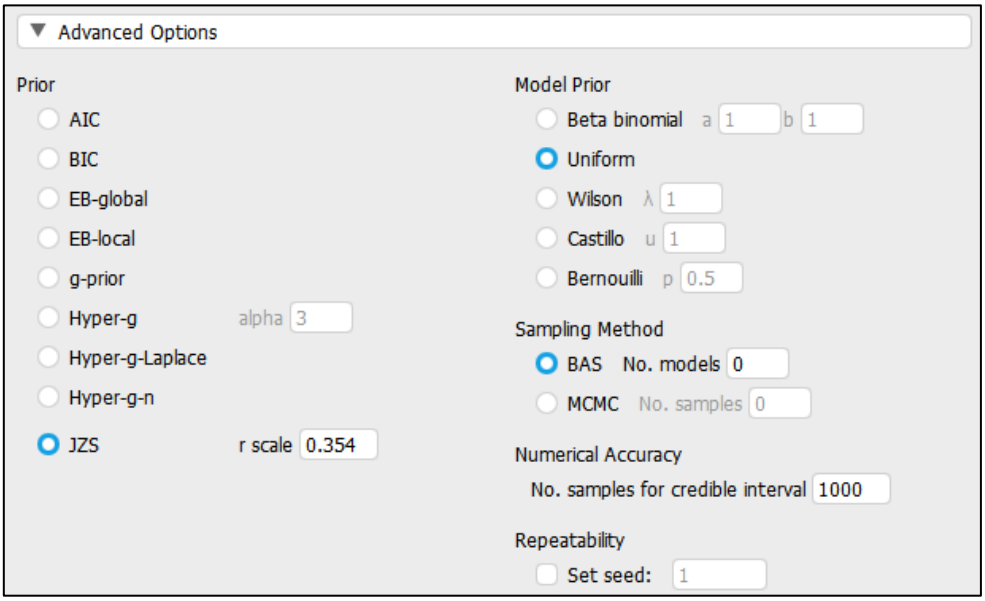
Open **Bayesian regression.csv**. that we used for simple regression. Go to Regression > Bayesian linear regression and put distance into the Dependent Variable (outcome) and now add R_strength, R_flexibility and Bilateral strength to the Covariates (Predictor) box.



In the statistical analysis options use the same options as used in the simple regression example.



In Advanced Options, under Model Prior, select Uniform which will assign equal prior probabilities for each possible model.



UNDERSTANDING THE OUTPUT

You will now get the following outputs:

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2
Null model	0.125	0.004	0.029	1.000	0.000
R_Strength + Bilateral Strength	0.125	0.267	2.553	63.856	0.782
R_Flexibility + Bilateral Strength	0.125	0.185	1.592	44.283	0.761
R_Flexibility	0.125	0.131	1.053	31.254	0.649
R_Strength + R_Flexibility + Bilateral Strength	0.125	0.122	0.973	29.154	0.805
Bilateral Strength	0.125	0.111	0.877	26.600	0.636
R_Strength + R_Flexibility	0.125	0.092	0.713	22.090	0.715
R_Strength	0.125	0.087	0.665	20.728	0.614

P(M) = prior model probability; P(M|data) = posterior model probability; BF_M = change from prior model odds to posterior model odds; BF_{10} = Bayes factor for each row (model) against the one on top (this is why the first BF = 1).

JASP models all possible predictor permutations and in this case, there are 8 possible models each of which has been assigned an equal uninformed prior i.e. each model has a probability of 0.125.
The largest posterior probability P(M I Data) and BF_M increases are seen in model 2 where R_strength and Bilateral Strength are used as predictors. This is associated with the largest BF_{10}

value of 63.86 which is very strong evidence for the alternative hypothesis (model). This is defined as the best model.
The R^2 value states that this can account for 78.2% of the variance in the model compared to the 61.4% seen in the simple regression model. Just to note, however, model 5 has a higher R^2 value.

In cases where there are many possible alternative models, it may be easier to change the Bayes factor to BF_{01} and Compare to the best model in the options.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{01}	R^2
R_Strength + Bilateral Strength	0.125	0.267	2.553	1.000	0.782
R_Flexibility + Bilateral Strength	0.125	0.185	1.592	1.442	0.761
R_Flexibility	0.125	0.131	1.053	2.043	0.649
R_Strength + R_Flexibility + Bilateral Strength	0.125	0.122	0.973	2.190	0.805
Bilateral Strength	0.125	0.111	0.877	2.401	0.636
R_Strength + R_Flexibility	0.125	0.092	0.713	2.891	0.715
R_Strength	0.125	0.087	0.665	3.081	0.614
Null model	0.125	0.004	0.029	63.856	0.000

Here the model containing right leg strength and bilateral strength has been selected as the best model (with a Bayes factor of 1). The BF_{01} i.e. favouring the null model allows comparison of the other models with the best one. For example, the best model is favoured 3 times more than one just including right leg strength and 64 times more than the null model.

The coefficients are shown for all the covariates included in the analysis:

Posterior Summaries of Coefficients

						95% Credible Interval	
Coefficient	Mean	SD	P(incl)	P(incl|data)	$BF_{inclusion}$	Lower	Upper
Intercept	486.077	13.406	1.000	1.000	1.000	458.790	520.189
R_Strength	1.931	2.201	0.500	0.568	1.317	−0.143	6.905
R_Flexibility	2.576	3.225	0.500	0.531	1.130	−0.430	10.178
Bilateral Strength	1.234	1.083	0.500	0.686	2.183	−0.012	3.314

My personal preference is to rerun the analysis using just the best model covariates and use those criteria.

Posterior Summaries of Coefficients

	Coefficient	Mean	SD	P(incl)	P(incl|data)	$BF_{inclusion}$	95% Credible Interval	
							Lower	Upper
b_0	Intercept	486.077	13.579	1.000	1.000	1.000	453.129	512.356
b_1	R_Strength	2.973	2.231	0.500	0.754	3.065	0.000	6.954
b_2	Bilateral Strength	1.648	1.080	0.500	0.806	4.163	0.000	3.534

Descriptives

	N	Mean	SD
Distance	13	486.077	85.250
R_Strength	13	66.769	10.402
Bilateral Strength	13	88.846	21.687

CONSTRUCTING THE REGRESSION EQUATION

Now there is one constant (b_0) and two regression coefficients (b_1 and b_2). These coefficients can be put into the linear equation.

$y = b_0 + b_1*x_1 + b_2*x_2$

y = estimated dependent outcome variable score,

c = constant **(mean value of the outcome variable)**

b_1 = regression coefficient (**R_strength)**

b2 = regression coefficient for Bilateral strength

x_1 = score difference for the R-strength variable **(= x – mean x)**

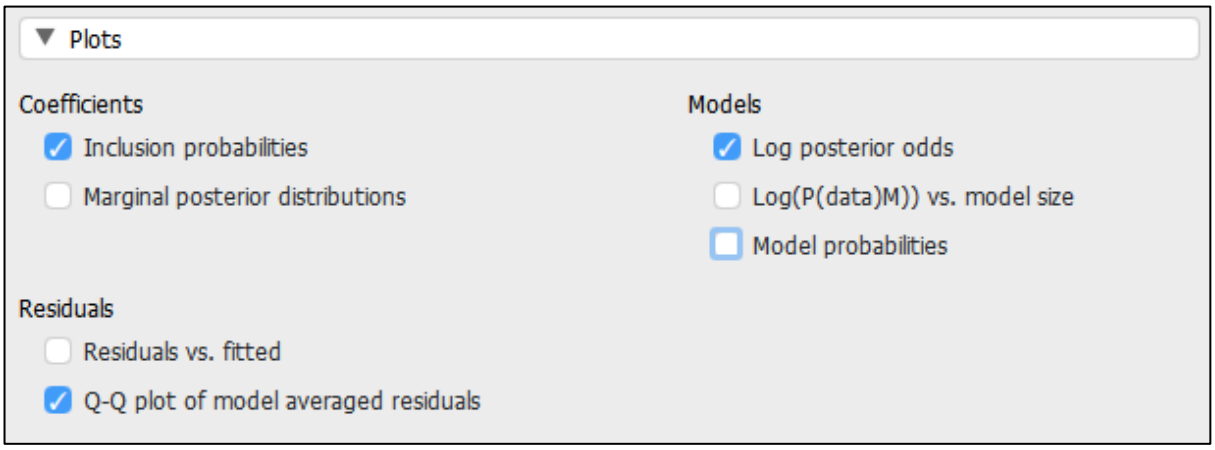
x_2= score difference for the Bilateral strength variable **(= x – mean x)**

For right leg strength of 77kg and bilateral leg strength of 121 kg, the predicted kick distance from the regression equation will be:

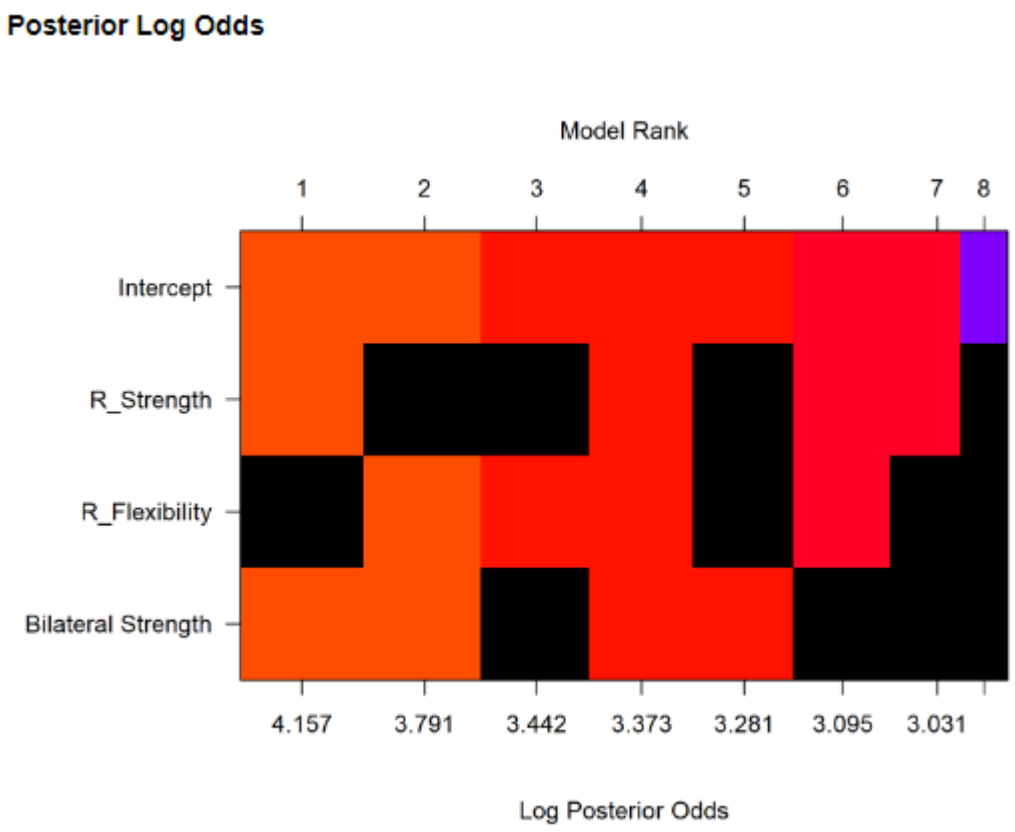
Distance = 486.077 + (2.973 * [77 – 66.769]) + (1.648 * [121 – 88.846]) = 578 feet

ADDITIONAL PLOTS

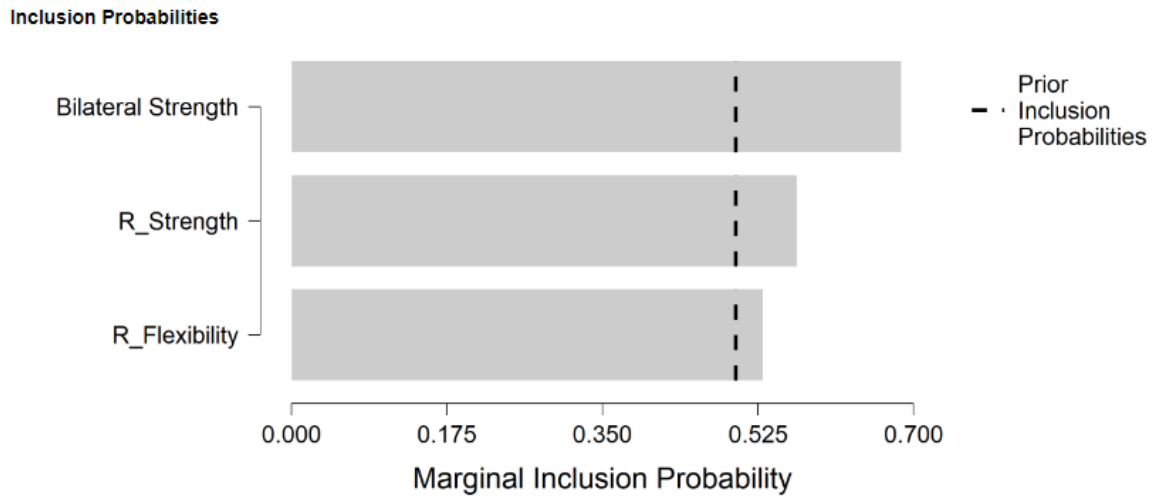
Tick the following options in Plots:



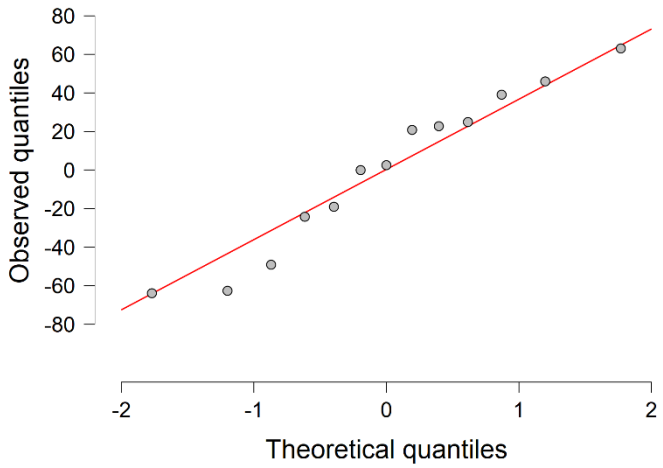
The first plot just enables each possible model to be visualised. The coloured squares are the included covariates (the null model being purple). Here the best model (ranked 1) includes the intercept, right leg and bilateral leg strength since it has the highest Log posterior odds.



The next graph shows how close each covariate is to the cut-off prior inclusion probabilities. Right leg flexibility is very close to this cut-off point and was not included in the best model.



The Q-Q plot shows that the standardized residuals fit fairly well along the diagonal suggesting that both assumptions or normality and linearity have also not been violated.



REPORTING THE RESULTS

A Bayesian multiple regression was carried out using right leg flexibility, right and bilateral leg strength, as predictors of rugby kicking distance. An uninformed uniform prior [P(M)] of 0.125 was set for each of the possible 8 models. There was strong evidence for a regression model including the right leg and bilateral leg strength (BF_{10} = 63.9) compared to the null model.

It is suggested that the Model Comparison and Posterior summaries of coefficients tables are also shown along with the regression equation.

BAYESIAN ANOVA

Whereas t-tests compare the means of two groups/conditions, one-way analysis of variance (**ANOVA**) compares the means of 3 or more groups/conditions. The Bayesian approach compares the predictive performance of different models. JASP features Bayesian versions of the between subjects, repeated measures, and mixed ANOVAs.

In these analyses, the following models are compared:

H_0 – Null hypothesis: predicts the overall mean

H_1 – Alternate hypothesis: predicts the means of the different levels of the fixed factor.

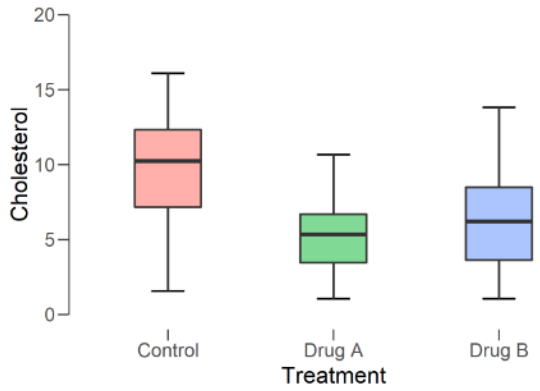
If the alternative hypothesis model outperforms the null model, this is an indication that the dependent variable differs between the levels of the independent variable. However, this does not say between which specific levels these differences occur. To determine where the group differences are, post hoc (From the Latin *post hoc*, "after this") tests can be conducted.

ASSUMPTIONS

1. The independent variable must be categorical, and the dependent variable must be continuous.
2. The groups should be independent of each other.
3. The dependent variable should be continuous and approximately normally distributed.
4. There should be no outliers.
5. There should be homogeneity of variance between the groups. The first 2 assumptions are usually controlled using an appropriate research method design.

RUNNING THE BAYESIAN ANOVA

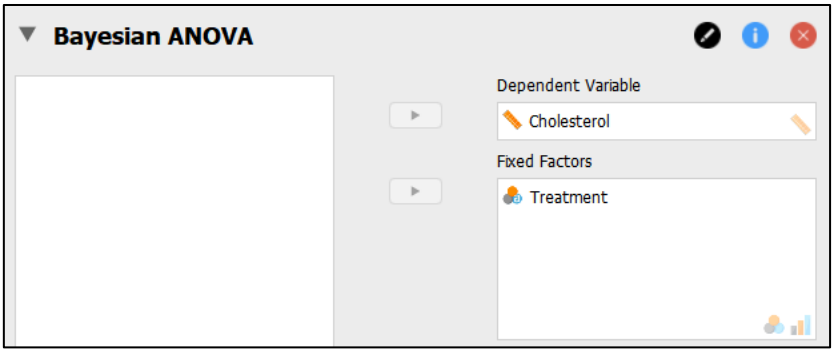
Load **Bayesian Independent ANOVA.csv**. This contains data showing blood cholesterol levels (mmol/L) in a control group and two groups taking different statin drugs. For good practice check the descriptive statistics and the boxplots for any extreme outliers.



There is no evidence that the response variable is consistently non-normal across all populations - each boxplot is approximately symmetrical. No extreme outliers are observed. There is no evidence that variance, as estimated by the height of the boxplots, differs between the groups.

NOTE: When running the ANOVA analysis using the included dataset the results are likely to be very slightly different to the ones in this presented chapter. This is because the analyses are based on numerical algorithms like Markov chain Monte Carlo (MCMC). The degree to which the results fluctuate is quantified by an error percentage. The higher the error percentage, the higher the fluctuation of the results.

Go to ANOVA > Bayesian ANOVA, put Cholesterol into the Dependent Variable and the treatment groupings into the Fixed Factors box.



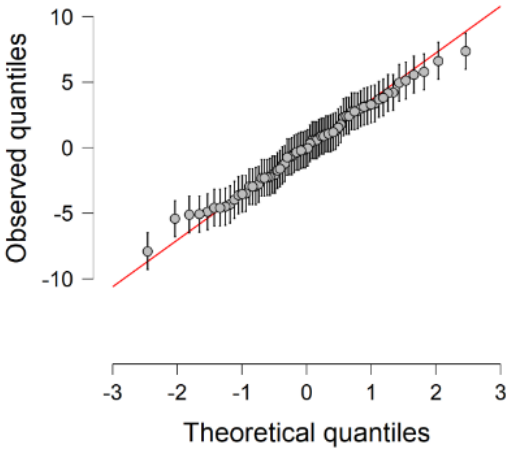
In the main analyses options

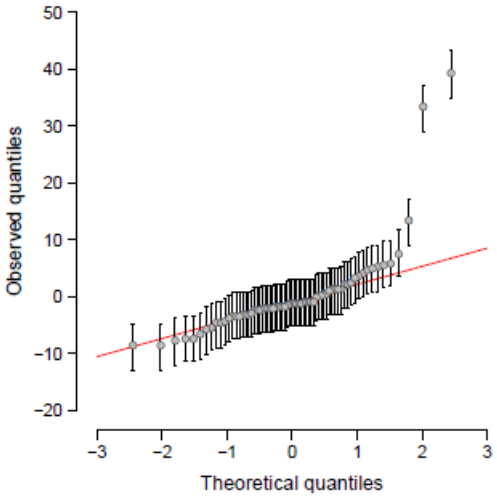
- ✓ Change the Order option to 'Compare to the null model.
- ✓ Plots – Q-Q plot of residuals

This will initially result in one table and one graph.

UNDERSTANDING THE OUTPUT

Firstly, it is important to test the assumption of normality, in this case, that the residuals are normally distributed. This can easily be done by looking at the Q-Q plot (below left). If the residuals are normally distributed, they should lie consistently along the diagonal line. Any obvious deviations along the line (as seen below on the right) would suggest that the assumption of normality has been violated.





The following table compares the competing models:

Model Comparison ▼

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.002	0.002	1.000	
Treatment	0.500	0.998	545.791	545.791	0.010

Models: shows the two models tested, null and treatment. The null model is shown first.

P(M): for the ANOVA, the analysis sets the prior probabilities of each model to be equal (i.e., prior model odds of 0.5)

P(M |data): shows the updated probabilities having now seen the data (i.e., posterior model probabilities).

BF_M: shows how much the data have changed the prior model odds

BF_{10}: shows the Bayes factors for each model. The first entry is always 1 since the null model is compared against itself. The BF_{10} for treatment, 546 suggests that the data are 546 times more likely under the model incorporating treatment, than under the null model.

Error %: is very small, 0.01%, indicating that the sensitivity to numerical fluctuations is minuscule.

If the evidence suggested that the data is best predicted by the null model or that the evidence for the alternative was inconsequential. Although evidence for a lack of an effect is still information – there is no point in following up with further analyses.

FURTHER ANALYSIS

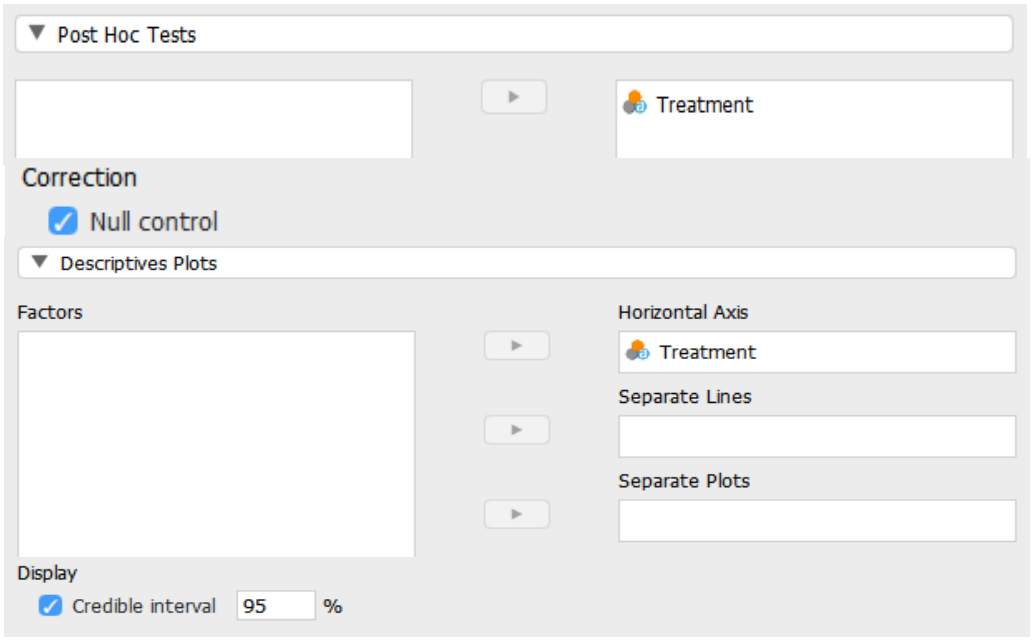
Select the following options for further analysis.

In the main analysis options:

- ✓ Tables – Descriptives
- ✓ Plots – Model averaged posteriors – Group levels in a single plot

Add treatment to Post hoc tests

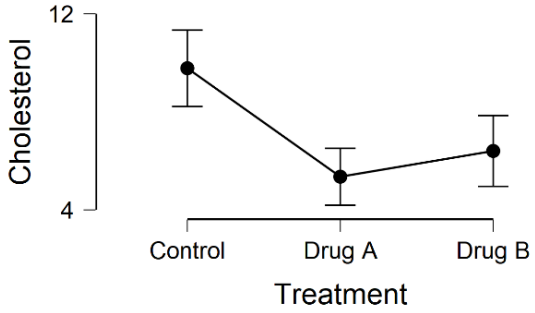
Add Treatment to the horizontal axis in Descriptive plots and display credible intervals



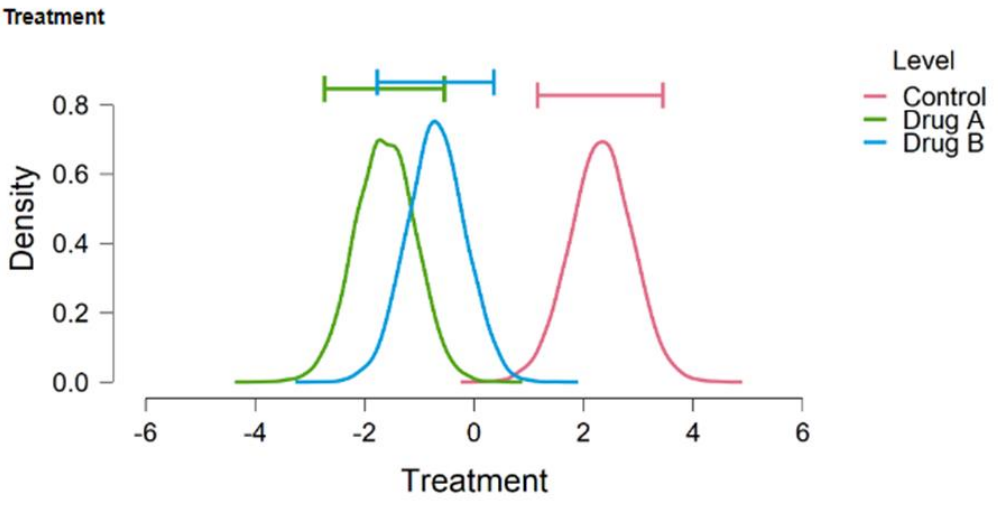
The descriptives and plot show that both drug groups have lower cholesterol than the control group.

Descriptives - Cholesterol

				95% Credible Interval	
Treatment	Mean	SD	N	Lower	Upper
Control	9.778	3.689	24.000	8.220	11.336
Drug A	5.352	2.748	24.000	4.192	6.512
Drug B	6.402	3.421	24.000	4.957	7.847



The model-averaged posterior distributions (horizontal bars show the 95% credible intervals around the median) are shown below. There is a clear difference between the two drug groups compared to the control group.



Bayesian post hoc testing is based on pairwise comparisons using Bayesian t-tests. As in frequentist analyses, multiple t-tests will increase familywise error. In JASP, methods are used to correct for multiplicity based on adjusting the prior odds. The post hoc comparisons are shown in the table below. The relative plausibility of each model is specified by the prior odds. If the odds are <1, there is some prior belief that there is no difference. The posterior odds are the result of multiplying the prior odds by the BF and represent the relative plausibility of the models *after* observing data.

Post Hoc Comparisons - Treatment

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
Control	Drug A	0.587	467.323	795.578	2.872e −7
	Drug B	0.587	10.587	18.023	5.066e −5
Drug A	Drug B	0.587	0.295	0.502	0.023

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holdsacross all comparisons (Westfall, Johnson, & Utts, 1997). Individualcomparisons are based on the default t-test with a Cauchy (0, r =1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Comparison of Drug A to the control: the posterior odds suggest that the alternative hypothesis (H_1) is 467 times more likely than the null hypothesis (H_0). The update from prior to posterior odds can be described as decisive evidence in favour of H_1.

Comparison of Drug B to the control: the posterior odds suggest that the alternative hypothesis (H_1) is 10.6 times more likely than the null hypothesis (H_0). The update from prior to posterior odds can be described as strong evidence in favour of H_1.

Comparison of Drug B to Drug C: the posterior odds suggest that the null hypothesis (H_0) is 3.4 (1 / 0.295) times more likely than the null hypothesis (H_1). The update from prior to posterior odds can be described as moderate evidence in favour of H_0.

REPORTING THE RESULTS

The Bayesian one-way ANOVA indicates that the data were 540 times more likely to occur under the model including the effect for treatment, compared to the model without the effect. In order to follow up on this result, we compared each level of the dependent variable. The cholesterol levels on drug A and drug B were 5.35 and 6.04 mmol/L respectively compared to the control group (9.79 mmol/l). Post hoc comparisons of Control .vs. Drug A and Control .vs. Drug B revealed posterior odds of 467 and 10.5, which indicates decisive and strong evidence respectively in favour of the alternative hypothesis, that is, a reduction in cholesterol levels.

BAYESIAN REPEATED MEASURES ANOVA

The Bayesian one-way repeated measures ANOVA (**RM ANOVA**) is used to assess if there is a difference in means between 3 or more groups, featuring the same set of participants tested multiple times or under different conditions. Such a research design, for example, could be that the same participants were tested for an outcome measure at 1, 2, and 3 weeks or that the outcome was tested under conditions 1, 2, and 3 (i.e., within each subject).
The independent variable should be categorical and the dependent variable needs to be a continuous measure. In this analysis, the independent categories are termed **levels** (i.e., these are the related groups). So, in the case where an outcome was measured at weeks 1, 2, and 3, the 3 levels would be week 1, week 2, and week 3.
The models under consideration are

H_0: the null model, where there are no differences between the levels: i.e., no effect ($\delta=0$)

H_1 the alternative model, where there are differences between the levels: i.e., there is an effect ($\delta\neq0$)

ASSUMPTIONS

The RM ANOVA makes the following assumptions:

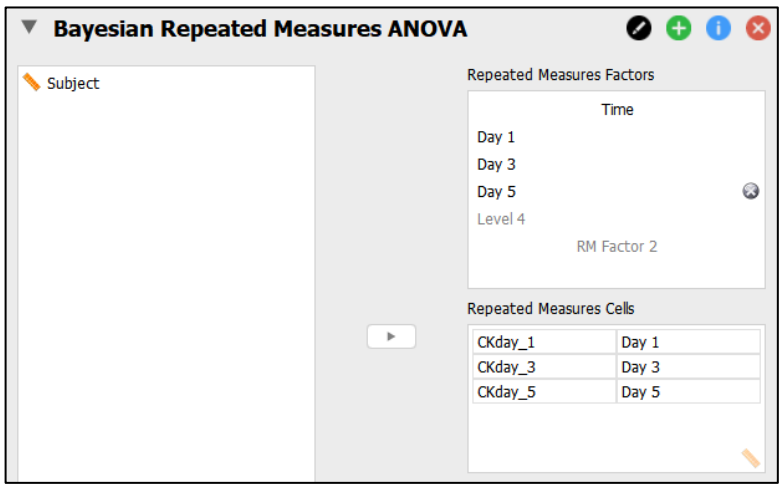
- The dependent variable and residual should be approximately normally distributed.
- There should be no outliers.
- Homogeneity of variances across the factor levels.

RUNNING THE BAYESIAN ANOVA

Load **Bayesian RMANOVA.csv**. This contains data showing creatine kinase (CK) levels (mmol/L) in blood taken over days 1, 3, and 5 following a muscle damage protocol. For good practice, check the descriptive statistics and the boxplots for any extreme outliers. It can be seen that there are no outliers.

NOTE: When running the ANOVA analysis using the included dataset the results are likely to be very slightly different to the ones in this presented chapter. This is because the analyses are based on numerical algorithms like Markov chain Monte Carlo (MCMC). The degree to which the results fluctuate is quantified by an error percentage. The higher the error percentage, the higher the fluctuation of the results.

Go to ANOVA > Bayesian Repeated Measures ANOVA. In Repeated measures factors, define the RMFactor 1 as Time and add days 1, 3, and 5 as levels. Then add the appropriate variables to the Repeated Measures cells.



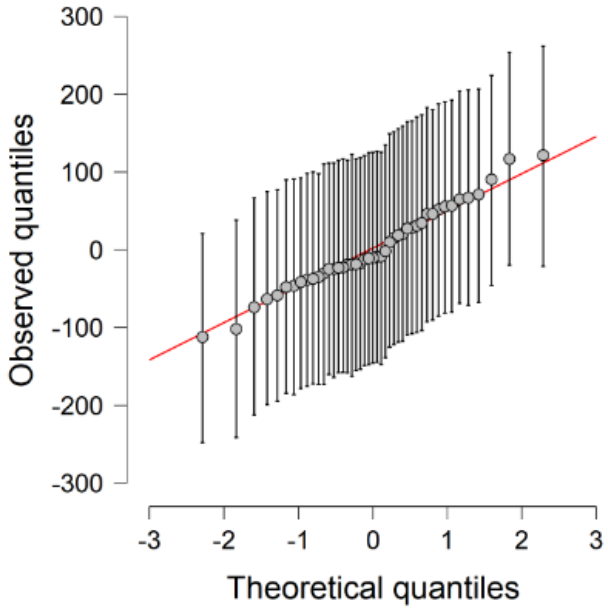
In the main options add the following:

- ✓ Order - 'Compare to null model'
- ✓ Tables – Descriptives
- ✓ Plots – Q-Q plots of residuals

In descriptive plots add Time to the horizontal axis

UNDERSTANDING THE OUTPUT

Firstly, it is important to test the assumption of normality, in this case, that the residuals are normally distributed. This can easily be done by looking at the Q-Q plot.



If the residuals are normally distributed, they should be positioned consistently along the diagonal line. Any obvious deviations along the line would suggest that the assumption of normality has been violated.

Bayesian Repeated Measures ANOVA

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.500	1.017e −9	1.017e −9	1.000	
Time	0.500	1.000	9.836e +8	9.836e +8	0.847

Note. All models include subject

Models: shows the two models tested, null and Time.

P(M): for the RMANOVA, the analysis sets the prior probabilities of each model is equal (i.e., 50:50).

P(M |data): shows the updated posterior probabilities having now seen the data.

BF_M: shows how much the data have changed the prior model odds

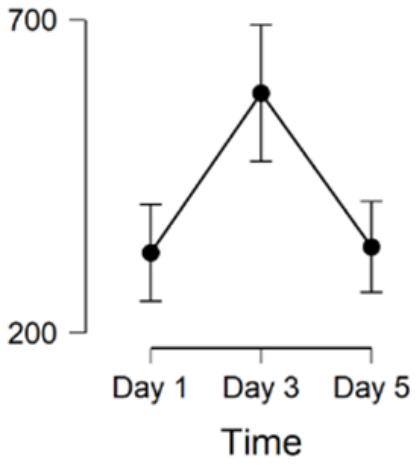
BF_{10}: shows the Bayes factors for each model. The first entry is always 1 since the null model is compared against itself. The BF_{10} for time, $9.83*10^8$ suggests that the model which includes Time predicts the observed data $9.83*10^8$ times better than the null.

Error %: is very small, 0.85%, and can be considered negligible.

The descriptive values and plots show that CK levels were higher on day 3 than days 1 and 5.

Descriptives

				95% Credible Interval	
Time	Mean	SD	N	Lower	Upper
Day 1	328.000	139.346	15.000	250.833	405.167
Day 3	583.000	196.602	15.000	474.125	691.875
Day 5	337.267	131.117	15.000	264.657	409.877



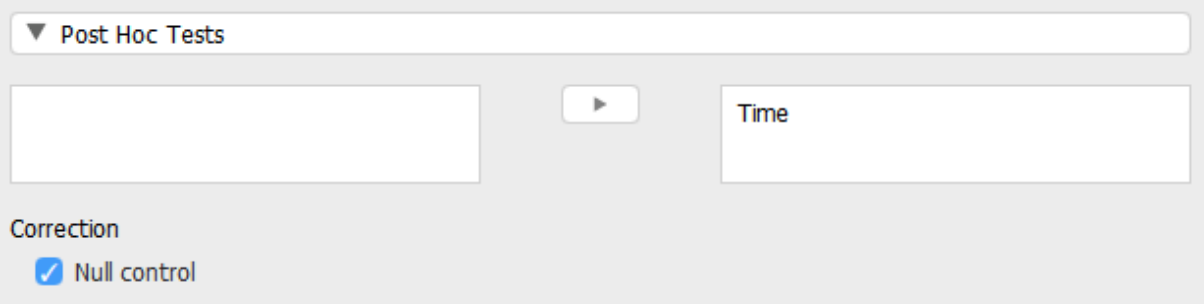
If the evidence suggested that the data is best predicted by the null model or that the evidence for the alternative was inconsequential. Although evidence for a lack of an effect is still information – there is no point in following up with further analyses.

FURTHER ANALYSIS

Select the following options for further analysis:

- ✓ Tables Estimates
- ✓ Plots Model averaged posteriors – Group levels in a single plot
- ✓ Plots Q-Q plots of residuals

Also, add the following:

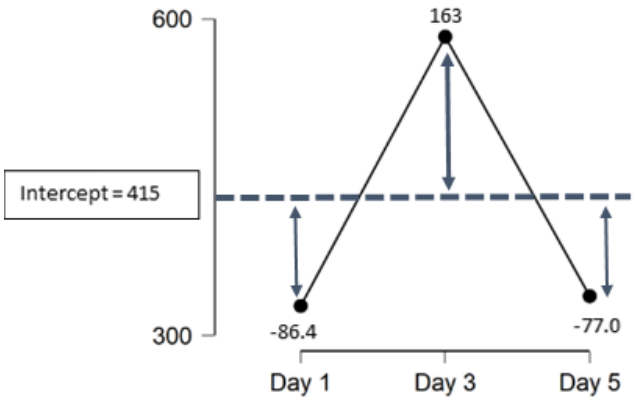


Estimates are shown in the Model averaged posterior summary table:

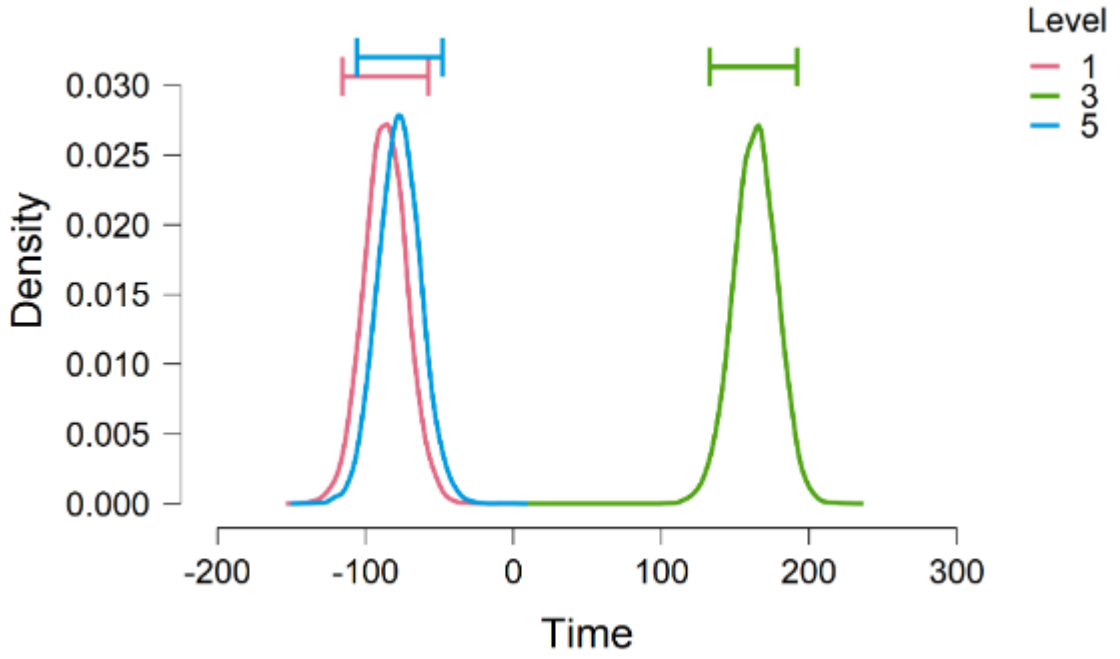
Model Averaged Posterior Summary

				95% Credible Interval	
Variable	Level	Mean	SD	Lower	Upper
Intercept		415.741	39.820	335.014	496.788
Time	Day 1	−86.438	14.430	−115.541	−57.469
	Day 3	163.450	14.691	132.996	191.848
	Day 5	−77.012	14.467	−105.637	−47.874

This table shows the mean differences and 95% credible intervals for each of the factor levels normalised to the intercept (mean value of all the data) and is explained graphically below.



The model-averaged posterior distributions (horizontal bars show the 95% credible intervals around the median) are shown below on the left. There is a clear separation between day 3 and days 1 and 5. Below is a visualisation of the model-averaged posterior summary table data.



The post hoc comparisons are shown in the table below. The relative plausibility of each model is specified by the prior odds, i.e., the relative probability of the models before observing data. If the odds are <1, there a prior belief that there is no difference. The posterior odds are the result of multiplying the prior odds by the BF (which is affected by the prior distribution) and represent the relative probability of the models after observing the data.

Post Hoc Comparisons - Time

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
Day 1	Day 3	0.587	30416.858	51782.097	9.499e −10
	Day 5	0.587	0.223	0.379	0.005
Day 3	Day 5	0.587	28576.282	48648.673	1.031e −9

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holdsacross all comparisons (Westfall, Johnson, & Utts, 1997). Individualcomparisons are based on the default t-test with a Cauchy (0, r =1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Comparison of CK levels on day 1 with day 3: The posterior odds indicate that the data is 30,416 times more likely to occur under the alternative hypothesis (H_1) than under the null hypothesis. This can be described as decisive evidence in favour of H_1.

Comparison of CK levels on day 1 with day 5: The posterior odds indicate that the data is 4.48 (1/ 0.223) times more likely to occur under the alternative hypothesis (H_0) than under the alternative hypothesis. This can be described as moderate evidence in favour of H_0.

Comparison of CK levels on day 3 with day 5: The posterior odds indicate that the data is 28576 times more likely to occur under the alternative hypothesis (H_1) than under the null hypothesis. This can be described as decisive evidence in favour of H_1

.

REPORTING THE RESULTS

Using a Bayesian RM ANOVA (specifying a multivariate Cauchy prior on the effects[6]), the Bayes factor indicates that the data are $9.72*10^{8}$ times more likely under the model that includes time as the predictor, compared to the null model. Post hoc comparisons of day 1 .vs. day 3 and day 3 .vs. day 5 revealed posterior odds of 30,416 and 28,576 against the null hypothesis, which indicates decisive evidence in favour of the alternative hypothesis. When comparing day 1 and 5, there was moderate evidence in favour of the null hypothesis.

[6] Rouder et al 2012, van den bergh 2019 https://psyarxiv.com/spreb

BAYESIAN MIXED FACTOR ANOVA

Mixed factor ANOVA (another two-way ANOVA) is a combination of both independent and repeated measures ANOVA involving more than 1 independent variable (known as factors). Below is a design with time as the within and group as the between factor:

Independent variable (Factor 2)	*Independent variable (Factor 1) = time or condition*		
	Time/condition 1	**Time/condition 2**	**Time/condition 3**
Group 1	Dependent variable	Dependent variable	Dependent variable
Group 2	Dependent variable	Dependent variable	Dependent variable

The factors are split into levels, therefore, in this case, Factor 1 has 3 levels and Factor 2 has 2 levels. This results in 6 possible combinations.

A "main effect" is the effect of one of the independent variables on the dependent variable, ignoring the effects of any other independent variables. There are 2 main effects tested: in this case comparing data across factor 1 (i.e., time) is known as the "**repeated measures**" factor while comparing differences between factor 2 (i.e., groups) is known as the "**between-subjects**" factor. I**nteraction** is where one factor influences the other factor.

The standard frequentist approach to ANOVA is to compare the variances between levels of a defined factor where the H_0 is that these variances are equal.

The Bayesian ANOVA compares the predictive performances of the possible competing models, i.e., how likely a set of data is under one model compared to another. In most cases, one model is the null model (H_0) suggesting that the data is purely random and the alternative model (H_1) that one or more of the factors have an effect. In the mixed factor analysis, multiple models are tested.

ASSUMPTIONS

Like all other analyses, mixed factor ANOVA makes a series of assumptions which should either be addressed in the research design or can the tested for.

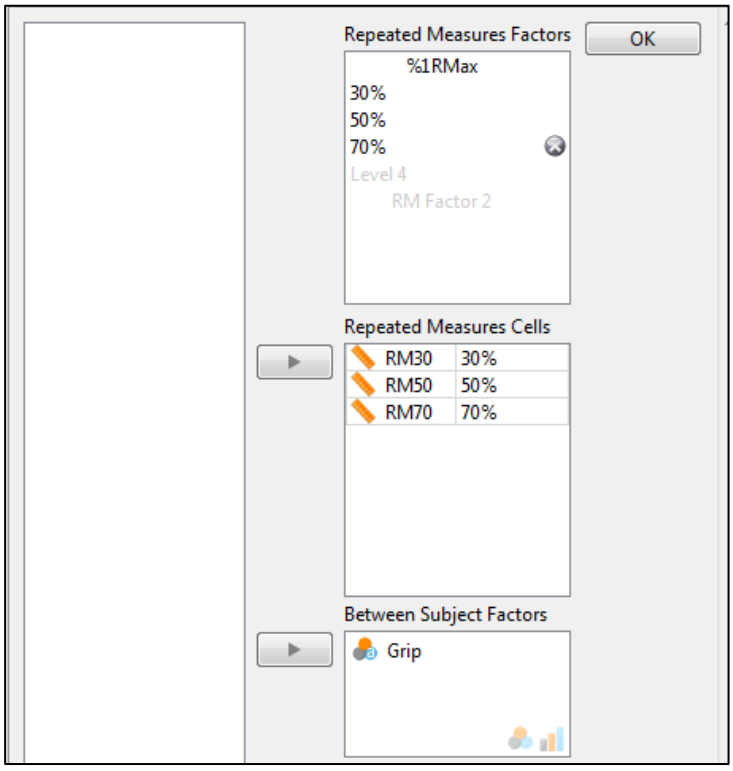
1. The "**Repeated measures**" factor should contain at least two related (repeated measures) categorical groups (levels).
2. The "**Between-subjects**" factor should have at least two categorical independent groups (levels).
3. The dependent variable should be continuous and approximately normally distributed for all combinations of factors.
4. There should be homogeneity of variance between the groups.
5. There should be no outliers.

RUNNING THE MIXED FACTOR BAYESIAN ANOVA

Open **Bayesian Mixed ANOVA.csv** in JASP. This contains 4 columns of data relating to the type of weightlifting grip and speed of the lift at 3 different loads (%1RM) for deadlifting. Column 1 contains the grip type, columns 2-4 contain the 3 repeated measures (30, 50 and70%). Check for outliers using boxplots.

NOTE: When running the analysis using the included dataset the results are always likely to be very slightly different to the ones in this chapter. This is because the analyses are based on numerical algorithms like Markov chain Monte Carlo (MCMC) which reports an error percentage. The higher the error percentage the higher the fluctuation of the results.

Go to ANOVA > Bayesian Repeated measures ANOVA. Define the Repeated Measures Factor, %1RMax, and add 3 levels (30, 50 and 70%). Add the appropriate variable to the Repeated measures Cells and add Grip to the Between-Subjects Factors:



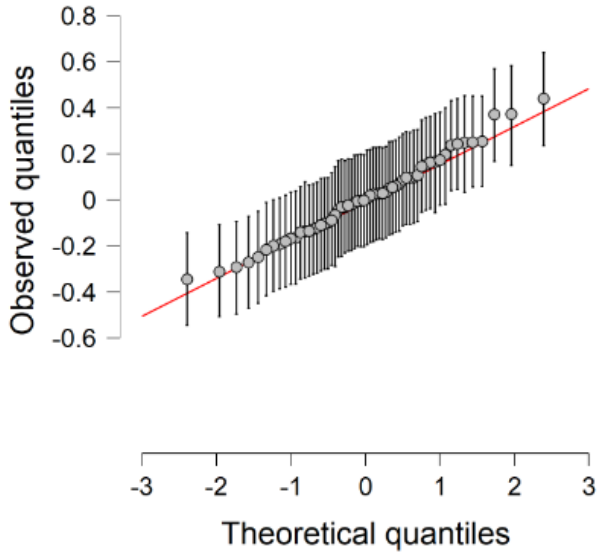
Then select the following options:

- ✓ Bayes Factor – BF_{10}
- ✓ Order – compare to best model
- ✓ Tables – Effects – Across all models
- ✓ Q-Q plots of residuals.
- ✓ Descriptives

In Descriptive plots move %1Rmax to the horizontal axis and Grip to 'Separate lines'
The output should initially comprise of 4 tables and 3 graphs.

UNDERSTANDING THE OUTPUT

Firstly, it is important to test the assumption of normality, in this case, that the residuals are normally distributed. This can easily be done by looking at the Q-Q plot.



If the residuals are normally distributed, they should lie consistently along the diagonal line. Any obvious deviations along the line would suggest that the assumption of normality has been violated. The assumption of homogeneity of variance can be assessed using Levene's test which is calculated as part of the classical ANOVA analysis.

Comparison of the competing models – Best model

The first column lists all models determined: four alternative models and one null model. The models

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
%1RMax + Grip + %1RMax ✻ Grip	0.200	0.997	1265.774	1.000	
%1RMax + Grip	0.200	0.003	0.012	0.003	2.680
%1RMax	0.200	6.193e −5	2.477e −4	6.212e −5	1.838
Grip	0.200	4.660e −17	1.864e −16	4.675e −17	1.893
Null model (incl. subject)	0.200	2.057e −17	8.228e −17	2.064e −17	1.700

Note. All models include subject

are ordered by their predictive performance relative to the best model in this case.

In the other columns, results are presented for:

P(M): for the ANOVA, the analysis sets the prior probabilities of each of the five models to be equal (i.e., 0.2).

P(M |data): shows the updated probabilities having now seen the data.

BFM: shows how much the data have changed the prior model odds

BF_{10}: shows the Bayes comparison with the best model; for the first row, it is always 1 since it is being compared to itself.

H(1):%1RM + Grip + %1RM*Grip
A model based on the alternative hypothesis that lift speed depends on %1RM, grip type and the interaction between these two factors. This is the best model and has a BF_{10}=1 since it is being compared to itself.

H(1):%1RM + Grip
A model based on the alternative hypothesis that lift speed depends on %1RM and grip type. This a BF_{10} of 0.003 or a BF_{01} of 322, suggesting that the data are 322 times more likely under the best model than under the model with main effects only.

H(1): %1RM, H(1): grip
Models based on the alternative hypothesis that lift speed depends on either %1RM or grip alone have extremely small BF_{10} values, as does the null model.

Comparison of the competing models – Null model

Alternatively, the data can be compared to the null model rather than the best model. In the options change the order to 'compare to the null model'. The model comparison has tested 5 models and compares the alternative models to the null model (H_0) which that states lift speed is not dependent on any other factors.

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	2.057e −17	8.228e −17	1.000	
%1RMax + Grip + %1RMax ✻ Grip	0.200	0.997	1265.774	4.846e +16	1.700
%1RMax + Grip	0.200	0.003	0.012	1.501e +14	2.071
%1RMax	0.200	6.193e −5	2.477e −4	3.011e +12	0.699
Grip	0.200	4.660e −17	1.864e −16	2.266	0.833

Note. All models include subject

H(1): grip
A model based on the alternative hypothesis that lift speed depends on grip type alone. This has a very small Bayes factor of 2.26 suggesting that there is very little evidence for this model, compared to the null model.

H(1): %1RM
A model based on the alternative hypothesis that lift speed depends on %1RM alone. This has an extremely large BF_{10} (i.e., $3.01*10^{12}$), decisively supporting this model over the null model.

H(1):%1RM + Grip

A model based on the alternative hypothesis that lift speed depends on %1RM and grip type. This also has an extremely large BF_{10} (i.e., $1.5*10^{14}$), decisively supporting this model over the null model.

H(1):%1RM + Grip + %1RM*Grip

A model based on the alternative hypothesis that lift speed depends on %1RM, grip type and the interaction between these two factors. This is the best model and has the largest BF_{10} (i.e., $4.86*10^{16}$), against the null model.

In order to compare the %1RM + Grip model against the %1RM + Grip + %1RM*Grip model, one can divide out the null hypothesis by computing $4.86*10^{16}$)/ $1.5*10^{14}$) = 324, which should give (approximately, due to rounding) the same result as the earlier 'compare to best model' analysis (i.e., BF = 322).

Whether one wants to compare to either the best or the null models is a matter of personal choice, the result is effectively the same.

Analysis of effects

This table shows the prior and posterior inclusion probability and the inclusion Bayes factor for each of the model's predictors. These data are based on all the models simultaneously.
%1Rmax and grip are considered as the main effects and the %1Rmax*Grip the interaction.

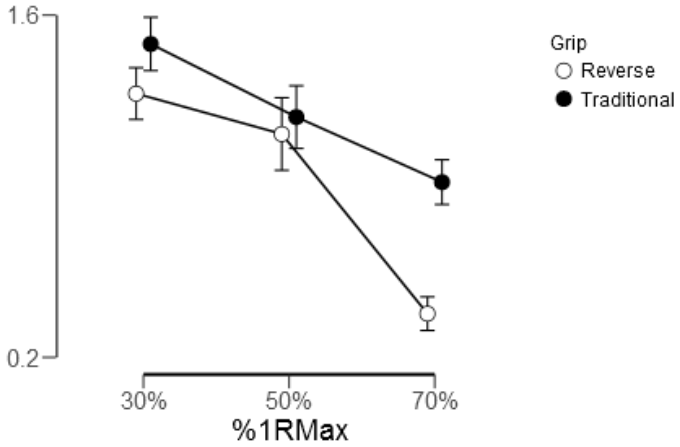
Analysis of Effects

Effects	P(incl)	P(incl|data)	BF_{incl}
%1RMax	0.600	1.000	∞
Grip	0.600	1.000	10764.591
%1RMax ✻ Grip	0.200	0.997	1265.774

The data suggests that there is infinite evidence for the inclusion of %1Rmax than a model without this predictor. (it is 'infinite' because of the computer's limited ability to present very small or very large numbers,). There is also decisive evidence for the inclusion of Grip and the interaction as predictors.
Descriptive data and plots are shown below.

Descriptives

%1RMax	Grip	Mean	SD	N
30%	Reverse	1.279	0.178	10.000
	Traditional	1.482	0.217	10.000
50%	Reverse	1.114	0.198	10.000
	Traditional	1.183	0.256	10.000
70%	Reverse	0.379	0.105	10.000
	Traditional	0.917	0.086	10.000



If the evidence suggested that the data is best predicted by the null model or that the evidence for the alternative was inconsequential. Although evidence for a lack of an effect is still information – there is no point in following up with further analyses.

POST HOC TESTING

If the ANOVA yields meaningful predictors (i.e., models outperforming the null model), post hoc testing can now be carried out. In Post Hoc Tests add %1RM to the analysis box on the right. Bayesian post hoc testing is based on pairwise comparisons using Bayesian t-tests. As in frequentist analyses, multiple t-tests will increase familywise error. In JASP, methods are used to correct for multiplicity based on adjusting the prior odds.

In the analysis options, now:

- ✓ Plots – Model averaged posteriors – Group levels in a single plot

Add %1Rmax and Grip to the right in 'Post Hoc tests'. Select Null control**.**

Post Hoc Comparisons - %1RMax

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
30%	50%	0.587	20.362	34.664	3.050e −4
	70%	0.587	5.135e +8	8.742e +8	5.573e −14
50%	70%	0.587	6918.088	11777.453	2.806e −8

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holdsacross all comparisons (Westfall, Johnson, & Utts, 1997). Individualcomparisons are based on the default t-test with a Cauchy (0, r =1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

The adjusted posterior odds show that there is strong evidence for a difference between 30% and 50% %1Rmax whereas there is decisive evidence for differences between 30 and 70% as well as 50 and 70%1Rmax.

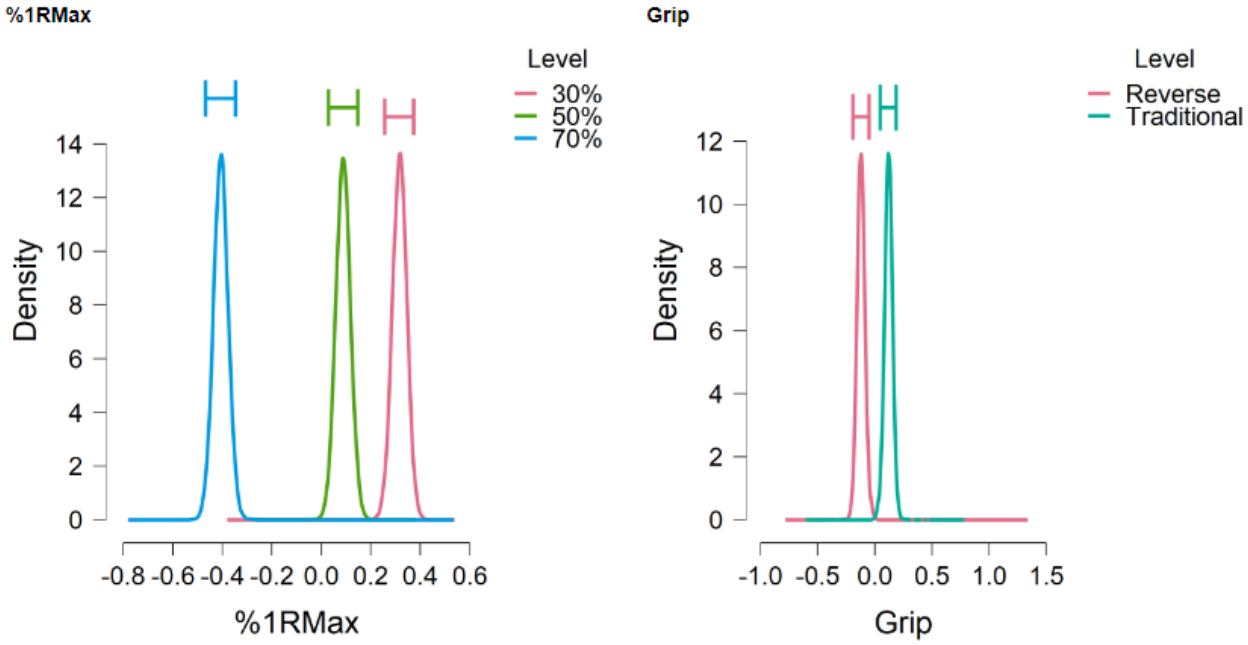
Post Hoc Comparisons - Grip

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
Reverse	Traditional	1.000	6.541	6.541	0.002

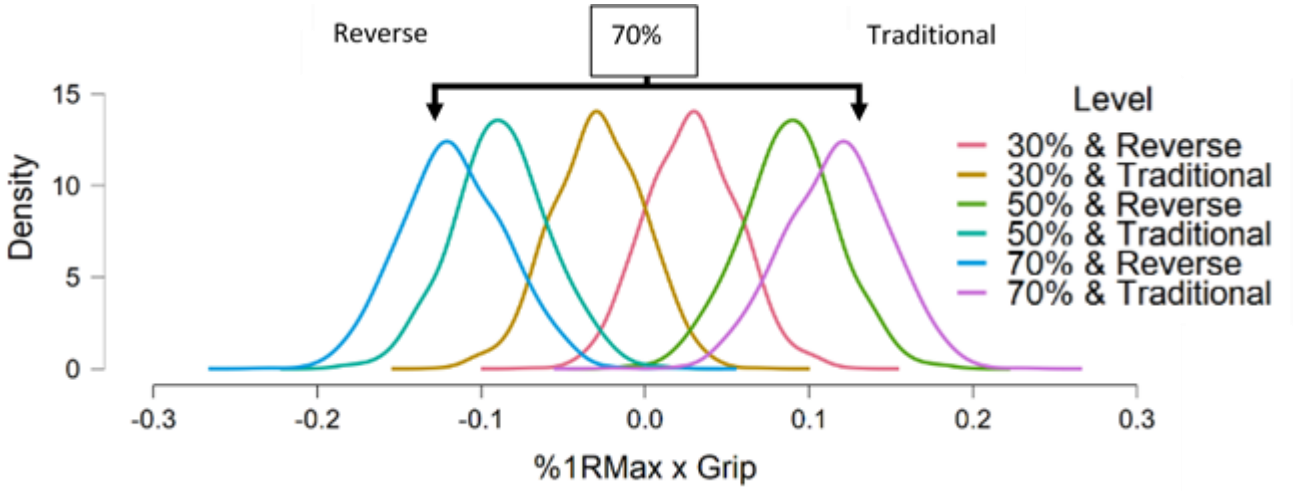
Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holdsacross all comparisons (Westfall, Johnson, & Utts, 1997). Individualcomparisons are based on the default t-test with a Cauchy (0, r =1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

There is also moderate evidence for a difference between reverse and traditional grips $BF_{10} = 6.54$.

The model average posterior distributions for the main effects are shown below. There is a clear separation between the %1Rmax levels with 30% having the highest lift velocity and 70% the lowest. For grip, the two distributions are closer but still separate without overlapping credible intervals, with the traditional grip exhibiting higher lift velocities than the reverse grip.



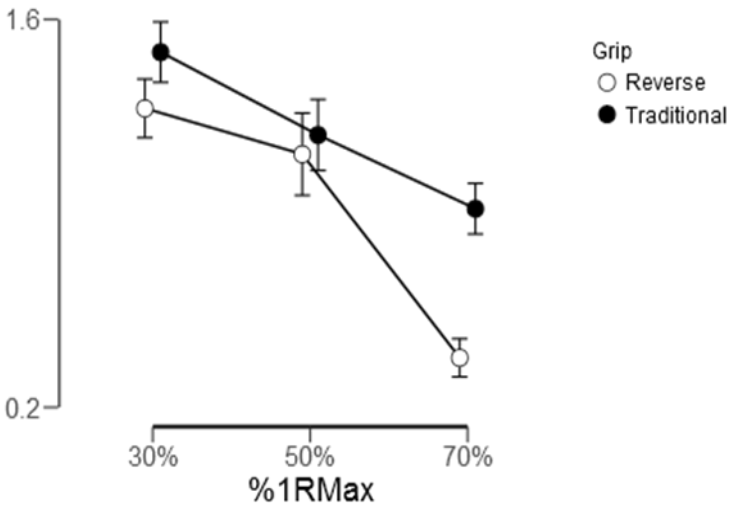
The model-averaged posterior distributions for the interactions are shown below. As can be seen, the largest separation is between 70% traditional and reverse lifts.



REPORTING THE RESULTS

This study determined the velocity of deadlifts using two different grips and 3 loads based on %1Rmax. Examination of the Q-Q plots suggested that the assumption of normality was not violated. A Bayesian mixed factor ANOVA determined that the data were best represented by a model that included both main factors, grip and load, and the grip*load interaction. The Bayes factor (BF_{10}) was $4.86*10^{16}$, indicating decisive evidence in favour of this model when compared to the null model. The BF_{10} in favour of indicating the interaction effect (on top of the two main effects) equalled 322.

Post hoc comparisons (Bayesian t-tests controlled for multiplicity) were subsequently performed. For the load, the adjusted posteriors show that there is strong evidence for a difference between 30% and 50% %1Rmax (20.6) whereas there is decisive evidence for differences between 30 and 70% as well as 50 and 70%1Rmax ($5.1*10^8$ and 6918 respectively).



BAYESIAN CONTINGENCY TABLES

This is the equivalent of the frequentist chi-square (χ^2) test for independence which can be used to determine if an association exists between two or more categorical variables. The test produces a contingency table, which displays the cross-grouping of the categorical variables.

The test compares two hypotheses:

H_0: that the categorical variables are independent of each other.

H_1: that the categorical variables are in some way dependent on each other.

The analysis requires two assumptions to be met:

- The two variables must be categorical data (nominal or ordinal)
- Each variable should comprise two or more independent categorical groups

There are 4 methods for determining the Bayes factors based on the sampling plan of the research design. Consider a researcher wants to collect data on tennis players referred to a physiotherapist for ankle injuries and is interested to see if there is a link between the player's gender and whether they had had a previous ankle injury.

- **Poisson sampling:**

The sampling scheme is to collect data for a six-month period. There is, therefore, no restriction on the cell counts, the cell and grand total counts will be random. Each cell count will have a Poisson distribution.

- **Joint multinomial sampling:**

In this case, data will only be collected for the first 100 players referred to the physiotherapist. This is like the Poisson scheme except that the grand total is now fixed.

- **Independent multinomial sampling**

In this case, data will be collected from 50 male and 50 female players. Therefore, either the rows or columns are fixed and therefore multinomially distributed.

- **Hypergeometric sampling**

Such a sampling system is rarely applied. In this case, data is collected such that BOTH columns AND rows are fixed. This can also be used when two continuous variables are split by their median values i.e. median split on age (old-young) and height (small-tall).

When running the Bayesian contingency table analysis, it is important that the correct sampling scheme is selected in the options.

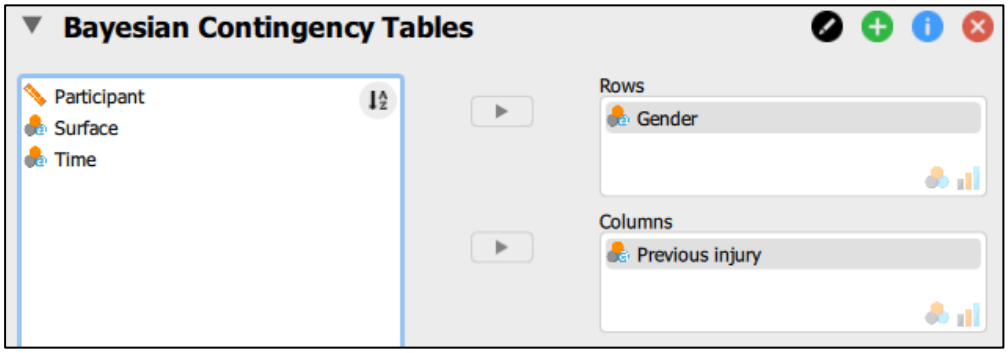
RUNNING THE ANALYSIS

Open Bayesian contingency.csv in JASP. This spreadsheet has data from 85 recreational tennis players referred to a physiotherapist practice with ankle injuries over a 6-month period. There are five columns of data:

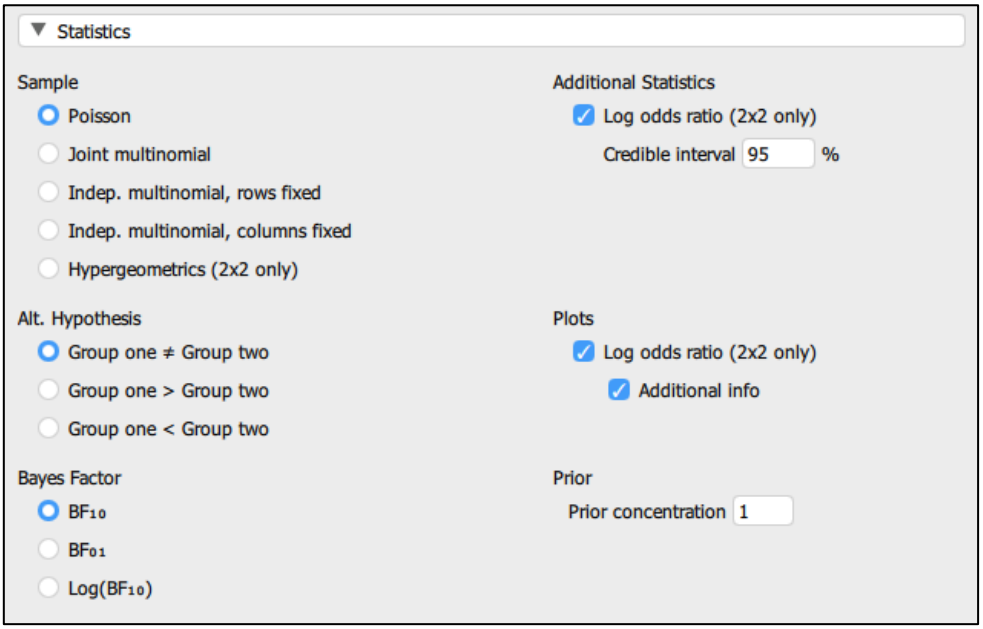
1. Subject ID
2. Gender
3. Type of playing surface
4. Time of day
5. Previous history of an ankle injury

Go to Frequencies > Bayesian Contingency tables. Is there an association between gender and the history of a previous ankle injury? By convention, the independent variable is usually placed in the contingency table columns and the dependent variable is placed in the rows.

Move gender to Rows and previous injury to Columns.



In Statistics, select the following options, noting that the sampling scheme used in this study was Poisson sampling



UNDERSTANDING THE OUTPUT

The output should comprise three tables and one figure. The contingency table shows the counts for each cell as well as the row and column totals. It can be seen that 33% of females had a history of a previous ankle injury while for the males it was approximately 25%

Contingency Tables

	Previous injury		
Gender	No	Yes	Total
Female	22	11	33
Male	41	11	52
Total	63	22	85

The Bayesian tests report the Bayes factor in support of the alternative hypothesis, where BF_{10} Poisson = 0.961 (the BF in support of the null hypothesis can be shown by selecting BF_{01} in the Statistics options and is 1.04). Therefore, there is no evidence supporting either of the hypotheses and the test is inconclusive.

In the other table the median log odds ratio and its calculated credible intervals. This works out as females being only 1.85 times more likely to have had a previous ankle injury compared to males.

Bayesian Contingency Tables Tests

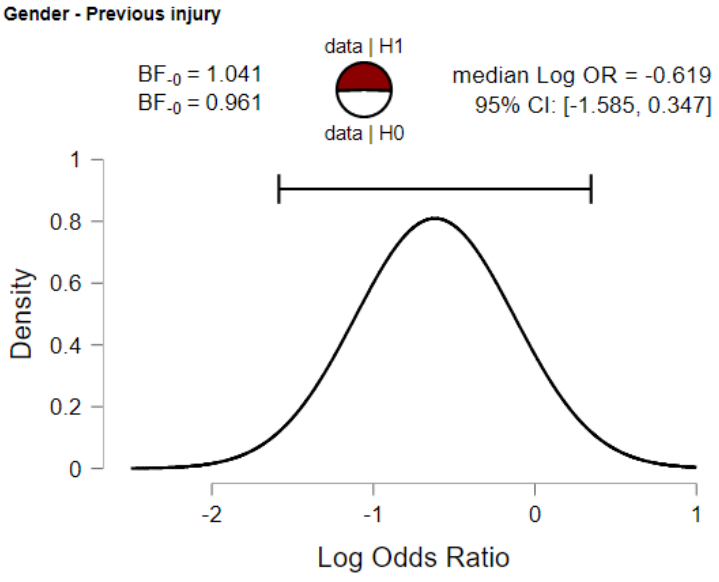
	Value
BF_{10} Poisson	0.961
N	85

Note. For all tests, the alternative hypothesis specifies that group *Female* is not equal to group *Male*.

Log Odds Ratio

	95% Credible Interval	
Log Odds Ratio	Lower	Upper
−0.619	−1.575	0.336

The Bayes factors and odds ratios are graphically visualised in the Log Odds Ratio plots.



THE BAIN MODULE

In frequentist statistics when performing null hypothesis significance testing (NHST) there is only a dichotomous decision for either rejecting or not rejecting the null hypothesis (H_0) based on the observed data. This does not allow for direct support for H_0 only that there is not enough evidence to reject it. The evidence in favour of H_0 itself cannot be quantified.

The null hypothesis is usually stated as H_0: the effect = 0

While the alternative hypothesis is H_1: ≠ H_0

The effect in question could be a correlation or difference between means.

For example, when comparing the means of two groups

H_0: the effect = 0 mean of group 1 = mean of group 2

H_1: H_1: ≠ H_0 mean of group 1 ≠ mean of group 2

When comparing the means of three groups

H_0: the effect = 0 mean of group 1 = mean of group 2 = mean group 3

H_1: H_1: ≠ H_0 differences between the groups now explicitly exclude H_0, i.e. the three group means are not equal to each other.

These H_1 alternatives are considered to be ***unconstrained*** and are denoted in JASP as **Hu**.

BAIN[7] is an abbreviation for **BA**yesian **IN**formative hypothesis evaluation. This uses the Bayes factor to evaluate the evidence for both the H_0 and multiple alternative hypotheses without having to account for multiple testing.

When null and alternative hypotheses are evaluated using the Bayes factor, all have equal standing, i.e. neither has the role of the traditional null or alternative hypotheses, they are simply different hypotheses. The probability of observing the data is computed given each hypothesis and translated into the Bayes factor from which the best hypothesis is selected.

BAIN allows alternative hypotheses by offering or entering model constraints. In a t-test, for example, the four possible hypotheses could be:

H_0: the effect = 0 mean of group 1 = mean of group 2

Hu: Hu: ≠ H_0 mean of group 1 ≠ mean of group 2

H_1: mean of group 1 > mean of group 2

H2: mean of group 1 < mean of group 2

[7] Hoijtink H et al (2019). A tutorial on testing hypotheses using the Bayes factor. Psychological Methods, 24, 539-556. DOI: 10.1037/met0000201

Independent t-test example

For comparisons of hypotheses, JASP uses Welch's t-test which does not assume that the variance of the dependent variable is the same in both groups. Open JASP and go to the **+** icon at the top right and tick the BAIN module. This will now add BAIN to the top menu.

Open Independent t-test.csv, click on BAIN 

and then select Welch's t-test. Add Weight gain to the dependent variable and Diet to the grouping variable. In the main options select:

Hypothesis test as Equal vs. not equal (i.e. H_0: the effect = 0 and unconstrained H_1: ≠ H_0). The analysis assigns equal prior probabilities for each hypothesis (0.5:0.5).

- ✓ Bayes factor: BF_{10}
- ✓ Other vs. equal
- ✓ Tables – Descriptives
- ✓ Plots – posterior probabilities and descriptive plots

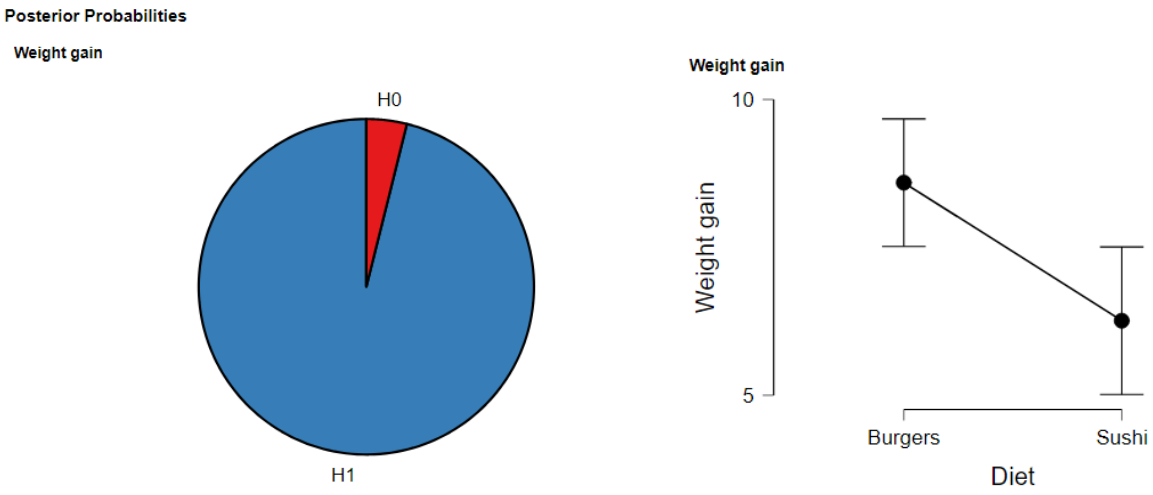
This should result in two tables and two plots.

Bain Independent Samples Welch's T-Test

	Hypothesis	BF	Posterior probability
Weight gain	H0: Equal		0.039
	H1: Not equal	24.703	0.961

Note. The alternative hypothesis H1 specifies that the mean of group 1 is unequal to the mean of group 2. The posterior probabilities are based on equal prior probabilities.

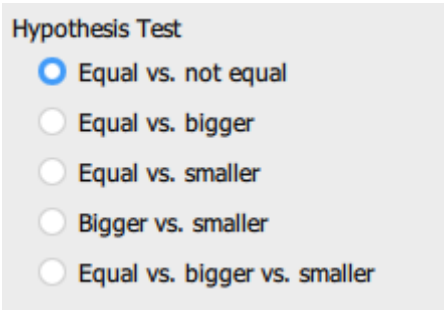
This table shows the evidence in support of the two competing hypotheses. This provides strong evidence in favour of the unconstrained alternative hypothesis which has a posterior probability of 96.1% compared to 3.9% for the null hypothesis. The Bayes factor is therefore 24.7 (0.961/0.039).



The posterior probabilities are also visualised on a pizza plot. The descriptive statistics and plot show that weight gain is higher on a burger diet.

Adding constrained alternative hypotheses

Return to the main analysis options where JASP offers a series of unconstrained and constrained models. Now select Equal vs, bigger vs. smaller.



Now three hypotheses have been tested:

H_0: mean of group 1 = mean of group 2

H_1: mean of group 1 > mean of group 2

H2: mean of group 1 < mean of group 2

By selecting BF_{10}, the alternative hypotheses are compared to the null hypothesis. Having seen the data there is strong evidence (BF_{10} = 49.4) in favour of H_1 with a posterior probability of 97.9% compared to 2% for H_0 and 0.1% for H2.

Bain Independent Samples Welch's T-Test

	Hypothesis	BF	Posterior probability
Weight gain	H0: Equal		0.020
	H1: Bigger	49.359	0.979
	H2: Smaller	0.047	0.001

Note. The null hypothesis H0 (equal group means) is tested against H1 (first mean larger than second mean) and H2 (first mean smaller than second mean). The posterior probabilities are based on equal prior probabilities.

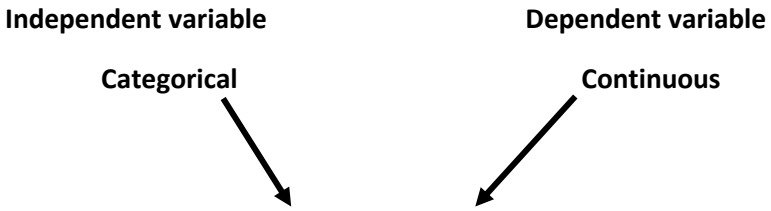
When comparing H2 with H_0, the BF_{10} = 0.047 or BF_{01} = 21.1 (1/0.047). If both hypotheses are deemed equally likely a priori, this means that the null hypothesis is now 21 times more likely than group 1 being smaller than group 2.

EXPERIMENTAL DESIGN AND DATA LAYOUT IN EXCEL FOR JASP IMPORT.

Independent t-test

Design example:

Independent variable	**Group 1**	**Group 2**
Dependent variable	Data	Data



	A	B
1	Group	Data
2	1	0
3	1	0
4	1	3.8
5	1	6
6	1	0.7
7	1	2.9
8	1	2.8
9	1	2
10	1	2
11	1	8.5
12	1	1.9
13	1	3.1
14	1	1.5
15	1	3
16	1	3.6
17	1	0.9
18	1	-2.1
19	2	2
20	2	1.7
21	2	4.3
22	2	7
23	2	0.6
24	2	2.7
25	2	3.6

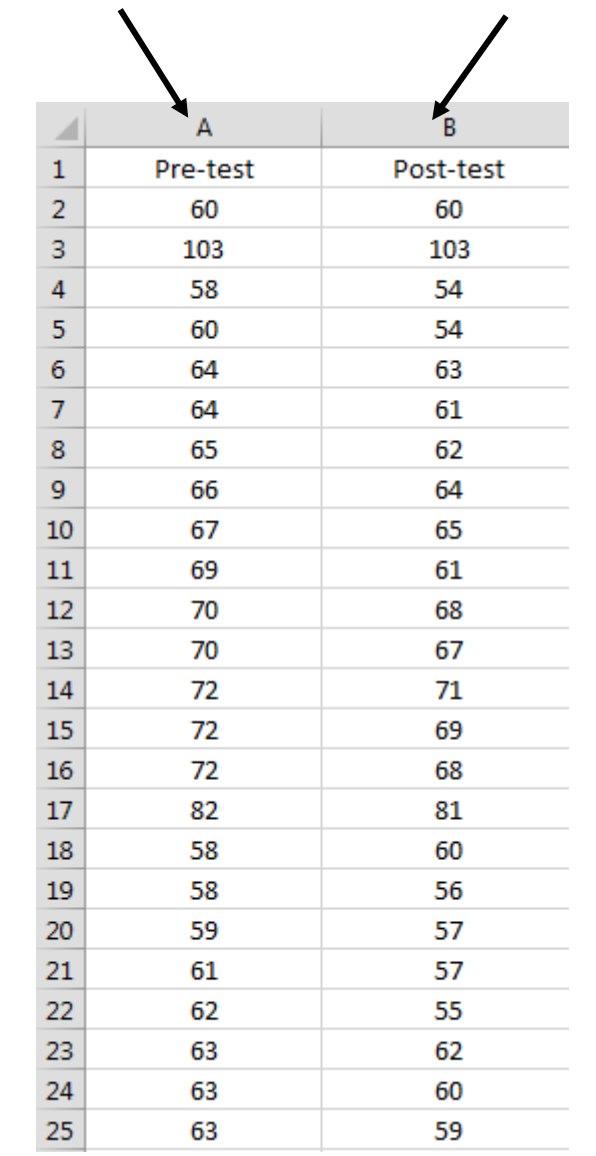
More dependent variables can be added if required

Paired samples t-test

Design example:

Independent variable	**Pre-test**	**Post-test**
Participant	**Dependent variable**	
1	Data	Data
2	Data	Data
3	Data	Data
..n	Data	Data

Pre-test **Post-test**



	A	B
1	Pre-test	Post-test
2	60	60
3	103	103
4	58	54
5	60	54
6	64	63
7	64	61
8	65	62
9	66	64
10	67	65
11	69	61
12	70	68
13	70	67
14	72	71
15	72	69
16	72	68
17	82	81
18	58	60
19	58	56
20	59	57
21	61	57
22	62	55
23	63	62
24	63	60
25	63	59

Correlation

Design example:

Simple correlation

Participant	Variable 1	Variable 2	Variable 3	Variable 4	Variable ..n
1	Data	Data	Data	Data	Data
2	Data	Data	Data	Data	Data
3	Data	Data	Data	Data	Data
…n	Data	Data	Data	Data	Data

Multiple correlation

	A	B	C	D	E	F
1	Participant	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
2	1	533	77	77	106	106
3	2	472	63	59	92	93
4	3	484	82	77	93	78
5	4	536	72	72	103	93
6	5	630	77	68	104	93
7	6	563	68	68	101	87
8	7	531	77	82	108	106
9	8	344	50	50	86	92
10	9	346	54	50	90	86
11	10	386	59	54	85	80
12	11	460	54	63	89	83
13	12	492	63	59	92	94

Regression.

Design example:

Simple Regression

Participant	Outcome	Predictor 1	Predictor 2	Predictor 3	Predictor ..n
1	Data	Data	Data	Data	Data
2	Data	Data	Data	Data	Data
3	Data	Data	Data	Data	Data
…n	Data	Data	Data	Data	Data

Multiple regression

	A	B	C	D	E	F
1	Participant	Outcome	Predictor 1	Predictor 2	Predictor 3	Predictor 4
2	1	533	77	77	106	106
3	2	472	63	59	92	93
4	3	484	82	77	93	78
5	4	536	72	72	103	93
6	5	630	77	68	104	93
7	6	563	68	68	101	87
8	7	531	77	82	108	106
9	8	344	50	50	86	92
10	9	346	54	50	90	86
11	10	386	59	54	85	80
12	11	460	54	63	89	83
13	12	492	63	59	92	94

More factors and covariates can be added if required

One-way Independent ANOVA

Design example:

Independent variable	**Group 1**	**Group 2**	**Group 3**	**Group…n**
Dependent variable	Data	Data	Data	Data

Independent variable **Dependent variable**

(Categorical) **(Continuous)**

	A	B
1	Group	Dependent variable
2	Group 1	3.8
3	Group 1	6
4	Group 1	0.7
5	Group 1	2.9
6	Group 1	2.8
7	Group 1	2
8	Group 1	2
9	Group 1	3.5
10	Group 2	1.9
11	Group 2	3.1
12	Group 2	1.5
13	Group 2	3
14	Group 2	3.6
15	Group 2	0.9
16	Group 2	-0.6
17	Group 3	1.1
18	Group 3	4.5
19	Group 3	6.1
20	Group 3	5
21	Group 3	2.4
22	Group 3	3.9
23	Group 3	3.5
24	Group 3	5.1
25	Group 3	3.5

More dependent variables can be added if required

One-way repeated measures ANOVA

Design example:

	Independent variable (Factor)			
Participant	**Level 1**	**Level 2**	**Level 3**	**Level..n**
1	Data	Data	Data	Data
2	Data	Data	Data	Data
3	Data	Data	Data	Data
4	Data	Data	Data	Data
..n	Data	Data	Data	Data

Factor (time)

	A	B	C	D
1	Participant	Week 0	Week 3	Week 6
2	1	6.42	5.83	5.75
3	2	6.76	6.2	6.13
4	3	6.56	5.83	5.71
5	4	4.8	4.27	4.15
6	5	8.43	7.71	7.67
7	6	7.49	7.12	7.05
8	7	8.05	7.25	7.1
9	8	5.05	4.63	4.67
10	9	5.77	5.31	5.33
11	10	3.91	3.7	3.66
12	11	6.77	6.15	5.96
13	12	6.44	5.59	5.64
14	13	6.17	5.56	5.51
15	14	7.67	7.11	6.96
16	15	7.34	6.84	6.82
17	16	6.85	6.4	6.29
18	17	5.13	4.52	4.45
19	18	5.73	5.13	5.17

Levels

(Related groups)

More levels can be added if required

Two-way Independent ANOVA

Design example:

Factor 1	**Supplement 1**			**Supplement 2**		
Factor 2	**Dose 1**	**Dose 2**	**Dose 3**	**Dose 1**	**Dose 2**	**Dose 3**
Dependent variable	Data	Data	Data	Data	Data	Data

Factor 1 **Factor 2** **Dependent variable**

	A	B	C
1	supp	dose	len
2	OJ	1000	19.7
3	OJ	1000	23.3
4	OJ	1000	23.6
5	OJ	1000	26.4
6	OJ	1000	20
7	OJ	1000	25.2
8	OJ	1000	25.8
9	OJ	1000	21.2
10	OJ	1000	14.5
11	OJ	1000	27.3
12	OJ	2000	25.5
13	OJ	2000	26.4
14	OJ	2000	22.4
15	OJ	2000	24.5
16	OJ	2000	24.8
17	OJ	2000	30.9
18	OJ	2000	26.4
19	OJ	2000	27.3
20	OJ	2000	29.4
21	OJ	2000	23
22	VC	1000	16.5
23	VC	1000	16.5
24	VC	1000	15.2
25	VC	1000	17.3

More factors and dependent variables can be added if required

Two-way Repeated measures ANOVA

Design example:

Factor 1 **Interventions**	**Level 1** **i.e. intervention 1**			**Level 2** **i.e. intervention 2**		
Factor 2 **Time**	**Level 1** **i.e. time 1**	**Level 2** **i.e. time 2**	**Level 3** **i.e. time 3**	**Level 1** **i.e. time 1**	**Level 2** **i.e. time 2**	**Level 3** **i.e. time 3**
1	Data	Data	Data	Data	Data	Data
2	Data	Data	Data	Data	Data	Data
3	Data	Data	Data	Data	Data	Data
..n	Data	Data	Data	Data	Data	Data

Factor 1 levels 1-n

Factor 2 levels 1-n

	A	B	C	D	E
1	Subject	Factor 1 level 1	Factor 1 level 2	Factor 2 level 1	Factor 2 level 2
2	A	7.38	6.52	9.27	14.32
3	B	7.71	10.83	11.48	16.38
4	C	6.19	10.42	9.77	15.45
5	D	9.27	11.78	15.45	16.96
6	E	11.41	9.52	11.65	15.64
7	F	5.29	5.82	9.22	13.01
8	G	8.54	9.43	10.92	17.35
9	H	7.89	8.43	8.26	12.57
10	I	5.49	6.64	11.39	14.02
11	J	9.26	9.36	13.03	16.24
12	K	6.9	7.09	9.02	14.7
13	L	8.57	9.64	8.33	13.71

Two-way Mixed Factor ANOVA

Design example:

Factor 1 **(Between subjects)**	**Group 1**			**Group 2**		
Factor 2 levels **(Repeated measures)**	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 1**	**Trial 2**	**Trial 3**
1	Data	Data	Data	Data	Data	Data
2	Data	Data	Data	Data	Data	Data
3	Data	Data	Data	Data	Data	Data
..n	Data	Data	Data	Data	Data	Data

Factor 1

(Categorical)

Factor 2 levels

(Continuous)

	A	B	C	D
1	Group	Level 1	Level 2	Level 3
2	Group 1	1.31	0.9	0.9
3	Group 1	1.29	0.89	0.72
4	Group 1	1.8	0.9	0.96
5	Group 1	1.4	1.26	0.97
6	Group 1	1.49	1.18	0.88
7	Group 1	1.35	1.15	0.92
8	Group 1	1.45	1.19	1
9	Group 1	1.21	1.2	0.85
10	Group 1	1.79	1.48	0.99
11	Group 1	1.73	1.68	0.98
12	Group 2	1.55	0.9	0.55
13	Group 2	1.27	0.95	0.41
14	Group 2	1.53	0.87	0.42
15	Group 2	1.26	1.15	0.44
16	Group 2	1.14	1.12	0.38
17	Group 2	1.11	1.08	0.34
18	Group 2	1.1	1.0758	0.18
19	Group 2	1.08	1.18	0.24
20	Group 2	1.3	1.26	0.39
21	Group 2	1.45	1.55	0.44

Contingency tables

Design example:

Participant	Response 1	Response 2	Response 3	Response…n
1	Data	Data	Data	Data
2	Data	Data	Data	Data
3	Data	Data	Data	Data
..n	Data	Data	Data	Data

All data should be categorical

	A	B	C	D	E
1	Respondant	Response 1	Response 2	Response 3	Response 4
2	1	Female	clay	Morning	yes
3	2	Male	astro	Morning	No
4	3	Female	grass	Evening	No
5	4	Male	clay	Afternoon	No
6	5	Male	clay	Morning	No
7	6	Male	grass	Evening	No
8	7	Female	grass	Evening	yes
9	8	Male	clay	Morning	yes
10	9	Female	grass	Morning	No
11	10	Male	clay	Afternoon	No
12	11	Female	clay	Afternoon	No
13	12	Male	astro	Afternoon	No
14	13	Male	astro	Afternoon	No
15	14	Male	astro	Afternoon	yes
16	15	Female	clay	Morning	No
17	16	Male	astro	Afternoon	yes
18	17	Female	astro	Afternoon	yes
19	18	Male	grass	Morning	No
20	19	Male	clay	Afternoon	No