Method Sheet 27

How to perform a one way ANOVA statistical test

Overview

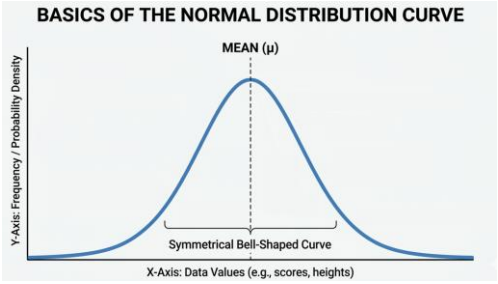
This method sheet explains how to perform a one way ANOVA statistical test using the Jamovi statistical analysis package. This type of test is helpful when analysing data from experiments where multiple different treatments, such as a control condition and several different natural compounds or extracts, were analysed in the same experiment. A one way ANOVA tests first whether any of the treatments differ significantly from each other. Post-hoc tests can then be added to the ANOVA results, yielding p-values for each pair-wise comparison between, for example, the effects of each extract or compound on the measured response, and the vehicle control.

How ANOVA overcomes the problem of multiple testing

One of the key problems in statistics is the danger of discovering false positive "significant" results through multiple testing. This can occur when you complete multiple statistical tests at the same time. For example, if you compared 20 different drug treatments to a control using 20 individual t-tests, it is likely that at least one would give a false positive "significant" p-value, even if none of the drugs have any real activity, since the chances of obtaining a p-value of less than 0.05 is one in twenty just by chance alone.

ANOVA overcomes this by evaluating the variance of the entire dataset in a single step. Instead of many individual comparisons, it initially tests whether any differences exist between the group means. If this initial test is not significant (i.e. if the p-value is greater than 0.05), you stop at that point, as it indicates there is no difference between the control substance and any of the tested compounds. If this test is significant ($p<0.05$), it tells you that at least some of the tested treatments give responses that are significantly different from the vehicle control. To find out which ones, and how significant these are, you must perform a post-hoc test - such as Tukey's or Dunnett's test. These post-hoc tests build on the earlier ANOVA analysis, automatically correcting for any impact of multiple testing, and tell you exactly which treatments and/or concentrations give significantly different results from the control condition and the p-values of these comparisons.

Testing for data normality in drug discovery data



ANOVAs are designed to analyse data that are **normally distributed**. That means, if you were to plot a curve showing how frequently each value appears in a large data set, it would appear as a bell shape curve. This curve is what we call the normal distribution (see image at right).

Although most forms of data from biological assays yield data that are normally distributed, we prefer not to assume that, and typically perform a statistical test, such as the **Shapiro-Wilk test**, to ensure the data are distributed in this way before choosing a statistical test. The problem with drug discovery data, and bioassay results generally, is that the number of independent experiments is typically small, perhaps three or four repeats of the same experiment (i.e. n=3 or 4).

This small number presents a difficulty for the Shapiro-Wilk test, as it can only work effectively with much larger 'n' values. A further problem common to drug discovery work is that a single screen can test hundreds or thousands of different compounds. Applying hundreds or thousands of normality tests would be both impractical and prone to type I errors. Therefore, when analysing screening data, we typically assume that the data are close enough to normal distribution for use in ANOVA on the basis of historical use of the assay and visual inspection to ensure there are no very large outliers within replicate measurements of the same treatments.

Statistical packages suitable for ANOVA testing

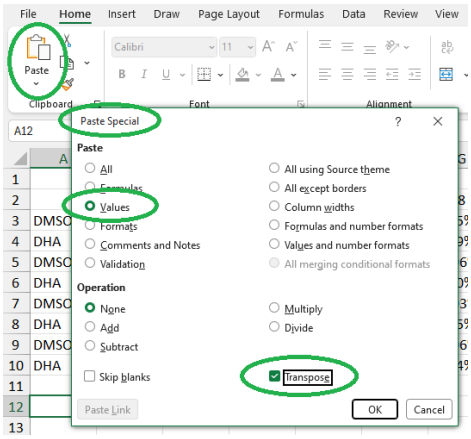
Microsoft Excel does not have the capability to perform ANOVA tests. Specialised statistical analysis software packages such as GraphPad PRISM and SPSS can be used to perform ANOVA, but require payment of a licence fee. Examples of software packages that are free to use and capable of performing ANOVA include R and Jamovi. As Jamovi uses a simple online web interface similar to commercial alternatives often used in industry, the method below explains how to perform a one way ANOVA using Jamovi.

Arranging your data for analysis in Jamovi

1) The arrangement of data in your Excel file is slightly different from that required by Jamovi for analysis by ANOVA.
2) First, you must bring together the results from each of the three or four separate experiments into the same table.
3) List the names of the different extracts or compounds used horizontally from the second cell of the top row, including the control condition (e.g. DMSO).
4) Then, in the first column of the new table, insert a heading to indicate the experiment numbers (e.g. Exp 1, Exp 2, etc.)
5) Next, paste as values only (not cell references) the results from each of the different experiments - your table should now look something like this:

	A	B	C	D	E	F	G
1							
2		DMSO	Extract 1	Extract 2	Extract 3	Extract 4	Extract 5
3	Exp 1	96%	26%	7%	6%	19%	19%
4	Exp 2	85%	22%	4%	0%	1%	13%
5	Exp 3	73%	16%	0%	1%	0%	0%
6	Exp 4	98%	22%	5%	3%	11%	11%

6) Now we have to rotate this table by 90° so that the orientation of the data are compatible with the Jamovi interface.
7) Highlight and copy the table, including all of the headings in the top row, but do not select the experiment number column, as we do not require this for the analysis.
8) Click on an empty cell below the selected table.
9) Click on 'Paste' then 'Paste Special' at the bottom of the menu options, and in the options box that appears, select both checkboxes for 'Values' and 'Transpose'



10) Your table of data values should now be rotated by 90º, as shown below:

8	DMSO	0.963	0.849	0.734	0.984
9	Extract 1	0.262	0.216	0.163	0.219
10	Extract 2	0.071	0.038	0.001	0.051
11	Extract 3	0.056	0.003	0.011	0.025
12	Extract 4	0.190	0.008	0.005	0.113
13	Extract 5	0.192	0.130	0.003	0.107

11) Format the cells so that they are in number format only, not percentages, and have at least three decimal places.

12) Prepare a new table of two columns below this one with two headings, ‘Treatment’ and ‘Response’

13) Copy all the treatment headings from the the first column of the previous table, then paste them 4 times (or 3 if only 3 experiments) one block below the other without any empty gaps so there is a single column of 4 (or 3) repeats of the same set of treatment names, starting just below the heading ‘Treatment’

14) Highlight then drag (or cut and paste) each column of result values from the previous table, and stack them below each other without any gaps into the second column of the new table just below the ‘Response’ heading.

15) Your new table should now have two columns - the first contains the names of the treatments, and the second all of the results, without any gaps.

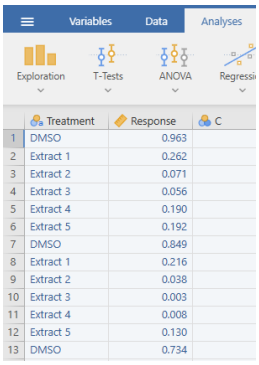
16) Ensure that every treatment name aligns correctly with the corresponding response in the neighbouring column - they must match up exactly or the test will not work.

17) Your data should now be in Jamovi format, looking something like the table shown at right:

16	Treatmen	Response
17	DMSO	0.963
18	Extract 1	0.262
19	Extract 2	0.071
20	Extract 3	0.056
21	Extract 4	0.190
22	Extract 5	0.192
23	DMSO	0.849
24	Extract 1	0.216
25	Extract 2	0.038
26	Extract 3	0.003
27	Extract 4	0.008

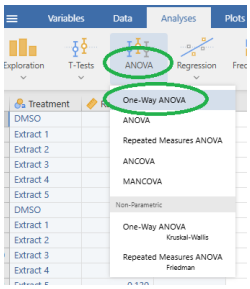
Using Jamovi for one way ANOVA

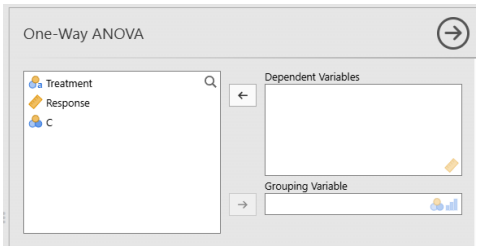
1) Jamovi is a free, open-source software application for statistical analysis, available at the following website: https://www.jamovi.org/
2) If you are able to download Jamovi to your own PC, it will run faster and be easier to use, but there is also a free to access cloud version available from the same website.
3) To use the cloud option, click on 'Jamovi Cloud' then 'Start' then choose a convenient log in method (e.g. with a Google account).
4) If you receive a message indicating that the server is busy, it is because too many users are accessing the free service presently, try again a little while later.
5) Highlight and copy your Excel data WITHOUT the headings.
6) Paste them from cell A1 of the Jamovi web sheet.
7) Double click on the heading letter 'A' of the first column and type 'Treatment' in the uppermost text box that appears.
8) Do the same for column 'B' and rename it with 'Response'
9) Click the up arrow in a circle to close the edit box.
10) Your data should now look something like this in Jamovi:

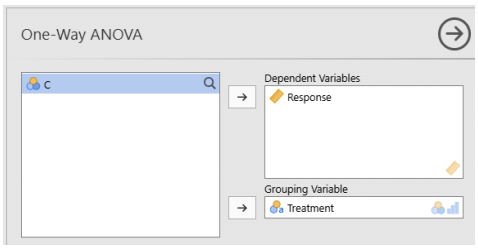


	Treatment	Response	C
1	DMSO	0.963	
2	Extract 1	0.262	
3	Extract 2	0.071	
4	Extract 3	0.056	
5	Extract 4	0.190	
6	Extract 5	0.192	
7	DMSO	0.849	
8	Extract 1	0.216	
9	Extract 2	0.038	
10	Extract 3	0.003	
11	Extract 4	0.008	
12	Extract 5	0.130	
13	DMSO	0.734	

11) Click on the large 'ANOVA' button in the top ribbon, then select the 'One way ANOVA' option.
12) A new dialogue box appears as shown below, use the mouse to click on the term 'Response' and drag it across to move it into the 'Dependent variable' box.
13) Move 'Treatment' into the 'Grouping variable' box, as shown below:







14) This tells Jamovi that the dependent variable (the response we are measuring) is present in the column with heading 'Response', and that we want to compare the means of the diffferent responses present in the 'Treatment' column.
15) As soon as you have done this, a table showing the results of the ANOVA appears, which should look something like this:

One-Way ANOVA

One-Way ANOVA (Welch's)

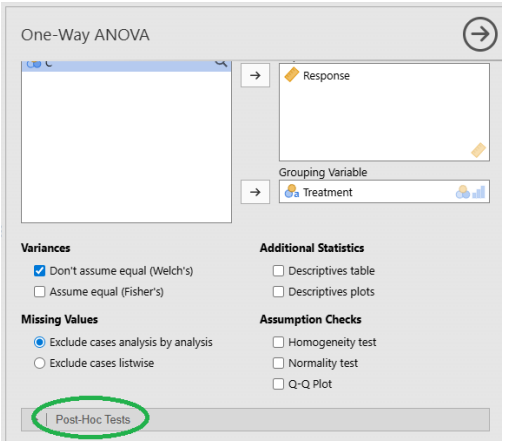
	F	df1	df2	p
Response	40.4	5	8.09	< .001

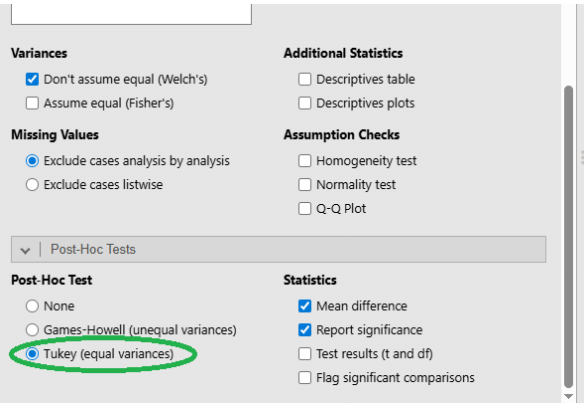
Interpreting the data from the initial ANOVA test

1) We have performed the ANOVA to answer the question of whether any of the treatments has a significant impact on the response in comparison with the control condition.
2) The right-most column of the table shows the p-value of the ANOVA test, in this example, the p-value is calculated to be less than 0.001.
3) We typically set the threshold for significance as being any p-value that is less than 0.05, which means we can be 95% sure that the difference seen is real and not just arising from chance.
4) If the p-value for the test is greater than 0.05, we would stop here and perform no further analysis, since that would indicate that none of the test substances had any inhibitory properties greater than the vehicle alone.
5) In this example, however, the p-value is well below the threshold, so we now want to ask an additional question, which treatments in particular gave a significant reduction in the response?
6) To answer this, we now have to extend the analysis with a post-hoc test.

Using Tukey's post-hoc test in Jamovi

1) In the ANOVA control pane, scroll down and click on 'Post hoc tests'
2) Select the check box for 'Tukey (equal variances)' to tell Jamovi which type of post-test to use.





3) A new table of results will appear on the right hand side of the screen.
4) This table contains far more information than you need for your dissertation - so you will have to focus only on those results, specifically the p-values, that matter the most.
5) In particular, you must focus on the comparison between the DMSO control (the negative control), and each of the different treatments, which may be different natural compounds or extracts of interest.
6) In this example, the appropriate p-values are given in the first row of the table, which shows the results of comparisons of DMSO against each of Extracts 1 to 5 (highlighted green in the example image below).

Post Hoc Tests

Tukey Post-Hoc Test – Response

		DMSO	Extract 1	Extract 2	Extract 3	Extract 4	Extract 5
DMSO	Mean difference	—	0.667	0.842	0.8587	0.8035	0.7760
	p-value	—	< .001	< .001	< .001	< .001	< .001
Extract 1	Mean difference		—	0.175	0.1913	0.1360	0.1085
	p-value		—	.030	.015	.129	.314
Extract 2	Mean difference			—	0.0165	−0.0388	−0.0663
	p-value			—	.999	.970	.779
Extract 3	Mean difference				—	−0.0553	−0.0827
	p-value				—	.880	.592
Extract 4	Mean difference					—	−0.0275
	p-value					—	.994
Extract 5	Mean difference						—
	p-value						—

7) In this example, all five of the comparisons against control give a very significant p-value (each is less than 0.001), which gives us strong confidence that all 5 extracts have a reliable and reproducible effect on the response (in this case, killing cancer cells).

8) You should now take a note of every p-value for these specific comparisons, since these will be necessary for your dissertation when you need to put the stars for significance on your chart of the effects of the different treatments.

Notes

- A much better post-hoc test to use in this type of analysis is Dunnett's post-hoc test, since this compares every treatment against the control, and ignores all the other possible comparisons, resulting in a much simpler and easier to read table of p-values.
- As Jamovi is unable to perform a Dunnett's test, we perform a Tukey test instead, but if you are able to perform ANOVA in another software package, you should choose the Dunnett's post-hoc test option.