

IASS Guide to ANOVA - Basic

After image data has been converted somehow into numerical data (e.g. the numbers of craters within an area, the number of fish etc) then it can be useful to make comparisons with other image data.

A common question may relate to whether or not there are actual differences in the numbers of something in two different areas, such as:

- Numbers of craters in the lunar highlands compared to the lunar lowlands
- Fish numbers in a fished area and a fishery protected area
- The number of barnacles on an exposed rock and a sheltered rock

The number of questions possible is related to the interests of the researcher.

It is a simple matter to work out the Mean, Median and Mode concentrations of these items, when counted in several different images. The Standard deviation can also be readily computed. These simple figures give some indication on whether or not there is really a 'significant difference' in numbers between the two locations. They do not however give any opinion as to whether or not the two sets of data are likely to be from different populations (i.e. from areas of the moon where crater numbers are indeed different, such as between the lunar highlands and lunar lowlands).

Say we measure the size of 5 snails in photos on the north and south side of a rock, and get the following measurements:

North side of rock	South side of rock
23 mm	40 mm
23 mm	41 mm
43 mm	54 mm
36 mm	56 mm
45 mm	76 mm

Now, by looking at the numbers, the snails on the South side are generally larger, though some snails on the North side are bigger than some on the South side. Are the two populations (north and south) distinct?

To determine the chances that this is the case, we can use an (Analysis of Variance) ANOVA test. This sort of test is having progressively more impact in biological sciences, and we will introduce a free online test here.

Running a simple ANOVA online

- 1) Please check the 'Resources' page of the IASS website to make sure that the test is still listed: (<http://imageanalysis.weebly.com/resources.html>). On that page there are some tutorial videos also outlining the procedures within this document, if you want to check them.

- Open the weblink to the Online ANOVA (<http://vassarstats.net/anova1u.html>)

One-Way Analysis of Variance for Independent or Correlated Samples

[Traducción en español]

The logic and computational details of the one-way ANOVA for independent and correlated samples are described in Chapters 13, 14, and 15 of [Concepts and Applications](#).

Procedure:

- Initial Setup:**

Enter the number of samples in your analysis (2, 3, 4, or 5) into the designated text field, then click the «Setup» button for either Independent Samples or Correlated Samples to indicate which version of the one-way ANOVA you wish to perform.

- Entering Data Directly into the Text Fields:**

After clicking the cursor into the scrollable text area for Sample 1, enter the values for that sample in sequence, pressing the carriage return key after each entry except the last. (On a Macintosh platform, the carriage return key is labeled "Return" on a "Windows

The above webpage should open, with some initial info about running the tests...

- Entering data is simple. For the example above the lower section of the webpage (The *Setup* and *Data entry*) sections should look like this:

Setup	
Number of samples in analysis = <input type="text" value="2"/>	
<input type="button" value="Independent Samples"/>	<input type="text" value="5"/>
<input type="button" value="Correlated Samples"/>	standard weighted-means analysis
<input type="button" value="Unweighted"/>	Click this button only if you wish to perform an unweighted-means analysis. Advice: do not perform an unweighted-means analysis unless you have a clear reason for doing so.
<input type="button" value="Weighted"/>	Click this button to return to a standard weighted-means analysis

Data Entry				
Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
23	40			
23	41			
43	54			
36	56			
45	76			
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

The 'Number of samples in analysis' should be set to two – representing 'North side' and 'South side' counts. The 'Independent samples' is the number of measurements taken on each side, in this case, 5 snails from each side.

- 4) After the data is entered as above, press 'Calculate'
- 5) This will produce a *Data Summary* and *Anova summary* at the bottom of the page, which should look like this:

<i>Data Summary</i>						
	Samples					
	1	2	3	4	5	Total
N	5	5				10
ΣX	170	267				437
Mean	34	53.4				43.7
ΣX^2	6228	15109				21337
Variance	112	212.8				248.9
Std.Dev.	10.583	14.5877				15.7766
Std.Err.	4.7329	6.5238				4.989

standard weighted-means analysis						
<i>ANOVA Summary</i> 5						
Source	SS	df	MS	F	P	
Treatment [between groups]	940.9	1	940.9	5.79	0.042758	
Error	1299.2	8	162.4			
Ss/Bl				Graph Maker		
Total	2240.1	9				

Ss/Bl = Subjects or Blocks depending on the design.
Applicable only to correlated-samples ANOVA.

Much of this table can be explained via Wikipedia or via further chapters of this online guide. For now, the most important number is the *P* value near the bottom right of the table. This value gives an indication as to whether or not the two measured populations are likely to be distinct from each other:

P value Percentage chance populations are distinct

0.1	90% chance of distinctness
0.05	95% chance of distinctness
0.001	99.9% chance of distinctness
<0.001	Greater than 99.9% chance of distinctness

So, the *P* value is not giving a **DEFINATE CONCLUSION** on whether or not the populations are different from one another. It is giving a percentage estimation based on the size of the samples, the spread in data and difference in averages of the compared sets.

For the data we tested, there is >95% probability that the populations are distinct. The test does not say whether the difference is related to the side of the rock, just that there is a difference. Care must be taken when making these tests to bear in mind what is actually being tested.

Comparing more than 2 sets of measurements

So, above we compared two sets of data. If we compare more, ANOVA will still only tell us whether **SOME** of the sets are different, not where these differences are. Here we add another example to that above:

North side of rock	South side of rock	Under rock
23 mm	40 mm	17 mm
23 mm	41 mm	26 mm
43 mm	54 mm	33 mm
36 mm	56 mm	35 mm
45 mm	76 mm	42 mm

- 1) The new column, for 'Under rock' is placed in the Sample 3 column of the online form and 'The number of samples in analysis' changed to 3.
- 2) Press 'calculate' and the following output is given:

Data Summary						
	Samples					
	1	2	3	4	5	Total
N	5	5	5			15
ΣX	170	267	153			590
Mean	34	53.4	30.6			39.3333
ΣX^2	6228	15109	5043			26380
Variance	112	212.8	90.3			226.6667
Std.Dev.	10.583	14.5877	9.5026			15.0555
Std.Err.	4.7329	6.5238	4.2497			3.8873

standard weighted-means analysis					
ANOVA Summary					
Source	SS	df	MS	F	P
Treatment [between groups]	1512.9333	2	756.4667	5.47	0.020489
Error	1660.4	12	138.3667		
Ss/Bl					Graph Maker
Total	3173.3333	14			

Ss/Bl = Subjects or Blocks depending on the design.
Applicable only to correlated-samples ANOVA.

- 3) So, the P value indicates here that there is ~98% chance that there is a difference in the slug lengths measured within at least two locations. But which locations?
- 4) Underneath the 'ANOVA summary' section of the output the following output is now presented:

Tukey HSD Test

HSD[.05]=19.86; HSD[.01]=26.55
M1 vs M2 nonsignificant
M1 vs M3 nonsignificant
M2 vs M3 P<.05

M1 = mean of Sample 1
M2 = mean of Sample 2
and so forth.

HSD = the absolute [unsigned] difference between any two sample means required for significance at the designated level. HSD[.05] for the .05 level; HSD[.01] for the .01 level.

[Print this Page](#)

This output is the 'Tukeys HSD Test', which will be explained in another .pdf on this website in the future, or can be looked up on Wikipedia. What the test basically does is compare all the columns in pairs with each other... M1 vs M2 , M1 vs M3 etc. Where there is a significant difference between the two columns, the difference is indicated.

In this current test, M2 vs M3 shows a $P < 0.05$ difference , meaning there is a 95% chance that the populations of snails under the stone are distinct from those on the South side of the stone. According to the Tukey test however , there is no indicated distinctiveness between the North and South snails.

---This result is NOT that indicated by the previous North / South ANOVA. This is because the HSD test is more conservative a test than ANOVA, and is more likely to err on the side of caution in drawing conclusions. We discussed earlier that there was some overlap in values between lengths of snails on the North and South of the rock, and to the HSD test, these render the distinction between these populations as not significant. The HSD test does however, despite some minor overlap, determine the under stone and South side populations to differ from each other.

Later .pdfs within this course will highlight these problems / concerns. The most important factor to consider when making these comparisons in research work is to always report what has been done and take care when considering 'significance'.