

Analyzing and Interpreting Data

Linguistics 390a
Florian Schwarz

1 Handling Your Data

Once your questionnaires have been filled out by participants, you should enter the results in a spreadsheet program like Excel.

You'll want to be able to compute averages for each condition. For statistical analyses, you will also need the averages for each subject in each condition (and each item) [i.e., what was the average score for subject 1 in condition A, etc.]. It's best if you enter your data in a table like the following:

Subject	Sentence	Condition	Rating Score
1	1	A	3
1	2	B	2
2	1	A	5
2	2	B	4
3	1	A	2
3	2	B	4
...			

However you choose to organize your data, you should make sure that you know for each data point which subject, which sentence, and which condition it belongs to!

In Excel, it's easy to automatically compute averages with the 'AVERAGE' function. Just enter '=average(COLUMNROW_{firstcell}:COLUMNROW_{lastcell})' in a cell, and it will display the average for the range of cells specified (you may need to sort the data by condition to easily compute condition averages).

The best way to get averages for subjects and items is by using Excel's Pivot-table function. (See <http://people.umass.edu/florian/excel.pdf> for a short introduction on pivot tables)

What if you're asking a yes/no question (or another binary question)?

For entering the scores, just record one of the answers (say, 'yes'), as '1's, and the other as '0's. When you use the average function for such data, you'll automatically get the percentage values (with 1 = 100%).

Graphing the data

Once you have entered the data, you may want to produce a graph to get a visual impression of the data. The mean is important, but one and the same mean can result from very different distributions, so it's helpful to see whether there is a clear pattern in the graph.

Good graphs to use are histograms and scatter plots.

2 Analyzing your data & Interpreting results

Let's say that in your experiment, condition A answers were judged to be cooperative 70% of the time, and condition B answers 30% of the time. For a pilot, that's a pretty good result. But you really want to know whether that is just something that happened by chance, or reflects reality.

Statistical tests can tell you how likely it is that the difference you found corresponds to an actual difference.

You can use a T-test in Excel to compare two conditions in this way. (If you end up having enough data for your project to do that, talk to me on the steps you need to take for this!)

If you confirm a statistical difference in a full study, you can go on to think about the theoretical implications of your hypothesis. This will most likely lead to new questions and hypotheses that you can test experimentally.

'No' means 'you don't know'

Two conditions will hardly ever be exactly the same in your data. But what if the difference is very small, and you don't find a statistically significant difference?

Can you conclude that there is no difference? NO!!!

Not finding a difference is a **null result** – it does not mean at all that there actually is no difference between the conditions!

You don't know any more than you did before (except that your study design and the amount of data you have collected are not enough to show that there is a difference)