

Lecture 22 Worksheet: Doing Statistical Analysis

We will start with everybody working with the same data set pulled fresh from the Social Intelligence Test on LabintheWild.org. The data set captures data from more than 50,000 participants. It should be fun to work with.

0. Get into pairs

If you launched your own visual search experiment, get together with your team. If there are three of you, you can all work together. If there are 4 of you, break up into two pairs.

Otherwise, pair up however you want. You will need at least one laptop per team, ideally with Excel or something similar.

1. Take the Social Intelligence Test

The test relies on some words that are not part of common speech. Before you proceed, make sure you know what these words mean:

- Dispirited
- Despondent
- Defiant

Please wait for the class to start before you proceed (I'll say a few words first and I don't want to interrupt you in the middle of the test)

Just so that you know what the data capture, take the [Social Intelligence Test](#). Don't overthink the answers -- it's best if you answer with your initial gut reaction.

Side note: This test measures a psychological concept known as the "[theory of mind](#)" -- the ability to attribute cognitive and emotional states to other people.

2. Download the data

The data are in this [GDrive folder](#). Look for a file named "L22 - social intelligence data - 2016.04.14.xlsx"

There are two worksheets inside. One contains all the data, the other explains what each of the columns means. Read through those explanations and quickly familiarize yourselves with the data.

3. Pick a question to answer (take no more than 2 minutes)

I suggest one of the following:

- Is there a gender difference?
- Is there a cultural difference? You can pick two countries and see if people who grew up in these countries differ substantially in how well they can read emotions of others.

- Is there an age difference? If you pursue this question, pick two substantially different age groups.

4. Which data should we exclude from analysis?

All of the above questions require between-subjects comparisons. There is a danger of both noise and confounds in the data. For example, the column "retake" captures people who reported having taken the test before. We should probably remove these people from analysis. Not only are their results not informative (they might have just memorized responses), there may also be unequal proportions of those people among men/women, different nationalities, different age groups.

What other variables should we pay attention to?

Side note: Data exclusion is a standard step in the analysis of any experimental results. You always need to disclose how many observations you excluded and what criteria you used. If you exclude too few observations, you risk drawing incorrect conclusions. If you exclude too many, people you present your results to may get suspicious that you kept massaging the data until you saw the story you liked.

Tool note: if you aren't yet, you should familiarize yourself with Excel's "Filter" feature (other software should have similar capabilities)

We will pause for a discussion here.

5. Analysis 0: Eyeball the data

Before you do any statistical analysis, you have to develop a quick sense of what the data may be telling you. Draw a few charts (if possible, include standard errors or confidence intervals).

Tool note: If you are using Excel, experiment with Pivot Tables (look in the "Data" menu).

6. Perform the statistical analysis

Start by performing Wald test by computing the means, the standard errors, the difference of the means, the standard error for the difference, the 95% confidence interval. If you need a reminder of how we did it, look back in this [GDrive folder](#) for the file called "L21 - Wald test data.xlsx". This is the spreadsheet I used to perform Wald test in class on Tuesday.

Now learn how to do the t-test. You want a two-tailed version. And given that you are performing a between-subjects analysis, you want the two-sample unequal variance (heteroscedastic) version.

7. Now do it again!

If you have your own data, perform an analysis on those. There is a good chance that this time you will be performing a within-subjects analysis so you may need to make some different choices along the way (e.g., for t-test, you will want the "paired" version)!

If you do not have your own data, pick another question to ask of the data set I have shared with you.