DATA ANALYSIS EXERCISE:

Did nanoparticles of CuO, MnO or ZnO, and/or Fusarium wilt, affect growth of Chrysanthemum?
We will do each activity once together, and then you can try it on your own with a different group of numbers.

Part 1: Review time! Do you remember what our initial question was, and what our hypotheses and null hypotheses were? Let's discuss how it went.

_____

Part 2: Getting Familiar With Summary Statistics
1. Open the Excel file labeled SampleData.xlsx, and look at the worksheet labeled Height_data. This has the height data from last week arranged in a way that Excel prefers.

2. In the options at the top of the Excel page, where it says HOME INSERT PAGE LAYOUT etc., click on DATA. If you were able to install the Data Analysis Toolpak, there should be an option to the far right labeled "Data Analysis". Click on that and a list of tools should appear.

3. The "Descriptive Statistics" option is a way to get all the summary numbers for a group of samples at once.
   - Click on Descriptive Statistics in the Data Analysis menu and hit OK. A box should open with options.

4. In the Descriptive Statistics Box, click the button next to input range and drag a box from cell B2 to cell F62 (all the rows of numbers plus their labels).
   - Under "Grouped By", select Columns.
   - Select the box that says "Labels in first row".
   - Under "Output options", select output range and then select where in the worksheet you want the output to appear. OR, you can just select New Worksheet Ply.
   - Select Summary Statistics Box.
   - Then click OK and the stats should appear!

5. **Let's Discuss together:** Which treatment had the largest mean value? Which had the largest median? Which had the shortest minimum height? What are all the different measures of variation you see?
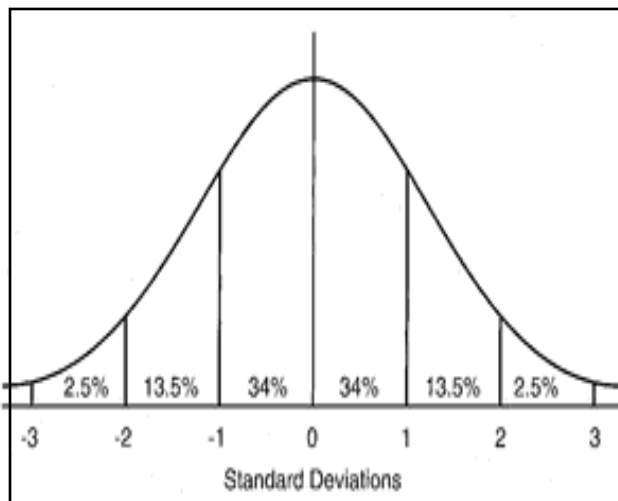
Quick reference definitions:
The **standard deviation** is an estimate of how much the individual measurements differ from the mean. If you have an average height of 18 cm with a standard deviation of 2, this tells you that 67% of the plants are between 16 and 20 cm, and 95% of the plants are between 14 and 22 cm.

The **standard error** is a similar estimate that takes into account your sample size, which makes it a better estimate of the reliability of the mean. The standard deviation divided by the square root of your sample number, in this case SD/5.47.

ON YOUR OWN: Repeat the above, but obtain the summary statistics for uninfected and infected treatments separately. Do you see any difference in means, median, or other statistics?

_____

Part 3: Performing a t-test

It may look like there are differences between the means that you just determined, but you can't really tell without a statistical test. **A t-test determines whether the mean of a group of numbers differs significantly from the mean of another group of numbers.**

Is the height of uninfected copper plants significantly different from uninfected untreated plants? Let's use a t-Test to find out!

1. Staying on the Height Data page, go to DATA, Data Analysis, and select t-test: Two-Sample Assuming Unequal Variances. A box should pop up when you click OK.

2. Click the button next to Variable 1 range, then select boxes 2B through 32B (the untreated measurements plus the label). Then click the Variable 2 range, and select boxes 2C-32C for copper.

3. We are interested in the P(T<=t) two-tail value. What is it? If the P-value is below 0.05, we can say there is a significant difference between the untreated and copper values. (Or to be more precise, that there is a 95% chance that the differences we saw were NOT due to random chance).

On your own: Pick a pair to compare. (examples: Uninfected-untreated vs. Mn, Zn, or Fungicide, infected untreated vs. Cu, Mn, Zn, or fungicide, or uninfected vs. infected within any treatment sample.) Make a hypothesis about the comparison, perform a t-test, and report your findings back to the group.

Part 4: TWO-WAY ANOVA

T-tests are great for determining whether two groups of numbers are different, but ANOVA (Analysis Of Variance) is a way to compare three or more groups of numbers. Here we tested two different factors, infection and nanoparticle treatment, so we will do a "two-way" or two-factor ANOVA test.

ANOVA will test three null hypotheses:
Null 1. There will be no difference in height between the infected and the uninfected plants.
Null 2. There will be no difference in height between the different groups of nano treatments.
Null 3. There is no interaction between the effect of nanoparticle treatment and the effect of infection. Put another way- the effect of Fusarium infection on plant height is not dependent on nanoparticles

1. In the Height data sheet, click on DATA, then Data Analysis, then ANOVA 2-factor with replication.
2. For Data Input Range, Select box 2A through 62F (the labels AND treatments for everything)
3. The Rows per sample should be 30. The "alpha" should be 0.05. Click OK.
4. This produces summary tables for Uninfected, Infected, and Total, but we are mostly interested in the ANOVA table at the bottom. We will look at F (Fisher value), the p-value, and F-crit. We can reject the null hypothesis if the F is greater than the F-crit AND if the P-value is below 0.05.

5. In the ANOVA table, look at the Sample row. This row examines differences between the Uninfected and Infected plants. Is the F value larger than the F-crit? Is the P-value below 0.05? Can we reject null hypothesis 1?

6. Now look at the Columns row. This examines the difference between Untreated, copper, manganese, zinc, and fungicide treatments. Is the F value larger than the F-crit? Is the P-value below 0.05? Can we reject null hypothesis 2?

7. Now look at the Interaction row. Can we reject null hypothesis 3? What does this mean?

The last part- Dot plots online!

1. Open your browser and navigate to http://statistika.mfub.bg.ac.rs/interactive-dotplot/

2. Under the "Data" menu on the left, click on "upload", then "choose file", and pick "ChrysanthemumHeights.csv" from the Desktop. You should see a dataframe pop up. Check to make sure the data looks the same as the instructor's.

3. Under the "Dotplot" menu on the left, click on "Build Dotplot". You should see a group of dots indicating the height measurements for Chrysanthemum under the five different treatments (here called "groups").

4. Spend a few minutes on your own checking the different options under "Graph type". For each graph type, try out the different options in the "options" box that appears below the "Graph type" box.
- What information does a dotplot show that a box plot does not, and vice-versa?
- What information do dotplots and box plots show that a bar graph does not?
- Which do you think is the clearest way to present the results of this experiment?

5. Now, try some different ways to separate the data by Fusarium inoculation status (these are subgroups labeled "infected" or "uninfected").
- With "Dotplot" checked under "Graph Type", go to the subgroup menu and click on "color code".
- Select the box to show subgroups side by side, and choose median or mean summary lines.
- Did Fusarium inoculation have an effect on Chrysanthemum height in any of the treatments? Which visualization makes the effects of inoculation clear?
- Is the mean line different from the median line? Which is more appropriate to put in a figure?