Descriptive and Inferential Statistics

Guidelines for project

Rubric for presentation and associated documents

Statistical Question	Valid statistical question that is clearly stated, focused, and interesting	5
Data Collection & Descriptive Statistics	Raw data and source are included Descriptive statistics	5
Data Display	Includes appropriate, well- labeled, accurate displays (graphs and tables) of the data	5
Data Analysis / Hypothesis Testing	Analysis of the data is accurate, thorough, and appropriate Conditions are checked correctly for any inference procedures	5
Conclusion/ Statement of Significance/Rejection of Null Hypothesis	Stated whether or not the data was statistically significant for each experimental group. Supported those statements with inferential statistics results. Stated whether the null hypothesis was rejected or could not be rejected.	5
Reflection on Process	Gives a good overall picture of the project—what went well and what didn't—and includes ideas for further study	5
Simulation	At least one statistical concept (CLT, Confidence Interval, Hypothesis testing, Anova treatments, correlation, etc)	5
Overall Presentation	Well-organized and well-designed presentation of content.	5
Total		40

Note: Each group can come up with their own/other datasets & problem statement as well!

Instructions: Using the data that you selected, perform both descriptive and inferential statistics for your presentation.

Descriptive Stats:

You need to calculate the following descriptive stats:

- Mean
- Range
- Variance
- Standard Deviation
- Count

Inferential Stats:

Perform hypothesis testing depending on the number of experimental groups contained within the data to determine the statistical significance of your data/results.

After calculating both your descriptive and inferential statistics, you will need to submit both a table and a graph, along with a statement of significance/rejection of null hypothesis.

Table --This is not a table of your **raw data**. It is a summary of the statistics done (see example provided)

--Columns and rows must be straight and neat (typed).

--Headings (with UNITS) are required on all columns and rows.

--Titles go above the table, typed in ALL CAPS (TABLE 1: PLANT HEIGHT (cm) VERSUS LIGHT EXPOSURE).

Graph of the Mean --Can be either a line graph or a bar graph. Line graphs show trends or relationships. Bar graphs are used for comparison. Only graph the MEAN values.

--The independent variable goes on the X-axis, the dependent variable goes on the Y-axis. Label axes with names and units. Include a key.

--Titles go **below** the graph, typed in ALL CAPS (FIGURE 1: PLANT HEIGHT (cm) VERSUS LIGHT EXPOSURE)

--Use software packages to create graphs when possible.

Statement of Significance and Whether or Not the Null Hypothesis Was Rejected You must include a statement detailing whether your data/results are significantly significant. If your project had more than one experimental group, you may have more than one statement of significance. You must use your results from your inferential statistics (again, for most of you, your t-test results) to justify your statement(s). Make sure to also include a statement as to whether or not the NULL HYPOTHESIS was rejected or was not rejected.

References

Datasets

1. https://github.com/thomas-haslwanter/statsintro_python/tree/master/ipynb/Data
2. https://www2.stetson.edu/~jrasp/data.htm
3. https://www.mat.univie.ac.at/~neum/statdat.html

Learning Material and code

4. https://leanpub.com/openintro-statistics
5. https://www.itl.nist.gov/div898/handbook/
6. https://kupdf.net/download/an-introduction-to-statistics-with-python-with-applications-in-the-life-sciences_59757881dc0d600a09043378_pdf
7. https://github.com/thomas-haslwanter/statsintro_python (github link to book above)
8. https://github.com/YikaiZhangskye/ML/blob/master/Unpingco%20J.%20-

%20Python%20for%20Probability%2C%20Statistics%2C%20and%20Machine%20Learning%20-%202016.pdf
9. https://github.com/unpingco/Python-for-Probability-Statistics-and-Machine-Learning/tree/master/chapters/statistics/notebooks (gitbhub link for book above)
10. http://ethen8181.github.io/machine-learning/ab_tests/frequentist_ab_test.html#Hypothesis-Testing
11. http://greenteapress.com/thinkstats2/thinkstats2.pdf
12. http://onlinestatbook.com/2/estimation/ci_sim.html
13. http://onlinestatbook.com/Online_Statistics_Education.pdf
14. http://home.ubalt.edu/ntsbarsh/Business-stat/opre504.htm
15. http://home.ubalt.edu/ntsbarsh/simulation/sim.htm
16. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219542
17. https://github.com/wkeilsohn/Anova-and-Graphing-in-Python/blob/master/Keilsohn_ANOVA1%20Project_2018.py
18. https://github.com/merveozdemir/Data_Manipulation_Visualization
19. http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/goodness.htm