Assignment 1: Introduction to Statistics

|  |  |
| --- | --- |
| Due: | Wednesday, May 29th by end of class |
| Assignment Type: | Group (Up to 6 – may be randomly assigned) |
| Assignment Title: | Introduction to Statistics |
| Style: | One problem at a time on Discord – WORK TOGETHER |
| External Sources: | You are allowed notes, books, and searches |
| Description: | For this lab, there are going to use Google Colab to complete a quick refresh of statistics using Python. |
| Points | 40 |
| Starting Files | Download the notebook file from Blackboard under “Course Materials” then “Meetings” then “Meeting 1 – May 29” Import it into Google Colab. |
| Checking | As you complete each task in the following pages, please post something in the general Discord channel like, “Group 1 Check”. The instructor or TA will score that problem and the group will move on to the next question. |

This lab was introduced by Jianhua Ruan @ utsa.edu

# Getting Started with Google Colab

1. We are going to be using Google Colab this semester. Colab notebooks are Jupyter notebooks that are hosted by Colab. We will be able to do Python and R in Colab this semester.

Start by watching this video on your own:

<https://www.youtube.com/watch?v=inN8seMm7UI&list=PLQY2H8rRoyvyK5aEDAI3wUUqC_F0oEroL&index=1>

Go to <https://colab.research.google.com/>

We first want to update our settings. To do this, go to “Tools” then “Settings” then “Editor” and click to enable both “Show line numbers” and “Show indentation guide”.

In Blackboard, under “Course Materials” then “Meetings” then “Meeting 1 – May 29” there should be a notebook file called, “cmsc462-lab1.ipnb”. Download the notebook file and save it on your computer. In Google Colab, you can now click “File” and then “Upload Notebook”. You should be able to browse to where you saved the notebook file, and it should open.

Do NOT use anything for coding in Python or R except for Google Colab. We do not want to use PyCharm or anything like that.

# Statistics using Python Programming

Please use the Google Colab notebook file from the previous section to complete this assignment.

1. Write a Python function to merge sort a list of numbers. Test your function on a small random integer array of 10 elements. Additionally, write merge which merges two sorted lists. You can write your code based on this: <https://www.geeksforgeeks.org/merge-sort/>

# Summary Statistics using Python Programming

1. Write Python functions to calculate the summary statistics of a list of numbers: max, min, mean, standard deviation, median, 75th percentile, and 25th percentile. (Must be calculated with functions defined by yourself. **Cannot use any existing module in Python except to check your math.**) Use the sets of random numbers as generated in the code skeleton to test your code and output the summary statistics for each data set. Here are some examples, <https://www.w3resource.com/python-exercises/math/python-math-exercise-57.php>

# Histograms using Python Programming

1. Write a Python function that accepts a list of numbers in any range, then scales the numbers to integers in [0, 9], and then count the number of occurrences of each integer in the dataset. Test your function on the three datasets used in 2. You can get some ideas for how to do this here: <https://realpython.com/python-histograms/>

# Histograms using Python Programming

1. Write a Python function based on the results of the functions written above to create a simple histogram as described here: <https://realpython.com/python-histograms/>

You can use these instructions to create some visualizations (although not on GL – you have to use PyCharm or similar). Don’t forget to use plt.show() to display your histograms.

Show both code and result view!!