Problem Set 04
Cumulative Distribution Plots
Problems

Instructions

1. This problem set contains paired programming problems. You are responsible for following the appropriate guidelines for each problem. Follow the instructions for each problem below. Create appropriate files as instructed.
2. Save all files to your Purdue career account in a folder specific to PS04.
3. Compress all deliverables into one zip folder named **PS04_yourlogin.zip**. Submit the zip file to the Blackboard drop box for PS04 before the due date.

Problem Set Deliverables

Item	Type	Deliverable
Problem 1: Golf Ball Distances	Paired	PS04_golfdistance_*yourlogin1_yourlogin2*_report.pdf PS04_golfdistance_*yourlogin1_yourlogin2*.m Any data file loaded into your code
Problem 2: Descriptive Statistics UDFs	Paired	PS04_stats_script_ *yourlogin1_yourlogin2*.m PS04_stats_nino_ *yourlogin1_yourlogin2*.m PS04_stats_io_ *yourlogin1_yourlogin2*.m PS04_report_stats_io_ *yourlogin1_yourlogin2*.pdf
Other Required Deliverables	--	histogramRight.p

Install histogramRight

Download the `histogramRight.p` file from the assignment to your PS04 folder. To call it from a script, it must be present in the same folder as the calling script. Also be sure to include the file in your deliverables zip folder when you submit the assignment.

Publishing User-Defined Functions

You must publish specified MATLAB user-defined functions to a PDF file. The process is slightly different than the process for publishing scripts. Read the instructions for publishing functions that is in the "Publishing MATLAB Code" item in the Problem Set folder on Blackboard to learn how to format your function and how to publish it.

Problem Set 04
Cumulative Distribution Plots
Problems

Problem 1: Golf Ball Distances

Paired Programming

Problem Background

Figure 1. "Iron Byron" machine.

The United States Golf Association (USGA) tests golf balls to make sure that they conform to the rules of golf. One criteria for acceptable golf balls is overall flight distance. To test flight distance, a machine nicknamed "Iron Byron" hits the balls with a driver. The machine outputs a uniform force for each hit so that each ball should fly the same distance.

Engineers have calibrated Iron Byron to use just enough force to hit a golf ball exactly 260 yards. They tested a batch of 100 prototype golf balls and recorded the distance (in yards) each ball travelled. The team needs at least 65% of the golf balls to be within 4 yards of the 260-yard target distance to move forward with the next phase of development. Perform a basic statistical analysis to determine how accurate the prototype golf ball is and make a recommendation as to whether the engineering team can move forward with development.

Problem Steps

1. Open *PS04_golfdistance_template.m* and complete the header information.
2. Save your program with the name **PS04_golfdistance_*yourlogin1_yourlogin2*.m**.
3. Open the file *Data_GolfDistance.txt*, which contains distances golf balls have flown in yards. Review the format of the data file and the information it contains.
4. Load the data into the `INITIALIZATION` section of your program.
5. In the `DESCRIPTIVE STATS & FORMATTED TEXT` section of the script file,
    a. Calculate descriptive statistics for the data: minimum, maximum, range, mean, median, standard deviation. *Note: See note at the beginning of PS03 regarding the range command.*
    b. Display all of the results from the descriptive statistics calculations.
6. In the `HISTOGRAM & CDP` section of the script file,
    a. Generate a histogram using the `histogramRight` command that uses an appropriate bin edge vector that makes the plot easy to interpret. Use the descriptive statistics and the engineers' requirements to help determine the vector.
    b. Determine the relative and cumulative fractional values of the data. Ensure your cumulative frequencies start from '0".
    c. Create a cumulative distribution plot (CDP) using the same number of bins and bin edges as in the histogram. Place the CDP as a subplot below the histogram in the same figure window.
    d. Format your plots for technical presentation.

Problem Set 04
Cumulative Distribution Plots
Problems

7. In the ANALYSIS section of your code, answer the following questions in **brief, complete sentences**:
    - Q1. How did you determine the number of bins for the histogram? Justify your answer.
    - Q2. Referring to your cumulative distribution plot, what is the approximate median of the data set? Explain how you arrived at your answer.
    - Q3. What is the shape and skew of the distribution? Justify your answer using the histogram, CDP, and descriptive statistics.
    - Q4. Referring to your cumulative distribution plot, what is the likelihood that the distance will be between than 256 and 264 yards? State the steps used to arrive at your answer.
    - Q5. Can the engineers move forward with the next step in development for the prototype golf ball? Explain your answer.
8. Once you have answered all the questions and debugged your code, publish the code to a word document. Save it as **PS04_golfdistance_*yourlogin1_yourlogin2*_report.pdf**.

Reference: http://www.golftipsmag.com/instruction/faults-and-fixes/lessons/become-a-swing-machine.html#.VRF0ahBYB8E

Problem 2: Descriptive Statistics UDFs

Paired Programming

Problem Setup

Descriptive statistics are calculations that are useful in many different applications. For this problem, you will calculate three descriptive statistics in a script and then in different types of user-defined functions.

You will need the following vectors to test your code. Use the names they have been given here.

```
vec1 = [8.800, 8.871, 8.813, 8.861, 8.842, 8.815, 8.755]
vec2 = [14.081, 14.175, 13.896, 13.633, 14.560, 13.478, 13.507]
```

Problem Steps

A. Write a script for descriptive statistics

1. Open **PS04_stats_script_template.m**. Complete the header information. Save the file as **PS04_stats_script_ *yourlogin1_yourlogin2*.m**.
2. In the INITIALIZATION section of your code, assign each vector.
3. In the STATISTICS CALCULATIONS section, compute the range, mean, and standard deviation for each vector. Ensure that all computation lines are suppressed.

Problem Set 04
Cumulative Distribution Plots
Problems

4. In the `FORMATTED TEXT DISPLAY` section, use `fprintf` statements to display the statistics for each vector. Show all values with two decimal places.
5. Run your script. In the `COMMAND WINDOW OUTPUTS`, paste as comments the text that is displayed in the Command Window.

B. Convert your script to a no-input, no-output user-defined function (stats_nino).

Your script calculates range, mean, and standard deviation for hardcoded vectors. Your UDF needs to also calculate those statistics for hardcoded vectors. It must:

- Have an operational function definition line
- Print the descriptive statistics for both vectors to the Command Window using `fprintf` statements

6. Open **PS04_stats_udf_template.m**. Complete the full header. Note that it contains new information that isn't present in the script header. Save the file as **PS04_stats_nino_ *yourlogin1_yourlogin2*.m**.
7. Create a function definition line for the UDF.
8. Copy relevant sections from your script into the UDF. Fix them as necessary to make them work within the UDF.
9. From the Command Window, clear your workspace and the Command Window. Then, run your function from the Command Window.
10. Paste as comments the function call and the displayed text to the `COMMAND WINDOW OUTPUTS` section of the code.

C. Convert stats_nino to an input, output UDF (stats_io).

This UDF requires input and output arguments. It must

- Have an operational function definition line
- Have one input argument, a generic vector of data
- Return **only** mean and standard deviation as output arguments
- Print the descriptive statistics for the input vector to the Command Window using `fprintf` statements

11. Save a copy of **PS04_stats_nino_ *yourlogin1_yourlogin2*.m** as **PS04_stats_io_ *yourlogin1_yourlogin2*.m** and revise the header.
12. Revise the function definition line.
13. Revise the UDF code as necessary.
14. Run your function from the Command Window using `vec1`. Rerun using `vec2`.
15. Paste as comments the two function calls and the displayed text to the `COMMAND WINDOW OUTPUTS` section of the code.

Problem Set 04
Cumulative Distribution Plots
Problems

D. Run the UDFs and answer questions

16. In the `ANALYSIS` section of **stats_IO**, answer the following questions.

    Q1: Follow these instructions, in order, and then answer the question below.

    1. Clear your workspace.
    2. Run your **script** file from the Command Window.
    3. Clear your workspace.
    4. Run your **stats_nino** function from the Command Window.

    What differences do you see between the results of the script and the UDF?

    Q2: Clear the workspace. Run your **stats_io** function from the Command Window using `vec1` as the input vector.

    How is the result different from the result generated by **stats_nino**?

    Q3: Type **`help PS04_stats_io_yourlogin1_yourlogin2`** into the Command Window and hit enter. What do you see? Why is this helpful?

Publish your stats_io function to a PDF file using a third vector `vec3`, defined as `[3.2, 6.9, 5.2, 4.7]`. Name the published file **PS04_stats_io_*yourlogin1_yourlogin2*_ report.pdf**.