Problem Set 03
Statistics and Histograms
Problems

Instructions:

1. This problem set contains one paired programming and one individual programming problem. Each problem has a set of deliverables that needs to be submitted. You are responsible for following the appropriate guidelines and instructions below. Create appropriately-named files as instructed.
2. Save all files to your Purdue career account in a folder specific to PS03.
3. Compress all deliverables into one zip folder named **PS03_yourlogin.zip**. Submit the zip file to the Blackboard drop box for PS03 before the due date.

Problem Set

Item	Type	Deliverable
Problem 1: Daylight Tournaments	Paired	PS03_daylight_*yourlogin1_yourlogin2*.m PS03_daylight_ *yourlogin1_yourlogin2*_report.pdf Any data set loaded into your m-file
Problem 2: Powder Analysis	Individual	PS03_metalpowder_*yourlogin*.m PS03_metalpowder_*yourlogin*_report.pdf Any data set loaded into your m-file

Range Command

The range command is part of the *MATLAB and Simulink Student Suite* and the MATLAB versions available on Software Remote and ITaP computers; however, it is **not** included in the basic MATLAB Student version.

If the range command is not available to you, then use the difference between the maximum value and minimum value to calculate the data's range.

Formatting Reminder

Always format your text, plots, and numerical outputs in a professional manner.

- Numerical values must have a reasonable number of decimal places and include units when necessary.
- Displayed text should be descriptive and professional. Use full sentences.

Problem 1: Daylight Tournaments

Paired Programming

Problem Setup

Your engineering firm is working with a sports club that wants to build an outdoor soccer complex in West Lafayette. The club would like to host youth tournaments and wants to set the tournament season to be

Problem Set 03
Statistics and Histograms
Problems

from April 1 to August 31, with the possibility of extending the season (earlier or later or both). Outdoor lighting is expensive and is not within the club's budget, so the club would like to ensure that daylight hours are sufficient to operate the tournaments. Daylight hours are assumed to be between sunrise and sunset. A tournament could happen on any day during their season and would need at least 12 hours of daylight per day to operate.

You have been provided with a data file, which is named **2016_sunrise_sunset_times.csv**, that contains the sunrise and sunset times for every day in 2016. Note the following information about the data set:

- The data use Day of Year in the following calendar date format: 1 Jan 2016 for Day 1.
- The times are in 24-hour clock time with the format hhmm: 811 is 8:11 am, 1758 is 5:58 pm.

Problem Steps

A. Initialize your script:

1. Open PS03_daylight_template.m and complete the header information. Save the file with the name **PS03_daylight_*yourlogin1*_*yourlogin2*.m**.
2. Open **2016_sunrise_sunset_times.csv** and review the information it contains.
3. In the `INTITIALIZATION` section of your script, write the code to import and extract the data into separate vectors.
4. Calculate the hours of daylight in each day of the year:

   Your data is in clock time. Convert clock time to decimal hours (e.g., 8:45am = 8.75 decimal hours). Then, calculate the hours of daylight in a day. ***Hint:*** the `fix` command rounds each element of a vector to the nearest integer toward zero.

   a. In the `DAYLIGHT CALCULATIONS` section of your script, write the code to:

      i. Calculate the number of hours of daylight per day. Round the final decimal time to the nearest hundredths. ***Hint***: the `round` command rounds toward the nearest decimal or integer.

      ii. Create a vector that contains only the daylight data for the proposed season of April 1 to August 31. ***Hint***: Recall months may have a different number of days.

5. In the `STATISTICS & FORMATTED DISPLAY` section of your script, write the code to:

   a. Calculate the minimum, maximum, range, mean, median, and standard deviation for

      i. The number of hours of daylight per day for the whole year

      ii. The number of hours of daylight per day for the proposed tournament season

   Note: See information on `range` command on the first page of the problem set.

   Reminder: Do not name your variables with built-in MATLAB function names.

   b. Use print commands to display all of the descriptive statistics results to the Command Window.

Problem Set 03
Statistics and Histograms
Problems

6. In the `HISTOGRAMS` section of your script:
    a. In Figure 1, use the `histogram` command to create a histogram of the number of hours of daylight per day for the whole year. Use the `histogram` command defaults (i.e., do not set the number of bins or the edge values).
    b. Use the resulting histogram properties to determine the bin edges and frequencies in each bin for this histogram. Name the bin edges vector `edges_all` and name the frequency vector `freq_all`.
    c. Create a vector of bin edges that can be used to create a histogram with 20 evenly-spaced bins that include the full range of daylight hours for the entire year. Name this vector `edges_rule`.
    d. In Figure 2, create a histogram that uses the `edges_rule` vector to set the bin edges.
    e. In Figure 3, create a histogram that shows the distribution of daylight hours during the proposed tournament season (April 1 to 31 August) with the number of bins set to 8.
    f. Ensure that all histograms are formatted for technical presentation.
7. In the `ANALYSIS` section of your code, answer the following questions:

    Q1: The use of 20 bins in the second histogram (Step 6.d) was a guideline based on the square root of the number of data points in the data. Do you think 20 bins an appropriate number of bins for the histogram or can you identify a better number of bins? Explain your answer.

    Q2: Is the data in the second histogram figure (Step 6.d) unimodal or bimodal? Explain your answer.

    Q3: Describe the skew of the distribution of the data in the third histogram (Step 6.e). Specifically, is it positive, negative, or undefined? Why?

    Note: Do not use the terms left or right to describe skew as this is mathematically inconsistent given certain scenarios.

    Q4: Use the third histogram to justify whether or not the proposed tournament season of April 1 to August 31 is possible.

    Q5: Use the **descriptive statistics** to explain whether or not the club can add 30 days to the proposed season.
8. Publish your script as a PDF file named **PS03_daylight_*yourlogin1*_*yourlogin2*_report.pdf**.

Reference: http://aa.usno.navy.mil/data/docs/RS_OneYear.php

Problem Set 03
Statistics and Histograms
Problems

Problem 2: Powder Analysis

Individual Programming

Problem Setup

Many metal parts, such as appliance gears, engine components, and surgical tools, are manufactured using powder metallurgy. Powder metallurgy is a metal-forming process that heats compressed metal powder to temperatures below the metal's melting point. A mechanical powder press can be used to shape metal powder. Metal powder is added to a die that is then pressed with enough force to fuse the metal particles together without liquefying them.

Using powders with predictable particle size can be important to obtaining high-quality parts. Too much variance in powder particle size can lead to unwanted voids when pressed into parts. A batch of powder can be sieved through multiple screens or through a particle-size analyzer to determine the particle size distribution of the metal grains.

Cornerstone Inc., a powder manufacturer, has asked you to perform a basic statistical analysis on two samples of manufactured stainless steel powder taken from different production runs. They want you to determine if each meets their quality control requirements for particle size. To ensure the manufacturing process is compliant, a powder must have:

- a size range of 9 – 80 microns (μm), where up to 8% of the total particles can be outside that range (i.e., smaller than 9 μm or larger than 80 μm),
- a distribution where the mean is within +/- 2.5 μm of the median, and
- a mean particle size that is within +/- 2.5 μm of a 41.5 μm baseline mean.

You have been provided with a data file, which is named **Data_metalpowder.csv**, that contains the particle size data for the two samples, CSIB-146 and CSIB-172, of stainless steel powder. Note that the particle sizes are integer values.

Problem Steps

1. Open **PS03_metalpowder_template.m** and complete the header information. Save your script with the name **PS03_metalpowder_*yourlogin*.m**.
2. Open **Data_metalpowder.csv** and review the information it contains.
3. In the `INTITIALIZATION` section of your script, write the code to import and extract the data.

Problem Set 03
Statistics and Histograms
Problems

4. In the `DESCRIPTIVE STATISTICS & FORMATTED TEXT DISPLAY` section of your script:
    a. Calculate the descriptive statistics for each powder sample: minimum, maximum, range, mean, median, and standard deviation.
    b. Use print commands to display all of the descriptive statistics to the Command Window.
5. In the `HISTOGRAMS` section of your script:
    a. Create a 2x1 (row x col) subplot with two histograms. Each histogram in the subplot should have the same number of bins and edges, which must be managed appropriately so the histograms can be used for meaningful comparisons.
        i. The top histogram will be the particle size distribution for CSIB-146.
        ii. The bottom histogram will be the particle size distribution for CSIB-172.

    Hint: Consider the within specification particle size range when determining the bin edges. Careful selection of the bin edges will enable you to use the histogram properties to determine the number of particles outside the required size range.

    b. Format the histograms for technical presentation.
6. In the `CALCULATIONS & FORMATTED TEXT DISPLAY` section, write the code to:
    a. For each of the two samples, calculate the following:
        i. The percent of particles that are outside the required size range. See the hint in Step 5.a. ***Note***: Do not report percentage as a decimal, e.g. 0.12; instead say 12%.
        ii. The difference between the sample median and sample mean and whether the mean is greater than the median or less than the median.
        iii. The difference between the sample mean and the baseline mean.
    b. For both samples, print the results of the above calculations to the Command Window.
7. In the `ANALYSIS` section, answer the following questions

    Q1: Are the particle size distributions of the samples similar or different? Justify your answer by comparing the two histograms, making reference to the shape and skew of the distributions.

    Q2: Is the powder meeting the quality control requirements? Justify your answer based on the powder specifications provided above.

8. Publish your script as a PDF file named **PS03_metalpowder_*yourlogin*_report.pdf**.

Resources:
http://thelibraryofmanufacturing.com/powder_processes.html
http://www.precision-powder-metallurgy.com/oil-pumprotor.htm