A12 · Linear Regression

Introduction

Assignment Goals

This assignment uses Excel to build on the least squares analysis you learned in ENGR 131 and then adds two MATLAB linear regression problems with engineering contexts.

Successful Completion

This assignment has **3** problems. All problems go with Classes A and B.

1. Read *Notes Before You Start*, on **Page 1**.
2. Read each problem carefully. You are responsible for following all instructions within each problem.
    a. The deliverables list within each problem contains everything you are expected to submit.
3. Complete the problems using the problem-specific templates when a template is provided in the assignment download.
4. For any file, replace *template* or *login* in the filename with your Purdue Career Account login.
5. Review your work using the learning objective evidences.
6. When your work is complete, confirm your deliverables are submitted to Gradescope.
    a. Note the three different assignments in Gradescope.
        i. **A12 – Skills Problems**: submit your deliverables for Problem 1. Help link.
        ii. **A12 – Context Problems**: submit your individual deliverables for Problems 2 and 3. Help link.
        iii. **A12 – Team Planning**: submit your team plan for Problems 2 and 3 as a team. Help link.
    b. You can resubmit your work as many times as you want; only the final submission will be graded.
    c. Do **NOT** upload any document not listed in the deliverables. Do not upload temporary versions of m-files (*.m~ or *.asv) – these files will be ignored by Gradescope.
7. Late submissions will be accepted up to 24 hours after the due date and will result in a 25% penalty.

Learning Objectives & Grading

This course uses learning objectives (LOs) to assess your work. You can find a full list of the course LOs here. Review the grading outline at the end of each problem in this assignment to see each problem's LOs.

Notes Before You Start

Helpful MATLAB Commands

Learn about the following built-in MATLAB commands, which might be useful in your solutions:

`polyfit, polyval`

Problem 1: Excel Linear Regression

Introduction

This problem has you using Excel to determine least squares model and goodness of fit for a data set. You will compare models and discuss their differences. You will use your knowledge of your data set to make reasonable predictions using your model.

Submission

Gradescope Assignment	A12 – Skills Problems	Assignment Type	Individual
Deliverables	☐ Requested results and information ☐ A12Prob1_bugReports_*login*.xlsx ☐ A12Prob1_figure_*login*.png		

Problem

A software bug is an unexpected result or failure in a released software package. All complex software has bugs in it. Most are minor, and some are even famous, like the Minus World in Super Mario Bros. But sometimes a software bug contributes to a catastrophic failure, as in the massive 2003 electrical grid blackout in the United States and Canada that affected approximately 50 million people.

To minimize negative effects, software producers track, manage, and fix bugs through new updates or new releases. A software firm has provided data that contains bug reports for one of their major software packages. The company previously determined a model to predict how the number of bug reports changes over time for a similar software package, $N = -5.6t_r + 235$, where N is the number of bugs reported and t_r is the time since release (in months).

You must perform least-squares regression to find a new model for the number of software bugs reported as a function of time since the new software's initial release. You will find the data in the CALCULATIONS sheet of the provided Excel template.

Instructions

Excel Calculations

Using the **CALCULATIONS** sheet in the Excel template, you must

1. Use the data and the least-squares equations to determine the least-square line for the data.
2. Plot the data and confirm your least-squares line using Excel's "add a trendline" feature. Show both the model equation and the r-squared value. Update the equation display using clear and appropriate variable names in place of x and y. Format the plot for technical presentation.
3. Determine the goodness of fit (SSE, SST, and r^2) for both the provided model and your least-squares model.

Save a plot image

After you have completed the calculations and created the plot, you need to save an image of your Excel plot that shows the data, the trendline, and the least-squares model (as described in Step 2 above). Save an image of your Excel plot with the trendline and r-squared value visible.

- If you have a Mac, right click the chart and select "Save as picture".
- If you have Windows, you may need to Copy the figure, paste into Powerpoint as an image, right click the image in Powerpoint and select "Save as picture".

*You can use any image file type (png, jpg, tiff, etc.) that will render in Gradescope but name the file appropriately.

Answer Questions & Submit to Gradescope

After you complete the calculations, open Gradescope and select the assignment **A12 – Skills Problems**. You will find 5 parts within **Q1**. Excel – Linear Regression. Answer all the questions and their parts. Save your answers periodically in case you lose internet connection or need to stop working on the assignment before it is complete.

a. Submit your XLSX file, with the appropriate file name.
b. Submit your plot image, with the appropriate file name.
c. Answer the questions presented in Gradescope.

Note: Failure to submit the XLSX worksheet will result in a 0 on this problem.

Grading

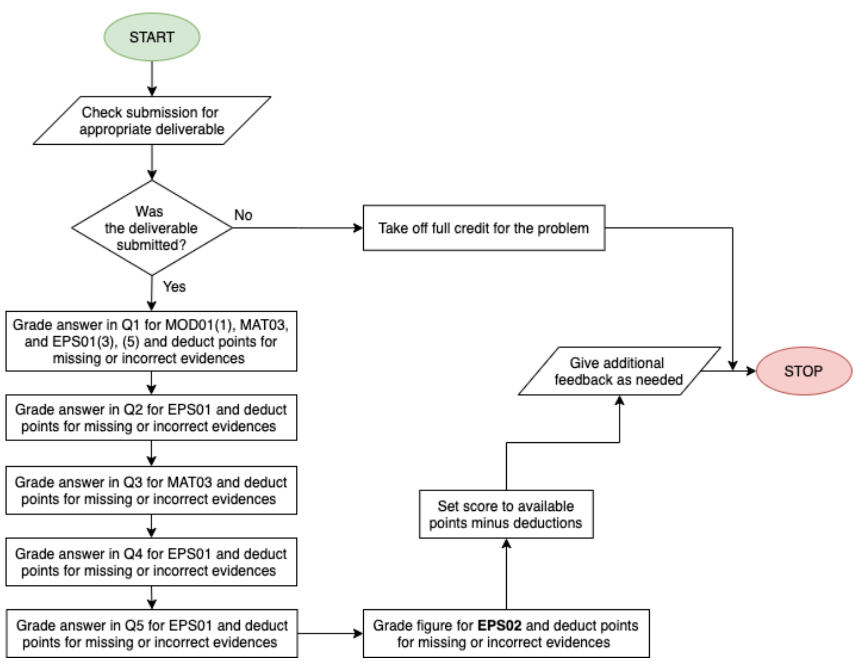
LOs: PC05, MAT03, MOD01, EPS01, EPS02

Point value: 10 points. Problem is graded by question. Partial credit is available and is based on evidences in MAT03, MOD01, EPS01, EPS02. If you do not meet the PC05 expectations, you will lose additional credit.

	Points
Q.1.(1)	2
Q.1.(2)	1
Q.1.(3)	1
Q.1.(4)	2
Q.1.(5)	2
Figure	2

	Points
PC05 (1)	-100%
PC05 (2)	-25%
PC05 (4)	-15%
PC05 (8)	-100%

Grading Process



Problem 2: Greenhouse Gas Analysis

Introduction

This problem will allow you to apply linear regression concepts to data sets that contain environmental data used in news, commercial, academic, and scientific settings. You will discuss what you know about the models using your knowledge of the data.

Submission

Individual

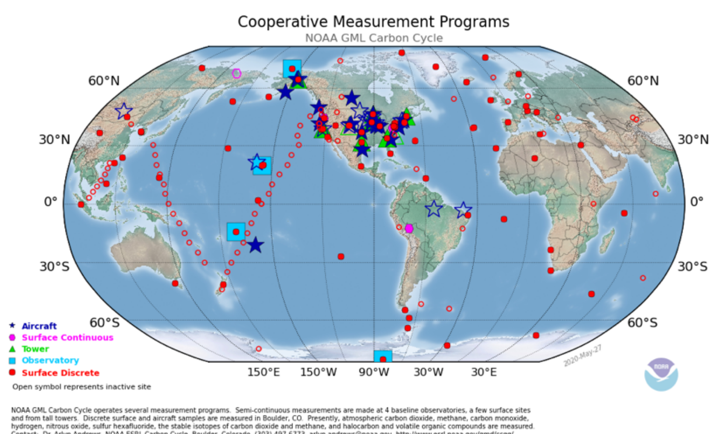
Gradescope Assignment	A12 – Context Problems	Assignment Type	Individual
Deliverables	☐ A12Prob2_airPollution_*login*.m Supporting files: ☐ A12Prob2_CO2figure_*login*.png ☐ A12Prob2_SF6figure_*login*.png ☐ Data_NOAA_ESRL_co2_trend_1980-2021.csv ☐ Data_NOAA_ESRL_sf6_trend_1997-2021.csv		

Team Plan

Gradescope Assignment	A12 – Team Planning	Assignment Type	Team
Deliverables	☐ Requested information		

Problem

Environmental engineers develop solutions to minimize air pollution. Carbon dioxide (CO2) and sulfur hexafluoride (SF6) emissions are significant greenhouse gases. The National Oceanic and Atmospheric Administration's Earth System Research Laboratory (ESRL) maintains records of atmospheric greenhouse gas concentrations.



For this problem, you must model the global trend in CO2 and SF6 atmospheric concentrations. You will find the CO2 data, with its complete NOAA header, in the file named **Data_NOAA_ESRL_co2_trend_1980-2021.txt** and the SF6 data in the file **Data_NOAA_ESRL_sf6_trend_1997-2021.txt**.

Write a MATLAB no-input, no-output function to perform least squares regression on the provided data. You will model average CO2 (in ppm) as a function of decimal year and then model average SF6 (in ppt) as a function of decimal year. Your function must

- Perform linear regression on the data to get the least-squares coefficients, first for CO2 and then for SF6.
- Determine the predicted values of each linear model.
- Calculate SSE, SST, and r^2 values for each model.

- Display the linear model equation (with clear variable names), SSE, SST, and r2 to the Command Window for each model. Make sure you can differentiate between the information for CO2 and SF6.
- Generate two figures: one figure that displays the data and the trend line on the same axes for CO2 and a second figure that displays the data and trend line on the same axes for SF6. Format both figures for technical presentation.

In the `ANALYSIS` section, answer the following questions:

Q1. From your analyses, can you draw a conclusion about the accuracy of the data measurements? Provide justification for your answer.

Q2. For which data set does a linear model best explain the variation that exists in the data? Clearly state the basis of your reasoning and provide justification for your answer.

Q3. You want to see the long-term trends in CO2 over a 100-year span. Predict the CO2 levels at year 1950.0, 2000.0, and 2050.0. Report your results and justify your response using your knowledge of the data.

You will need to submit an image file for each of your figure windows. From the figure window, select **File > Save As**. Set the file format to PNG (see help if needed). Properly name the file.

Instructions

1. Read through the entire problem statement.
2. **With your teammates**: develop and document a plan to solve this problem.
    a. Understand the expectations of the problem.
    b. Discuss strategies for solving the problem. This can include citing examples from class notes, drawing pictures, outlining a plan using text or pseudocode, etc. **DO NOT SHARE CODING SOLUTIONS.**
    c. Submit your plan to the team assignment in Gradescope
        1. Open the Gradescope assignment for this assignment's team plan (see the submission list at the beginning of this problem).
        2. In the area for this problem:
            a. Enter the names of your teammates who participated in the planning.
            b. Enter a brief description of your team's plan to solve the problem. The plan should be connected to the problem and have at least 2-3 steps. It should not be a detailed explanation of every step necessary to solve the problem.
            c. If you have image files, etc., that you would prefer to share, then you may add them in the *Optional* file submission area.
        3. Save your results.
    d. Add your teammates to the submission. Select 1 team member to submit the plan. **Work together** to make sure it is done correctly.
        1. Click **Submit & View Submission** at the bottom of the assignment
        2. Add all teammates to the group (Gradescope instruction link)
        3. All teammates confirm that you get a submission email and verify that you can see the submission in your Gradescope.
        4. You only need to add teammates one time (regardless of the number of problems in the assignment or the number of resubmissions your team makes).

3. **Individually**:
    a. Complete your m-file and run it to get your results.
        - The team plan is an initial start on the problem. It may not be completely correct, and you may find flaws in the plan once you start coding. You should make any individual changes that are necessary to obtain the best solution. You will be assessed on your individual solution to the problem.
    b. Cite any peers you worked with in your script header if their help changed how you decided to solve the problem. Make sure you also completed the rest of the script header.
    c. Submit your properly named m-file, data file, and image file to the appropriate problem in the individual Gradescope assignment (see the submission list at the beginning of this problem).
        - Submit your deliverables once all your context problems are complete. Click here for help.
        - Do not submit any other files.

References

Ed Dlugokencky, NOAA/GML (www.esrl.noaa.gov/gmd/ccgg/trends_sf6/)

Ed Dlugokencky and Pieter Tans, NOAA/GML (www.esrl.noaa.gov/gmd/ccgg/trends/)

Grading

LOs: PC05, MAT03, MOD01, EPS01, EPS02

Point value: 8 points. Problem is graded by question. Partial credit is available and is based on evidences in MAT03, MOD01, EPS01, EPS02. If you do not meet the PC05 expectations, you will lose additional credit.

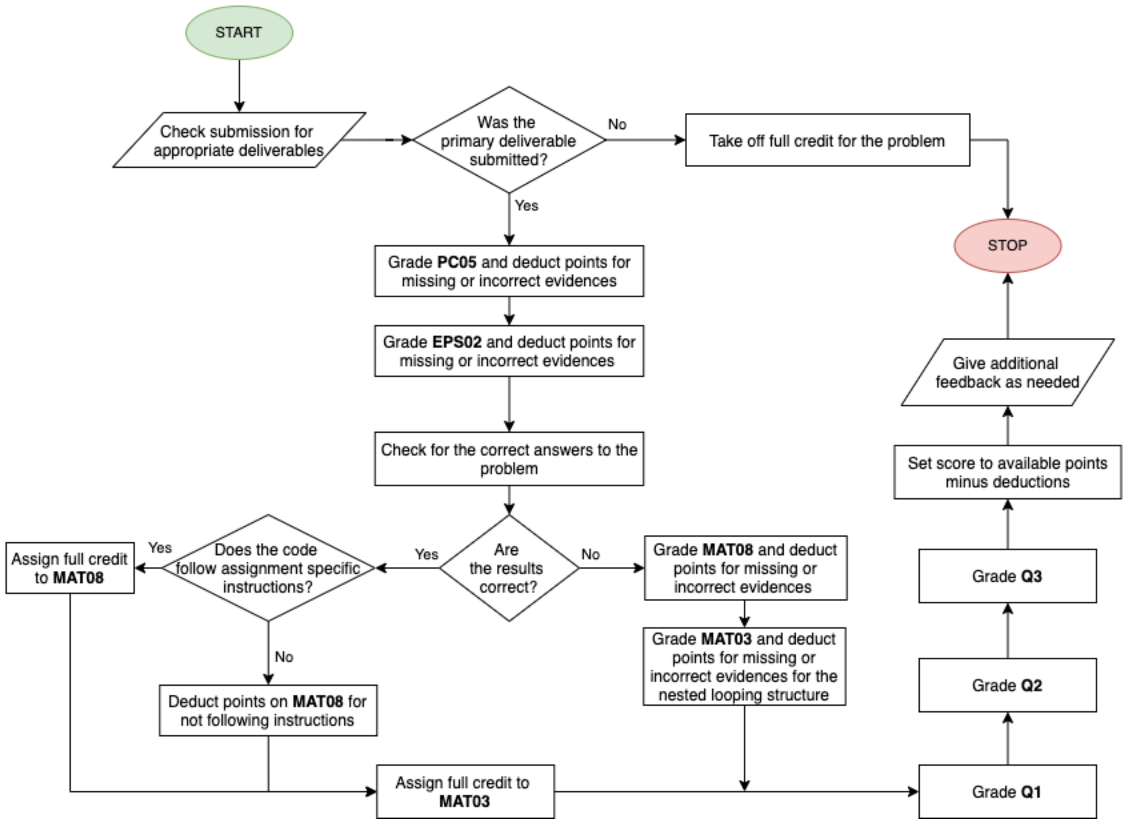
LO Table

	PC05	EPS02	MAT08	MAT03
(1)	-100%	0.4	0.5	0.6
(2)	-25%	0.4	1.5	0.6
(3)	-10%	0.4	0.5	0.6
(4)	-15%	0.4	0	0.6
(5)	0	0.4	0	0
(6)	0	0.4	0	0
(7)	0	0.4	0	0
(8)	0	0	0	0

Analysis	Points*
Q1	1
Q2	1.2
Q3	1.2

* Assessment of analysis questions is based on LOs in EPS01

Grading Process



Problem 3: Solar Panel Output

Introduction

You will combine linear regression, selection structures, and user-defined functions in this problem. Use your knowledge of functions to allow a user to predict values using the model you develop within the program. Use your knowledge of your data to design the function to make reasonable predictions using the model.

Submission

Individual

Gradescope Assignment	A12 – Context Problems	Assignment Type	Individual
Deliverables	☐ A12Prob3_panelOutput_*login*.m Supporting files: ☐ A12Prob3_figure_*login*.png ☐ Data_panelX5_output_measurements.csv		

Team Plan

Gradescope Assignment	A12 – Team Planning	Assignment Type	Team
Deliverables	☐ Requested information		

Problem

You work for a solar energy company and are interested in the electricity production of a new residential solar panel your team is developing. Your teammates ran an experiment where they installed a test panel and monitored the electric output at different sunlight levels. The experiment results are in the data file **Data_panelX5_output_measurements.csv**. The amount of sunlight is measured in peak sun hours (hr) and the electrical output is measured in kilowatt-hours (kWh).

You must create a model of the data provided that can be used to estimate a panel's output for a range of peak sun hours. Write a user-defined function that does two things. First, it will find the least squares model for the data and second, it will accept one peak sun value and return the predicted electrical output. Your function must do the following:

- Use the data to determine the least-squares model for the data.
- Plot the data and model on the same axes, formatted for technical presentation.
- Display the linear model equation (with clear variable names), SSE, SST, and r2 to the Command Window.
- Accept one input: a value for peak sun hours
- Return one output: a predicted electrical output value for the input peak sun value.

- Display the predicted electrical output to the Command Window along with the peak sun time used to make the prediction.
  - Use your knowledge of the data to determine if there needs to be an explanation of the reliability of the prediction. Use a selection structure to display different justifications, if necessary.
  - Your function must determine a value for the predicted electrical output so that the program does not generate an error.

Once your function is working properly, call it from the Command Window. First, use a peak sun value of 4.8 hours and then use a peak sun value of 7.5 hours. Copy your function calls and the text displays and paste them as comments into the `RESULTS` section of your function.

You will need to submit an image file for each of your figure windows. From the figure window, select **File > Save As**. Set the file format to PNG. Name the files using the format given in the assignment instructions.

Instructions

1. Read through the entire problem statement.
2. **With your teammates**: develop and document a plan to solve this problem.
   a. Understand the expectations of the problem.
   b. Discuss strategies for solving the problem. This can include citing examples from class notes, drawing pictures, outlining a plan using text or pseudocode, etc. **DO NOT SHARE CODING SOLUTIONS.**
   c. Submit your plan to the team assignment in Gradescope
      1. Open the Gradescope assignment for this assignment's team plan (see the submission list at the beginning of this problem).
      2. In the area for this problem:
         a. Enter the names of your teammates who participated in the planning.
         b. Enter a brief description of your team's plan to solve the problem. The plan should be connected to the problem and have at least 2-3 steps. It should not be a detailed explanation of every step necessary to solve the problem.
         c. If you have image files, etc., that you would prefer to share, then you may add them in the *Optional* file submission area.
      3. Save your results.
   d. Add your teammates to the submission. Select 1 team member to submit the plan. **Work together** to make sure it is done correctly.
      1. Click **Submit & View Submission** at the bottom of the assignment
      2. Add all teammates to the group (Gradescope instruction link)
      3. All teammates confirm that you get a submission email and verify that you can see the submission in your Gradescope.
      4. You only need to add teammates one time (regardless of the number of problems in the assignment or the number of resubmissions your team makes).
3. **Individually**:
   a. Complete your m-file and run it to get your results.

  - The team plan is an initial start on the problem. It may not be completely correct, and you may find flaws in the plan once you start coding. You should make any individual changes that are necessary to obtain the best solution. You will be assessed on your individual solution to the problem.

b. Cite any peers you worked with in your script header if their help changed how you decided to solve the problem. Make sure you also completed the rest of the script header.

c. Submit your properly named m-file, data file, and image file to the appropriate problem in the individual Gradescope assignment (see the submission list at the beginning of this problem).
  - Submit your deliverables once all your context problems are complete. Click here for help.
  - Do not submit any other files.

References

https://rtc.sandia.gov/

Grading

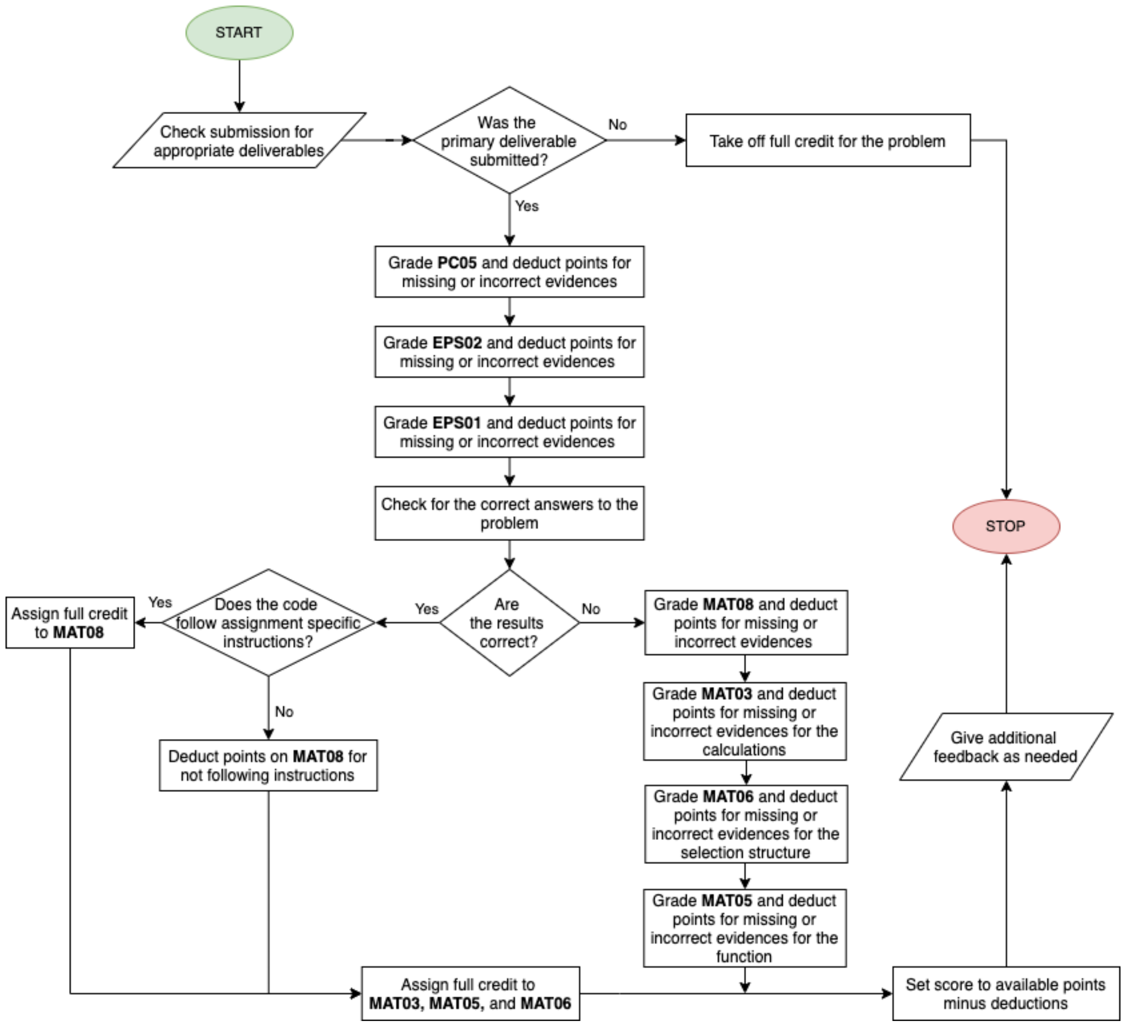
LOs: PC05, MAT03, MOD01, EPS01, EPS02

Point value: 8 points. Problem is graded by question. Partial credit is available and is based on evidences in MAT03, MOD01, EPS01, EPS02. If you do not meet the PC05 expectations, you will lose additional credit.

LO Table

	PC05	EPS02	EPS01	MAT08	MAT03	MAT05	MAT06
(1)	-100%	0.2	0.4	0.75	0.2	0.4	0
(2)	-25%	0.2	1.2	1.5	0.2	0.4	0.3
(3)	-10%	0.2	0.4	0.75	0.2	0	0.3
(4)	-15%	0.2	0.4	0	0.2	0	0.3
(5)	0	0.2	0.4	0	0	0	0.3
(6)	0	0.2	0	0	0	0	0
(7)	0	0.2	0	0	0	0	0
(8)	0	0	0	0	0	0	0

Grading Process


START
Check submission for appropriate deliverables
Was the primary deliverable submitted?
No
Take off full credit for the problem
Yes
Grade **PC05** and deduct points for missing or incorrect evidences
Grade **EPS02** and deduct points for missing or incorrect evidences
Grade **EPS01** and deduct points for missing or incorrect evidences
Check for the correct answers to the problem
STOP
Are the results correct?
Yes
No
Does the code follow assignment specific instructions?
Yes
Assign full credit to **MAT08**
No
Deduct points on **MAT08** for not following instructions
Grade **MAT08** and deduct points for missing or incorrect evidences
Grade **MAT03** and deduct points for missing or incorrect evidences for the calculations
Grade **MAT06** and deduct points for missing or incorrect evidences for the selection structure
Grade **MAT05** and deduct points for missing or incorrect evidences for the function
Give additional feedback as needed
Assign full credit to **MAT03, MAT05,** and **MAT06**
Set score to available points minus deductions