Math 118 Spring 2021

Technology Project

Prescription Drug Costs

**\*\*\*INSTRUCTIONS\*\*\***

For this assignment, you will use a data set with 990 rows and 2 columns. The full data set of over 1,000,000,000 (yes, billion) records is available publicly on cms.gov ([link to Prescription Drug Profiles data page](https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/BSAPUFS/Prescription_Drug_Profiles)). The first 1,000 rows were taken from the file and reduced to the included 3 columns.

* Install Word/Excel on the computer you will use to do your project.
  + They are available for free in Office 365. Links are on the Outlook page where you check your student email, in the drop down, upper left corner of the page.
  + If you cannot install the software, you can use the browser version – the links are in the same drop-down menu.
* Perform the relevant calculations and generate the relevant graphs.
  + Necessary calculations and graphs are indicated below, numbered C1-C8 in the first section.
  + Do your work in the same Excel spreadsheet where you find the data. Use any blank space, or multiple tabs if you like.
  + Save your work with a different file name. You do not have to put your calculation results in the Word document.
* Type your responses to the reflection questions, numbered R1 – R5, in the Word document.
  + When you are done, save your work with a different file name.
* Submit both files to the assignment drop box.

**Calculations and Visualizations (11 pts)**

Distribution: Shape, Center, and Spread (7 pts):

Use the data in Column B to complete the following:

1. Construct a frequency distribution table and relative frequencies using the classes in the table provided in the data dictionary. It is copied below for convenience. You can check that it matches the table in the Data Dictionary, available from the link in the instructions. (2 pts)

|  |  |
| --- | --- |
| Lower limit | Upper limit |
| $0.00 | $13.00 |
| $13.01 | $25.00 |
| $25.01 | $95.00 |
| $95.01 | $47,512.00 |

1. Construct the histogram for the frequency distribution table in question 1 (1 pts)
2. Calculate the mean and median values. (1 pts)
3. Find the five-number summary (min, max, median, Q1 and Q3). (1 pts)
4. Use these values to draw a box plot. (1 pts)
5. Calculate the standard deviation. Find the 68% and 95% ranges using the empirical rule. (1 pts)

Correlation (4 pts):

Use the data in columns A and B to complete the following:

1. What is the correlation coefficient for the relationship between the mean HCC score (Column A) and average total drug cost (Column B)? (1 pts)
2. Draw the scatter plot diagram for the pair (A,B) with the regression line. (3 pts)

**Reflections (10 pts)**

1. How similar are the relative frequencies in the table you made for the provided data set, compared with the relative frequencies in the data dictionary? Access the data dictionary as a pdf file using the link in the instructions. (2 pts)
2. Why do you think the table in the data dictionary uses classes with unequal widths? What effect does that have on how you perceive the shape of the distribution? (2 pts)
3. Which do you think gives the better visualization of the shape of the distribution: the histogram, or the box plot? Why? (2 pts)
4. Do you expect the ranges given by the empirical rule you found in problem 5 will be accurate in the entire data set? Why or why not? (2 pts)
5. How strong is the correlation you found in question 6? What does that tell you about the relationship between HCC score and average total drug cost? (2 pts)

**Grading Rubric (21 pts total)**

|  |  |  |
| --- | --- | --- |
| Total Score | Grade Category Feedback | Gradebook Entry |
| 1 – 5 pts | Substantial revision needed | 1 (F) |
| 6 – 8 pts | Revisions needed to pass | 2 (D) |
| 9 – 11 pts | Minimum passing | 3 (C) |
| 12 – 14 pts | Above minimum standards | 4 (B) |
| 15 – 17 pts: | Strong demonstration of mastery | 5 (B+/A-) |
| 18 – 20 pts: | Adaptation, Original Ideas, Extra Independent Effort | 6 (A) |
| 21 pts: | Exemplary work above course level | 7 (A+) |

You may collaborate by discussing the assignment with peers, tutors, or the instructor. You may not copy and paste work, or type word-for-word or symbol-for-symbol, from another student’s work.

Each student will submit their work individually in Brightspace, and standard rules for academic integrity will apply. If you submit before the draft deadline, you will receive feedback. This gives you an opportunity to revise and resubmit to improve your score.

Draft Deadline 1: Saturday 5/1, 11:59pm

Draft Deadline 2: Saturday 5/8, 11:59 pm

Final Deadline: Saturday 5/15, 11:59 pm