**Homework 2**

**Your student #**

|  |
| --- |
| **Homework Guidance**  To ensure the integrity and educational value of your work, please adhere to the following guidelines as you complete Homework 1:  **1) Originality and Collaboration**: You are encouraged to work independently and ensure that the submissions are your own. While collaboration with your peers for understanding concepts and discussing problems is allowed, directly copying work from your colleagues is strictly prohibited. Your submission should reflect your understanding and your ability to apply what you have learned.  **2) Use of GPT-like Platforms**: You are permitted to use GPT-like platforms for assistance with your homework. However, this tool should only be used when you fully understand the answers it provides. The purpose of using such platforms is to enhance your learning, not to bypass the learning process. ***Keep in mind that midterm and final exams will be conducted without internet access***. If there is a significant discrepancy between the code you submit for homework and your ability to write similar code during an exam, ***it will be considered cheating***. Such instances will result in a score of zero for the involved exam component.  **3) Submission Quality**: Your focus should be on submitting code that you comprehend thoroughly. Fancy or complex code that goes beyond your level of understanding is not the goal. We value honesty and genuine effort. Make sure that you can explain and justify every line of code you submit. This approach will not only help you in your homework but also prepare you for the no-internet exams. |

Your coding HWs (check week02 pdf files) from 1 to 4 should be answered here:

1.

2.

3.

4.

5. Levels of lead contamination in soil samples from various urban playgrounds were documented in a study by Johnson (1985). The data below represent the concentration of lead (in mg/kg) at different depths. Calculate the following statistical measures to evaluate the central tendency and variability of lead contamination. Then, discuss why these estimates might differ.

A. Mean

B. Trimmed mean (at 10-percent trimmed and 20-percent trimmed)

C. Geometric mean

D. Median

|  |
| --- |
| **Lead concentrations (in mg/kg)** |
| 0.001 0.030 0.10 0.003 0.040 0.454  0.007 0.51 0.49 0.020 0.077 1.02 |

6. Using the lead concentration data from Exercise 1, compute the following statistical terms to further understand the distribution and variability of the data. Discuss the differences in the values from A through D and their implications for environmental health.

A. Standard deviation

B. Interquartile range

C. Median absolute deviation

D. Skew and quartile skew

7. Total Nitrogen (TN) (mg/L) was measured in samples of precipitation by Oltmann and Shulters (1989). Some of their data are presented below. Compute summary statistics for these data. Which observation might be considered an outlier? How should this value affect the choice of summary statistics used?

A. To compute the mass of TN falling per square mile.

B. To compute a typical concentration and variability for these data?

|  |
| --- |
| **Total Nitrogen (TN) (mg/L)** |
| 0.3 0.9 0.36 0.92 0.5 1.0 0.7 9.7 0.7 1.3 |

A) Assuming the outling observation is accurate, representing some high-nitrogen location which is important to have sampled, the mean must be used to compute the mass of nitrogen falling per unit area. It would be best if it computed by weighting the concentrations by the area at each one represents. The median would under-represent this mass loading.

B) The median would be a better “typical” concentration, and the IQR a better measure of the “typical” variability than the mean and standard deviation. This is due to the strong effect of the one unusual point on these traditional measures of location and variability.