

Assignment #2

(Due on: December 11, 2018 at mid-night)

Problem 1

Implement a function that takes a vector of numbers as an input and outputs the values needed to plot the boxplot. Specifically, the function should output the 1st quartile (q1), 2nd quartile (q2), 3rd quartile (q3), the inter-quartile range, a vector containing the outlier points and the vector containing the extreme outlier points as defined in the lecture. Apply your function to the data given in the file Data1.txt.

Deliverables:

- Your code.
- The values of q1, q2, q3, inter-quartile range, outlier points, and extreme outlier points.

Problem 2

Implement a function to verify the Central Limit Theorem using throwing a die example. The steps of the function should be as follows:

1. Simulate the outcome of throwing a die 1000 times: Use a population of discrete values (1, 2, 3, 4, 5, 6) with uniform distribution.
2. Plot the histogram (100 bins).
3. Simulate the outcome of throwing TWO dice 1000 times: by generating two samples as in step 1.
4. Calculate the average of the resulting values of the two dice in each time.
5. Plot the histogram of the average value (100 bins).
6. Calculate the mean and variance.
7. Simulate the outcome of throwing TEN dice 1000 times: by generating ten samples as in step 1.
8. Calculate the average of the resulting values of the ten dice in each time.

Assignment #2

(Due on: December 11, 2018 at mid-night)

9. Plot the histogram of the average value (100 bins).
10. Calculate the mean and variance.

Deliverables:

- Your code.
- The histograms obtained in steps 2, 5 and 9.
- The mean and variance values obtained in steps 6 and 10.

Problem 3

Implement a function that computes the p-value to test the hypothesis that the mean of one dataset is equal to the mean of another dataset. The function should take as inputs the two datasets and the significance level, and outputs the p-value in addition to the boundaries of the acceptance region. The significance level is 0.05.

Deliverables:

- Your code.
- Apply the function to the two datasets “Data3-1.txt” and “Data3-2.txt”. What is the p-value in this case? Plot the histogram of both datasets.
- Apply the function to the two datasets “Data3-1.txt” and “Data3-3.txt”. What is the p-value in this case? Plot the histogram of both datasets.

Important Notes:

- **Do not use Python built-in functions for computing q1, q2 and q3 in Problem 1. However, you can use the sorting and median functions. You are also allowed to use the random number generators needed for Problem 2. In Problem 3, you can use functions that compute the inverse of the Gaussian cumulative distribution function.**
- This is an individual assignment. It is not a team assignment.