

Lab 3 (100 pts.) - Assessing the Normality of Data**Objectives: Creating and Interpreting Normal Probability Plots (QQ plots)**

In addition to the regular graphs, code, and interpretations, please submit the data for parts B, C and D ONLY (clearly labeled) **at the end of the lab report** in an Appendix to your individual submission. Make sure each section is clearly labeled. Please DO NOT include the data for parts E or F. Remember that you are only supposed to include information that is asked for. In this lab, we are not interested in the mean and standard deviations of the distributions.

A. (10 points) Online Prelab

B. (10 points) Normal random numbers (no data file required) Use software to generate 10 observations from a normal distribution with mean, $\mu = 9$ and standard deviation, $\sigma = 4.5$.

(You must submit your data for this question. No credit will be given without data.)

1. (3 pts.) Code:
2. (3 pts.) Make a histogram of these observations. How does the shape of the histogram compare with a normal density curve?
3. (4 pts.) Make a normal probability plot of the data. Does the plot suggest any important deviations from normality? Please provide specifics to explain your answer.

C. (10 points) Normal random numbers (no data file required) Use software to generate 100 observations from a normal distribution with mean, $\mu = 9$ and standard deviation, $\sigma = 4.5$.

(You must submit your data for this question. No credit will be given without data.)

1. (2 pts.) Code:
2. (3 pts.) Make a histogram of these observations. How does the shape of the histogram compare with a normal density curve?
3. (3 pts.) Make a normal probability plot of the data. Does the plot suggest any important deviations from normality? Please provide specifics to explain your answer.
4. (2 pts.) Compare and contrast the plots from part B and part C. Remember both of these parts are from the same normal distribution.

D. (40 points) Random numbers from other distributions (no data file required)

Use software to generate 100 observations from the following distributions (10 points for each of the distributions) (I) right skewed (Exponential), (II) left skewed (Beta), (III) short tailed (Uniform) and (IV) long tailed (t-distribution). **The code for how to generate the functions is in the tutorial.** You are required to use the exact code given in the tutorials including the parameters. The format of the lab report should be: a) a section title which indicates the distribution under consideration (I, II, III, or IV), b) the code for that section c) the histogram with answer to 2 below, d) the normal probability plot with answer to 3 below.

(You must submit your data for this question. No credit will be given without data.)

1. (2 pts. for each distribution) Code:
2. (4 pts. for each distribution) Make a histogram of these observations. Please describe the shape of the distribution. How does the histogram compare with a normal density curve?
3. (4 pts. for each distribution) Make a normal probability plot of your data. Please describe the shape of the plot. Does the plot suggest any important deviations from normality? Please provide specifics to explain your answer. You should be able to use your explanation to determine which of the four types of distributions generated the plot when you encounter plots later in the semester.

E. (15 points) Comparison of data – GROUP (submitted separately on Blackboard)

Each group must consist of 3-4 people and will submit only one combined report. Even though it is due at the same time as the rest of the lab report, please submit this part as a separate pdf file under the “Lab 3 Group” link in Blackboard. Be sure that the names of all group members with their section time(s) are at the top of the page. If your grade is not on Blackboard (give us a week to grade the assignment), submit a regrade request with the names and sections of your group members. In addition, please indicate which person submitted the report so that we can find the grade.

Note that you still need to submit the answers to Parts B, C, D, and F as your individual report for Lab 3 in Blackboard.

1. (5 pts.) Present all of the graphs for Parts B, C and D from all of the members of the group where the graphs for the same distribution and number of data points are grouped together. Therefore, there will be 3-4 histograms and 3-4 normal probability plots for each type of random number simulation (6 sets in total: 1 from Part B, 1 from Part C, and 4 from Part D).
2. (10 pts.) For each of the six sets (histogram and normal probability) of plots (1 from Part B, 1 from Part C and 4 from Part D), please answer the following question: Do these plots look reasonably similar or do they look different? If they look different, please propose a possible explanation.

F. (15 points) The distribution of Assaults Per Population (data file: Clean US Data)

We are interested in the normality of the Number of Assaults per 100,000 people (AssaultsPerPopulation).

1. (5 pts.) Code:
2. (5 pts.) Make a histogram of these observations. Which distribution do you think this data is (normal, right skewed, left skewed, short tailed or long tailed)? Please explain your answer.
3. (5 pts.) Make a normal probability plot of these data. Which distribution do you think this data is (normal, right skewed, left skewed, short tailed or long tailed)? Please explain your answer.