

Statistics Project—Distribution Comparison by Quantiles

Background Thinking: The Five-Number Summary (FNS)--minimum, first quartile (Q1), median (M), third quartile (Q3) and maximum--is a robust¹ means of describing the distribution of univariate quantitative data. If we suspect that our data is from a normally distributed population then it seems natural to compare the FSN of the data set to the FSN of a normal distribution. The idea is that in some sense Q1, M, and Q3 should correspond to z-scores of -0.67, 0, and 0.67. If the data is normally distributed then the line $y = \mu + \sigma z$ should go through (-0.67,Q1), (0,M), and (0.67,Q3).

The Question: Suppose we reverse the process and use a line fitted to the points (-0.67,Q1), (0,M), and (0.67,Q3) as a proxy for the $y = \mu + \sigma z$ line? We can call this the qq-line. Here are some interesting possibilities:

1. Since FSN is robust do we get a robust estimate of σ from the slope of the qq-line?
2. Is the discrepancy between our σ estimate and the sample standard deviation a measure of un-normalness? (Hint: Use sample data from a population that is NOT Normally distributed.)
3. Can we extend this idea beyond quartiles to quintiles, deciles, percentiles, etc?
4. Can this process be used for distributions other than the Normal distribution?

Getting Started: Have your instructor assist you with creating an Excel sheet that will generate sample data from a Normal distribution. Add to your Excel sheet to generate the qq-line (you can use the Excel functions SLOPE and INTERCEPT). Experiment with this Excel sheet until you have some intuition as to how this will go.

First Meeting: Prior to the meeting do the “Getting Started” exercise. Meet with your instructor and explain your ideas. Determine the following for your project—what questions to test, what resolution to use, the response variable, the explanatory variable(s), how many tests to run, how to tabulate the data, and what could go wrong. Write a summary of this meeting and turn it in to the instructor.

Excel Platform (Second Meeting): Develop an Excel sheet to run your experiments. Document how your excel sheet works. Review this test setup with your instructor.

Conduct Tests and Compute Statistics: Your team should end up with a table of test results and a table of statistics based on the test results.

Third Meeting: Present your results. If there are problems with the testing procedures resolve them and run the tests again. Discuss what conclusions are justified. Discuss every section of the project report and what should be there.

Write Project Report Draft: Use the guidance from the Third Meeting to write your draft report.

¹ “Robust” is used to mean insensitive to outliers.

Fourth Meeting: Present your draft report to the instructor. Use your instructor's critique to write the final report.

Write the Project Report.

Additional Project Guidelines:

Due Dates

[Report Format](#)

[Report Writing Cautions](#).