

Scripting Assignment 2: Intervals and distributions

Points: 30

Due: Before class, Thursday, March 2nd

Evaluate the following three questions using R and the provided datasets. Submit your script, outputs, and discussion of your answer in a single R notebook.

- Email both your rmd file and the html file to me.
- Add what you determine to be your best graphical output from this assignment as a single figure to the Best_Graphs_1DistFitting.pptx on the google drive. Include a short caption below the plot but don't include your name.

1. (5 points) A well yield of 0.85 gallons/min/foot was measured in a well in Virginia. Is this yield likely to belong to the same distribution as the previous well yield measurement provided below, or does it represent something larger? Answer by computing 95% parametric and nonparametric intervals. Which interval is more appropriate for these data?

Well Yields in Virginia (gallons/min/foot)

0.001	0.03	0.1	0.003	0.04	0.454
0.007	0.04	0.49	0.02	0.077	1.02

HINT: Use the qt function in R to get quantiles from the t-distribution (See tutorial here for more info: <http://www.r-tutor.com/elementary-statistics/probability-distributions/student-t-distribution>)

2. (10 points) LeeFerry-mon-data.txt contains monthly streamflow observations in acre-feet at the Lee's Ferry station on the Colorado River.
 - a. Fit several parametric distributions to each month of data
 - b. Summarize your results visually using histograms, q-q and p-p plots
 - c. Calculate K-S statistics for each distribution
 - d. Discuss what the best distributions are and whether you think different distributions are warranted for different months of data.
3. (10 points) Daily_Rain.csv contains daily precipitation data for the period 1889-2005 for a station in Southwest Australia (available from <http://www.isse.ucar.edu/extremevalues/projects.html>, project 1, station 9619).
 - a. Obtain winter season (May – Oct) maximum rainfall – i.e. 'block maxima' resulting in one value per year.
 - b. To this annual maximum precipitation compute 50-yr, 100-yr and 500-yr return period rainfall by fitting extreme value distributions to this data – (a) Gumbell EV-1, (b) Log-Pearson Type III, (c) Log Normal and (d) Generalized Extreme Value distribution. Use the 'fExtremes' or the 'Extremes' toolkit library in R.
 - c. Plot the histogram of the data and overlay the four fitted PDFs and develop four Q-Q plots - on the distribution.
 - d. Discuss which distribution you think fits best and explain why.

Grading:

25 points divided by question: Complete answers, appropriate methods used, and thoughtful analysis of findings

5 points: Clarity of script and quality of visuals