Statistics 3080
Homework 8
Due: Wednesday, March 28

Complete the following problems in a R Markdown file and submit your compiled PDF.

Before you begin, you should set the random seed to your birthday month, day, and year. For example, March 28, 1992 would be 3281992. You should set the seed only once in your entire document.

Problem 1: The file *weather_full.txt* contains the high temperature in Charlottesville for December days over the past two years.

(a) Create a histogram of the data with white bars outlined in black using 10 degree bins. Make sure that the x-axis is descriptively labeled. Describe the shape of the population distribution.

(b) Use a Monte Carlo simulation with 10,000 repetitions to estimate the mean and the standard deviation of the sampling distribution of the median of a sample of 15 December days.

(c) Create a histogram of the estimated sampling distribution of the median of a sample of 15 December days with white bars outlined in black using 5 degree bins. Describe the shape of the estimated sampling distribution.

(d) Create a histogram of the estimated sampling distribution of the median of a sample of 15 December days with white bars outlined in black using 3 degree bins. Describe the shape of the estimated sampling distribution.

(e) Based on the shape of the estimated sampling distribution of the median of a sample of 15 December days, does using a methodology based on the Central Limit Theoreom seem appropriate to estimate or test the median?

Problem 2: The file *weather_samp.txt* contains the high temperature in Charlottesville for 15 randomly selected December days over the past two years.

(a) Create a histogram of the sample with white bars outlined in black using 10 degree bins. Make sure that the x-axis is descriptively labeled. Describe the shape of the sample distribution and indicate whether the assumptions for using bootstrapping hold. Explain.

(b) Determine the bootstrap 95% confidence interval for the median high temperature in Charlottesville in December.

(c) Use the bootstrapped medians to estimate the standard deviation of the sampling distribution of the median high temperature of a sample of 15 December days. Use this value to determine the one-sample 95% z confidence interval around the median.

(d) Which of the bootstrap 95% confidence intervals is wider? Is this consistent with your answer to Problem 1 part (e)?

Problem 3: Airlines oversell seats on their flights to adjust for purchased tickets that are not used. Each airline uses a proprietary algorithm to determine how many seats to sell for a given flight. Suppose that a particular algorithm is designed such that only 5% of flights have more passengers arrive than seats on the plane with adjustment for variation by using the conservative limit of a 99% confidence interval. Suppose that one of the planes that flies to Charlottesville has 55 passenger seats that are sold according to this algorithm. The file *flights.txt* contains the number ofticketed passengers that arrived for a flight on this plane for a random sample of flights on normal flying days (no major weather system, no major airport delays, etc.).

(a) Determine the 99% bootstrap confidence interval for the 95th percentile of the number of arriving ticketed passengers using 10,000 repetitions.

(b) Identify the conservative endpoint of the resulting bootstrap confidence interval. Explain.

(c) Has the algorithm worked according to this sample of flights?