Exam 2 take home questions (ThE2)

Stat 422

Sections 1, 10

Fall 2018

Instructions:

Since this is part of exam 2, you are to work individually on these problems. Do not discuss these questions with anyone except Renae. Work each problem on a *separate page*, bring the take home problems (and all intermediate calculations) with you to the second exam. One or more of these problems will be handed in with the in-class exam, and information from others will be used on the in-class questions. [**General rule**: if N is given or known in a problem, use it to calculate the fpc regardless if $fpc \geq 0.95$; if N is not given or known, then there can be no fpc]

(1) A guitar store only sells guitars made by three companies: Gibson, Ibanez, and Taylor. A separate random sample was taken of the guitars for sale from each company. The data file called `guitars.csv` contains the selling prices for each type of guitar in the sample.
  (a) Calculate an estimate of the average guitar price at the store, and an error bound for your estimate.
  (b) What sample size should we take in a future sample (and how would we allocate it among companies) to achieve an error bound of $200?
(2) A random sample of 59 houses was collected from the houses listed for sale in the Boise, Idaho area. The data file called `boisehouses.csv` contains the listed price for each house as well as the square footage of the house. The mean square footage of all 20,000 houses for sale is 2524 square feet. From this data, estimate the average listing price in the Boise area and the bound using the following methods:
  (a) SRS
  (b) Ratio
  (c) Regression
  (d) Create a scatterplot of the data (just one graph)
(3) A systematic random sample was taken from the set of all Presidents of the United States. The data file `potus heights.csv` random sample includes the height (in inches) of each sampled President.
  (a) From this data, estimate the average height of United States Presidents. Calculate two error bounds for your estimate, one using the usual SRS formula, and one using the successive difference variance estimator.
  (b) Which variance estimator is more appropriate for these data? Briefly explain.
(4) For the housing data in problem 2, can you think of a way to design a future sampling study to have a lower error bound? (For example, how could more information be used?)