Statistics 411/511
Homework 3
Due Tuesday, October 20 by midnight Pacific time

- **Instructions:** Upload homework to Gradescope via Canvas (access specific homework assignments from the Assignments link on the Canvas course page). Your file must be a pdf document. Please see the end of the syllabus for formatting guidelines.
- The problems are assigned from the **third edition** of the textbook. If you have another edition, consult the copy on reserve at the library website for the homework problems.
- **Academic Integrity** You are encouraged to *discuss* the homework with other students, but what you turn in must be your own work in your own words. **DO NOT** copy someone else's homework. You may share ideas and R code, but do not share R output or written language. The syllabus contains details and links to OSU's Student Conduct Code and procedure for reporting suspected academic misconduct.

1. This exercise is intended to give you practice log-transforming data and reporting result of an analysis on log-transformed data. You will work with the out of state tuition data from the college tuition data of exercise 32 on page 83. The data frame is `ex0332`.
   (a) Examine the structure of the data frame using `head()`. Turn in your R code and output.
   (b) Obtain "summaries" of the out of state tuition for public universities and for private universities using `summary()`. See item 11 of Lab 1 for example code. Turn in your R code and output.
   (c) Use the example code on page 5 of Outline 3 to produce histograms for the two college types. Turn in your R code and graph. [The argument `xlim` is the minimum and maximum values for the horizontal axis. If you omit this argument, R will try to pick sensible values.]
   (d) Log-transform the out of state tuitions. Obtain summaries as in b and histograms as in c using the logged data. Turn in your R code, summary output, and graph. [Remember: "log" means natural log in Statistics.]
   (e) State the three assumptions needed to use the t-tools. For each assumption, state your opinion whether it is reasonable for the untransformed data, then for the transformed data. Give a reason for your opinion. You may be very brief here. One or two sentences per assumption will be enough.
   (f) The t-tools are robust to many departures from the assumptions, so even if you don't believe the assumptions are met, perform a two-sample t-test using R on the logged data to answer the research question, "is out of state tuition higher at private universities?" Submit your R code but not output.
   (g) Give a "statistical conclusion" reporting the results of your hypothesis test in part f. A statistical conclusion should be on the original scale, not the log scale.

(Problem 1 continued on next page)

(h) Obtain a two-sided 95% confidence interval for the difference in population means, using the logged data, then back-transform the endpoints of the interval. Submit your R code and the resulting back-transformed interval.

(i) Give a "statistical conclusion" reporting your back-transformed confidence interval from h