Statistics 411/511
Homework 2
Due Tuesday, October 12 by midnight

- **Instructions:** Please see the end of the syllabus for guidelines. Upload your homework to Gradescope via Canvas (access specific homework assignments from the "Assignments" link at the left of the Canvas course page). Your file must be a pdf document. **There will be a one-point deduction if you don't assign pages** (see this Gradescope help video).
- Do the computational part of the homework shortly after completing the week's lab activity.
- The problems are assigned from the **third edition** of the textbook. If you have another edition, consult the copy on one-hour 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. In item 7 of Lab 2, you downloaded the data file cats.csv. This spreadsheet contains data from a study comparing average daily water drunk by cats exposed to either still or flowing water. The data are from a study reported in C. Pachel and J. Neilson (2010), "Comparison of Feline Water Consumption Between Still and Flowing Water Sources: A Pilot Study," Journal of Veterinary Behavior, Vol. 5, pp. 130-133.

   (a) Calculate the difference between still water and flowing water consumption for each cat. Include your R code. Note that if you shut down R since you did the lab, you will need to read in the data again.

   (b) Produce a histogram of the differences you calculated in part (a). Include your R code and the histogram.

   (c) Calculate the sample mean, sample standard deviation, and sample size of the differences you calculated in part (a). Include your R code and results.

   (d) Calculate the t-statistic for a paired t-test of the null hypothesis that the mean difference is 0. See bubble 3 in Display 2.6 for a formula. Include your R code and result.

   (e) Check your work by performing the same paired t-test described in part (d), this time using R's `t.test()` function. Include your R code and output.

2. Consider the rat data in Exercise 25 on page 25 (`ex0125`). Assume that the researchers randomly assigned the rats to the two groups, and that they are interested in the following research questions.

   i Does calcium supplementation reduce blood zinc levels in rats?
   ii By how much does calcium supplementation reduce blood zinc levels in rats?

   (a) Produce side-by-side boxplots of the data. Include your R code and your plot.

(Problem 2 continued on next page)

(b) Write a statistical conclusion to answer the above two research questions.

3. The *sampling distribution* of a statistic is an important concept in statistical inference. Page 15 of the textbook describes what is meant by the sampling distribution of the statistic $\overline{Y}_2 - \overline{Y}_1$. Display 2.3 illustrates the sampling distribution of the sample mean $\overline{Y}$ in a one-sample (paired) situation. You may recall that the *Central Limit Theorem* says that the sampling distribution of the sample mean is approximately normal, as long as the sample size is large enough.

   This homework problem aims to give you the opportunity to ponder the concept of a sampling distribution. There is no need to do any computing or calculation. Display 2.3 might be helpful. Each part can be answered in one or two sentences.

   (a) Describe what is meant by the sampling distribution of the sample mean.

   (b) Consider the sampling distribution of the sample standard deviation. Will this sampling distribution be the same as the sampling distribution of the sample mean? Why or why not?