**HOMEWORK 2**

*Inference for one mean and the difference of two means*

For this reading and assignment, use the free online textbook, *OpenIntro Statistics, 4th Edition*. To see the textbook, go to <https://www.openintro.org/book/os/>. Click on *Free – OpenIntro Statistics PDF*, and then click on *Read Free Sample* on the left. A pdf version of the textbook should open.

Reading: This assignment focuses on content from Sections 2.1, 7.1 and 7.3. Read those sections. OpenIntroStats introduces sampling distributions using sample proportions instead of sample means. That material is in section 5.1, but much of that material concerns details for proportions that don’t apply to sample means.

Notes:

* This assignment requires the use of JMP. See the JMP information posted on Canvas if you don’t already have it installed on your personal computer. Problems resulting from not getting JMP working will not allow you to submit your assignment late. Please plan to work ahead and email your instructor questions if needed.
* See the JMP information posted on Canvas on how to save JMP output into files or copy it into a Word document.
* Round all numbers to 2 decimal places unless otherwise specified.
* In Canvas, the assignment is set up as a quiz. Please either type or copy and paste your answers into the appropriate places.

Complete the following questions from the textbook.

1. 7.2 (page 259) T distributions (answer each with solid, dashed, or dotted)
   1. normal
      1. Solid
   2. T with 5 df
      1. Dashed
   3. T with 1 df
      1. Dotted
2. 7.12 (page 261) lead exposure, parts a and b only. Treat the “suburb data” as if the average blood level there were known to be exactly 35 μg/l.
   1. The blood lead concentration of police officers is not different from the average of 35ug/l.
   2. Conditions for inference
      1. Random sample: Random sampling not stated; cannot be sure
      2. N = 52 > 30.
      3. Independent sample: an officer’s blood lead concentration is independent of other officer’s concentration.
   3. Z = (124.32 – 35) / (37.74 / sqrt(52)) = 17.07, p < .00001. The blood lead concentration 124.32 ug/l of police officers who are subject to constant inhalation of automobile exhaust fumes is significantly higher than the average concentration of 35 ug/l in the suburb with no history of lead exposure.
3. Lead exposure. The lead.csv file contains simulated data for the 52 urban police officers. For this problem, assume that these 52 officers are a random sample from the population of all police officers in this urban area. Use JMP to:
   1. Calculate a 95% confidence interval for the mean blood lead concentration for police officers in this urban area. Report the:
      1. 113.81215
      2. 134.82593
   2. Test the null hypothesis that the “urban” mean is the same as the suburban mean of 35.
      1. P < .0001
   3. Write a one-sentence conclusion from this hypothesis test. Use a scale of evidence to write your conclusion.
      1. There is convincing evidence that the mean blood lead concentration is significantly higher than the average of 35.
4. 7.50 (page 299) College credits. Note: “point estimate” is OpenIntroStats’s name for what I have called the “estimate”. It is a single number. The reason for “point estimate” is that a confidence interval is sometimes called an “interval estimate”.
   1. Mean 13.65
   2. Median 14
   3. Standard deviation 1.91
   4. IQR 2
   5. 16 credits, Surprised? (yes/no) Note: In this histogram, values of exactly 16 are included in the 15-16 bar.
      1. No, because it’s included in the histogram, and isn’t even an outlier.
   6. 18 credits, Surprised? (yes/no)
      1. Yes. 18 is the max in that sample and is very rare. It’s got a P-value of less than .00001, approaching 0. (Z = 22.77)
   7. Measure of variability of the estimate.
      1. Standard Error of the Mean
      2. SE = s/(sqrt(n)) = 1.91/sqrt(100) = .191
5. 7.58 – Age at first marriage. Find two errors.
   1. First error (free answer):
      1. Hypothesis should be in terms of population mean, not sample average
   2. Second error (free answer):
      1. null and alternate hypothesis are swapped.

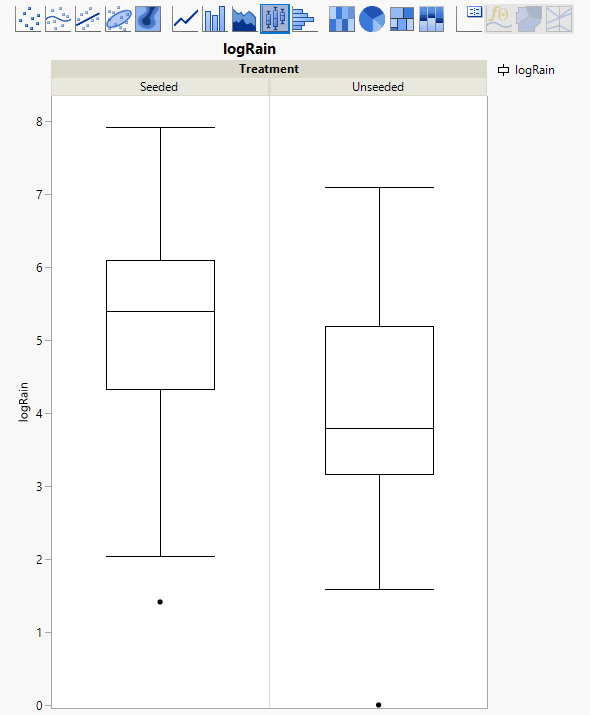
Answer questions 6 - 9 based on the following scenario and dataset posted on Canvas.

A simple random sample of students at a university (not ISU) were surveyed. The variable we will look at first is *Exercise* which refers to the number of hours of exercise per week.

1. Research Question 1: What is the population mean number of hours exercised for students at the university where the data was collected?
   1. Why would it be appropriate to use a confidence interval to answer Research Question 1? (free answer)
      1. We don’t have data for the population, only a sample, so this average would just be an estimate of the parameter.
   2. Are the conditions met to create a 95% confidence interval for the *Exercise* variable?
      1. Yes
   3. briefly explain (free answer)
      1. Random sample: SRS
      2. Independent samples: measuring one person’s exercise doesn’t affect other measurements
      3. N > 30 OR Normal Distribution: N > 30
2. Use JMP to create a 95% confidence interval for the *Exercise* variable. Note that you will likely use the *Analyze, Distribution* analysis in JMP, however there are other ways to get the same results.
   1. lower bound of the confidence interval: 8.4598
   2. upper bound of the confidence interval: 9.6482
3. Write a one sentence conclusion about the *Exercise* variable that includes the confidence interval.
   1. The population mean of time students at this university spend exercising is 95% likely to be between 8.4598 and 9.6482.

The data in cloud.csv are from a randomized experiment to evaluate the effectiveness of cloud seeding. Cloud seeding is the practice of spraying (from airplanes) large amounts of silver iodide into clouds with the intent of stimulating formation of rain drops and increasing the amount of rain that falls from the cloud. On 52 days, meteorologists identified clouds that were considered suitable for seeding. On each day an airplane was dispatched to fly through the clouds. A random mechanism was used to decide whether that airplane sprayed water (unseeded, control treatment) or a silver iodide solution (seeded, active treatment). The total amount of rainfall produced by that cloud was measured by radar. For reasons we will discuss later, the appropriate measure is the natural logarithm of the rainfall. This is the logRain variable.

NOTE: there is also a Rainfall variable in this data set. Do not use Rainfall. Use logRain.

1. Draw side-by-side vertical box plots of the logRain values in the two treatments. Your answer is the graph.
2. Looking at the box plots, does there appear to be an effect of seeding on the average of the logRain variable?
   1. Yes
   2. Explain what in the box plot suggests an effect or lack of an effect (free answer).
      1. The centers visually differ significantly. Additionally, the quartiles differ, too.
3. Looking at the box plots, do the two groups (seeded and unseeded) appear to have similar variability?
   1. Yes
   2. Explain what in the box plot suggests similar (or not similar) variability.
      1. The quartiles are roughly the same size
4. Are these data paired (answer yes) or two independent samples (answer no)?
   1. No
   2. Explain your choice (free answer)
      1. Each individual cloud did not receive both treatments. Random clouds were selected, and treatment was assigned randomly. These are independent clouds.
5. Using JMP, calculate a 95% confidence interval for the difference in logRain means.
   1. lower bound of the confidence interval: -2.0467
   2. upper bound of the confidence interval: -0.2409
6. Using JMP, test the null hypothesis that cloud seeding has no effect on the amount of rain produced by a cloud, as quantified by the logRain variable. Report the p-value for this test.
   1. p = .0141
7. Use your confidence interval and p-value to write a one or two sentence conclusion about the effect of cloud seeding.
   1. We can be 95% confident that seeded clouds rain between .2409 and 2.0467 (unit) more than unseeded clouds. This is a significant affect because p = .0141 < .05, reject null hypothesis.
   2. There is significant evidence that seeded clouds have a different amount of rain than unseeded. (p-value = .0141). The estimated effect is -1.1438 with a 95% confidence interval of [-0.2409, 2.0467].