ST511 HW #4

Reading: Review Chapters 5 and 6 of Ott & Longnecker.
See Canvas Calendar for due date.

42 points total, 2 points per problem part unless otherwise noted.

1. The housing department in a large city wants to estimate the average rent for rent-controlled apartments in the city. They need to determine the number of renters to include in a survey in order to estimate the average rent to within $80 using a 95% confidence interval. From past surveys, the monthly charge for rent-controlled apartments ranged from $1600-$3200.
    A. Suppose that based on the previous survey, almost all (>99%) apartment rents fell within $1600-$3200. Use this information to "estimate" the standard deviation.
    B. Using the standard deviation from above, find the (minimum) sample size required to achieve a 95% ME < $80.

2. A national agency sets recommended daily allowances for many supplements. In particular, the allowance for zinc for adult men is 15 mg/day. The agency would like to determine if the average intake of zinc for adult men is <u>greater than</u> 15 mg/day. Suppose from a previous study they estimate the standard deviation to be 2 mg/day and they conjecture that the true population mean is 15.4 mg/day. The investigators plan to use a one-sample t-test with $\alpha=0.05$.
    A. Find the power with n=120 for the scenario above.
    B. If the standard deviation was smaller (less than 2) would the power be higher or lower than that calculated in part A?
    C. If the sample size was larger (more than 120) would the power be higher or lower than that calculated in part A?
    D. If we used $\alpha=0.01$ (instead of 0.05), would the power be higher or lower than that calculated in part A?
    E. Using a conjectured mean of 16 mg/day (instead of 15.4), would the power be higher or lower than that calculated in part A?
    F. Return to the original scenario and find the sample size required to achieve 80% power. Remember to "round" up to an integer value.

3. Use the data from Problem 5.27 which deals with lead concentrations in estuarine creeks.
    A. Construct a histogram, qqplot and run SW test of normality. What do you conclude about the normality of the data based on each of the criteria? Do the various plots and tests agree? (**4 pts**)
    B. Give the sample mean and median for this data.
    C. Use the sign test to test the null hypothesis that the <u>median</u> is equal to 30. Give the p-value and make a conclusion.
    D. Give a 95% confidence interval for the <u>median</u>. Note: For consistency, please report the "Upper Achieved CI".

E. Give a (standard) 95% confidence interval for the mean.
F. It should be clear from the diagnostics in part A that the assumption of normality is not met. Hence the CI from previous question is suspect. Give a 95% bootstrap studentized confidence interval for the mean. Hint: See "Boot Example2", but use a different value for set.seed.
G. Assuming that cumulative lead exposure is of interest, would the mean or the median be of more interest.

4. Use the data from problem 6.6 which concerns dissolved oxygen readings for Above and Below town sites. Note: The values for the Below site do not match what is shown in the textbook.
   A. Construct the side-by-side boxplots and include them in your assignment.
   B. Give the sample means and standard deviations for each site (Above and Below).
   C. Considering the summary statistics from above, is the pooled variance t-test or Welch-Satterthwaite t-test preferred here? Justify your response using the rule of thumb from the notes.
   D. Without assuming equal variances, give the 95% confidence interval for the difference between the means. Based on this interval, can we conclude that there is a difference between the population means? Explain.
   E. Run the Welch-Satterthwaite t-test to test H0: $H_0: \mu_1 - \mu_2 = 0$ versus a two-sided alternative. Give the p-value and conclusion.

Note for Q4: The data is in "wide" format. All questions can be answered using the current format. An alternative is to "transpose" the data to "long format". This is NOT required, but may be handy. For example:

```
library(tidyverse)
OxygenTr <- Oxygen %>%
   gather(key = "Site", value = "y")
str(OxygenTr)
```