Stat 311 Homework 6

This assignment requires the ice cream, birthweight, and cholesterol data sets provided with the assignment. See the data description documents for the ice cream and birthweight data sets for keys to coded variables and applicable units.

- Problems 1 and 2 use the ice cream data set. Problems 3 and 4 use the birthweight data set, problem 5 uses the cholesterol data set and problem 6 does not require a data set.
- Use the *T* distribution for all problems involving means, even if the sample size is large.
- Make sure to interpret all confidence intervals in the context of the problem.
- For all hypothesis tests, be sure to state the null and alternative hypotheses using symbols, the value of the test statistic (including *df* if applicable), the *p*-value, the decision, and a conclusion in the context of the problem.
- You will need to install the BSDA library before you start. You need this library for problem 6.

Ice Cream Data Set

1. Estimation
   a. Find and interpret the 90% confidence interval for the population mean puzzle score.
   b. Find and interpret the 90% confidence interval for the difference in mean puzzle scores for males and females (use males – females)
2. Testing a claim
   a. Create two subsets of video scores, one for students that prefer chocolate ice cream and one for students that prefer vanilla ice cream. Calculate the sample standard deviations.
   b. Test the claim that students with a preference for chocolate ice cream have mean video scores that are different that students that prefer vanilla ice cream. Use vanilla minus chocolate and a 10% significance level. Assume the population variances are not equal.
   c. Do you think the statistical test results from part (b) have practical significance?

Birthweight Data Set

3. Find the 99% confidence interval for the proportion of smokers in the birthweight data set. Assume large sample conditions are met.

Smoker:
Mean = 0.227
SD = 0.42
$SE(\bar{x}) = \frac{0.42}{\sqrt{42}} =$

Non-Smoker
p - 1.68
df = 41

99% CI = $0.227 \pm (1.68)($

4. Consider birthweights for mothers that are smokers and nonsmokers.
   a. By-hand, test the claim that the proportion of low birthweight babies is higher for mothers that smoked (use smoked – did not smoke) at the 1% significance level. Show your work, but you may use R to get the *p*-value. Assume large sample conditions are met and round all intermediate calculations to four decimal places.
   b. Repeat the test in part a) using `prop.test` in R. You do not need to restate the hypotheses and other information, as you should get comparable results and the same conclusion. Show that the square root of the chi-square test statistic from `prop.test` is equal to the *z*-score you got in part (a), within rounding error.
   c. Are the large sample conditions met for the tests in (a/b)? Note, you can just look at the table values do decide/no calculations are necessary. Do you think the hypothesis test results in (a/b) are valid?

Cholesterol Data Set

5. This problem was modified from here. This study used a cross-over trial experiment to investigate whether eating oat bran lowered serum cholesterol levels. Twelve individuals were randomly assigned a diet that included either oat bran or corn flakes. After two weeks on the initial diet, serum cholesterol (mmol/L) was measured and then participants were "crossed-over" to the other diet. After two-weeks on the second diet, cholesterol levels were measured again.
   a. Using a 5% significance level, test the claim that a diet that includes oat bran decreases serum cholesterol. [Use OatBran - Cornflake]
   b. Construct an appropriate confidence interval that is equivalent to the test in part (a). [Use OatBran - Cornflake]. Explain your choice of interval, and report and interpret the interval.

No Data Set

6. In clinical experiments involving distinct groups of independent samples, it is important that the groups be similar in the important ways that affect the experiment. In an experiment designed to test the effectiveness of paroxetine for treating bipolar depression, subjects were measured using the Hamilton depression scale with the summary results given below (based on data from a "Double-Blind, Placebo-Controlled Comparison of Imipramine and Paroxetine in the Treatment of Bipolar Depression," by Nemeroff et al., *American Journal of Psychiatry*, Vol. 158, No. 6).

	n	$\bar{x}$	s
Placebo	20	21.57	3.87
Treatment	15	20.38	3.91

   a. Use a 0.05 significance level to test the claim that the treatment and placebo groups come from populations with the same mean. Assume equal population variances. [Use Treatment – Placebo]
   b. Do you agree with the assumption of equal population variances used in the test in (a)? Explain.