Stat 311 R Assignment 4

Use `rmarkdown` to create HTML file output that you then convert to pdf. Your final product will be a pdf file uploaded to Canvas.

The R4 Tutorial introduces all the functions you will need to complete this assignment. Create a new .Rmd file for this assignment. To get you started, we have provided a template with the title section, setup chunk and code to read in all data sets. Copy, paste and edit code from the tutorial as needed for the assignment. Remember to include section and subsection headers as needed to make your assignment readable. Only include the output needed to answer the questions.

This assignment requires the ice cream, birthweight, and cholesterol data sets that are provided with the assignment. Data description documents are also included for the ice cream and birthweight data sets.

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. **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, the value of the test statistic (including *df* if applicable), the *p*-value, the decision, and a conclusion in the context of the problem.**

Since we have custom functions, we need to also make use of a .R file with the functions so that the .Rmd file knits. You will notice the line, `source(Functions.R)` in the setup chunk. You will need this file. When you write your custom function for part 4a, you will also need to copy it to the bottom of Functions.R—we have put a placeholder in the file for you to add your code. Also, notice the `library(BSDA)` line in the setup. You will need to install the BSDA library before you get started. You need this library for problem 6.

1. Find and interpret the 92% confidence interval for the population mean puzzle score.

2. Hypothesis test for the difference between mean puzzle scores by ice cream type.

   a. Create two subsets of puzzle scores, one for students that favor strawberry ice cream and one for students that favor chocolate ice cream.

   b. Test the claim that students with a preference for strawberry ice cream have higher puzzle scores than students that prefer chocolate ice cream. Use strawberry minus chocolate and a 1% significance level. Assume the population variances are not equal.

   c. Do you think the statistical test results from part (b) have practical significance?

3. Using the oneprop.CI function in the R4 Tutorial, find the 98% confidence interval for the proportion of smokers in the birthweight data set. Assume large sample conditions are met.

Stat 311 R Assignment 4

4. Consider birthweights for mothers that are smokers and nonsmokers.

   a. Create a function called twoprop.HT that uses a two-sample z test to test for the difference between two population proportions assuming independent samples.

   b. Test the claim that the proportion of low birthweight babies is higher for mothers that smoked (use smoked – did not smoke). Use a 5% significance level. Assume large sample conditions are met.

   c. Are the large sample conditions met for the test in (b)? Note, you can just look at the table values instead of doing any calculations. Do you think the hypothesis test results in (b) are valid?

5. This problem was modified from here. A cross-over trial experiment was used 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 once again recorded.

   a. Using a 5% significance level, test the claim that a diet that includes oat bran decreases serum cholesterol. [Use Cornflk – OatBran]

   b. Construct an appropriate confidence interval that could be used as an equivalent to the test in part (a). [Use Cornflake – OatBran]. Explain your choice and interpret the interval.

6. In clinical experiments involving different 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 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). 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]

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