Homework 3

Instructions: This homework is due in class on Friday Apr 26.

Please read the following guidelines for presenting your work and follow them diligently.

- Write your full name clearly on the top right of the first page. **Staple** pages on the left hand corner. Write neatly in complete sentences.
- You are required to work all the problems, however, only 5 will be graded. The page numbers below refer to the fifth edition of the text.
- Answer the questions in the order in which they are posed. Clearly number the questions as I have.
- You must first work independently on the homework. Please post questions on the discussion board or come to office hours once you have tried the problems.
- Be sure to show/explain your work thoughtfully. How you write your answers is important.
- If you use R to make plots or as a calculator, it is enough to simply include the output (e.g., appropriately labeled plot) in the main part of your homework without the R code.

---

1. A rat poison is described by its lethal dose d. However, the true lethal dose for a specific rat, call it X, is a random varable. If $X < d$, the rat dies, otherwise the rat survives. Suppose $Y = \ln(X)$ is normally distributed with mean μ and variance σ^2 (both unknown).

   A manufacturer of the poison claims that if the dose $d = 1$, 60% of the rats will die, while if it equals 4, about 95% of the rats will die. In an experiment 150 rats are given the dose $d = 2$ and 95 of them die. Does this provide evidence that the manufacturer has overestimated the effect of the poison? (Most of the problem is about figuring out the null hypothesis value for a binomial probability!)

2. Recall the UMP test for testing $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$ you derived in problem 3 on HW 2.

   (a) Invert the test to obtain a $100(1 - \alpha)\%$ confidence bound for σ^2.

   (b) Make a sketch like in our notes to illustrate the duality between the hypothesis test and confidence interval for this problem.

3. The dataset Birth weight 1 in https://www.stat.berkeley.edu/~statlabs/labs.html shows birthweights for babies born to women who smoked during pregnancy and for those

who didn't.

(a) Load the dataset in R (or other data analysis program) and make exploratory plots of the birthweight distribution for the two groups of mothers. Make a summary statement which includes a measure of center, spread and shape for each distribution. Also comment on whether or there are any outliers and why you think so. (Be sure to exclude mothers with unknown smoking status - 9)

(b) Conduct a t-test to investigate whether the evidence suggests that mothers who smoke have babies who are lighter. You may use t.test in R but be sure to give formulas for the test statistic as well as the p-value also. Also be sure to write a conclusion summarizing your conclusion.

4. A psychological experiment was done to investigate the effect of anxiety on a person's desire to be alone or in company. A group of 30 subjects was randomly divided into two groups of sizes 13 and 17. The subjects were told that they would be subjected to some electric shocks, but one group was told the shocks would be quite painful and the other group was told that they would be mild and painless. The former group was the 'high anxiety" group and the latter was the "low anxiety" group. Both groups were told there would be a 10 minute wait before the experiment began, and each subject was given the choice to wait alone or with other subjects.

   The following are the results: of the 17 in the high anxiety group, 5 chose to wait alone. Of the 13 in the low anxiety group, 9 chose to wait alone.

   Does the data provide evidence that anxiety provokes solitude? Do an *appropriate* analysis. Be sure to set up the model, assumptions etc. in your answer. Do not just dive into calculations.

5. Wassermann, L. Prove Theorem 11.3 on page 188. Reflect on this theorem and comment on what you learn about the Type 1 error of the rule that rejects H_0 when the p-value < 0.05.

   (*Hint:* Let V denote the P-value of the test and $T(X)$ the test statistic with CDF $F_\theta(.)$. Begin by showing that

   (a) $V = 1 - F_{\theta_0}(T(X))$. (yes, V is a random variable)

   (b) Now find the CDF of V: $P_{\theta_0}(V \leq v)$.

)