Stat 139: Homework 2

Due: Friday, September 19, before noon

- **Incorporate the relevant R output in the HW write-up**. Choose the output wisely, no need to print-out dozens of pages. No more than **two pages** of output should be displayed for each problem and the relevant parts should be **highlighted**.
- **A hard copy of your solutions is due at noon on Friday, in your TF's drop-box outside the Science Center 300 suite**. The solutions should not be submitted by email unless arrangements are made with a TF prior to the deadline (or unless late, see below).
- **Your code should be attached to the end of your write-up.** Use comments to indicate the code associated with each question.
- See syllabus for details on homework and collaboration policies: **acknowledge your collaborators**; your lowest homework score will be dropped; solutions submitted electronically within 24 hours after deadline will be graded with a penalty; solutions more than 24 hours late will receive no credit.

Related Reading: Ramsey and Schafer, Chapter 1.

Suggested review sources:

- *Introduction to the Practice of Statistics*, Moore, McCabe, [and Craig if 5th ed.], Sections 3.4, 4.3-4.4;
- Appendix B.1 in A. Sen and M. Srivastava, "Regression Analysis: Theory, Methods, and Applications" (or any other textbook used in math courses that you took).
- Review Sheet under "Readings" on the course web-site.

1. (5 points) Describe a situation in which you had to make a choice, and there was a correct answer that was not available as you made your choice (e.g., "take an umbrella or not", "walk or take a T", etc.). Think about what information you used to help you decide, and then briefly answer the questions below:
   (a) What was the decision about? Identify a parameter of interest that you based your decision on.
   (b) What were your H_0 and H_1 about this parameter?
   (c) Based on what quantitative or qualitative information were you deciding?
   (d) What was your guess of the p-value? Did you reject your H_0?
   (e) What were the possible consequences of making the wrong decision?
2. (10 points) Exercise 17 in Chapter 1 (2nd or 3rd ed.).
3. (20 points) **Environmental Voting of Democrats and Republicans in the U.S. House of Representatives**. Use the data provided in Problem 26, Chapter 1 (3rd ed.) to answer the following questions. Note that all calculations have to be performed in R. (First two chapters of the 3rd edition have been scanned and uploaded on the course web-site. The data file has been posted online as well.)

(a) Which test is more appropriate to address the research question in this problem, permutation test or randomization test?

(b) Based on your answer in (a), set up corresponding null and alternative hypotheses.

(c) Specify the test statistic that you are using and calculate its observed value.

(d) Draw the reference distribution of your test statistic based on 10,000 simulations. You may reuse the R code given here in Section 3.3. by adapting it to this problem.

(e) What is your p-value?

(f) Comment on assumptions of the chosen test.

(g) Write a brief summary of your findings and comment on the scope of inference.

4. (15 points) Let $X_1, \ldots, X_i, \ldots, X_n$ be independent random variables drawn from a population with mean μ and variance σ^2. Let $\bar{X}$ be a sample average.

(a) (5 points) Show that $Var(\bar{X}) = \sigma^2/n$.

(b) (10 points) $Var(\bar{X})$ can be estimated by s^2/n, where s^2 is a sample variance,

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right).$$

i. Show that $E(X_i^2) = \sigma^2 + \mu^2$, using the fact that $\sigma^2 = E\left((X_i - \mu)^2\right)$.

ii. Show that $E(s^2) = \sigma^2$, i.e., s^2 is an unbiased estimator of the population variance.

Therefore, s^2/n is an unbiased estimator of $Var(\bar{X})$.