Homework 3: Due on Feb. 17

Guideline

- Homework should be submitted via Gradescope by Friday midnight (11:59 pm. CDT).
- Homework answers to Simulations and data analysis should be written in R Markdown.
- Please find the following exercise question in [HMC] (Hogg, Mckean, and Craig 2018).
- Each question is worth of 10 points and the total is 50 points.

1. Ex 4.7.9
2. Ex 4.7.7
3. In 2020, Pfizer and BioNTech conducted a randomized double-blinded clinical trial (phase 3) to investigate the effectiveness of their new mRNA-based COVID-19 vaccine candidate, BNT162b2. Half of the 43,000 enrolled participants were randomly assigned to receive BNT162b2, and the other half were assigned to receive a placebo. It turned out that only 8 cases of COVID-19 were observed in the new vaccine group (BNT162b2) while 162 cases were in the placebo group. The relative risk of the new vaccine is only 0.05 ($= 8/162$), hence the efficacy is 95% ($= 1 - 8/162$). However, with such a small group of cases (due to a small rate of infection), could the efficacy of the new vaccine be caused by random noise?
   (a) Please draw a 2x2 contingency table to summarize the clinical trial and test whether COVID-19 infection is the same for the new vaccine group and placebo group.
   (b) Please manually calculate the test statistic, specify the distribution of the test statistic under H_0, and find the p-value of the test.
   (c) The above comparison can be done by two-sample proportion test. Please construct a z-test and compare it with your answer in part (b).
   (d) Please use "chisq.test" and "prop.test" (correct=FALSE) function in R to check your answer in part (b) and (c).
4. In a genetic inheritance study discussed by Margolin [1988], samples of individuals from several ethnic groups were taken. Blood samples were collected from each individual and several variables were measured. We shall compare the groups labeled "Native American" and "Caucasian" with respect to the variable mean sister chromatid exchange (MSCE). The data is as follows:

   Native American: 8.50 9.48 8.65 8.16 8.83 7.76 8.63
   Caucasian: 8.27 8.20 8.25 8.14 9.00 8.10 7.20 8.32 7.70

The dots plot suggests a difference with Native American MSCE being larger on average. Please construct a statistical test to investigate the hypothesis. Note that the sample sizes are very small here.

(a) Please specify the type of the test and explain why you choose such a test.

(b) Manually calculate the test statistics.

(c) Find the p-value of your test and draw a conclusion.

5. Comparison of the two-sample z-test, classical two-sample t-test and the Satterthwaite-Welch t-test. Please simulate 1000 Monte Carlo simulations, and count the percentage of rejections, in each of the following settings. Please make sure that all three tests are performed on the same simulation. (Do not generate new data to do a different test.)

   (a) H_0 equal variance setting. Generate two normal random samples of size 100 and 200 from $N(0,1)$ and $N(0,1)$, respectively. Compare the two samples using all three tests: two-sample z-test, classical two-sample t-test and the Satterthwaite-Welch t-test. For each test, how often do you reject H_0 in 1000 Monte Carlo simulations? (Type I error)

   (b) H_0 unequal variance setting. Generate two normal random samples of size 100 and 200 from $N(0,1)$ and $N(0,2)$, respectively. Compare the two samples using all three tests: two-sample z-test, classical two-sample t-test and the Satterthwaite-Welch t-test. For each test, how often do you reject H_0 in 1000 Monte Carlo simulations? (Type I error)

   (c) H_1 equal variance setting. Generate two normal random samples of size 100 and 200 from $N(0,1)$ and $N(0.3,1)$, respectively. Compare the two samples using all three tests: two-sample z-test, classical two-sample t-test and the Satterthwaite-Welch t-test. For each test, how often do you reject H_0 in 1000 Monte Carlo simulations? (Power)

   (d) H_1 unequal variance setting. Generate two normal random samples of size 100 and 200 from $N(0,1)$ and $N(0,2)$, respectively. Compare the two samples using all three tests: two-sample z-test, classical two-sample t-test and the Satterthwaite-Welch t-test. For each test, how often do you reject H_0 in 1000 Monte Carlo simulations? (Power)

   (e) Please redo part (a)-(b), but with sample size 5 and 10 instead of 200 and 300. Which test have a significant higher type I error than 5%?

   (f) Please redo part (a), but generate two random samples of size 5 and 10 both from $Exp(1)$. Which test is better to use in this situation? Why?