MSAN 504 — Probability/Statistics — Summer 2016
Homework Two

Instructions: For every hypothesis test on this homework assignment, you must (a) state the null and alternate hypotheses in both symbols and words; (b) compute an appropriate test statistic by showing the substitutions into the correct formula for that test statistic; (c) use R to convert that test statistic into a p-value; and then (d) make a correct statistical decision and interpret that statistical decision in the context of the particular problem. For every confidence interval on this homework assignment, you must (a) show the formula for the particular confidence interval; (b) show the substitutions into the formula; and (c) explicitly interpret the confidence interval that you've computed.

1. Psychologists studied the size of the tip in a restaurant when a message indicating that the next day's weather would be good was written on the bill. Here are tips from 20 patrons, measured in percent of the total bill: 20.8, 18.7, 19.9, 20.6, 21.9, 23.4, 22.8, 24.9, 22.2, 20.3, 24.9, 22.3, 27.0, 20.3, 22.2, 24.0, 21.1, 22.1, 22.0, and 22.7. Does a weather-inspired tip exceed 20 percent? Use a significance level equal to $\alpha = 0.06$.

2. Download the data set related to body fat percentages and other body measurements. The sample for this data set is taken from Brigham Young University, which is largely populated with Mormon students and faculty. (More specifically, these are male BYU students.) Some, though not all, Mormons fast one day per week and regular fasting has been connected with various positive health outcomes. Compute (by hand) an 85% confidence interval for the true mean body fat percentage of a male Mormon college student. (You may compute supporting quantities like $\overline{x}$ and s in R, but show the appropriate substitutions into a confidence interval for a single mean.) **Data Set:** `BodyFatPercentage.csv`

3. Research the concept of a "paired t test" using an introductory statistics textbook or some other online resource. Download the data set related to the performance of the Dow Jones Industrial Average versus a group of expert stock pickers chosen by the staff at the *Wall Street Journal*. Take the monthly returns for the stock experts and average them across stock experts. Then test the null hypothesis that experts do not outperform the Dow Jones Industrial Average. **Data Set:** `DartsVersusExperts.csv`

4. Download the data set related to Plano and Plus lenses, two different types of lenses that can be placed into traditional eyeglasses. Participants in a study were asked to complete a reading comprehension test. Compute a 99% confidence interval for the true mean difference between Plano and Plus reading comprehension scores. Why is a paired approach appropriate in this situation? **Data Set:** `PlanoVersusPlusLenses.csv`

5. Consider a coin that is flipped 34 times and comes up heads 15 times. You do not know what the true proportion of the time it is that this coin will come up heads. Build a 95% confidence interval for the true proportion of times the coin will come up heads.

6. Using the confidence interval that you constructed for problem #5, would you reject $H_0 : \pi = 0.75$ in favor of its two-sided alternative at the 5% level of significance? Why or why not?

7. For the scores on an achievement test given to a certain population of students, the expected value is 500 and the standard deviation is 100. Let $\overline{X}$ be the mean of the scores of a random sample of 35 students from the population. What does the central limit say the approximate distribution of $\overline{X}$ is? Then, compute the probability that $\overline{X}$ is between 460 and 540. (**Hint:** If $Z = (X - \mu)/\sigma$ is approximately standard normal, then X is approximately normal with mean μ and standard deviation σ.)

8. A random sample of size 24 is taken from a probability distribution with density function $f(x)$ such that $f(x) = \frac{1}{9}\left(x + \frac{5}{2}\right)$ when $1 < x < 3$ and zero otherwise. Let $\overline{X}$ be the sample mean. Use the Central Limit Theorem to compute the probability that $\overline{X}$ is between 2 and 2.15.

9. One Pareto random variable has the density function $f(x) = \alpha/x^{\alpha+1}$ for $\alpha > 0$ and $x > 1$. If $x < 1$, $f(x) = 0$. Use the method of maximum likelihood estimation to generate a formula for estimating α.

10. A **Poisson distribution** allocates strictly positive probability to integers greater than or equal to zero. Look up the probability mass function of a Poisson distribution online or in the readings provided on the Canvas page, and then derive a maximum likelihood estimator for λ, the parameter than controls it.

11. Suppose that $X_1, X_2, ..., X_n$ is a random sample governed by the probability density function $f(x; \theta)$, where $f(x; \theta) = \frac{1}{\theta^2} x e^{-x/\theta}$ if $x > 0$ and zero otherwise. Compute a maximum likelihood estimator for θ, the parameter that controls the density function.

12. Download the data set `BodyFatPercentage.csv`. Compute the linear correlation coefficient between age and body fat percentage, and then test the null hypothesis $H_0 : \rho = 0$ against the appropriate one-sided alternative at the 1% of significance. Use both the perspective of Pearson and the perspective of Fisher.

13. Compute an approximate 80% confidence interval for the true correlation between age and body fat percentage using the same data set as in problem 12. Take the approach suggested by Fisher, i.e., use the Fisher transformation.

14. Download the data set `AgrestiFinlayCrime.csv` and test whether or not the true population variance of the rate of crimes (per 100,000 individuals) is equal to 185000 at the 10% level of significance.

15. Download the data set `Abalone.csv`. Assuming that the lengths of the abalone are normally distributed, compute a 90% confidence interval for the true variance of the lengths of abalone.