

# Homework 7

October 31, 2019

## 1. Identifying Tests

**For the following prompts include:**

- the name of the test
- the parameter being tested,  $H_0$  and  $H_a$ .

Example: What test would you use to determine whether the starting salaries for statisticians are greater than \$80,000?

Ans: One-sample t-test for population mean,  $H_0 : \mu = \$80,000$  vs.  $H_a : \mu > \$80,000$ .

a. The 2010 Census found that the average family size in Minnesota was 3.05. What statistical procedure would you use to test whether the average family size in Minnesota is greater than 3.05 in 2019?

b. According to the 2017 American Community Survey, 5.7% of all workers in Minneapolis/St. Paul urbanized area commute to work by public transportation. What statistical procedure would you use to test whether the proportion of workers in Minneapolis/St. Paul urbanized area commute to work by public transportation in 2019 has changed from than in 2015?

c. According to the 2017 American Community Survey, 4.3% of all workers in Minneapolis/St. Paul urbanized area work in construction. What statistical procedure would you use to test whether the proportion of the workforce in the Minneapolis/St. Paul urbanized area is less than in 2017?

- d. In 2016-2017 NBA season, the average salary per team was \$97.6 million. What statistical procedure would you use to test whether the average salary of teams in the NBA in the 2019-2020 season is the same as in 2016-2017 season?
2. Management claims the mean income for all senior-level assembly line workers at 3M equals \$700 a week. An employee decided to test this claim believing it is different from \$700.
- a. State the hypotheses for this test, and make a conclusion to the test based upon the confidence interval below. What level is this test?  
99% CI of  $\mu$ : (631.354, 650.868)
- b. Suppose management claimed that 60% of senior-level assembly line workers at 3M is greater than \$650 a week. An employee created a 95% confidence interval of the proportion of senior-level assembly line workers using a random sample of 50 workers. Their confidence interval is below. Based off the confidence interval, is it valid to conclude that less than 60% of workers get paid more than \$650 a week? Why or why not?(Hint: what is the standard error of a z-test and what is the standard error of a confidence interval for proportions?)  
95% CI for p: (0.522,0.776)
3. Textbook 9.28 "Which t has p-value = 0.05", page 416.
4. Textbook Problem 9.54 "Practical significance", page 429.
5. According to the 2010 US Census, the average number of residents per housing unit for the n=87 counties in Minnesota was 2.10, and the standard deviation was 0.38. Test whether the true mean number of residents per housing unit in Minnesota in 2010 is less than the national value of 2.34 at the level  $\alpha = 0.05$ .
- a. Show all five steps of this test.
- b. What type of error could we be making in this context?

- c. What is the minimum average number of residents per household needed in order to fail to reject  $H_0$ ? Assume the sample standard deviation is the same.
- d. Suppose the true number of residents per household in Minnesota is normally distributed with a mean of 2.0 and standard deviation of 0.4. Suppose we reject null hypothesis if the sample mean number of residents is less than 2.27. What is the probability of making a type II error?

## R Problems

1. Using daily winter temperatures recorded at the Minneapolis St. Paul Airport from December 21, 2018 to March 19, 2019, we want to test if the mean maximum temperature of a given winter day for 2018 to 2019 was the same as 33 degrees Fahrenheit (the mean temperature of December 21, 2016 to March 19 2017). To do this, 30 days were randomly chosen between December 21, 2018 and March 19, 2019 and Maximum Temperature and Minimum Temperature were recorded.

```
dailyTemps = read.csv("2018-19DailyTempSamps.csv",header = TRUE)
```

- a. Create a histogram and boxplot of Maximum Temperature(column named MaxT) over this set of days, and describe the shape, center and spread of the distribution of this sample.
  - b. Complete a one-sample hypothesis test of means using R, you can skip step 3 because R will do this for you in the t.test function. Test this at level 0.01.
  - c. What type of error could you be making, why?
2. A lot of pitchers have been complaining that since because Major League Baseball changed the balls, more homeruns are hit now than in the past. We took a sample of 15 teams from 2019 and will compare the number of homeruns to the average number of homeruns hit by team in the 2010-2018 seasons which was 167.333.

```
teamSamp = read.csv("mlbTeamSamples.csv")
```

- a. Create a qqplot of the homeruns, do the data appear normal (Column for homeruns is HR)?
- b. Complete a one-sample hypothesis test of means using R, you can skip step 3 because R will do this for you in the t.test function. Test this at level 0.05.
- c. What kind of error could you be making, why? Should these results be trusted, why or why not?(Hint: remember the assumptions)