OPIM 5603-B14 — Statistics in Business Analytics
Fall 2019, University of Connecticut

Homework 8 - v1

Instructions: Please complete the following questions and submit them as an RNotebook (as an `Rmd` file) via the submission link on HuskyCT. You must submit the assignment by the time and due date listed on the course syllabus. Failure to submit a file by the deadline will result in a score of 0 on the assignment.

Set the heading of the RNotebook as an `html_document`, with a table of contents and without numbered sections. Add your name and a date to the header as well. The solution to each problem should be a separate section (specified by #), and each subproblem should be set as a subsection (specified as ##). For example, for Problem 2, you should have a section titled Problem 2, specified by:

```
# Problem 2
```

in your RNotebook. Also, for subproblem b in Problem 2, you should have a subsection, specified by:

```
## Problem 2b
```

As with all course material, the problems appearing in this homework assignment are taken from the instructor's real-world experiences, from other courses taught at the University of Connecticut, and from the sources listed in the course syllabus.

Note that R code submitted should work independent of the data that sits in the data structure. For example, suppose there was a vector `r_vec` with the values $(1, 2, 6)$ and the problem asks for you to create R code to create a vector `answer` which doubles each element of `r_vec`. The answer

$$\texttt{answer} \leftarrow \mathrm{c}(2, 4, 12)$$

would be given no credit. The answer

$$\texttt{answer} \leftarrow 2^{*}\texttt{r\_vec}$$

would be an appropriate answer.

You must show all steps in your solution. For example, if a problem asks for the expected value of a random variable that is binomially distributed with $n = 10$ and $\pi = 0.3$, and you simply write

$$3,$$

this will be given no credit. However,

$$10 * 0.3$$

would be given credit.

If you have any questions, please submit them via email to the instructor and/or the teaching assistant prior to submitting your solution.

Problem 1 (25 points)

Suppose you take a random sample of the waiting times for 13 patients at an emergency room before first being seen by a doctor. The waiting times, in minutes, are 13.2, 19, 2, 3, 7.4, 32.1, 8, 1, 3.4, 7.4, 8.9, 30.2, an 17.2.

a. Suppose you know that the population waiting times is normally distributed. Build a 85% confidence interval for the mean waiting time.

b. Suppose you know that the population standard deviation is 10 minutes. Build a 85% confidence interval for the mean waiting time.

c. Suppose you don't know that the population waiting time is normally distributed. Build a 85% confidence interval for the mean waiting time.

d. Suppose you know that the population waiting time is exponentially distributed. Build a 85% confidence interval for the mean waiting time.

Problem 2 (15 points)

Suppose you take a random sample of 260 manufactured power drills.

a. Suppose 253 of the sampled power drills work. Build a 95% confidence interval for the proportion of power drills that don't work.

b. Suppose 200 of the sampled power drills work. Build a 95% confidence interval for the proportion of power drills that don't work.

Problem 3 (30 points)

A hardware store is considering restructuring their shelving and the proposed plan will only have space for 162 propane tanks to be on display each day. Assume that the hardware store does not have worker capacity to restock shelves during a day, but will always restock up to the capacity for the start of the business day. In a random sample of 32 days, the mean number of tanks sold is 150.8 and the standard deviation is 50.3. The company drafting the plans claims that the hardware store has enough shelving to satisfy all daily demand for propane tanks. We will conduct a statistical test to evaluate this claim.

a. Formulate the null hypothesis (in words) to evaluate this claim.

b. Formulate the alternative hypothesis (in words) to evaluate this claim.

c. What is μ_0 in the statistical test?

d. What is the population parameter that the statistical test is concerned with?

e. Which sample statistic can be used to estimate the population parameter that the statistical tests is testing?

f. Is using the parametric tests discussed in class applicable? Explain why or why not.

g. What values of $\bar{x}$ would lead you to reject the null hypothesis at a confidence level of 90%? Calculate this based on the t-test and z-test.

h. What values of $\bar{x}$ would lead you to reject the null hypothesis at a confidence level of 95%? Calculate this based on the t-test and z-test.

i. What values of $\bar{x}$ would lead you to reject the null hypothesis at a confidence level of 99%? Calculate this based on the t-test and z-test.

j. Suppose we transform $\bar{x}$ into the test statistic $T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$. What values of T would lead you to reject the null hypothesis at a confidence level of 90%? Calculate this based on the t-test and z-test.

k. Suppose we transform $\bar{x}$ into the test statistic $T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$. What values of T would lead you to reject the null hypothesis at a confidence level of 95%? Calculate this based on the t-test and z-test.

l. Suppose we transform $\bar{x}$ into the test statistic $T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$. What values of T would lead you to reject the null hypothesis at a confidence level of 99%? Calculate this based on the t-test and z-test.

m. What is the p-value of this test? Calculate it based on $\bar{x}$ and also based on the transformed statistic, using both the t and normal distributions.

n. What is the minimum confidence level for which you would reject the null hypothesis? Answer this using the p-values for both the t and normal distribution.

o. Compare the solution you arrived at for the previous part with the rejection region defined in the early subsections. Explain, in detail, how you can use either the p-value or the definition of rejection regions to conduct this hypothesis test at a given confidence level.

Problem 4 (30 points)

You have been asked to create a confidence interval for the amount by which properties in a particular region are listed above their assessed values. From previous experience, you know that the standard deviation of the difference is between \$5,700 and \$20,000, and that the distribution is approximately normal.

a. Suppose you take a random sample of 30 properties. Given the range of possible values for the standard deviation, what is the maximum width of a 90% confidence interval?

b. A company is willing to pay you for a 90% confidence interval, but the amount that they are wiling to pay you depends on how wide the confidence interval is. The company is willing to pay you \$20,000 plus the margin of error times \$100. Each assessment costs \$1,950. Given the range of possible values for the standard deviation, how many houses should you have assessed?