Stat 260 R Assignment 3 - Spring 2025

Dr. J. Yu

Due Date: April 11, 2025

INSTRUCTION:

1. **If you complete your assignment in R or RStudio, please copy and paste the commands and their output into a Word document, and then print it to PDF.**
2. **Execute each line of code separately to ensure that it works properly.**
3. **Submit the PDF file to the Crowdmark in the R Assignment 3 activity.**
4. **Total marks [10].**

Instruction: In this assignment, be creating confidence intervals and testing hypotheses for a single mean or for a single proportion.

The two commands we will use are as follows:

- For a mean: t.test(x, alternative, mu, conf.level)
- For a proportion: binom.test(x, n, p, alternative, conf.level)

Using t.test for confidence intervals: If we are using the t.test command to create a confidence interval, we only need to specify x, our vector of observations, and conf.level, our desired confidence level.

Example: Suppose we have a set of five steel bolts, and we measure their weights in grams:

12.3, 12.5, 12.7, 11.5, 15.3

To create a 97% confidence interval for μ, the mean bolt weight. First, create a vector containing the data. (see the introduction to assignment 1 for review, if needed.)

> bolt.weight = c(12.3, 12.5, 12.7, 12.1, 12.6)

Next, I call for my 97% confidence interval.

> t.test(bolt.weight, conf.level = 0.97)

The output is as follows:

One Sample t-test

data: bolt.weight
t = 115.5025, df = 4, p-value = 3.37e-08
alternative hypothesis: true mean is not equal to 0
97 percent confidence interval: 12.08483 12.79517
sample estimates:
mean of x
12.44

Since we are creating a confidence interval, only the second half of the output is relevant to us. The 97% confidence interval is (12.08483, 12.79517), and the sample mean is $\bar{x} = 12.44$.

Using t.test for hypothesis tests: If we are using the t.test command to test a hypothesis, we must specify x, our vector of observations, **mu**, our hypothesized value for μ, and **alternative**, the form of our alternative hypothesis.

For **alternative**, we either specify that **alternative = "two.sided"**, or **alternative = "less"**, or **alternative = "greater"**.

Example: Suppose we want to test the alternative hypothesis that the true mean of the bolts is less than 13. That is, we want to test the hypotheses $H_0 : \mu = 13$, $H_1 : \mu < 13$.

We have already created a vector containing the data, so now we call for a hypothesis test.

>t.test(bolt.weight, mu = 13, alternative = "less")

The output is as follows:

One Sample t-test

data: bolt.weight
t = -5.1995, df = 4, p-value = 0.003259
alternative hypothesis: true mean is less than 13
95 percent confidence interval:
-Inf 12.66961
sample estimates:
mean of x
12.44

For a hypothesis test, the first half of the output is of interest.

- **t = -5.1995** tells us the observed value of the test statistic.

- **df = 4** tells us how many degrees of freedom were used for the t-distribution.

- **p-value = 0.003259** tells us the p-value for our hypothesis test.

Important: The next line, **alternative hypothesis: true mean is less than 13**, is just a statement of the alternative hypothesis being tested. R is **not** telling you what your conclusion should be; that is up to you to determine.

Here, since the p-value is less than 0.01, we conclude that there is very strong evidence against the null hypothesis.

Using binom.test for confidence intervals: If we are using the binom.test command to create a confidence interval, we must specify **x**, the number of successes, **n**, the number of trials, and **conf.level**, our desired confidence level.

Example: In a sample of 2000 items produced in a factory, 73 are of poor quality. Construct a 98% confidence interval for p, the true proportion of items of poor quality.

We identify that there are x=73 successes out of n = 2000 trials. We call for a confidence interval:

> binom.test(x=73,n=2000,conf.level=0.98)

The output is as follows:

Exact binomial test

data: 73 and 2000
number of successes = 73, number of trials = 2000, p-value $< 2.2e-16$
alternative hypothesis: true
probability of success is not equal to 0.5
98 percent confidence interval: 0.02742238 0.04746495
sample estimates:
probability of success
0.0365

Again, the information we need is in the last half of the output. The 98% confidence interval is (0.02742238, 0.04746495) and our estimate for p is $\hat{p} = 0.0365$.

Using binom.test for hypothesis tests: If we are using the binom.test command to test a hypothesis, we must specify **x**, the number of successes, **n**, the number of trials, **p**, our hypothesized population proportion, and **alternative**, the form the alternative hypothesis takes.

Example: Suppose for the factory example, we want to test $H_0 : p = 0.06$, $H_1 : p \neq 0.06$. We call for a hypothesis test.

> binom.test(x=73,n=2000, p = 0.06, alternative = "two.sided")

The output is as follows:

Exact binomial test

data: 73 and 2000
number of successes = 73, number of trials = 2000, p-value = 2.919e-06
alternative hypothesis: true
probability of success is not equal to 0.06
95 percent confidence interval:
0.02871716 0.04567648
sample estimates:
probability of success
0.0365

As before, the relevant hypothesis testing information is near the top. In particular, **p-value= 2.919e-06** tells us that the p-value is $2.919 * 10^{-6}$. Again, the line that follows is just a statement of the alternative hypothesis; it is **not** a conclusion given by R.

For our hypothesis test, we have found very strong evidence against H_0.

Important: For **binom.test**, R is using a different procedure than we are using in class. If you were to create a confidence interval by hand and compare it to the one R creates, they will not be identical.

Assignment: For each of the following questions, carry out all calculations using R . Calculations done by hand, by using the tables/formulas in the text, or by using other software will not be awarded marks.

1. In a soup factory, we take a random sample of 8 cans of tomato soup, and measure their sodium content (in mg). The following are our observations.

   510 520 515 516 517 519 522 510

   (a) (1 mark) Give the command and output to create a 96% confidence interval for the mean sodium content.
   (b) (1 mark) Using your confidence interval, decide if 515 is a reasonable estimate for μ.
   (c) (1 mark) Give the command and output to test the alternative hypothesis that the mean sodium content is less than 520 mg.
   (d) (1 mark) What is the observed value of the test statistic?
   (e) (1 mark) What is the p-value for our test?
   (f) (1 mark) If we were testing at a significance level of α= 0.01, what would the conclusion be?

2. From a random sample of 673 items made by a particular manufacturing process, it is found that 27 are defective.

   (a) (1 mark) Find a 99.5% confidence interval for the proportion of defective items made by the process. (Also include the commands and output.)
   (b) (1 mark) Give the command and output to test the alternative hypothesis that the proportion of defective items made by the process is greater than 0.03.
   (c) (1 mark) What is the p-value for our test?
   (d) (1 mark) What is the strength of evidence we have found against H_0?