MTH 324, Lab #6, Fall 2022   Name ____

Instructions: Follow along with the tutorial portion of the lab. Replicate the code examples in R on your own, along with the demonstration. Then use those examples as a model to answer the questions/perform the tasks that follow. Copy and paste the results of your code to answer questions where directed. Submit your response file and the code used (both for the tutorial and part two). Your code file and your lab response file should each include your name inside.

Hypothesis Testing with one Sample

We can conduct one-sample tests in R using built in functions, or our formulas. We'll take you through an example of doing the test using formulas, and then some examples for means and proportions using the built-in test functions.

Suppose that we want to determine if the mean gas mileage of cars at the time the mtcars dataset was collected was greater than 20 miles per gallon. Our hypotheses are then $H_0: \mu = 20, H_a: \mu > 20$. Let us assume that the population standard deviation is about 6 miles per gallon.

Recall that our test statistic is $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$. We can calculate the test values we need and the test statistic in R.

```
xbar<-mean(mtcars$mpg)
sigma<-6
mu0<-20
n<-length(mtcars$mpg)
z<-(xbar-mu0)/(sigma/sqrt(n))

```

Then we can find the P-value using the pnorm() function.

```
pnorm(z,lower.tail=FALSE)
```

In this case, we get a P-value of 0.465955 >> 0.05, so we fail to reject the null hypothesis. We don't have the evidence to assert that miles per gallon is on average over 20 mpg.

For one-sample tests, this procedure works okay for z-tests, as well as t-tests. But as our tests become more complex, we want to be able to use whatever built-in tests are available, and R has them for one-sample tests.

The z.test() function is in the BSDA package.

```
library(BSDA)
z.test(mtcars$mpg,alternative="greater", mu=20, sigma.x=6)
```

The output summarizes the analysis.

```
        One-sample z-Test

data:  mtcars$mpg
z = 0.085442, p-value = 0.466
alternative hypothesis: true mean is greater than 20
95 percent confidence interval:
 18.34599       NA
sample estimates:
mean of x
 20.09062
```

The analysis gives us the z-statistic and p-value, we can also specify a confidence interval. It also reminds us of the alternative hypothesis we are testing, and the mean of our sample.

The documentation for this test is linked in the references below for setting additional options.

To conduct a proportion test, we need counts. Some problems will give you these counts, but let's look at how to obtain them from data. In the mtcars dataset, the vs column in already in the form of 0s and 1s. If your data is not in this form, you can use an if() statement to select one of the observations to be 1 (success) and the others to be 0 (false).

Let's test this data to see if the proportion of successes is different than 50%.

```
x<-sum(mtcars$vs)
n<-length(mtcars$vs)

prop.test(x, n, p = 0.5, alternative = "two.sided", correct = TRUE)
```

The output here is similar to the z-test output.

```
        1-sample proportions test with continuity correction

data:  x out of n, null probability 0.5
X-squared = 0.28125, df = 1, p-value = 0.5959
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.2683511 0.6211967
sample estimates:
     p
0.4375
```

Given the p-value, we also fail to reject the null hypothesis. We saw from the example above that we have only 32 observations, that's why correct=TRUE is in our list of options. This is a very small sample for a proportion test.

There is also a binom.test() that can perform this test for situations that fail to meet our normal assumption tests.

```
binom.test(x, n, p = 0.5, alternative = "two.sided")
```

The output is shown below.

```
        Exact binomial test

data:  x and n
number of successes = 14, number of trials = 32, p-value = 0.5966
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2636381 0.6233743
sample estimates:
probability of success
                0.4375
```

The exact binomial (or using the correction factor) is appropriate for a proportion like this until the sample size exceeds 40 or so. The smaller the proportion, the larger the sample size should be.

The test we did for the z-test of means required us to make several assumptions, such as the standard deviation of the population, and the sample size is close to the rule of thumb. So, it would be better to conduct the t-test on this data instead. Our hypotheses remain the same.

```
x<-mtcars$mpg

t.test(x, mu = 20, alternative = "greater")

```

The output we get is shown below.

```
        One Sample t-test

data:  x
t = 0.08506, df = 31, p-value = 0.4664
alternative hypothesis: true mean is greater than 20
95 percent confidence interval:
 18.28418      Inf
sample estimates:
mean of x
 20.09062
```

This is similar to our previous results, but we now include the degrees of freedom, and out P-value is a little different (the sample size is large enough that it’s not that far off).

Don’t forget to test your assumptions for each test. For the proportion test, we want $npq \geq 10$. For the z-test and t-test requires that the data be normally distributed. You may want to construct a boxplot or a qqplot to test for approximate normality.

```
qqnorm(mtcars$mpg, pch = 1, frame = FALSE)
qqline(mtcars$mpg, col = "steelblue", lwd = 2)
```

The one-sample tests we've considered here are not the only possible tests. As we see in the class lecture, there are also test with the Poisson distribution, tests of variance (standard deviations), and so on. We can conduct these using the formulas.

Tasks

1. Conduct the hypothesis test examples below. For each test, clearly state your null and alternative hypotheses and show how you tested the assumptions of the test. Are there any potential problems with the assumptions for either test?
    - Researchers are interested in whether the pulse rate of long-distance runners differs from that of other athletes
    - They randomly sample 8 long-distance runners, measure their resting pulse, and obtain the following data:

      45, 42, 64, 54, 58, 49, 48, 56
    - The average resting pulse of athletes in the general population is 60 beats per minute
    - Test the null hypothesis at the 0.05 level of significance

    a.
    b. A political candidate conducts a poll to see how they are doing the week before the election. They survey 205 registered voters and find 117 people say they are going to vote for the candidate. If this good evidence to suggest the candidate is going to win the election?

References:

1. Discovering Statistics Using R. Andy Field, Jeremy Miles, Zoe Field. (2012)
2. https://book.stat420.org/applied_statistics.pdf
3. https://scholarworks.montana.edu/xmlui/handle/1/2999
4. https://www.rstudio.com/resources/cheatsheets/
5. http://www.sthda.com/english/wiki/qq-plots-quantile-quantile-plots-r-base-graphs
6. https://www.rdocumentation.org/packages/BSDA/versions/1.2.1/topics/z.test
7. http://www.sthda.com/english/wiki/one-proportion-z-test-in-r
8. https://statstutorial.com/how-to-perform-a-one-sample-z-test-in-r-with-examples/
9. http://www.sthda.com/english/wiki/one-sample-t-test-in-r
10. https://slideplayer.com/slide/8367889/
11. http://betsymccall.net/edu/CDSE/coding/R/normal_distributions.pdf