3818 R Homework 4

*** Student Name ***

8/20/2020

You may work in groups of three (but no more than three!). Please put the name of all group members at the top of the text file. ## Question 1 A light bulb manufacturer claims their light bulbs last for 1,000 hours. You sample 50 light bulbs and you find that the sample average light bulb lifetime duration is 955 hours. Assume that you know the population standard deviation, and that it is 220 hours. That is $n = 50$, $\bar{x} = 955$, $\sigma = 220$. Test the null hypothesis $H_0 : \mu = 1000$ against the alternative hypothesis $H_a : \mu < 1000$ at the 0.05 significance level. - Calculate the critical statistic using the `qnorm()` function. Compare this to the test statistic, found using the sample information. Do you reject, or fail to reject the null hypothesis? - Find the p-value associated with the sample statistic under the null hypothesis. You can do this using the `pnorm()` function. Report the p-value. Is your conclusion consistent with the first part of this question?

```
# Code for Question 1 goes here

z_score <- (955-1000)/sqrt(220^2/50)
```
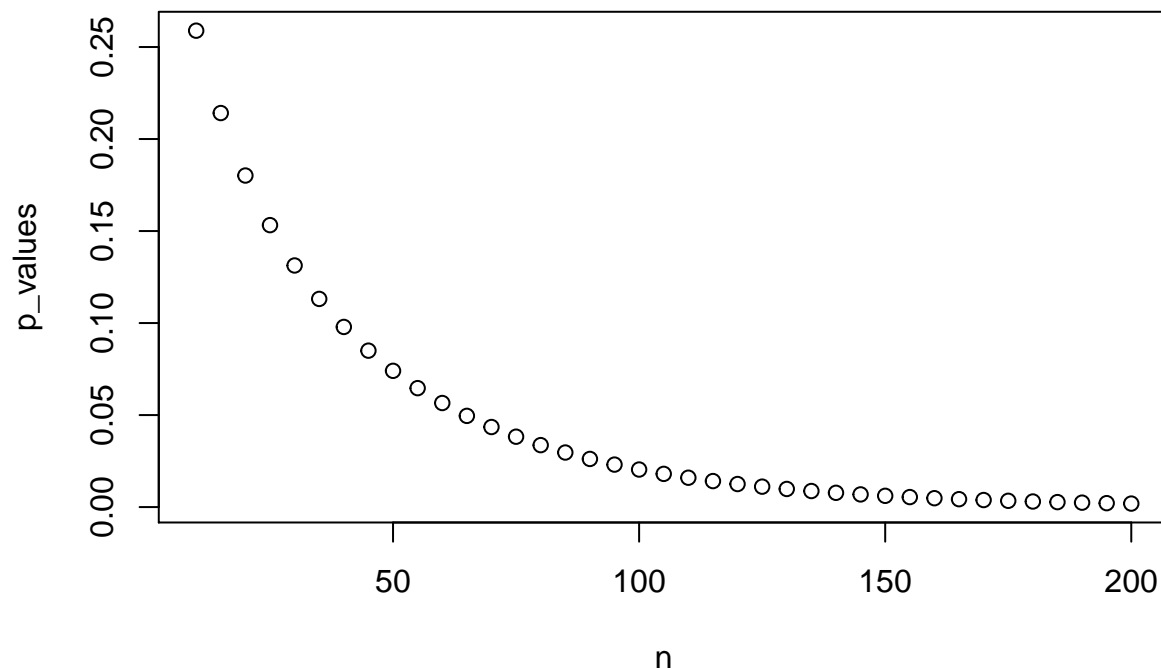
Answer:

Question 2

Consider the scenario presented in question 1. Here we will consider how the p-value will change with the sample size, keeping all of the other information the same. Create a vector, called **n**, of different sample sizes, from 10 to 200 at an increment of 5. You can do this using the seq() command. Use this vector of sample sizes, along with the information in question 1 and the `pnorm()` function, to create a vector of p-values that is the same length as the vector **n**. Plot the vector of sample sizes on the horizontal axis and the vector of p-values on the vertical axis using the `plot()` function. Label the axes. - What happens to the p-value as the sample size increases? - What range of sample sizes will give you a statistically significant result, holding the population standard deviation and sample average constant?

```
# Code for Question 2 goes here

n <- seq(10, 200, by= 5)
z_scores <- (955-1000)/sqrt(220^2/n)
p_values <- pnorm(z_scores)

plot(n, p_values)
```

Answer:

Question 3

Now we will do something similar to see how the sample average impacts the p-value. Create a vector, called **xbar**, of different observed sample averages, from 920 to 1000 at an increment of 5 (use the **seq()** function). Use this vector of sample averages along with the information in question 1 and the **pnorm()** function, to create a vector of p-values that is the same length as **xbar**. Plot the vector of sample averages on the horizontal axis and the vector of p-values on the vertical axis using the plot function. Label the axes (use **help(plot)** to learn how to make axis labels). - What happens to the p-value as the sample average increases? - What range of sample averages will give you a statistically significant result, holding the population standard deviation and sample size constant?

```
# Code for Question 3 goes here
n <- 100
```

Answer:

Question 4

Now, we'll go through the same procedure keeping sample size and population standard constant, but changing the population standard deviation. Create a numerical vector, called **sigma**, of different population standard deviations, from 120 to 250 at an increment of 5. You can do this using the **seq()** command. Use this vector of population standard deviations, along with the information in question 1 and the **pnorm()** function, to create a vector of p-values.

Plot the vector of population standard deviations on the horizontal axis and the vector of p-values on the vertical axis using the plot function. Label the axes. - What happens to the p-value as the population standard deviation increases? - What range of population standard deviations will give you a statistically significant result, holding the sample average and sample size constant?

```
# Code for Question 4 goes here
sample_mean <- 955
```

Answer:

Question 5

Now we'll do power analysis for question 1. For the hypothesis $H_0 : \mu = 1000$ vs. $H_a : \mu < 1000$, with $n = 50$, $\bar{x} = 50$, $\sigma = 220$, with $\alpha = 0.05$, there is a range of values of that will result in us rejecting the null hypothesis. This is called the rejection region, and the figure you created in question three hints at this. - Find the rejection region in terms of $\bar{x}$. - Now consider alternative true values of ranging from 800 to 1000 at intervals of ten. Create a vector of these values use the `seq()` function, and call this vector `true_mu`. - Use the `pnorm()` to calculate the probability of observing the critical sample average or a smaller value of $\bar{x}$ using the vector of true mean values. This is the power of the test for alternative true values of mu. - Plot the vector of alternative true values of μ on the horizontal axis and the vector a power on the vertical axis. Paste the figure below. - For what range of true values of μ does our test have a power of 80% or more?

```
# Code for Question 5 goes here
```

Answer:

Question 6

You started taking the bus to work. The local transit authority says that a bus should arrive at your bus stop every five minutes. After a while, you notice you spend a lot more than five minutes waiting for the bus, so you start to keep a record. You spend the next two months recording how long it takes for the bus to arrive to the bus stop. This give a total of sixty observations that denote the number of minutes it took for the bus to arrive (rounded to the nearest minute). These observations are hosted at https://mattbutner.github.io/data/bus_stop_time.csv - Use the t.test() command to test the hypothesis $H_0 : \mu = 5$ vs. $H_a : \mu \neq 5$ with $\alpha = 0.05$. - Report the exact p-value - Do you reject or fail to reject the null hypothesis? Interpret your conclusion.

```
# Code for Question 6 goes here
```

Answer: