ST 352

**Lab Assignment 1 SOLUTIONS**

***19 points***

***Due 11:59 PM on Friday, October 4***

**Reminder of the honor code:**

**lab assignments are to be completed individually!**

**Number your answers. Put the questions in order! Once completed, upload your assignment in Gradescope. After uploading your assignment in Gradescope, remember to link each question to the page on which that question appears in your document. (One-point deduction if you do not link questions to a page.)**

**Example**

At the beginning of Week 1, you participated in a one-question survey which asked how many minutes of TV you typically watch in a day. The number of minutes of TV watched per day for all who participated in the survey are posted in the **television** data set on Canvas. For the purpose of this assignment, assume that the sample data are representative of the number of minutes of TV watched per day for all OSU students (the population). Is there evidence that OSU students watch more than 90 minutes of TV per day, on average?

1. ***(1 point)*** What is the variable of interest? Is it categorical or quantitative?

***Time spent watching TV each day (minutes). It is quantitative***

***+1/2 for correct variable***

***+1/2 for correct type of variable (i.e. quantitative)***

* ***Note: if student had an incorrect variable but had the correct type of variable for how they defined the variable, they should receive ½ point. (That is, a student may incorrectly state the variable of interest as “whether or not a student watched more than 90 minutes of TV per day”, but correctly state that their variable is categorical. This student would receive ½ point.)***

1. ***(1 point)*** Do you think the sample is representative of the population of all OSU students (in terms of amount of TV watched per day)? Why or why not?

***Answers may vary, but should focus on the idea that a random sample of all OSU students (the population) was not taken. Possible arguments: 1) students argue that the sample is not representative of the population because it was not a random sample and TV viewing habits of those in the ST 352 class may be different than TV viewing habits of other OSU students. Or, 2) even though a random sample was not taken, students may argue that the sample is still representative of all OSU students because they may feel that TV viewing habits are not different between ST 352 students and any other OSU student. Look for support and accept a reasonable answer.***

***A “yes” or “no” response without an explanation: 1/4 point***

***Support is not quite correct but at least give an explanation: 1/2 to 3/4 point***

## Exploring the Sample Data

One reason for exploring the sample data is to determine if it is appropriate to use the t-methods to perform inference. The other is to take an initial guess if the null hypothesis will be rejected.

The following are needed to explore the sample data:

1) summary (or “descriptive”) statistics

* Use **favstats()** in R (if you have the MOSAIC package – if not, use **summary()**, **length()**, and **sd()** commands in R) – see pages 3 and 4 in the Lab 1 Notes.

2) obtain a dot plot or histogram

* Use the **dopPlot()** command in R (if you do not have the MOSAIC package, use the **hist()** command) – see page 3 in the Lab 1 Notes.

3) obtain a normal probability plot

* Use the **qqnorm()** and **qqline()** commands – see page 3 in the Lab 1 Notes.

1. ***(1 point)*** Provide a table of summary statistics. Note: do *not* copy and paste the table obtained in R. Rather, create your own table of summary statistics similar to the one provided below:

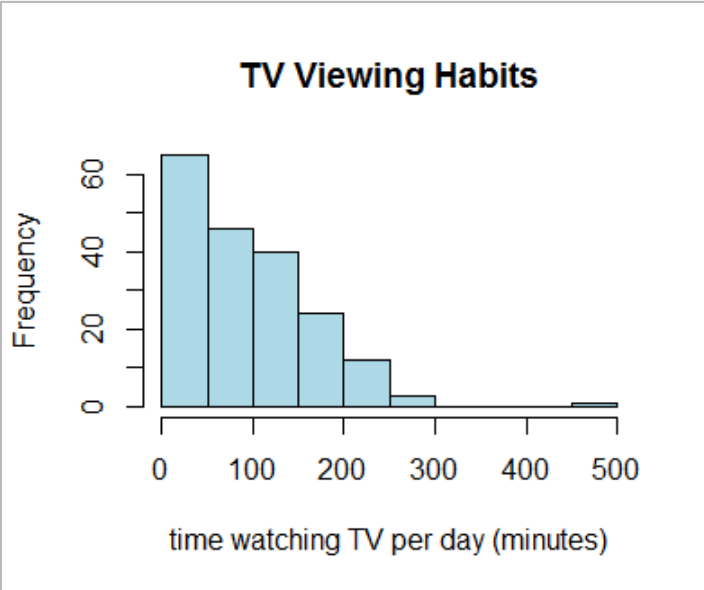
|  |  |
| --- | --- |
|  | **Time watching TV per day (minutes)** |
| **Sample size** | **191** |
| **Minimum** | **0** |
| **1st quartile** | **30** |
| **Median** | **66** |
| **mean** | **94.30366** |
| **3rd quartile** | **125** |
| **Maximum** | **480** |
| **Standard deviation** | **78.33035** |

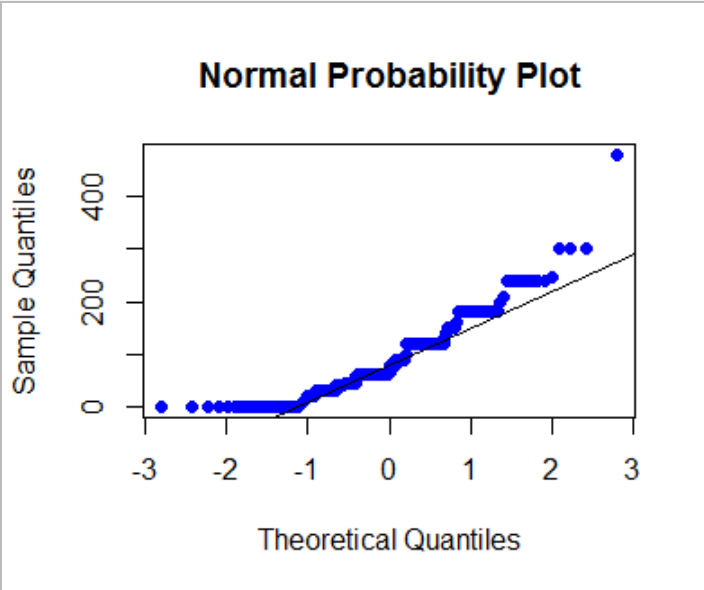
***-1/4 point if do not indicate units somewhere in their table***

***-3/4 points if copy and paste R output***

1. ***(2 points)*** Include your properly labeled histogram or dot plot and your normal probability plot here.

***I just have a histogram, but students could have a dot plot instead.***





***Only one of the dot plot or histogram needs to be included***

***+ 1 for histogram or dot plot***

***+ 0.25 for proper labels of dot plot or histogram***

***+ 0.75 for normal probability plot***

1. ***(2 points)*** Based on the normal probability plot, do you feel the t-methods can be used for inference? Briefly explain why or why not. (Note: recall that the t-methods should only be used if the distribution of sample means is approximately normal. Therefore, the answer to this question should focus on whether you believe the distribution of sample means will be approximately normal and why or why not.)

***Answers may vary somewhat. Possible answers:***

* ***T-methods can be used since the distribution of sample means will be approximately normal due to the large sample size even though the sample data are skewed***
  + ***Students must include a reference to the normal probability plot. If a student believes the t-methods can be used, they must justify why they can be used given the normal probability plot shows skewed sample data.***
* ***T-methods cannot be used. Even though the sample size is large, it may not be large enough for the distribution of sample means to be normal since the sample data are heavily skewed.***

***Proper support should include***

* ***Reference to the distribution of sample means being normal or not***
* ***Reference to the shape of the sample data (even if they said the sample size is large enough for the distribution of sample means to be approximately normal)***

***If student is missing the above in their support, deduct 1/2 to 2 points depending on the rest of their argument.***

1. ***(1 point)*** Using the summary statistics and/or your histogram or dot plot, do you feel the null hypothesis will be rejected? Briefly explain.

***Grade based on support as this is a bit subjective. I would imagine full credit will be given with most (if not all) answers unless the student doesn’t use the summary statistics or graph to support their answer. I can see some students saying “no” because a majority of the responses were close to 90 minutes (or the center of the histogram is close to 90 minutes). I can also see some students saying “yes” because there were a number of responses much larger than 90 minutes resulting in a center greater than 90 minutes. Again, look for support.***

# Determine the p-value

1. ***(3 points)*** State the null and alternative hypotheses in notation. Define the parameter used in the notation in the context of the problem.

***H0: μ = 90 minutes***

***HA: μ > 90 minutes***

***μ = mean amount of time watching TV per day among OSU students***

***+ 1/2 for correct parameter***

***+ 1/2 for “=” in null hypothesis***

***+ 1/2 for > in alternative hypothesis***

***+ 1/2 for correct “hypothesized value”***

***+ 1 for defining the parameter in context (1/2 point for “mean time watching TV” and 1/2 point for population)***

In R, use the **t.test()** command to obtain the t-statistic, degrees of freedom, and the p-value – see page 4 in the Lab 1 Notes.

1. ***(1 point)*** Report the t-statistic with degrees of freedom.

***t190 = 0.75932***

**1/2 point for each**

1. ***(3 points)*** Based on the p-value, state a conclusion in the context of the problem. (Include the p-value in parentheses at the end of your conclusion.)

***There is not enough evidence to indicate that the mean amount of time spent watching TV among OSU students is more than 90 minutes per day (p-value = 0.2243).***

***+1.5 for correct adjective based on student’s p-value (“not enough” or “no evidence”)***

***+ 1/2 for “mean”***

***+ 1/4 for including variable***

***+ 1/4 for including population***

***+ 1/2 for including p-value (if incorrect p-value, take off 1/2 point)***

# Construct a confidence interval for the population mean

Use the **t.test()** command or the **confint()** command in R to obtain a 90% confidence interval for μ, the mean amount of TV watched per day for all OSU students – see page 4 in the Lab 1 Notes.

10. ***(1 point)*** Report the 90% confidence interval for μ in proper syntax.

***(84.94 minutes, 103.67 minutes)***

***+1/2 for correct bounds***

***+1/2 for correct syntax (in parentheses, lower bound first, bounds separated by comma, and units included). If missing some of these, take off 1/4 point.***

11. ***(3 points)*** Interpret the 90% confidence interval for μ in the context of the problem.

***We’re 90% confident the mean amount of time OSU students spend watching TV is between 84.94 and 103.67 minutes per day.***

***+1/2 for “90% confident”***

***+1 for “mean”***

***+1/2 for variable (time watching TV per day)***

***+1/2 for including population (ALL OSU students)***

***+1/2 for including the bounds with units.***