

R Tutorial for STAT 350 Lab 6

Author: Leonore Findsen, Chunyan Sun, Sarah H. Sellke, Jeremy Troisi

1. T Procedure for One Population

The same function, `t.test()`, is used for both the confidence interval and hypothesis test. In this tutorial, there will be some additions from what was mentioned in Lab 5. When you are doing the analysis, the same function should be used for both the confidence interval and hypothesis test for a particular problem. Points will be taken off if there are two `t.test()` functions for one problem unless more than one confidence interval or more than one hypothesis test is requested.

The normality assumption is not always valid. However, if your data is right skewed, you can sometimes make it normal by taking a log transformation using the natural logarithm)

Example (DATA SET: `yoga.txt`) Some people claim children who practice yoga are more physically fit, self-confident, and self-aware. A random sample of pre-teens (ages 10–12) practicing yoga was obtained and their meditation (or quiet breathing) times (in minutes) per day were recorded.

- Make a boxplot, histogram, and Normal quantile plot to determine if there any systematic departures from normality.
- Make a log transformation of the data. Make a boxplot, histogram and Normal quantile plot to verify that the log transformation of the distribution is roughly symmetric with no outliers and to confirm normality..
- From your observations in parts b), is it appropriate to use the t- procedure for the transformed data?
- Calculate and interpret the 95% lower bound for the mean number of minutes the pre-teens spent practicing yoga.
- Are you convinced that the mean number of minutes that the pre-teens spent practicing yoga is more than 1.4 log min (4.055 mins)? Carry out a significance test to justify your answer. Your significance level should be consistent with what was given in part d).
- Compare the results from parts (d) and (e). Are they the same or different? Please explain your answer.

Solution:

```
yoga=read.table(file="yoga.txt",header=T,sep="\t")
attach(yoga)

# The code for parts a) and c) are omitted. Please see
# Labs 2 (boxplot) and 3 (histogram) for details.
# Transformation
logMin = log(Time..min.)

# Parts d) and e) the same code s
# Parameters for t.test
# mu: mu_0
# You always indicate confidence level, alpha = 1 - C
# possibilities for alternative are "two.sided" (confidence interval),
# "less" (upper confidence bound for one-sided test), "greater"
# (lower confidence bound for one-sided test)
t.test(logMin, conf.level=0.95, mu = 1.4, alternative = "greater")
```

1

STAT 350: Introduction to Statistics

Department of Statistics, Purdue University, West Lafayette, IN 47907

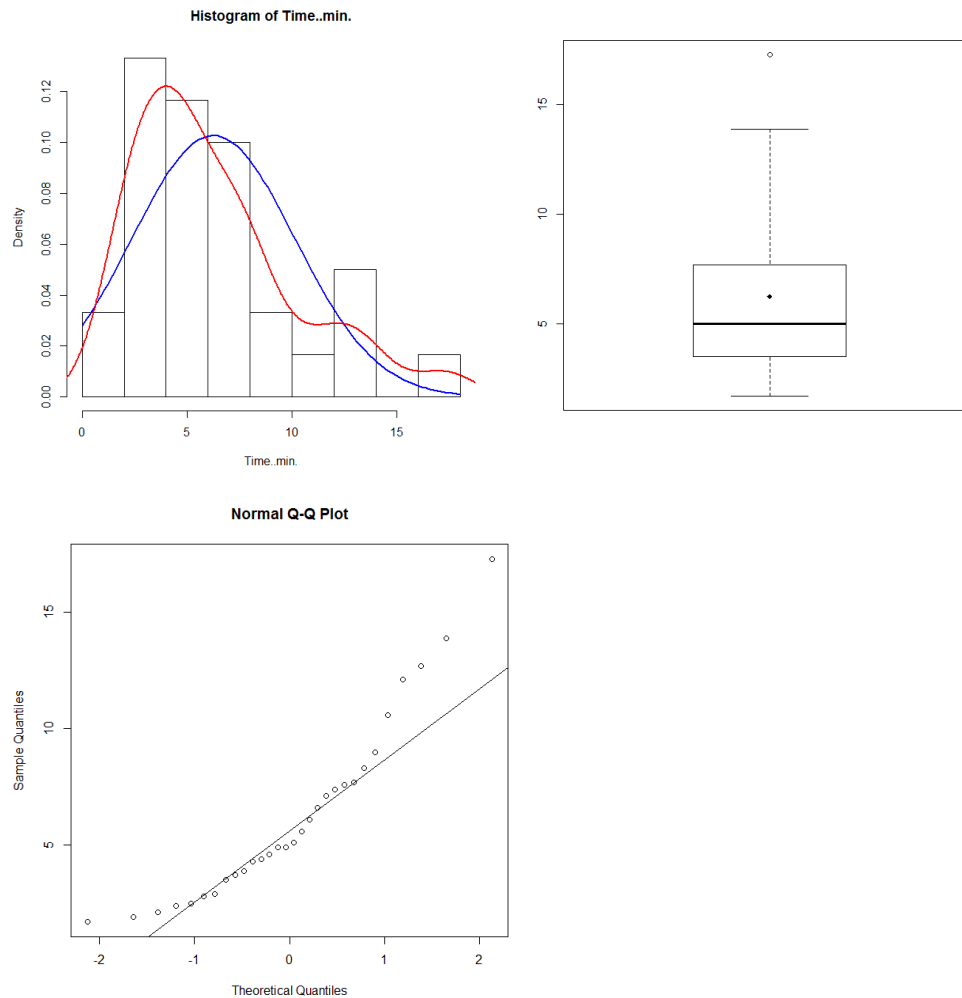
R Tutorial for STAT 350 Lab 6

Author: Leonore Findsen, Chunyan Sun, Sarah H. Sellke, Jeremy Troisi

a) Make a boxplot, histogram, and Normal quantile plot to determine if there any systematic departures from normality.

Solution:

See Labs 2 and 3 for details.



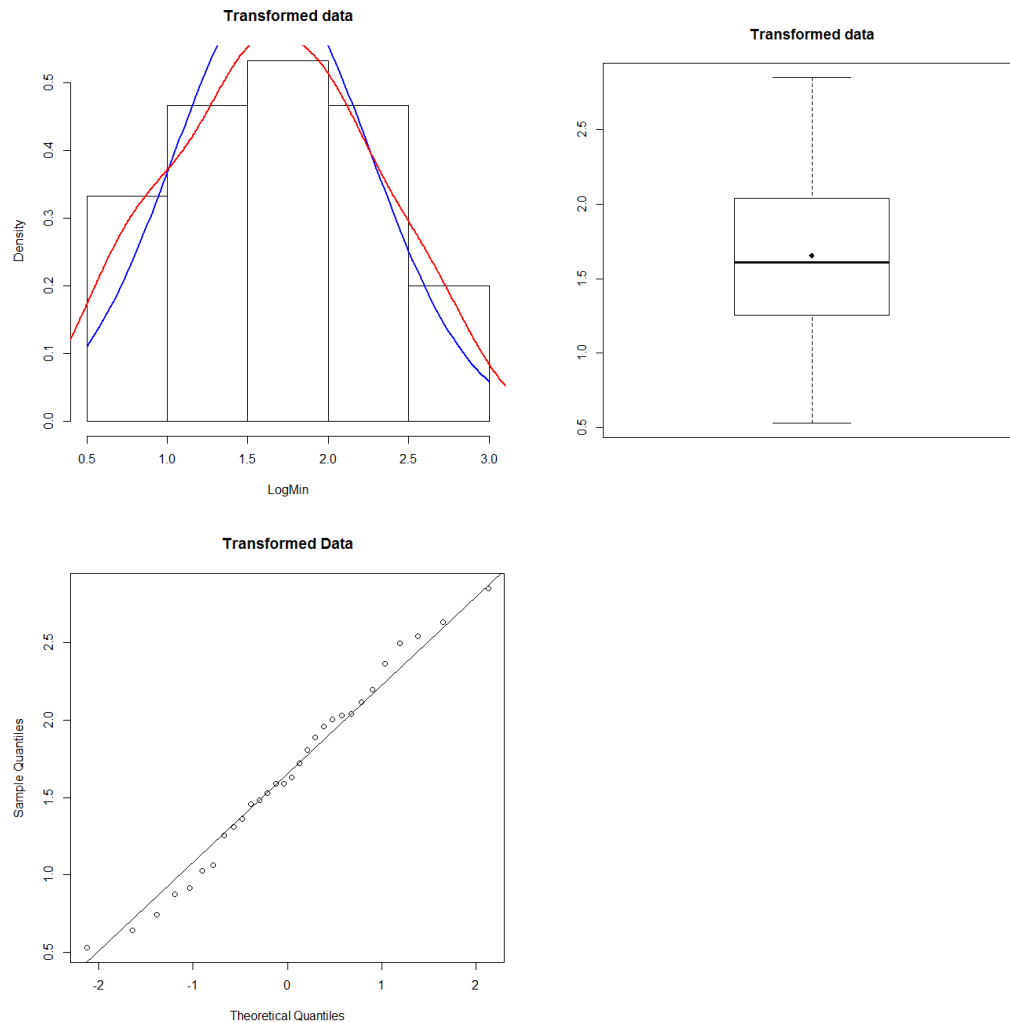
The distribution is right skewed with an outlier at high values. This is not a normal distribution.

R Tutorial for STAT 350 Lab 6

Author: Leonore Findsen, Chunyan Sun, Sarah H. Sellke, Jeremy Troisi

b) Make a log transformation of the data. Make a boxplot, histogram and Normal quantile plot to verify that the log transformation of the distribution is roughly symmetric with no outliers and to confirm normality.

Solution:



Now, the distribution is symmetric and the points on the probability plot roughly follow a straight line. This indicates that the distribution is approximately normal.

c) From your observations in parts b), is it appropriate to use the t- procedure for the transformed data?

Solution:

Assuming that the sample is SRS, the only other assumption that is necessary is to be sure that the distribution is normal. From the transformed data, the data is approximately normal. Therefore, this assumption is met.

R Tutorial for STAT 350 Lab 6

Author: Leonore Findsen, Chunyan Sun, Sarah H. Sellke, Jeremy Troisi

d) Calculate and interpret the 95% lower bound for the mean number of minutes the pre-teens spent practicing yoga.

Solution:

The same code (and output) should be used for both parts d) and e).

```
One Sample t-test

data: logMin
t = 2.2701, df = 29, p-value = 0.01541
alternative hypothesis: true mean is greater than 1.4
95 percent confidence interval:
 1.464032      Inf
sample estimates:
mean of x
 1.654579
```

The output for this part is highlighted in yellow.
The 95% lower bound is 1.464032 (1.464032, ∞)

We are 95% confident that the mean number of minutes that the pre-teens spent on yoga is more than 1.464032 log mins (4.3234 mins)

e) Are you convinced that the mean number of minutes that the pre-teens spent practicing yoga is more than 1.4 log min (4.055 mins)? Carry out a significance test to justify your answer. Your significance level should be consistent with what was given in part d).

The output for this part is highlighted in green above.

Step 1: Definition of the terms

μ is the population mean log number of minutes that pre-teens are practicing yoga..

Step 2: State the hypotheses

$$H_0: \mu = 1.4$$

$$H_a: \mu > 1.4$$

Step 3: Find the *Test Statistic*, report *DF*, find the *p - value*

$$t_{ts} = 2.2701$$

$$DF = 29$$

$$P\text{-value} = 0.01541$$

R Tutorial for STAT 350 Lab 6

Author: Leonore Findsen, Chunyan Sun, Sarah H. Sellke, Jeremy Troisi

Step 4: Conclusion:

$$\alpha = 1 - C = 1 - 0.95 = 0.05$$

Since $0.01541 \leq 0.05$, we should reject H_0

The data provides strong evidence (P-value = 0.01541) to the claim that the mean log number of minutes that pre-teens spend practicing yoga is at least 1.4 logmins (4.055 mins)

**f) Compare the results from parts (d) and (e). Are they the same or different?
Please explain your answer.**

Solution:

Since 1.464032 is greater than 1.4, we should reject the null hypothesis. That is, this data is consistent with the H_a ($\mu > 1.4$). However, we are also interested in the practical significance of our results. From the confidence bound, you would have to determine if 1.464032 (4.3234 mins) is 'close' to 1.4 (4.055 mins). In this case, you need to consider if the pre-teens are measuring in units less than a minute. If not, then there is no difference between the numbers.