SAS Tutorial for STAT 350 Lab 6

Author: Leonore Findsen, Cheng Li

1. T Procedure for One Population

The same procedure, proc ttest, is used for both the confidence interval and hypothesis In this tutorial, there will be some additions from what was mentioned in Lab 5. When you are doing the analysis, the same code should be used for both the confidence interval and hypothesis test for a particular problem. Points will be taken off if there are two proc ttest 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.

a) Make a boxplot, histogram, and Normal quantile plot to determine if there any systematic departures from normality.
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..
c) From your observations in parts b), is it appropriate to use the t- procedure for the transformed data?
d) Calculate and interpret the 95% lower bound for the mean number of minutes the pre-teens spent practicing yoga.
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).
f) Compare the results from parts (d) and (e). Are they the same or different? Please explain your answer.

Solution:

```
data yoga;
  infile "W:\yoga.txt" delimiter = '09'x firstobs = 2;
  input min;
  run;

*used for the first set of diagnostic tests;
proc ttest data = yoga H0 = 5 sides = U alpha = 0.05;
  var min;
  run;

* transformed data;
data transformed;
  set yoga;
  logmin = log(min); *this is the natural log, not common log;
```

SAS Tutorial for STAT 350 Lab 6

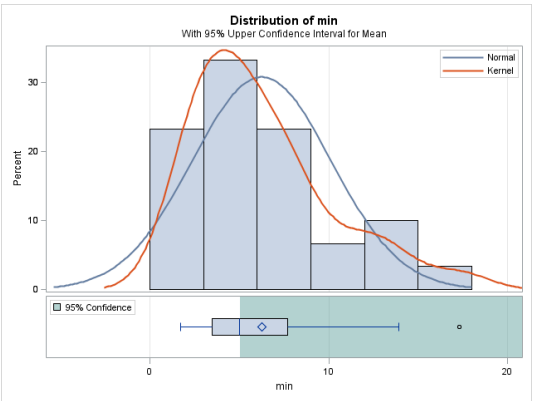
Author: Leonore Findsen, Cheng Li

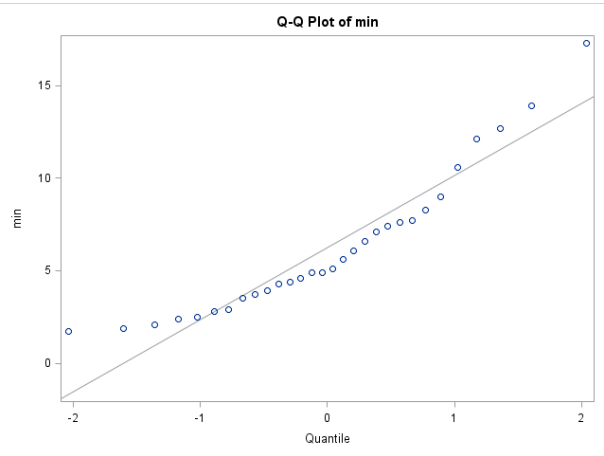
```
proc ttest data = transformed H0 = 1.4 sides = U alpha = 0.05;
* H0: indicates the value of the null value mu_0 for
    hypothesis tests only.
  Possibilities for sides are (significance tests, CI)
    2: 2-sided (not equal to, confidence interval)
    L: lower (<, upper confidence bound)
    U: upper (>, lower confidence bound)
  alpha = 1 - C: alpha is always specified for both CI and
    hypothesis tests;
    var logmin;
      run;
```

Remember: Only include the specific tables and or plots that are required. In addition, please only include the output in the sections where they are used.

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

Solution:



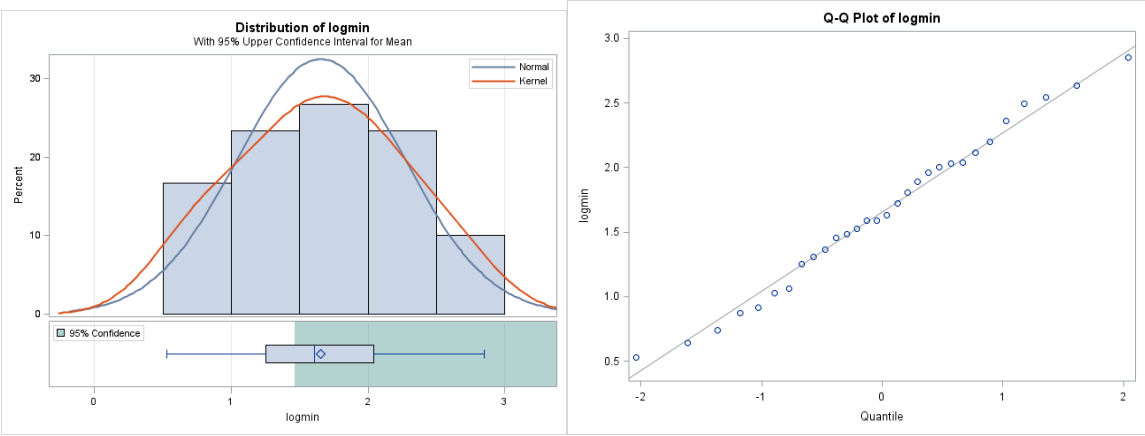


The distribution is right skewed with an outlier at high values. This is not a normal distribution.

Author: Leonore Findsen, Cheng Li

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.

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

Solution:

The same code should be used for both parts d) and e). I am including all of the output below; however, you should only include enough output to answer the question that is asked.

Author: Leonore Findsen, Cheng Li

The SAS System

The TTEST Procedure

Variable: logmin

N	Mean	Std Dev	Std Err	Minimum	Maximum
30	1.6546	0.6142	0.1121	0.5306	2.8507

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1.6546	1.4640	Infty	0.6142	0.4892	0.8257

DF	t Value	Pr > t
29	2.27	0.0154

To answer this question, only the red box is required.

The 95% lower bound is 1.4640 (1.4640, ∞)

We are 95% confident that the mean number of minutes that the pre-teens spent on yoga is more than 1.4640 log mins (4.3232 mins)

In the blue box: Std Dev stands for standard deviation, Std. Err stands for standard error.

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).

DF	t Value	Pr > t
29	2.27	0.0154

This is the same as the green box 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: $\mu > 1.4$

Author: Leonore Findsen, Cheng Li

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

t_{ts} = 2.27
DF = 29
P-value = 0.0154

Step 4: Conclusion:

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

Since 0.0154 ≤ 0.05 , we should reject H_0

The data provides strong evidence (P-value = 0.0154) 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 1.464 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 or our results. From the confidence bound, you would have to determine if 1.464 (4.3232 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.