SAS Tutorial for Stat 350 Lab 5

Author: Leonore Findsen, Cheng Li

1. Confidence Interval for One Population

To have SAS calculate a confidence interval (and perform a significance test), we use "proc ttest." There is procedure in SAS that performs a z test so I have generated the equations by hand. In SAS, we specify α not the confidence level; α = 1 - C. Therefore, in SAS, you indicate a 95% confidence level with α = 0.05.

Example (DATA SET: DMS.txt) Many food products contain small quantities of substances that would give an undesirable taste or smell if they were present in large amounts. An example is the "off-odors" caused by sulfur compounds in wine. Oenologists (wine experts) have determined the odor threshold, the lowest concentration of a compound that the human nose can detect. For example, the odor threshold for dimethyl sulfide (DMS) is given in the oenology literature as 25 micrograms per liter of wine (μg/l). Untrained noses may be less sensitive, however. Here are the DMS odor thresholds for 10 beginning students of oenology:

31 31 43 36 23 34 32 30 20 24

a) Make a boxplot and histogram to verify that the distribution is roughly symmetric with no outliers.
b) Make a Normal quantile plot to confirm that there are no systematic departures from Normality.
c) From your observations in parts a) and b), is it appropriate to use the t- procedure?
d) Generate a 95% confidence interval for the mean DMS odor threshold among all beginning oenology students (t test).

Solution:

```
data wine;
  infile "W:\PC-Text\DMS.txt" delimiter = '09'x firstobs = 2;
  input DMS group;
  run;

proc ttest data = wine sides = U alpha = 0.05;
* 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;
    var DMS;
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.

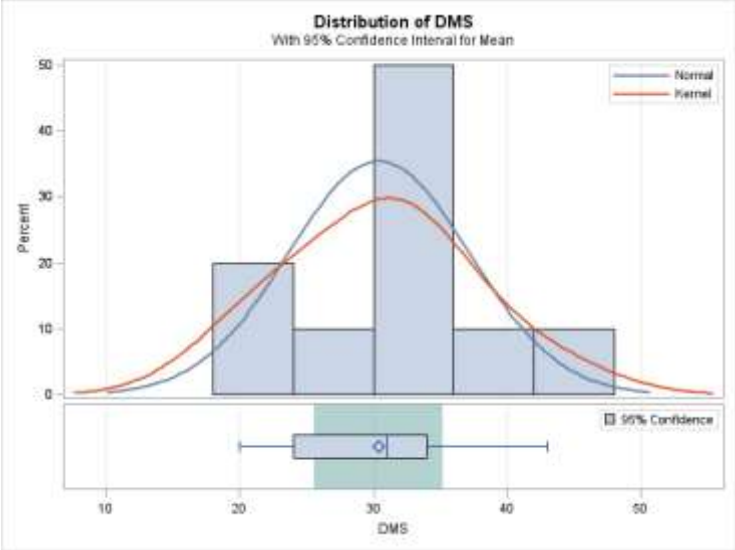
SAS Tutorial for Stat 350 Lab 5

Author: Leonore Findsen, Cheng Li

a) Make a boxplot and histogram to verify that the distribution is roughly symmetric with no outliers.

Solution:

No code is necessary because this is in the diagnostic plots.



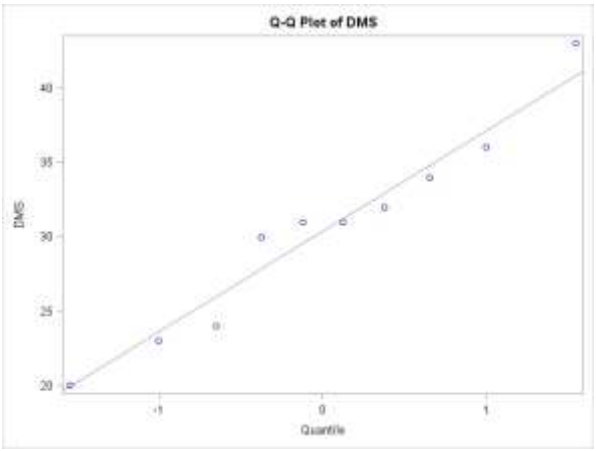
The median is close to the mean in the boxplot so the distribution is symmetrical. The histogram indicates that it is close to normal and symmetric also. I can determine the symmetry and normality better on the histogram than the boxplot because I can relate the blue and red curves easier. I see no outliers in the boxplot or the histogram; a modified boxplot is always generated.

b) Make a Normal quantile plot to confirm that there are no systematic departures from Normality.

Solution:

No code is necessary because this is in the diagnostic plots.

Author: Leonore Findsen, Cheng Li



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 a) and b), is it appropriate to use the t-procedure?

Solution:

Assuming that the sample is SRS, the only other assumption that is necessary is to be sure that the distribution is normal. Since the sample size is 10, we can not use CLT. However, from the normal quantile plot and the histogram, we can see that the distribution is approximately normal. Therefore, this assumption is met.

SAS Tutorial for Stat 350 Lab 5

Author: Leonore Findsen, Cheng Li

d) Generate a 95% confidence interval for the mean DMS odor threshold among all beginning oenology students (t test).

Solution:

The SAS System

The TTEST Procedure

Variable: DMS

N	Mean	Std Dev	Std Err	Minimum	Maximum
10	30.4000	6.7528	2.1354	20.0000	43.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
30.4000	25.5694	35.2306	6.7528	4.6448	12.3279

DF	t Value	Pr > |t|
9	14.24	<.0001

The 95% confidence interval is (25.5694, 35.2306)

Remember, only include the minimum number of required tables in your lab report. If you only include the part of the table including the confidence interval, be sure that you also include the labels.

2. Calculating the confidence interval for a z test.

Except as a classroom exercise, you should never use a z test. Therefore, SAS does not include any procedures for calculating the z test. The code that I have provided will do the necessary calculations.

SAS Tutorial for Stat 350 Lab 5

Author: Leonore Findsen, Cheng Li

e) Generate a 95% confidence interval for the mean DMS odor threshold among all beginning oenology students (z test).

Solution:

```
%LET alpha = 0.05; *value of alpha;
data wine;
  infile "W:\PC-Text\DMS.txt" delimiter = '09'x firstobs = 2;
  input DMS group;
  run;

*determins the value of n, xbar and sd (we are going to use the same
   standard deviation for both tests) and
   places the values in a file called zdata;

proc means data=wine;
var DMS;
output out=zdata mean=Mean n=n stddev=Std_Dev;
run;

*calculates the confidence interval;
data ztest;
  set zdata;
  CL = (1 - &alpha)*100;
  ZLower = Mean - probit(1-&alpha/2)*Std_Dev/sqrt(n);
  ZUpper = Mean + probit(1-&alpha/2)*Std_Dev/sqrt(n);

proc print data=ztest noobs;
   var n xbar std CL ZLower ZUpper;
run;
```

n	xbar	std	CL	zmin	zmax
10	30.4	6.75278	95	26.2147	34.5853

The 95% confidence interval for the z test is (26.2147, 34.5853).