

R Tutorial for STAT 350 Lab 5

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

1. Confidence Interval for One Population: t test

To have R calculate a confidence interval, we use the “t.test()” command. There is a special code to use if you are going to do a z test. This will be explained in part 2

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 ($\mu\text{g/l}$). 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

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

Solution:

```
DMS=read.table(file="DMS.txt",header=T)

# The code for parts a) and b) are omitted. Please see
#   Labs 2 (boxplot) and 3 (QQ plot and histogram) for details.

# Parameters for t.test
# 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(DMS$DMS, conf.level=0.95, alternative = "two.sided")
```

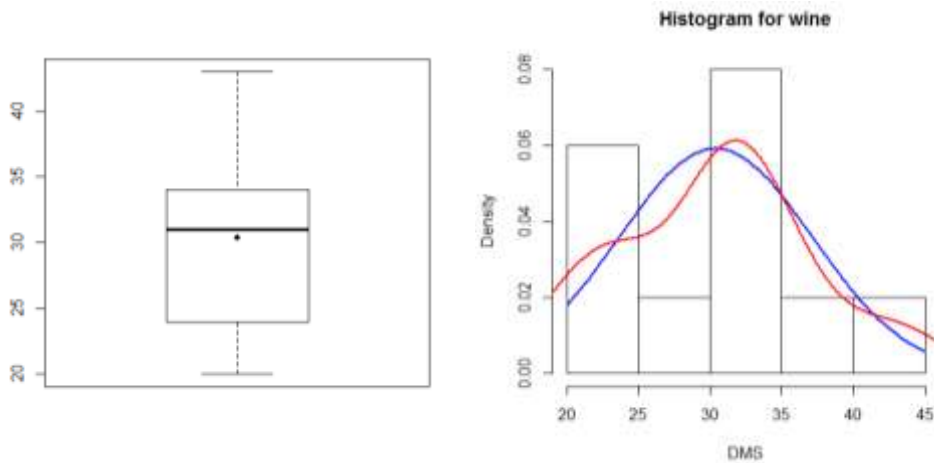
- Make a boxplot and histogram to verify that the distribution is roughly symmetric with no outliers.

Solution:

See the tutorials for Labs 2 and 3 for details.

R Tutorial for STAT 350 Lab 5

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

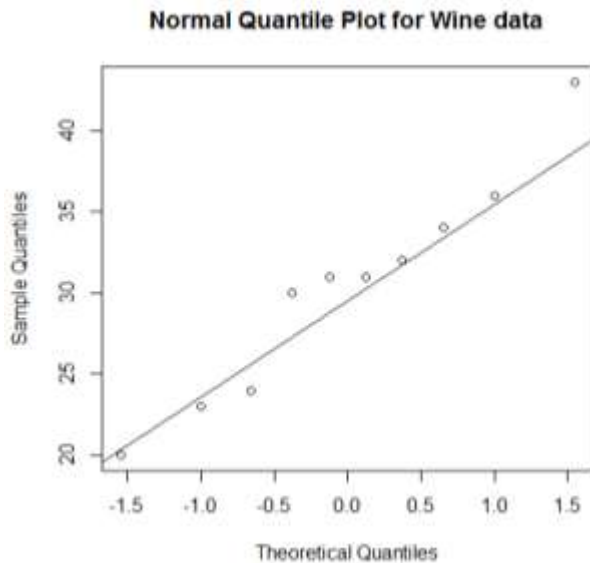


The median is close to the mean from the boxplot so the distribution is symmetrical. The histogram indicates that it is close to normal and symmetric also. I can determine the symmetry and normality better on the histogram than the boxplot because I can relate the blue and red curves easier. I see no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

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

Solution:

See the tutorial for Lab 3 for details.



The points on the probability plot roughly follow a straight line. This indicates that the distribution is approximately normal.

R Tutorial for STAT 350 Lab 5

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

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.

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

Solution:

```
One Sample t-test

data: DMS$DMS
t = 14.2361, df = 9, p-value = 1.775e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 25.56935 35.23065
sample estimates:
mean of x
 30.4
```

The 95% confidence interval is (25.56935, 35.23065)

2. Calculation of a z test

Except as a classroom exercise, you should never use a z test. Therefore, the default methodology in R does not include this test. However, there is a special library which does contain the z test which we will be using.

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

R Tutorial for STAT 350 Lab 5

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

Solution:

```
# the z test is not default in R, so we will be loading an
additional library to run this test
# You only need to run the following lines once.
install.packages('TeachingDemos')
library(TeachingDemos)

#the parameters are the same for the z.test as the t.test except
# that now, you need to specify the population standard
# deviation.
# z.test (x, stdev, alternative = "two.sided", conf.level = 0.95)
#   x contains the data, stdev is the known standard deviation,
#   alternative = "two.sided" indicates that this is a
#       confidence interval vs. a bound
# we are going to use the same standard deviation for both tests.

stdev =sd(DMS$DMS)
stdev

z.test(DMS$DMS, conf.level = 0.95, alternative="two.sided",
sd=stdev)
```

One Sample z-test

```
data:  DMS$DMS
z = 14.2361, n = 10.000, Std. Dev. = 6.753, Std. Dev. of the sample
mean = 2.135, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 26.21466 34.58534
sample estimates:
mean of DMS$DMS
      30.4
```

The 95% confidence interval is (26.21466, 34.58534)