Author: Min Ren

Lab 5 (75 pts.) - One Sample *T* Confidence Interval and Test
## Objectives: Confidence interval and significance tests

A. (25 points) Number of Friends on Facebook (Data Set: facebookfriends.txt - Website) Facebook provides a variety of statistics on their Web site that detail the growth and popularity of the site. One such statistic is that the average user has 130 friends. Consider the following SRS of *n* = 30 Facebook users from a large university.

99	148	158	126	118	112	103	111	154	85
120	127	137	74	85	104	106	72	119	160
83	110	97	193	96	152	105	119	171	128

1.   (5 pts) Do you think these data are Normally distributed? Use graphical methods to examine the distribution. Write a short summary of your findings.

Solution:
```
> friends<-read.table("facebookfriends.txt",header=T)
> friends
> attach (friends)
> hist(Friends,main="Histogram of facebook friends",freq=F)
> curve(dnorm(x,mean=mean(Friends),sd=sd(Friends)),col="blue",lwd=2,
  add=T)
> lines(density(Friends), col = "red", lwd=2)
> qqnorm(Friends)
> qqline(Friends)
```
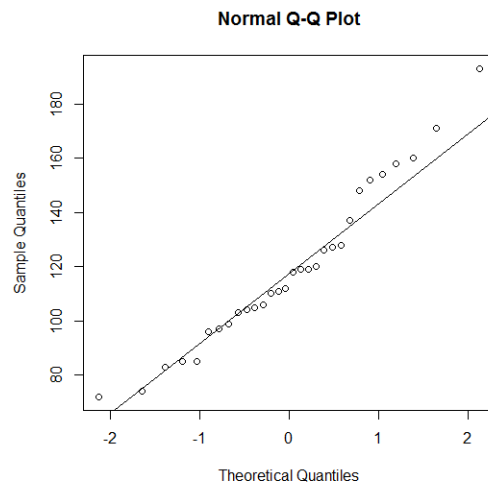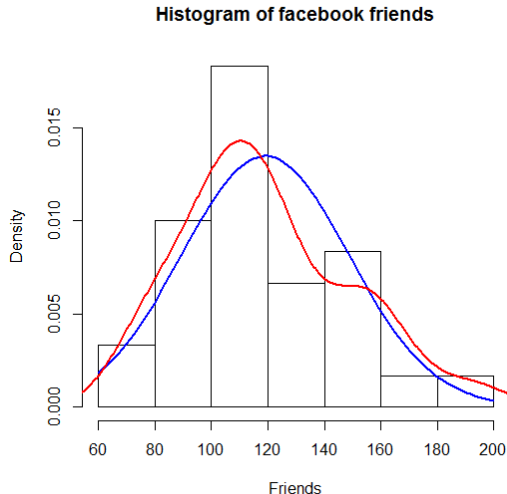

The data are roughly normal, since all the points in the QQplot are close to the straight line and the empirical density curve is close to the theoretical normal density curve.

2. (5 pts) Is it appropriate to use the *t* methods of this section to compute a 95% confidence interval for the mean number of Facebook users at this large university? Explain why or why not.

Solution:
Since the data are roughly normal with no outliers, it is appropriate to use the t methods for the mean number of Facebook users at this large university.

Author: Min Ren

3. (5 pts) Find the mean, standard deviation, standard error, and margin of error for 95% confidence. From those values, compute the 95% CI for μ. The CI is NOT to be calculated from the software package though the values in the first sentence maybe calculated via software. If the numbers are obtained via a software, please include the appropriate output. If the numbers are calculated by hand, please show your work. Work is required for the calculation of the CI.

Solution:

```
> #the mean
> avg<-mean(Friends)
> avg
[1] 119.0667
> #the standard deviation
> sdev<-sd(Friends)
> sdev
[1] 29.56691
> #the standard error
> se <- sd(Friends)/sqrt(length(Friends))
> se
[1] 5.398155
> #the margin error
> marginerror<-qt(0.975,length(Friends)-1)*se
> marginerror
[1] 11.04047
> #upper limit
> avg+marginerror
[1] 130.1071
> #lower limit
> avg-marginerror
[1] 108.0262
```

95% CI for μ is (108.0262,130.1071)

You have to use R to calculate the mean and standard deviation. The following can be used to calculate the rest of the numbers by hand.

$$standard\ error = \frac{s}{\sqrt{n}} = \frac{29.5669}{\sqrt{30}} = 5.39815$$

margin of error = (t*(29)) (standard error) = (2.045) (5.3982) = 11.039
CI: $\bar{x} \pm$ m = 119.1 ± 11.039 ==> (108.06, 130.139)

The difference between the values is that the exact critical value from R is 2.04523 and there is no round off error.

4   (5 pts) Report the 95% confidence interval for μ, the average number of friends for Facebook users at this large university. This answer is obtained from the software package so the output needs to be reported. Compare with your answer in part 3.

Solution:
> t.test(Friends,conf.level=0.95)

    One Sample t-test

data: Friends
t = 22.0569, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 108.0262 130.1071
sample estimates:
mean of x
 119.0667

95% CI for μ is (108.0262,130.1071)

The result is identical to what is calculated via R and is close to what is calculated by hand.

5 (5 pts) Interpret your 95% confidence interval for μ obtained in part 4.

Solution:
We are 95% confident that the population (true) mean number of friends for Facebook users at this large university falls in the interval (108.0262, 130.1071).

B (50 points) Counts of Picks in a 1-lb bag (Data Set: pickcount.txt -Website)
A guitar supply company must maintain strict oversight on the number of picks they package for sale to customers. Their current advertisement specifies between 900 and 1000 picks in every bag. An SRS of thirty-six 1-pound bags of picks were collected as part of a Six Sigma Quality Improvement effort within the company. The number of picks in each bag are shown in the following table.

924	925	967	909	959	937	970	936	952
919	965	921	913	886	956	962	916	945
957	912	961	950	923	935	969	916	952
917	977	940	924	957	920	986	895	923

1.   (12 pts) Create a histogram, boxplot, and a Normal quantile plot of these counts.

Solution:
```
> count<-read.table("pickcount.txt", header=T)
> count
> attach(count)
> hist(PickCount,main="Histogram of the number of picks",freq=F)
> curve(dnorm(x,mean=mean(PickCount),sd=sd(PickCount)),col="blue",
  lwd=2,add=T)
> lines(density(PickCount),col="red",lwd=2)
> boxplot(PickCount,main="Boxplot of the number of picks")
> points(mean(PickCount),pch=18)
> qqnorm(PickCount)
> qqline(PickCount)
```

Author: Min Ren

Boxplot of the number of picks

Histogram of the number of picks

Normal Q-Q Plot
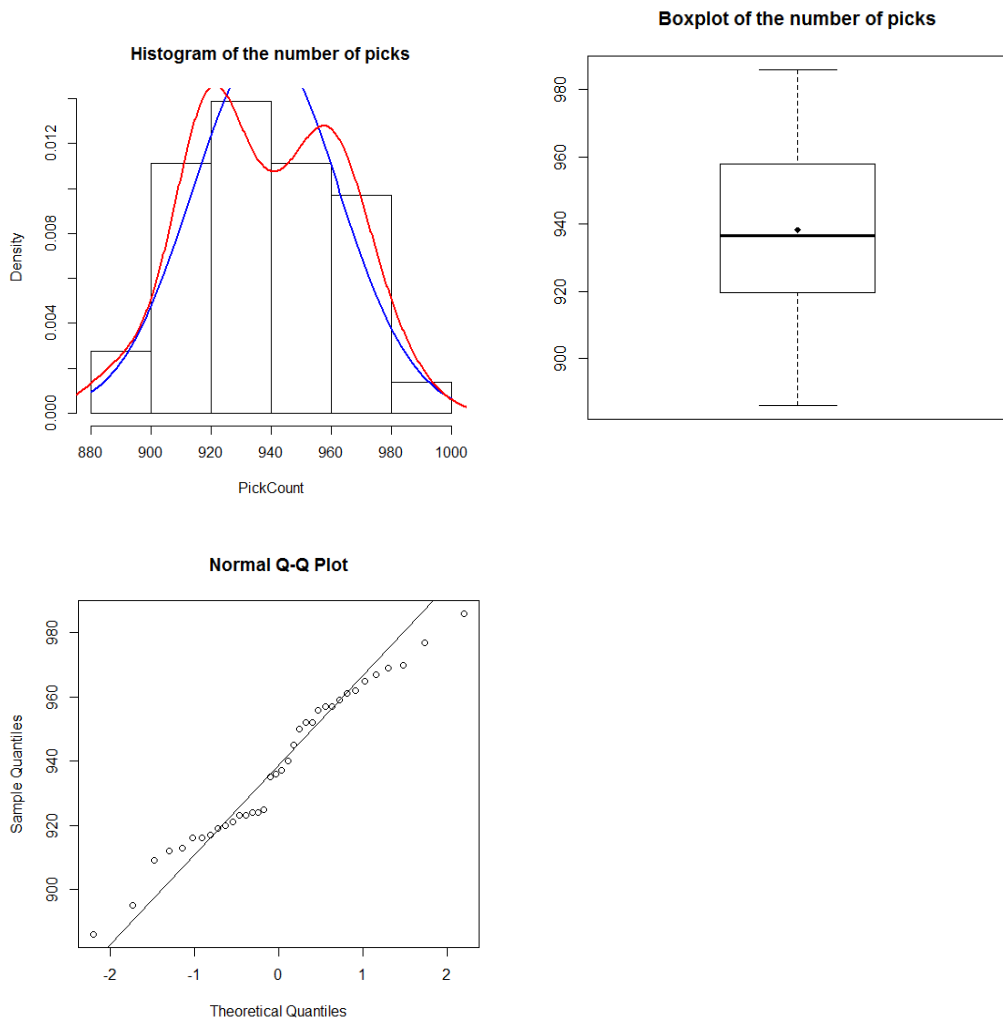
2. (4 pts) Write a description of the distribution. Comment on the skewness and Normality of the data. Note if there are any outliers.

Solution:
From the graphical summary above, we can find that the data is not skewed and is approximately normally distributed. There are no outliers

3. (5 pts) Based on your observations in part (1), is it appropriate to analyze these data using the *t* procedures? Briefly explain your response.

Solution:
It is appropriate to analyze these data using the t procedure since the data is symmetric and roughly normal without any outliers.

4. (3 pts) Find the mean, the standard deviation, and the standard error of the mean for this sample.

Solution:
```
> avg<-mean(PickCount)
> avg
[1] 938.2222
> sdev<-sd(PickCount)
> sdev
[1] 24.2971
> se <- sdev/sqrt(length(PickCount))
> se
[1] 4.049517
```

You may calculate the standard error by hand

$standard\ error = \frac{s}{\sqrt{n}} = \frac{24.2971}{\sqrt{36}} = 4.049517$

5. (5 pts) Find the 95% lower confidence bound for the mean number of picks in a 1-pound bag.

Solution (for parts 5 and 6):
```
> t.test(PickCount,conf.level=0.95, mu=925, alternative="greater")

        One Sample t-test

data:  PickCount
t = 3.2651, df = 35, p-value = 0.001225
alternative hypothesis: true mean is greater than 925
95 percent confidence interval:
 931.3803        Inf
sample estimates:
mean of x
 938.2222
```

The output for this part is highlighted in yellow.

The 95% lower bound is 931.3803 (931.3803, ∞)

6. (8 pts) Do these data provide evidence that the average number of picks in a 1-pound bag is greater than 925? Carry out a test of significance using the four-step procedure, with a significance level of 5%, state your hypotheses, the value of test statistic, the *P*-value, and your conclusions. Please provide the relevant output required for the steps and include all four steps written out by hand.

Solution:
The output for this part is highlighted in green above.

Step 0: Definition of the terms

μ is the population mean number of picks in a 1-pound bag.

Step 1: State the hypotheses

H_0: μ = 925
H_a: μ > 925

Author: Min Ren

Step 2: Find the *Test Statistic*.

t_t = 3.2651

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

DF = 35
P-value = 0.001225

Step 4: Conclusion:

$\alpha = 1 - C = 1 - 0.95 = 0.05$
Since $0.001225 \leq 0.05$, we should reject H_0

The data provides strong evidence (P-value =0.001225) to the claim that the average number of picks in a 1-pound bag is greater than 925.

7. (8 pts) Do these data provide evidence that the average number of picks in a 1-pound bag is greater than 935? Carry out a test of significance using the four-step procedure, with a significance level of 5%. Please see the directions for Part 6.

Solution:
```
> t.test(PickCount,conf.level=0.95, mu=935, alternative="greater")

        One Sample t-test

data:  PickCount
t = 0.7957, df = 35, p-value = 0.2158
alternative hypothesis: true mean is greater than 935
95 percent confidence interval:
 931.3803       Inf
sample estimates:
mean of x
 938.2222
```

The output for this part is highlighted in green above.

Step 0: Definition of the terms

μ is the population mean number of picks in a 1-pound bag.

Step 1: State the hypotheses

H_0: μ = 935
H_a: μ > 935

Step 2: Find the *Test Statistic*.

t_t = 0.7957

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

DF = 35
P-value = 0.2158

Step 4: Conclusion:

$\alpha = 1 - C = 1 - 0.95 = 0.05$
Since $0.2158 > 0.05$, we failed to reject H_0

Author: Min Ren

The data does not provide evidence (P-value =0.2158) to the claim that the average number of picks in a 1-pound bag is less than 935.

8. (5 pts) Compare your conclusions for parts (5), (6) and (7). Are they the same or different?

Solution:
In part 5, we find the lower bound is around 931. In part 6, we find that the mean is greater than 925. Thus 5 and 6 are consistent because if μ is greater than 931, it is greater than 925. However, in 7, we find that the mean is not greater than 935. This is also consistent with part 5 because if μ is greater than 931, it is not necessarily greater than 935.