Sarah Higgins

Probability & Statistics

Professor Federico

12/1/2017

Final Project

1. For AC and PP in the 2017 stats, perform the following computations:
   a. Find the mean, median, mode, variance, and standard deviation S of this sample:

      To find the mean of the assembly cost (AC), I used Excel. See the following screenshots: =AVERAGE(A4:A39) $335.63

      The mean of the assembly cost(AC) is $335.63. The dollar sign is missing in the excel spreadsheet because I scrubbed the data for importing the data into R in later steps.

      To find the mean of the percent of profit(PP), I used Excel. See the following screenshots: =AVERAGE(B4:B39) 35.44444

      The mean of the percent of profit(PP) is 35.444.

      To find the median of the assembly cost(AC), I used Excel. See the following screenshots: =MEDIAN(A4:A39) $337.83

      The median of the assembly cost(AC) is $337.83.

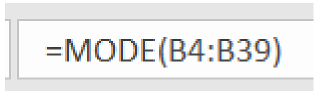
      To find the median of the percent of profit(PP), I used Excel. See the following screenshots: =MEDIAN(B4:B39) 36

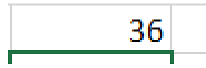
      The median of the percent of profit(PP) is 36.

      To find the mode of the assembly cost(AC), I used Excel. See the following screenshots: =MODE(A4:A39) 341.25

The mode of the assembly cost(AC) is $341.25.

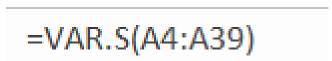
To find the mode of the percent of profit(PP), I used Excel. See the following

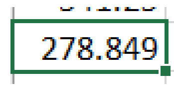
screenshots: 



The mode of the percent of profit(PP) is 36.

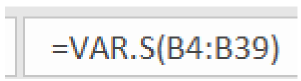
To find the variance of the assembly cost(AC), I used Excel. Because this is a SAMPLE of size 36, I used the command VAR.S() rather than VAR(). See the

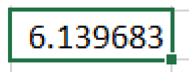
following screenshots: 



The variance of the assembly cost(AC) is $278.849.

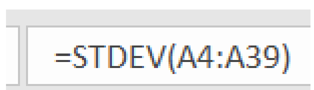
To find the variance of the percent of profit(PP), I used Excel. See the following

screenshots: 



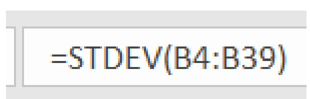
The variance of the percent of profit(PP) is 6.139683.

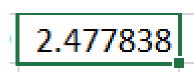
To find the standard deviation S of the assembly cost(AC), I used Excel. See the

following screenshots: 

The standard deviation S of the assembly cost(AC) is $16.6988.

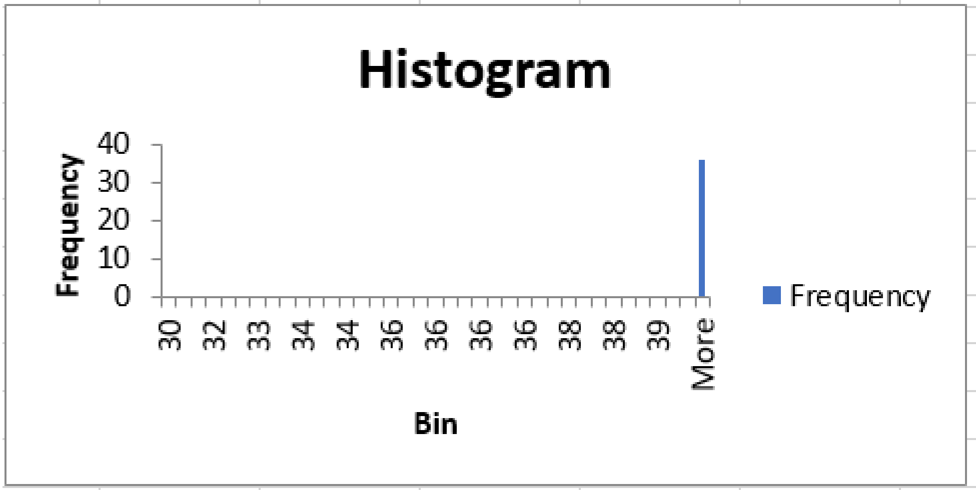
To find the standard deviation S of the percent of profit(PP), I used Excel. See the

following screenshots: 



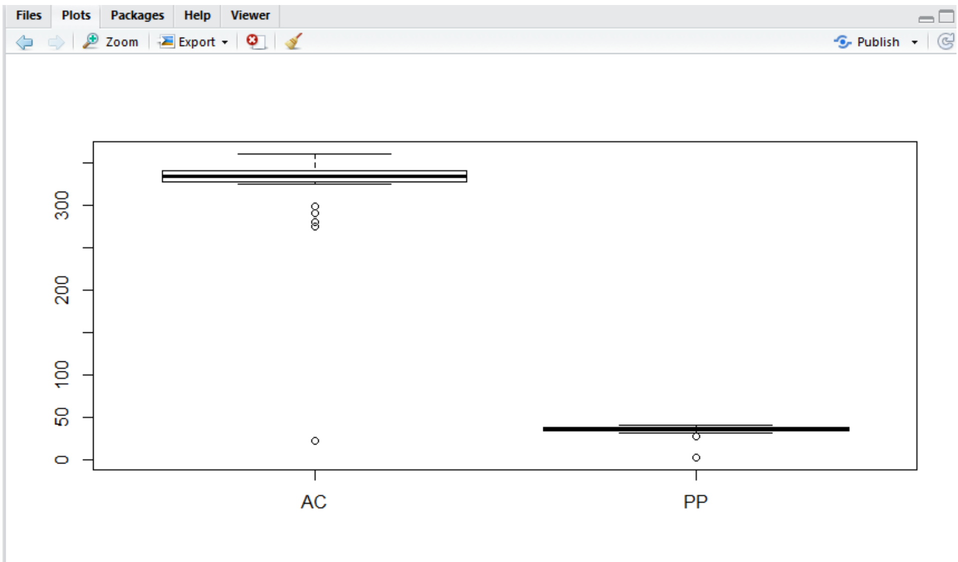
The standard deviation S of the percent of profit(PP) is 2.477838.

b. Create a histogram using your frequency table.



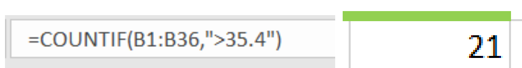
Above is a Histogram of assembly cost(AC) and percent of profit(PP) for 2017 Statistics.

c. Draw a boxplot of the assembly costs.



Above is a Boxplot of assembly cost(AC) and percent of profit(PP) for 2017 Statistics. Created with command boxplot(x); in R.

d. If 35.4% or more is considered a good percentage of profit, what proportion of the 36 days' PP are 35.4% or more?

Using Excel, we can count the number of cells that are greater than 35.4% with (see the following screenshots): 

This tells us that 21 of the 36 cells are greater than 35.4%, which is 21/36. Simplified gives us 7/12, or .58333. This means that the proportion of the 36 days' PP are 35.4% or more is 7/12 or .58333.

e. Of those days with PP of 35.4% or more, what percentage of those days were accomplished with an assembly cost of $330 or less?

Of those days with PP of 35.4% or more, the percentage of those days were accomplished with an assembly cost of $330.0 or less is 19.05%. This is because the total number of days this is true is 4/21, 21 being the number of days out of the 36 that have a (PP) of more than 35.4%. So, 4/21 days = .19047, which gives us a percentage value of 19.05%.

2. Assuming that the mean AC for the first 36 months is normally distributed, do the following:

a. Find the mean and the standard deviation, and then find what proportion of AC was between $320 and $350 per month.

Because the mean and standard deviation for the Normal Distribution are mean and standard deviation respectively, we can calculate these simply in Excel:

Using Excel, we find the mean in the Normal Distribution to be $328.70 (See the following screenshots): =AVERAGE(A1:A37) 328.70

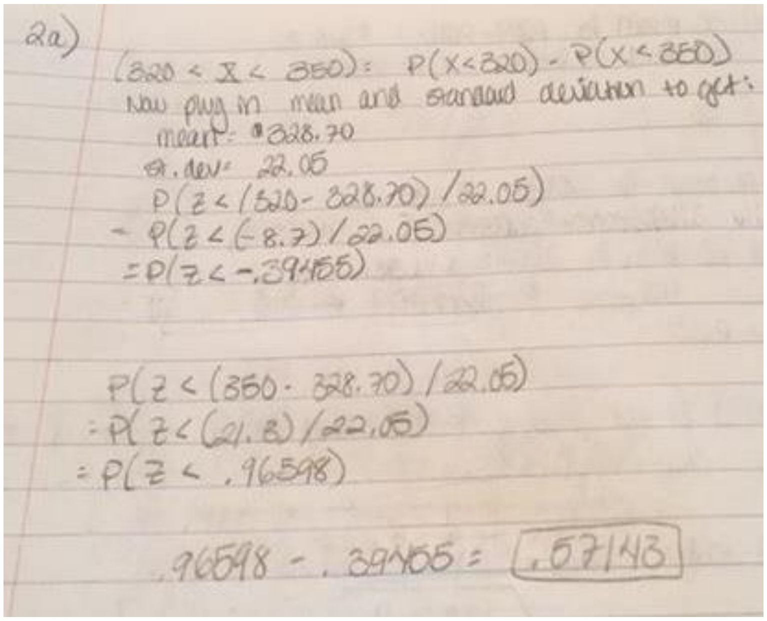
This the mean for the AC. Using Excel, we can also find the standard deviation in the Normal Distribution. See the following screenshots: =STDEV(A1:A37) 22.05

The standard deviation of the AC in the 36 months for the 2014-2016-Statistics_Monthly is $22.05.

The proportion of AC between $320 and $350 is 27/36, or .75. This can be found in Excel with the following commands (See following screenshots):

=COUNTIFS(A1:A37, ">320", A1:A37,"<350") 27.00

However, I also did it out by hand, and got a different answer. Because of that, I am going to leave both answers here because I am unsure of which one is correct, and hopefully this will help:



2a)

(320 < X < 350): P(X<320) - P(X<350)

Now plug in mean and standard deviation to get:

mean: 328.70

st. dev: 22.05

P(Z < (320 - 328.70) / 22.05)

= P(Z < (-8.7) / 22.05)

= P(Z < -.39455)

P(Z < (350 - 328.70) / 22.05)

= P(Z < (21.3) / 22.05)

= P(Z < .96598)

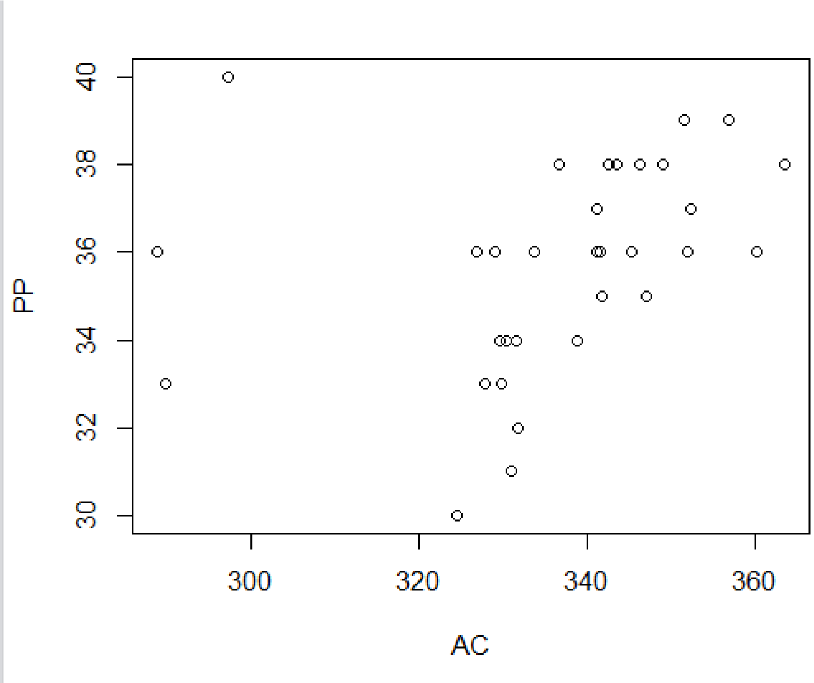
.96598 - .39455 = $\boxed{.57143}$

b. Use the 36 months of statistics to create a scatterplot, with AC on the X-axis and PP on the y-axis.

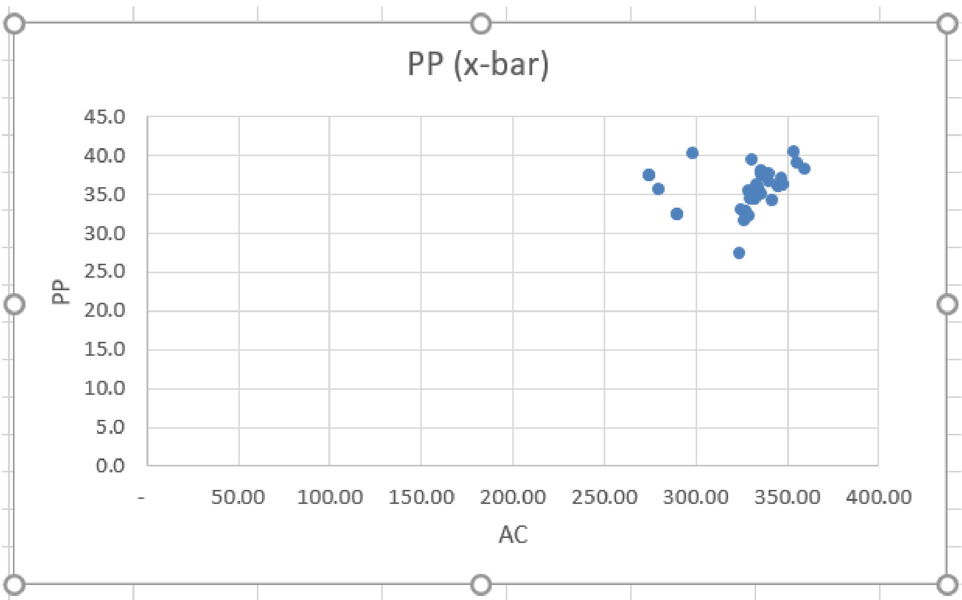
Commands used to create scatter plot:

```
> setwd("~/Prob&Stats/Final Project")
> x=("2014-2016Scrubbed.txt", header=T)
Error: unexpected ',' in "x=("2014-2016Scrubbed.txt","
> x=read.table("2014-2016Scrubbed.txt", header=T)
Error in read.table("2014-2016Scrubbed.txt", header = T) :
  more columns than column names
> x=read.table("2014-2016Scrubbed.txt", header=T)
> attach(x)
> plot(x)
> 
```

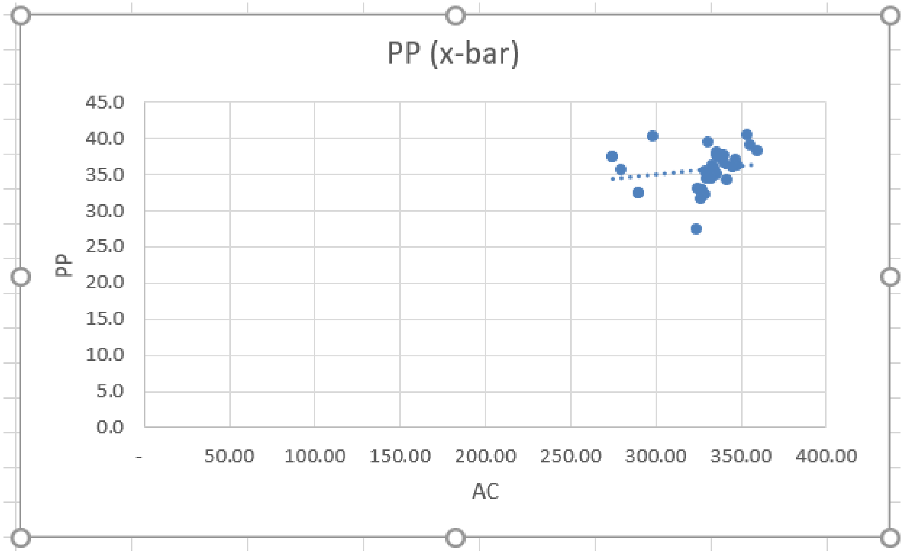
Scatter plot with AC on the X-axis and PP on the y-axis:



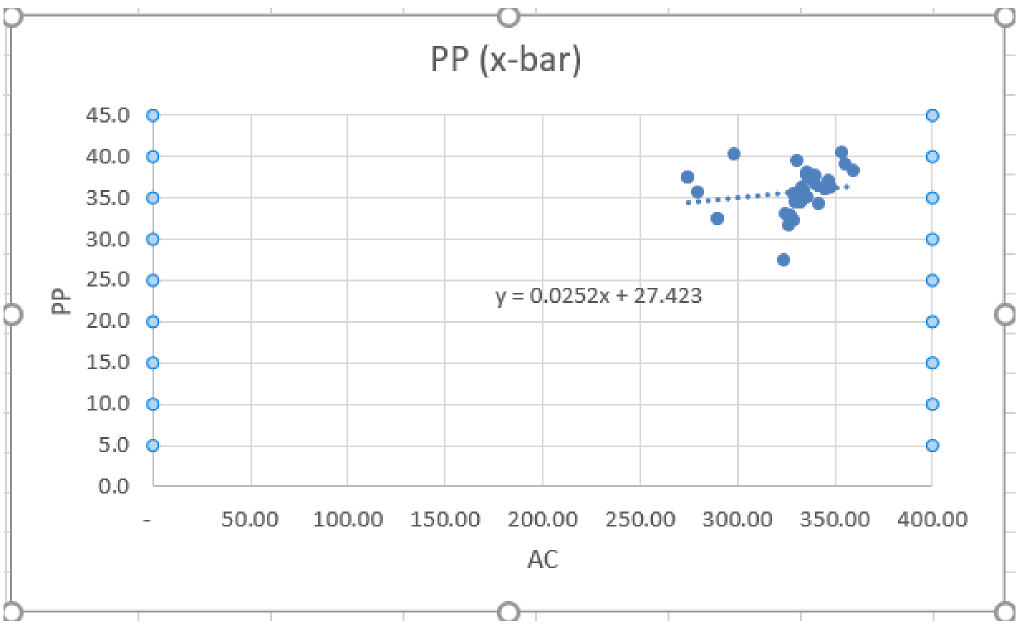
Scatterplot in Excel:



c. Draw in a trend line and calculate the equation for the line of regression.



With the Line of regression equation:



d. Using this equation from part c), approximate to the nearest whole number the percent of profit that could result from requiring an average assembly cost of $340 during the 36 months. Also approximate to the nearest cent the assembly cost that would result in producing an average of 34.7% per month during that same 36 months.

Using this equation, we find that y = 0.0252($340) + 27.423 = 35.991 during the 36 months. To approximate to the nearest cent, we do:

34.7 = 0.0252x + 27.423, find the value of x, and that is our answer:

$$34.7 = 0.0252x + 27.423$$

$$-27.423 \qquad\qquad - 27.423$$

$$7.277 = 0.0252x$$

Divide each side by 0.0252:

288.7698

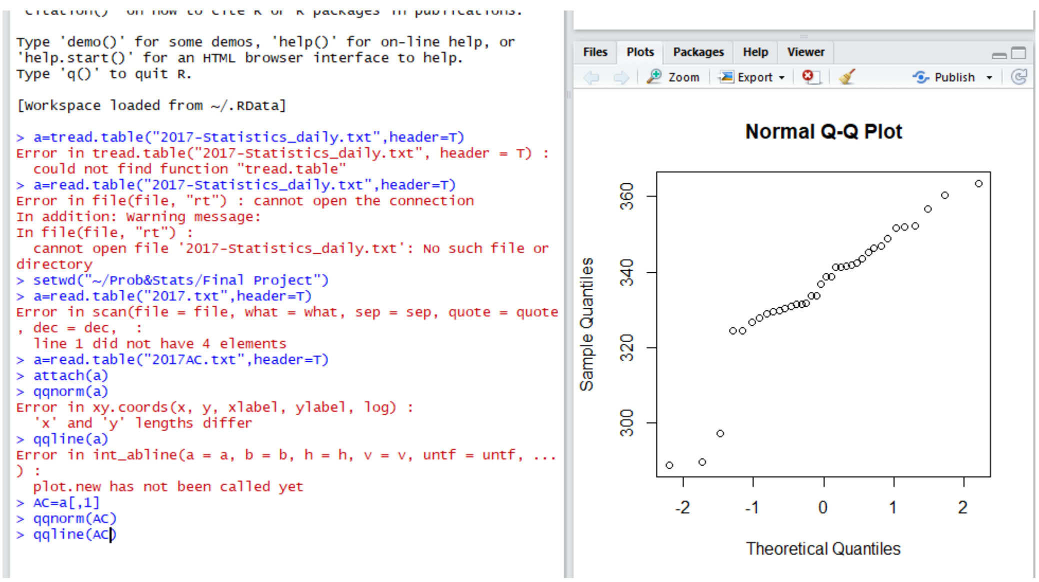
Round to the nearest cent, we get:

$288.77

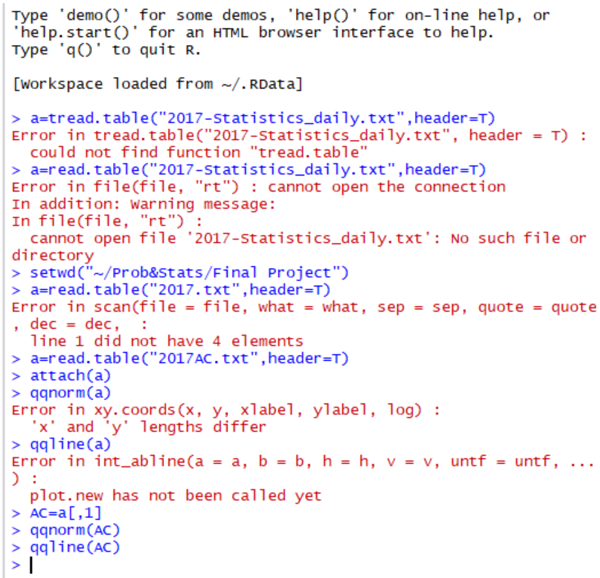
The nearest cent the assembly cost that would result in producing an average of 34.7% is $288.77.

e. Using R, show that the data you used to establish your sample of sample means for the first semester is normally distributed by creating a Normal Quantile Plot (qqnorm).

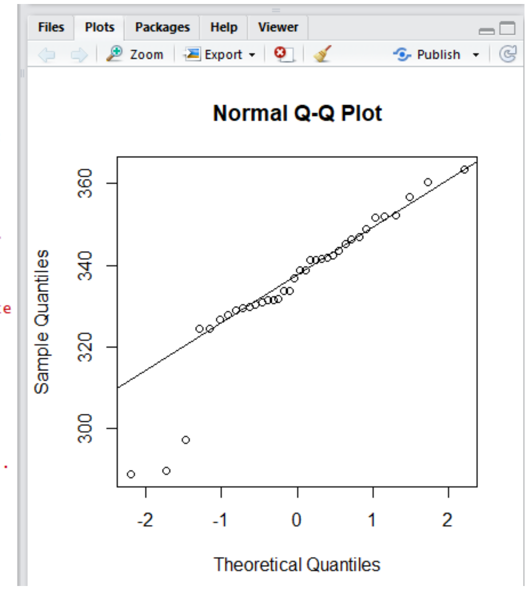
"The First Semester" implies that it is only asking for the 2017 Statistics, because logically that makes the most sense, seeing as that is a collection of data over the course of 3 months. So, using qqnorm() after converting the data in a as a=AC[,1], we get (see following screenshot below):



After running the qqline(AC) code on the bottom of that to insert the line, we get (see following screenshot below):

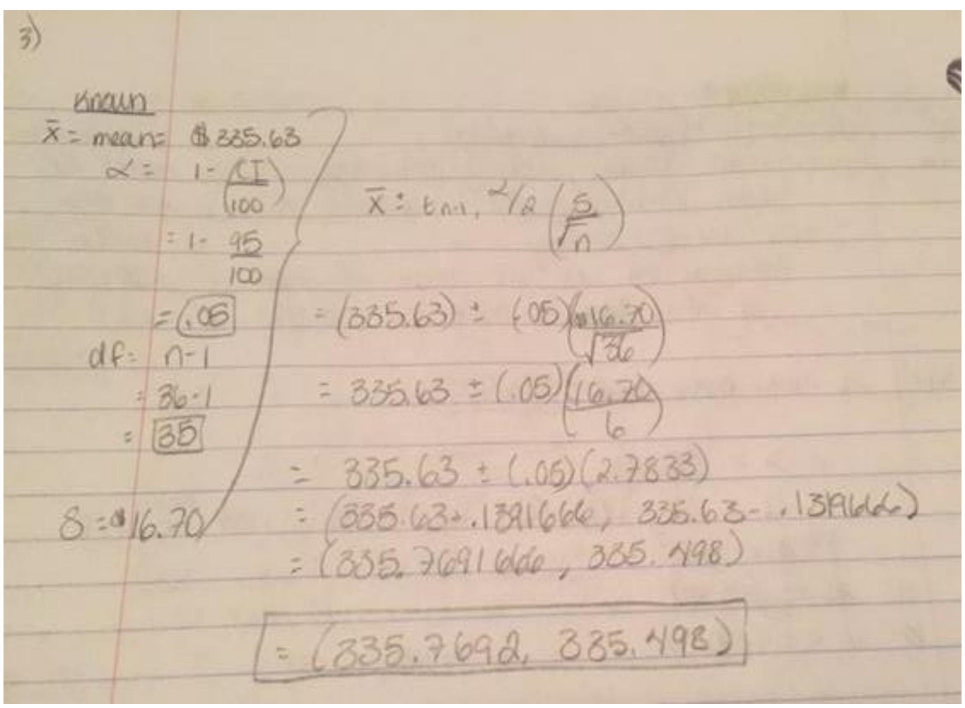


```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> a=tread.table("2017-Statistics_daily.txt",header=T)
Error in tread.table("2017-Statistics_daily.txt", header = T) :
  could not find function "tread.table"
> a=read.table("2017-Statistics_daily.txt",header=T)
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file '2017-Statistics_daily.txt': No such file or
directory
> setwd("~/Prob&Stats/Final Project")
> a=read.table("2017.txt",header=T)
Error in scan(file = file, what = what, sep = sep, quote = quote
, dec = dec,  :
  line 1 did not have 4 elements
> a=read.table("2017AC.txt",header=T)
> attach(a)
> qqnorm(a)
Error in xy.coords(x, y, xlabel, ylabel, log) :
  'x' and 'y' lengths differ
> qqline(a)
Error in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...
) :
  plot.new has not been called yet
> AC=a[,1]
> qqnorm(AC)
> qqline(AC)
> |
```



3. Assuming that the 36 days in 2017 are a representative sample and that the 36 days' AC fit a Normal Distribution Curve, do an estimation of the Mean by establishing a margin of error and a confidence interval of the mean assembly cost recorded during the 36 days. Use a 95% confidence-interval.

To find the 95% confidence interval CI, I did the calculations out by hand and attached a picture of my work (please see picture below) :



3)

Known

$\bar{x}$ = mean = \$335.63

$\alpha = 1 - \left(\frac{CI}{100}\right)$

$= 1 - \frac{95}{100}$

$= \boxed{.05}$

df = n-1

= 36-1

= $\boxed{35}$

s = \$16.70

$\bar{x} \pm t_{n-1, \alpha/2}\left(\frac{s}{\sqrt{n}}\right)$

$= (335.63) \pm (.05)\left(\frac{16.70}{\sqrt{36}}\right)$

$= 335.63 \pm (.05)\left(\frac{16.70}{6}\right)$

$= 335.63 \pm (.05)(2.7833)$

= (335.63 + .1391666, 335.63 - .1391666)

= (335.7691666, 335.498)

$\boxed{= (335.7692,\ 335.498)}$

The 95% Confidence Interval CI for the AC is (335.7692, 335.498).

To find the margin of error,

4. For the 36 months from 2014-2016, you calculated the mean monthly AC for the company as well as the standard deviation. Use these as the basis for the mean hypothesis claim, H0.

   a) Has this changed during the 36 days at the start of 2017?

      Yes, it has changed during the 36 days at the start of 2017. The mean obtained from the 2014-2016 statistics is $328.70, and the mean obtained from the 2017 statistics is $335.63. While it's not much different from the 2014-2016 mean, it's still a $6.93 difference. The standard deviation has also changed. The standard deviation obtained from the 2014-2016 statistics is $22.047824, and the standard deviation obtained from the 2017 statistics is $16.70. This is a difference of $5.347824. So,

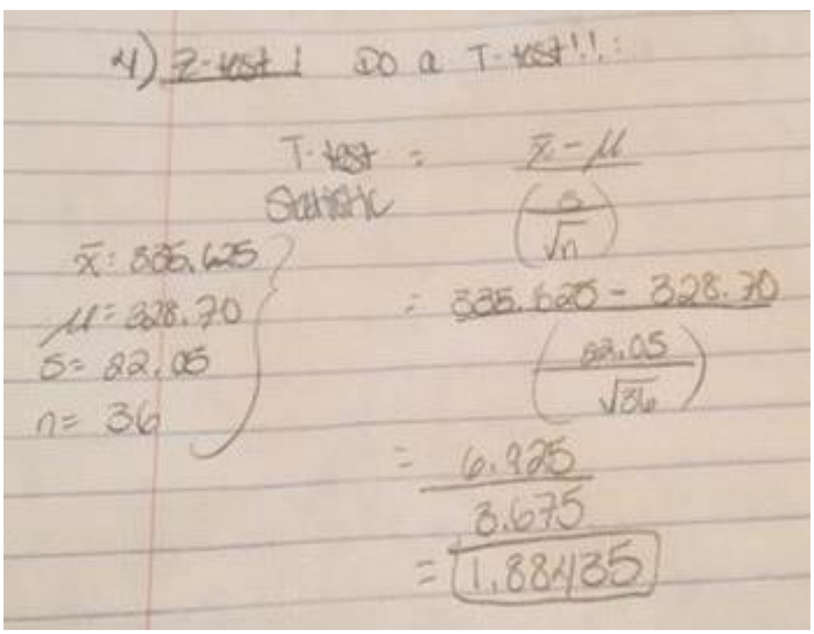
      H(0): The mean of AC is **not going to** equal 328.70.

   b) State your Alternative Hypothesis.

      H(A): The mean of AC will be **exactly equal to** 328.70.

   c) Use a significance level of to find a Z-test and a P-value that either supports or rejects the null hypothesis.

      To do the T-test, I wrote it out (see attached picture below):



Because 1.88435 is greater than 0.05%, we ACCEPT the null hypothesis at 0.05%.

d) Using the sample of the 36 days in 2017 for the company's AC, has there been a significant statistical change in the mean AC since 2014? State your conclusion.

   There has not been a significant statistical change in the mean AC since 2014, even though it is not the same. It's still very close, but still a different total value.

5. On the last page, describe in your own words what computational methods you learned in Math2100 that were used to complete this project.

There were many probability and statistics methods used to complete this project. First, obtaining the mean, median, mode, sample variance, and standard deviation of the sample. This was covered in the very beginning of the class on week 3 – the first three being a review: mean, median, mode – and the latter two: sample variance and standard deviation, were brand new. Second, the use of R being enforced throughout the project to find the boxplot and qqnorm() and qqline() were used just like in previous homework assignments. Third, the use of a Histogram, which is NOT the same as a bar graph – a histogram bulks the data into bins of how frequently the data shows up within each specific bin, as opposed to a bar graph which gives a qualitative visual of each single data value. Fourth, confidence intervals and the use of T-tests and Z-tests were also touched upon in this project, which we covered extensively toward the end of the semester. Fifth, hypothesis testing is touched upon in the last question of the project, which is also covered in great detail in the latter part of the semester. The class Probability & Statistics adequately prepared me for this project, because we covered all of the techniques and formulas needed to complete the project.