Probability and Statistics for Engineers	Math 2100	
Due 12-09-2021 **Via email**		**Final Project**

Name: Duy Vo

This project should be done in groups of 1, 2, or 3 people (maximum).

Rubric   In order to earn full credit for this project, you must submit all of the data indicated below.

Spread out your work to include multiple pages. Be sure that each individual chart, graph or diagram is not split between pages. Use **R** for all diagrams, tables, graphs, and computations.

Your report must be turned in on or before 12-09-2021 for full credit. Be sure to include your names, and all pertinent tables, diagrams, graphs and computations. On the last page, **describe in your own words what computational methods you learned in Math 2100 that were used to complete this project**.

Data

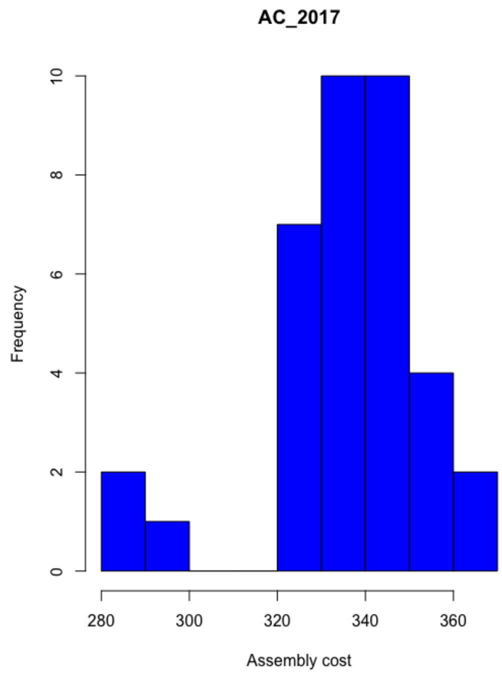
The spreadsheet entitled “Computer_Stats.xlsx” consists of the mean monthly assembly cost (AC) and the mean monthly percent of profit (PP) earned by a company for the 36 months from 2014 through 2016, columns A to C. It also contains the assembly cost and the percent of profit for the first 36 days of production during of 2017 following a change in ownership, columns E to G.

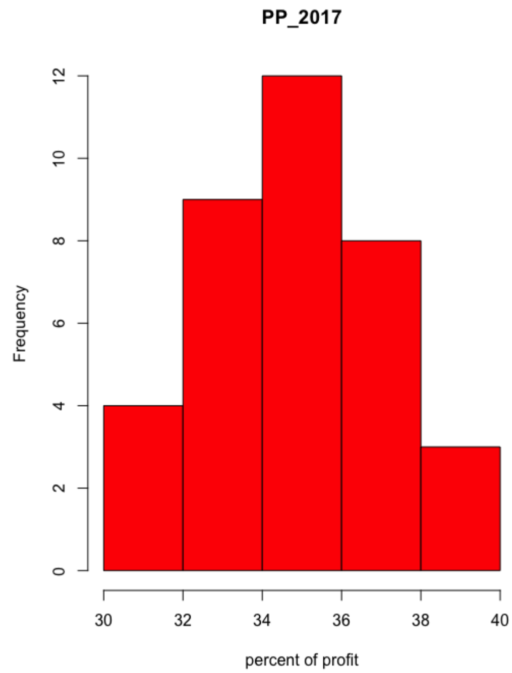
1. For **AC** and **PP** in the 2017 stats, perform the following computations:
    a. Find the mean, median, mode, and standard deviation ***s*** of this sample.

       mean_AC2017 = 335.625
       median_AC2017 = 337.825
       mode (AC_2017) = numeric
       sd_AC2017 = 16.69877

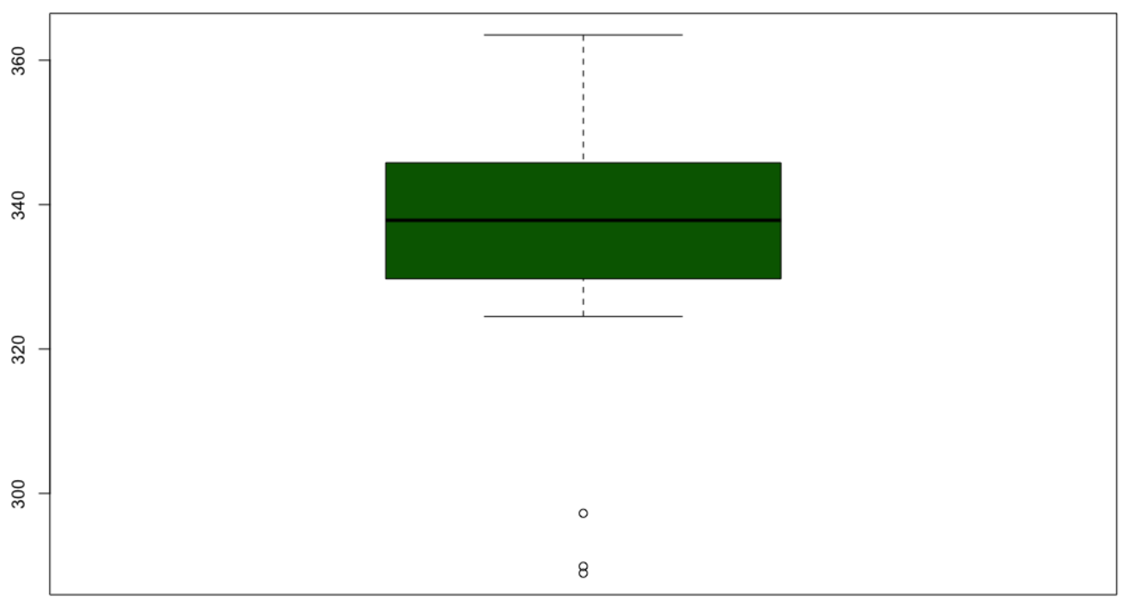
       mean_PP2017 = 35.44444
       median_PP2017 = 36
       mode (PP_2017) = numeric
       sd_PP2017 = 2.477838

    b. Create a histogram using your frequency table.





c. Draw a boxplot of the assembly costs.



d. Use the results from part (a) to find what percentage of the 36 days' AC is within $\pm 1\sigma$ of the mean. (Hint: Assume this sample was NOT taken from a Bell curve.)

   percentage_AC2017_within +/- 1sd = 80.55%

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

   proportion_PP2017 = 7/12

f. 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?

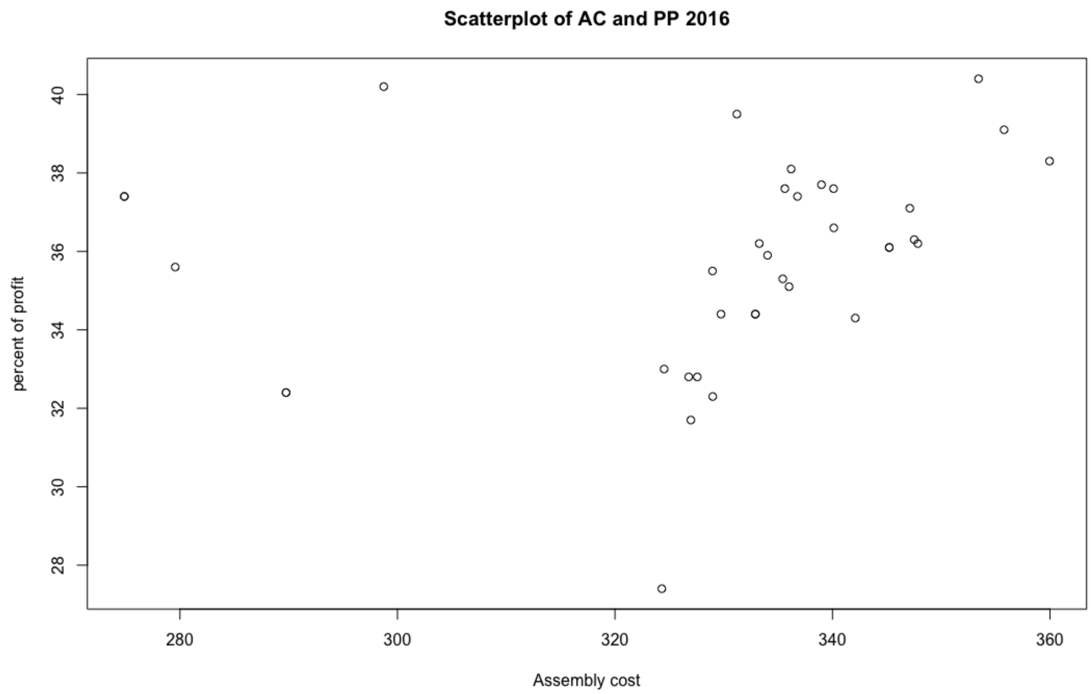
   P (PP_2017 >= 35.4 | AC_2017 <= 330) = 19.04762

2. Assuming that the **mean AC** for the first 36 months is normally distributed, do the following:
   a. Find the mean μ and the standard deviation σ, then find what proportion of AC was between $320 and $350 per month.

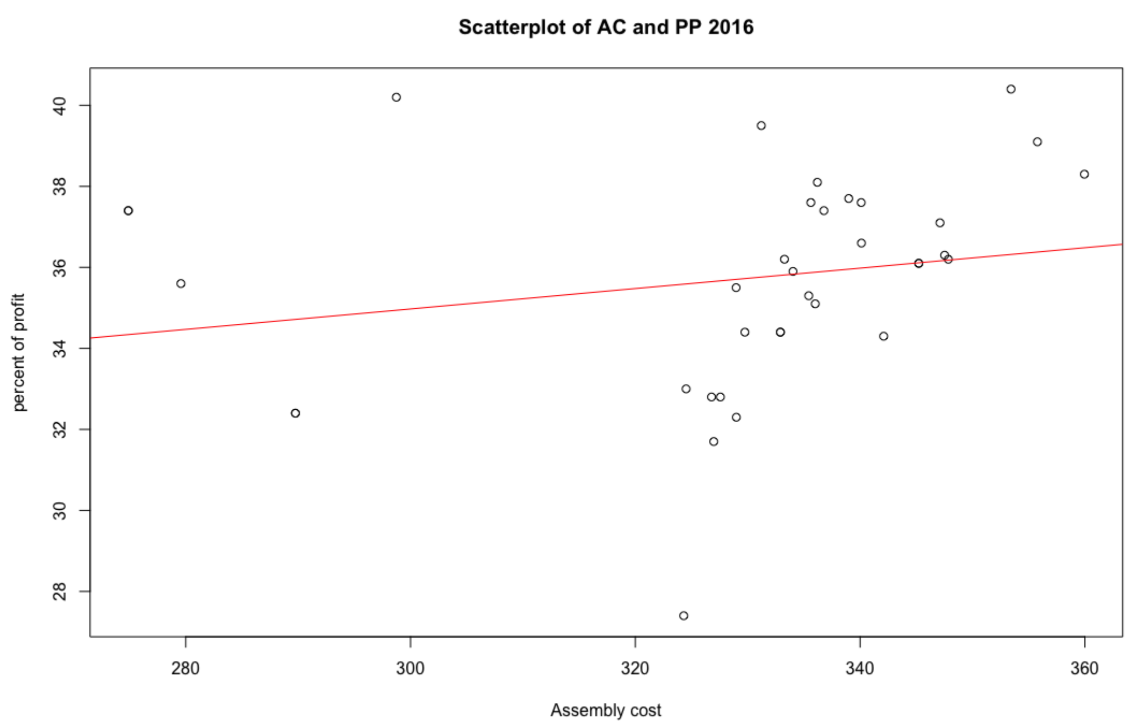
      mean_AC2016 = 328.7014
      sd_AC2016 = 22.04782

      proportion_AC2016 = 3/4

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



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



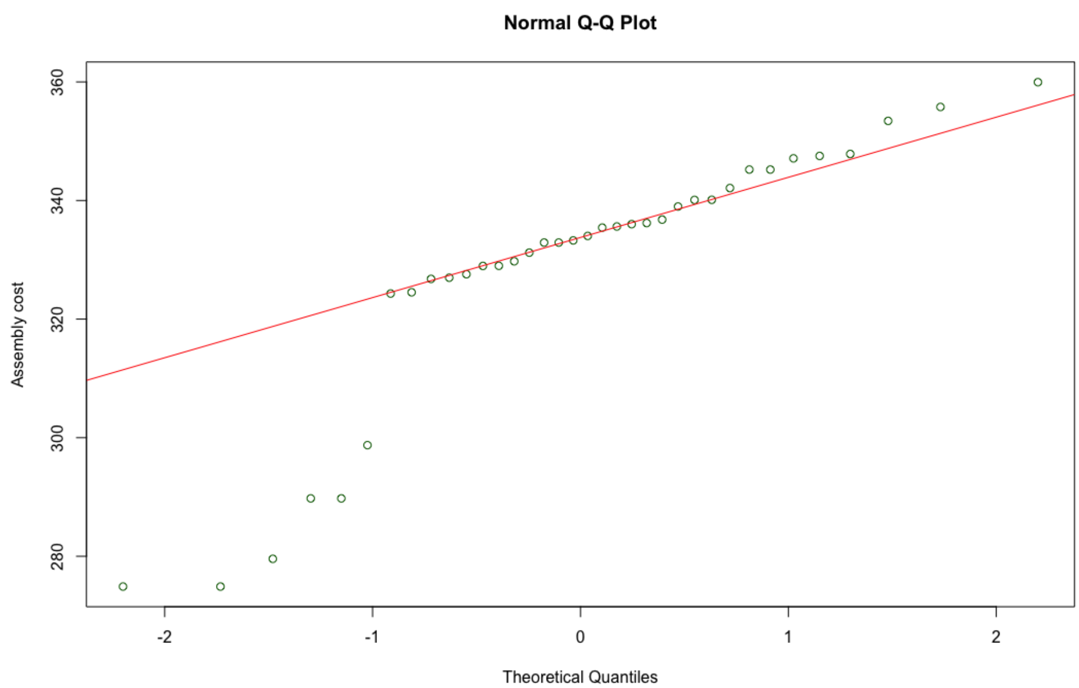
Equation for the line of regression: y = 27.42308 + 0.02516*x

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.

percentOfprofit = 36

assemblyCost = 289

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).



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-level.

   AC_2017 ~ N(335.625,16.69877/sqrt(36))
   n = 36 > 30
   alpha = 0.05
   Z-alpha/2 = 1.960
   (335.625-1.960*(16.69877/sqrt(36)), 335.625+1.960*(16.69877/sqrt(36)))

   Result: (330.1701, 341.0799)

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, H_0.

   Null Hypothesis - Ho : mean = 328.7014

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

      Yes, it changed

   b) State your Alternative Hypothesis.

      Alternative Hypothesis - Ha : mean # 328.7014

   c) Use a significance level of $\alpha = 0.05$ to find a Z-test and a P-value that either supports or rejects the null hypothesis.

      Xbar = mean_AC2016 = 328.7014
      mean = mean_AC2017 = 335.625
      Sd = sd_2016 = 22.04782
      n = 36

      Test Statistic
      Z-Ho = (328.7014 - 335.625)/(22.04782/sqrt(36))

      2 tail test
      Critical Value
      alpha = 0.05 => Z-alpha/2 = 1.960

Z_nullHypo = (328.7014 - 335.625)/(22.04782/sqrt(36))
Z_nullHypo = -1.884159

Conclusion
(-Z-alpha/2 < Z_nullHypo < Z-alpha/2)
(-1.960 < -1.884159 < 1.960) (True)
=> Fail to reject null hypothesis

P_value = [1-P(Z <= 1.88)]*2
P_value = (1 - pnorm(1.88,0,1))*2
P_value = 0.06010808

P_value = 0.06 > alpha = 0.05
=> also fail to reject null hypothesis

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 in one descriptive sentence.

There has not been a significant statistical change in the mean AC since 2014. As you can see the mean of 36 days AC in 2017 doesn't have a big difference from the mean of total 36 months AC since 2014-2016 as well as the mean of each month since 2014-2016.

Data and Value table:

cs_dataset

Filter

	AC_2016	PP_2016	AC_2017	PP_2017
1	347.52	36.3	345.35	36
2	347.11	37.1	346.25	38
3	353.43	40.4	351.65	39
4	342.10	34.3	341.75	35
5	336.01	35.1	347.00	35
6	340.10	37.6	288.95	36
7	336.20	38.1	342.50	38
8	328.99	32.3	330.35	34
9	331.21	39.5	326.85	36
10	340.12	36.6	341.25	36
11	326.78	32.8	329.55	34
12	324.52	33.0	327.85	33
13	279.57	35.6	329.00	36
14	336.78	37.4	363.50	38
15	334.03	35.9	341.25	37
16	327.57	32.8	329.90	33
17	326.98	31.7	324.50	30
18	335.43	35.3	338.90	34
19	329.75	34.4	331.60	34
20	328.97	35.5	333.80	36
21	338.99	37.7	343.45	38
22	274.89	37.4	352.35	37
23	345.22	36.1	351.95	36
24	332.91	34.4	289.90	33
25	289.76	32.4	331.75	32
26	335.62	37.6	336.75	38
27	333.27	36.2	341.65	36
28	298.74	40.2	356.75	39
29	324.31	27.4	330.95	31
30	359.96	38.3	360.25	36
31	355.78	39.1	297.25	40
32	347.85	36.2	348.95	38
33	274.89	37.4	324.50	30
34	345.22	36.1	338.90	34
35	332.91	34.4	331.60	34
36	289.76	32.4	333.80	36

assemblyCost_2d	289
counter1	29
counter2	21
counter3	4
counter4	27
i	289.76
mean_AC2016	328.701388888889
mean_AC2017	335.625
mean_PP2017	35.4444444444444
median_AC2017	337.825
median_PP2017	36
p_value	0.0601080779223997
percentage_AC2017_1d	80.5555555555556
percentage_AC2017_1f	19.047619047619
percentOfprofit_2d	36
proportion_AC2016_2a	0.75
proportion_PP2017_1e	0.583333333333333
sd_AC2016	22.0478237543207
sd_AC2017	16.6987702712334
sd_PP2017	2.47783827956599
value_AC2016	num [1:36] 348 347 353 342 336 ...
value_AC2017	num [1:36] 345 346 352 342 347 ...
value_PP2016	num [1:36] 36.3 37.1 40.4 34.3 35.1 37.6 38.1 32.3 39.5 36.6…
value_PP2017	num [1:36] 36 38 39 35 35 36 38 34 36 36 ...
X	330.170068466667
Y	341.079931533333
Z_nullHypo	-1.88415906878776

Computational methods you learned in Math 2100 that were used to complete this project.

- Mean()
- Median()
- Mode()
- Sd()
- Par()
- Hist()
- Boxplot()
- For loop
- Plot()
- Abline ()
- Summary()
- Qqnorm()
- QQline()
- Sqrt()
- Pnorm()