# STT 4660
# Homework #5

Timothy Stubblefield

October 14, 2020

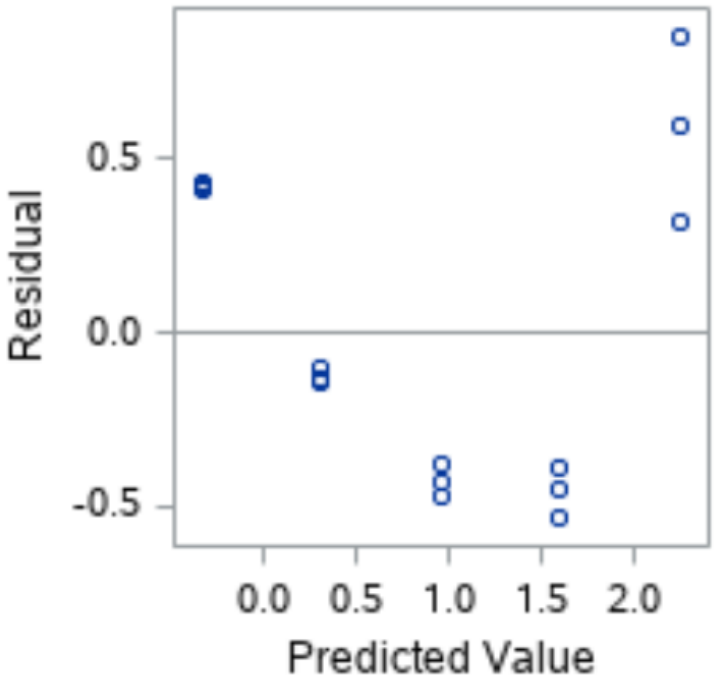
## 3.16 (e)

1) Referring to the Solution concentration problem from 3.15, X = Time and Y = Concentration of solution.

2) The residuals for Solution Concentration are as follows.

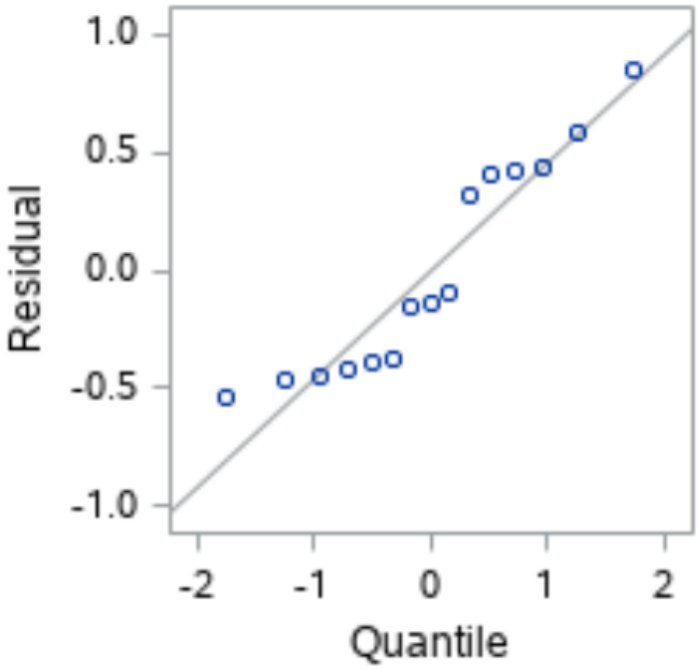
| Obs | y | x | predict | res |
|---|---|---|---|---|
| 1 | 0.07 | 9 | -0.34067 | 0.41067 |
| 2 | 0.09 | 9 | -0.34067 | 0.43067 |
| 3 | 0.08 | 9 | -0.34067 | 0.42067 |
| 4 | 0.16 | 7 | 0.30733 | -0.14733 |
| 5 | 0.17 | 7 | 0.30733 | -0.13733 |
| 6 | 0.21 | 7 | 0.30733 | -0.09733 |
| 7 | 0.49 | 5 | 0.95533 | -0.46533 |
| 8 | 0.58 | 5 | 0.95533 | -0.37533 |
| 9 | 0.53 | 5 | 0.95533 | -0.42533 |
| 10 | 1.22 | 3 | 1.60333 | -0.38333 |
| 11 | 1.15 | 3 | 1.60333 | -0.45333 |
| 12 | 1.07 | 3 | 1.60333 | -0.53333 |
| 13 | 2.84 | 1 | 2.25133 | 0.58867 |
| 14 | 2.57 | 1 | 2.25133 | 0.31867 |
| 15 | 3.10 | 1 | 2.25133 | 0.84867 |

3) Now, the plot of Residuals vs Predicted Values is as follows.




4) Now, that we have the necessary plots, we can make some conclusions. Looking at Residuals vs Predicted Values graph, the points appear to form a U shape. This implies that the relationship between X(Time) and Y(Solution Concentration) likely has a non-linear relationship instead of a linear relationship. Due to the U-shape of the graph, perhaps the relationship is quadratic.

5) Finally, the Normal Probability Plot or QQ plot is below.




6) Looking at the Normal Probability Plot and consulting the notes/textbook, it seems that Solution Concentration data is underdispersed compared to data that follows a normal distribution. As a result, there are less outliers in the data compared to data with a normal distribution.

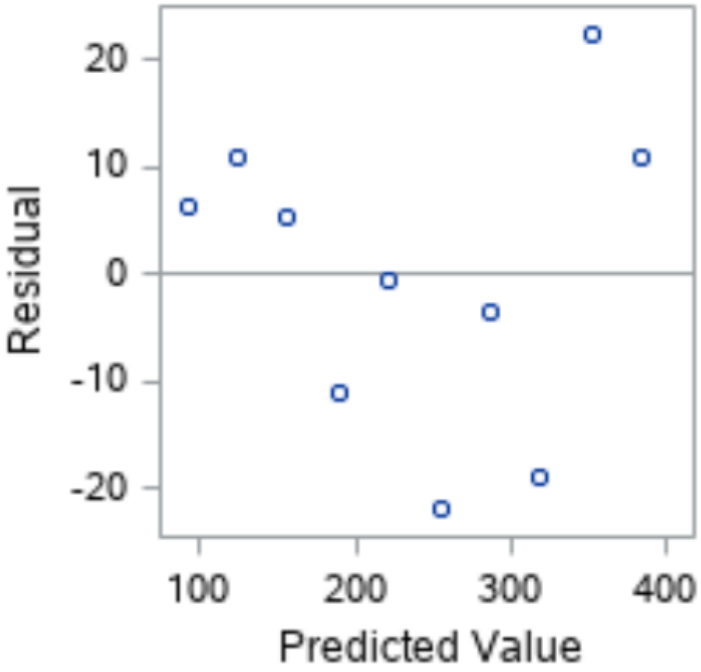
## 3.17 (e)

1) In this problem, we are looking at Sales Growth, that is X = Year and Y = Thousands of Units Sold.

2) The residuals for Sales Growth as as follows.

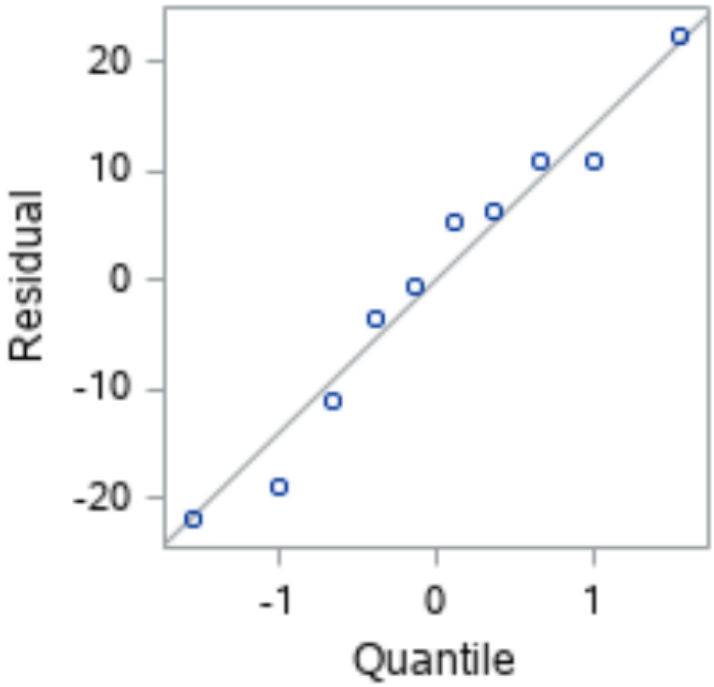
| Obs | y | x | predict | res |
|---|---|---|---|---|
| 1 | 98 | 0 | 91.564 | 6.4364 |
| 2 | 135 | 1 | 124.061 | 10.9394 |
| 3 | 162 | 2 | 156.558 | 5.4424 |
| 4 | 178 | 3 | 189.055 | -11.0545 |
| 5 | 221 | 4 | 221.552 | -0.5515 |
| 6 | 232 | 5 | 254.048 | -22.0485 |
| 7 | 283 | 6 | 286.545 | -3.5455 |
| 8 | 300 | 7 | 319.042 | -19.0424 |
| 9 | 374 | 8 | 351.539 | 22.4606 |
| 10 | 395 | 9 | 384.036 | 10.9636 |

3) Now, the plot of Residuals vs Predicted Values is as follows.




4) Time to make conclusions some conclusions with the plot. Upon viewing the Residuals vs Predicted Values plot, the points do not create any visible pattern or shape. This is the ideal case and means there is no relationship between Residuals and Predicted Value. Therefore, it is likely that there is a linear relationship between X(Year) and Y(Thousands of Units Sold).

5) Lastly, the Normal Probability Plot or QQ plot is below.



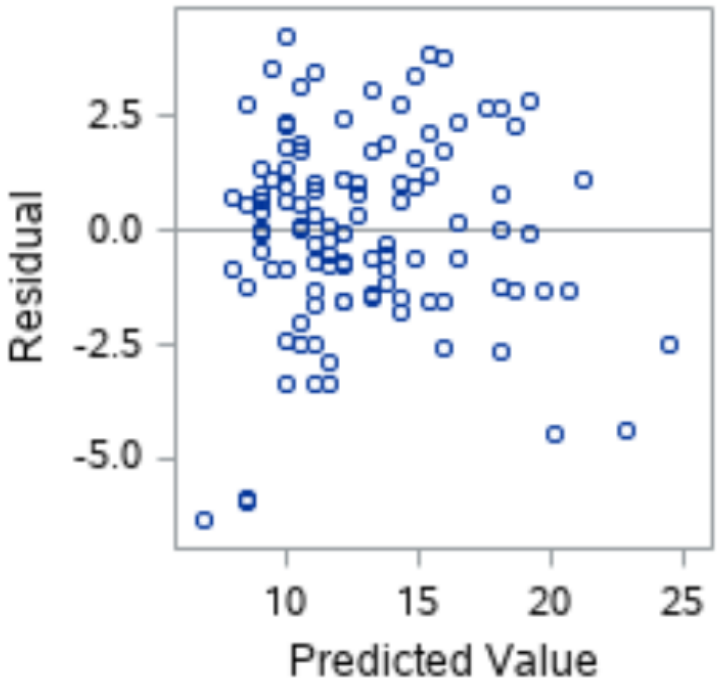

6) Looking at the Normal Probability plot and comparing the result to the notes, it appears that Sales Growth data is overdispersed compared to data that follows a normal distribution. Therefore, there are more outliers in this data compared to data with a normal distribution.

## 3.18 (d)

1) In this problem, we are looking at Production Time, where X = Production Lot Size and Y = Production Time in Hours.

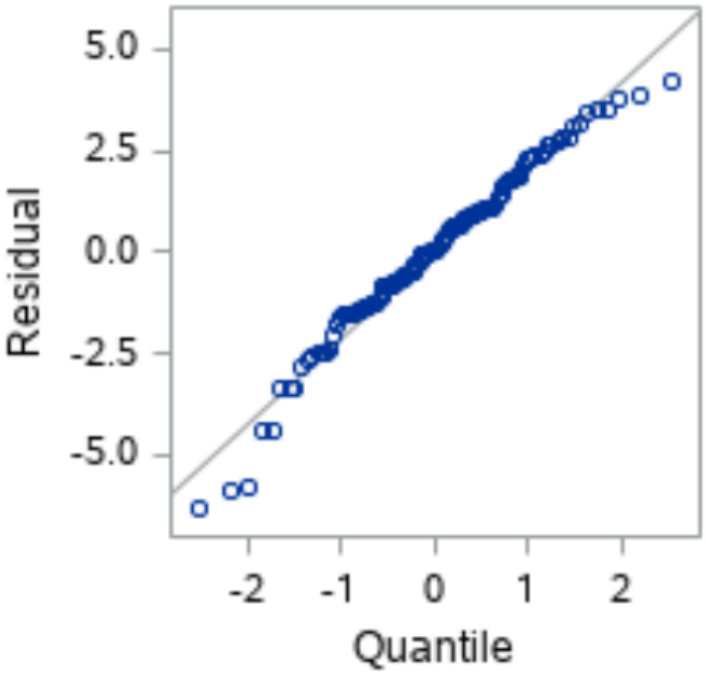
2) There are 111 residuals for Production Time and thus it is too large to save with the snipping tool and also quite cumbersome to include it the pdf.

3) So, the plot of Residuals vs Predicted Values is as follows.




4) With the data plotted successfully, we can look at the Residual vs Predicted Values plot and compare with the lecture notes. It seems there is no relationship between the two variables since the points appear to be scattered about the plot evenly. Thus, it is likely that there is a linear relationship between X(Production Lot Size) and Y(Production Time).

5) Finally, the Normal Probability Plot or QQ Plot is below.




6) When viewing the Normal Probability plot and comparing to the notes, the Probability Plot from 3.18 is the closest to a straight line (Ideal case) out of all the problems we have done. However, we can see that there is a slight left skew in the QQ Plot. Due to the left skewness, that means the left tail is heavier, and thus, there are more observations at lower values than at higher values.