Binnan Zhao

Renhua Guo

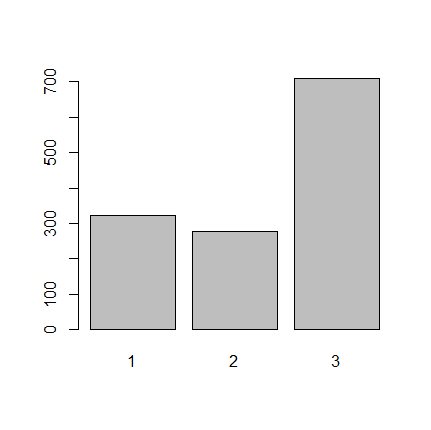
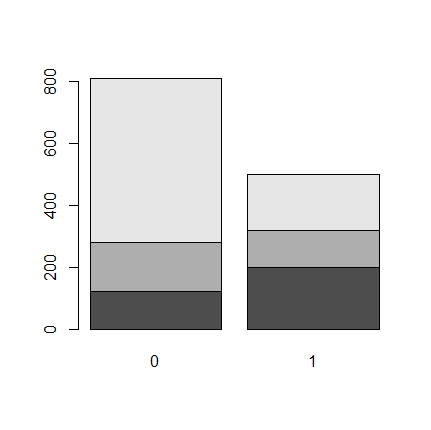
Kun Yang

Preston Turner

Ebow Bruce-Mensah

Project 1

Question 1.

1. From the flowing graph, it is interpreted that first class has 323 passengers, second class has 277 passengers, and third class has 709 passengers. 
2. From the flowing graph, it is interpreted that third class has the largest un-survival number, where first class has the greatest chance of surviving. 
3. dead survived

first 123 200

second 158 119

third 528 181

explanatory variable will be the passenger class(first, second, third), response variable will be the passenger survival numbers(dead, survived).

1. There are 2 marginal distributions; one is Sex with features female, and male. The other one is class with first class, second class, and third class.

Sex: Female = .3559, male = .644

Class: firstClass = .246, secondClass =.211, thirdClass = .541

1. There are 323 total passengers in first class did not survive, the conditional distribution for this class would be 123/323 = 0.3808; the survival rate of this class is 1-0.3808 = .6192, second class conditional distribution 119/277 = .4296, third class 181 / 709= .2552. Second class has the highest survival rate
2. survived

sex 0 1

female 127 339

male 682 161

gender verse survival, the total survivors of female are 339, the conditional distribution of female verse survival is 339 / 500 = .678, therefore the conditional distribution of male verse survival will be 1 - .678 = .322, so the female had a higher chance of survival than the male.

1. Female:

survived

pclass 0 1

1 5 139

2 12 94

3 110 106

Male:

survived

pclass 0 1

1 118 61

2 146 25

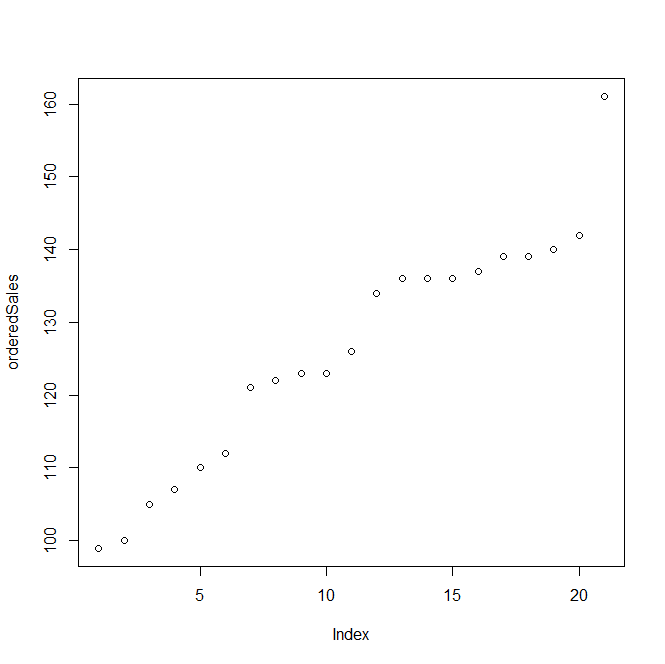
3 418 75

First class female survival conditional distribution 139/144=.9652, second class female survival conditional distribution 94 / 106= .8867, third class survival conditional distribution 106 / 216= .4907.

First class male survival conditional distribution 61/179=.3407, second class male survival conditional distribution 25 / 171 = .1461, third class male survival conditional distribution = 75 / 493=.1521.

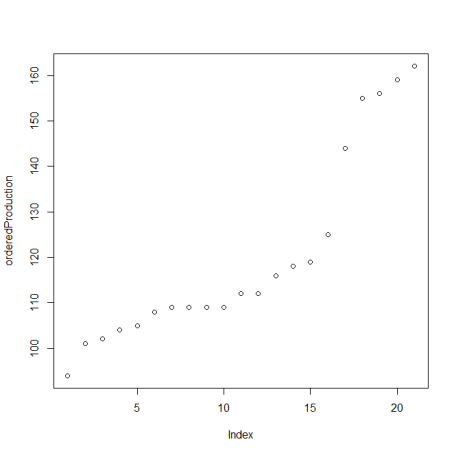
1. After the comparison between data g) and data e), it is concluded that such scenario is a Simpson’s paradox, because the population difference between classes became the lurking variable, therefore in conclusion the first class has the higher survival rate than second class as female and male.

Question 2

Sales Scatter Plot

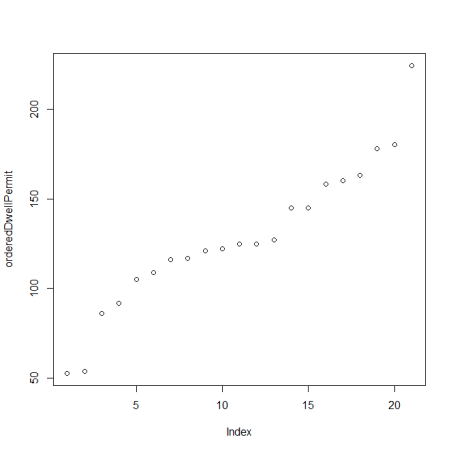
5 Points Summary: Min: 99.0 1st qtr: 112.0 Median: 126.0 3rd qtr: 137.0 Max: 161.0

Outlier boundary [74.5,174.5] No outliers

Production Scatter Plot

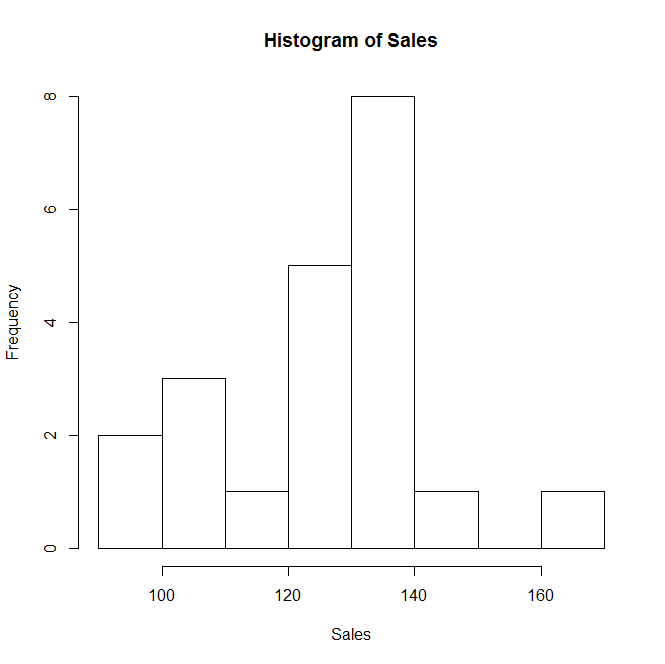
5 Points Summary:Min:94.0 1st qtr: 108.0 Median: 112.0 3rd qtr: 125.0 Max:162.0

Outlier boundary [82.5,150.5] yes there are higher bound outliers, 155 156 159 162

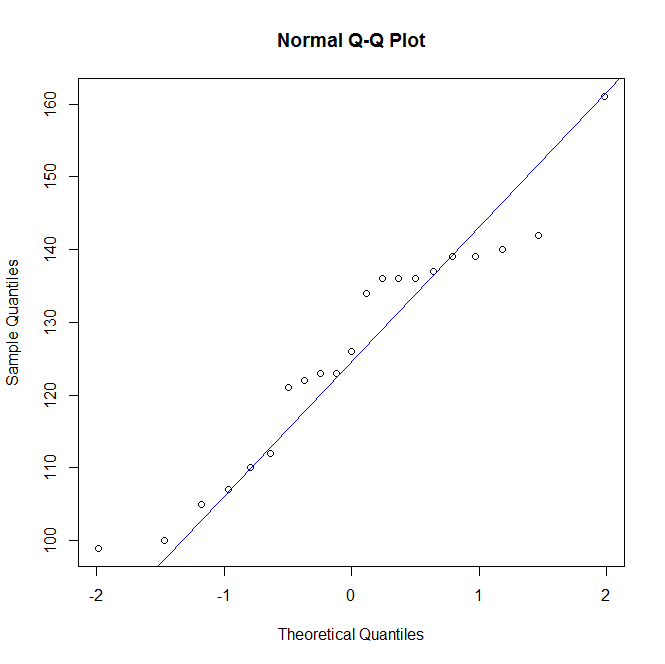
DwellPermit Scatter Plot

5 Points Summary:Min:53.0 1st qtr:109.0 Median:125.0 3rd qtr:158.0 Max:224.0

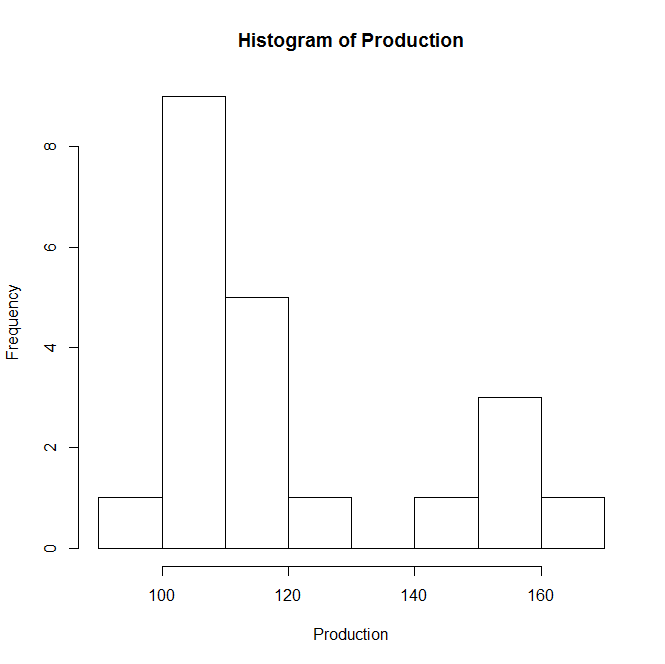
Outlier boundary: [35.5,231.5] no outliers

Sales Histogram: 

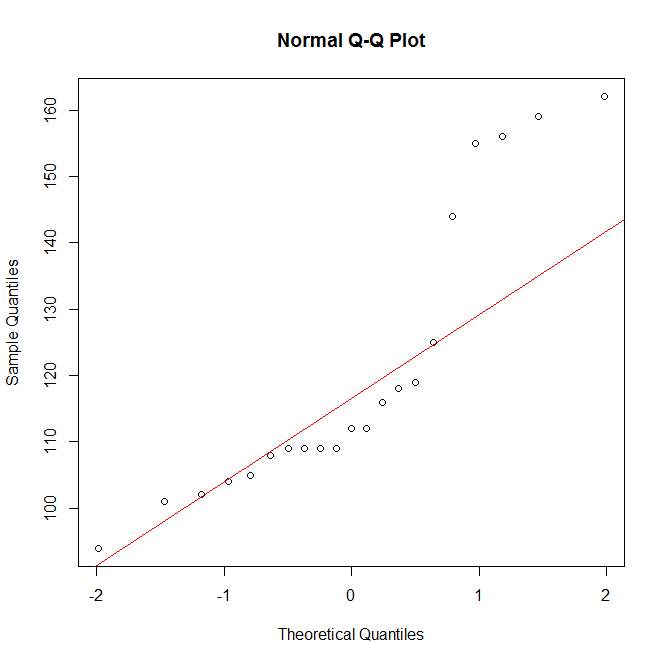
Does not follow normal distribution, it is skewed to the right.

Sales Normal QQ Plot: 

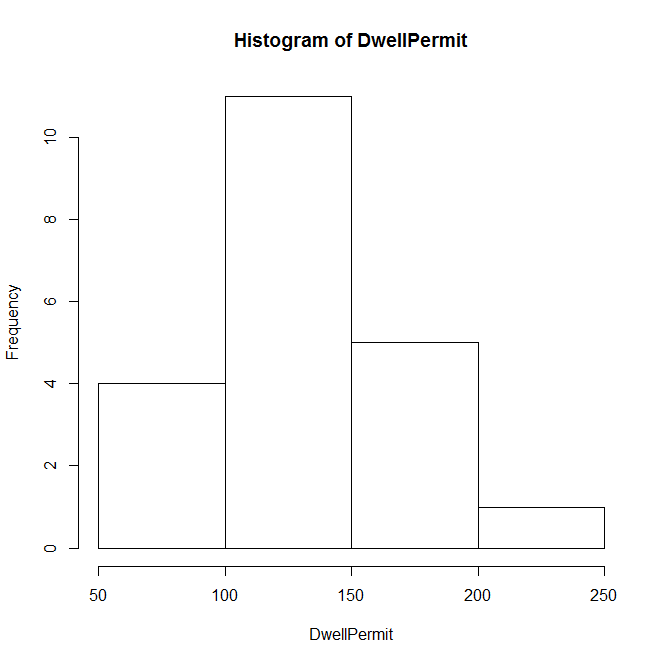
Does not follow the normal QQ plot, because the QQ line does not lie on all the points.

Production Histogram: 

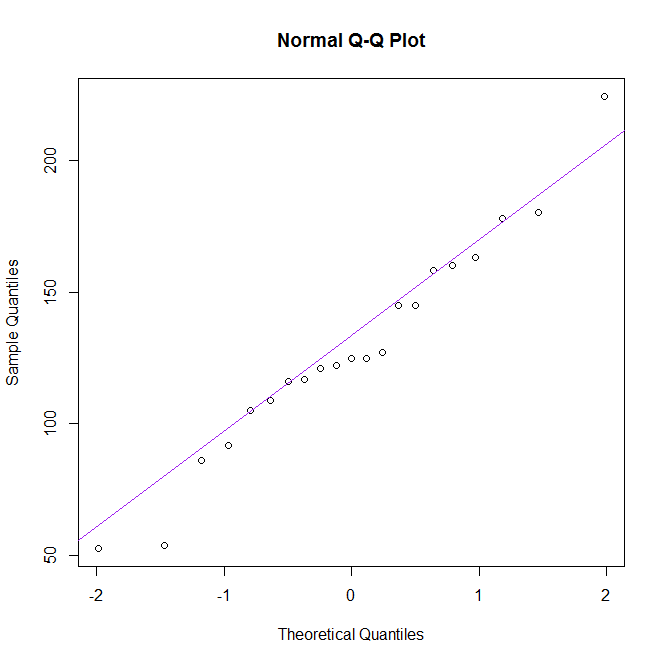
It does not follow the normal histogram, it is skewed to the right

Production QQ : 

It does not follow the normal QQ plot, because the normal QQ line does not lie on all the points of production.

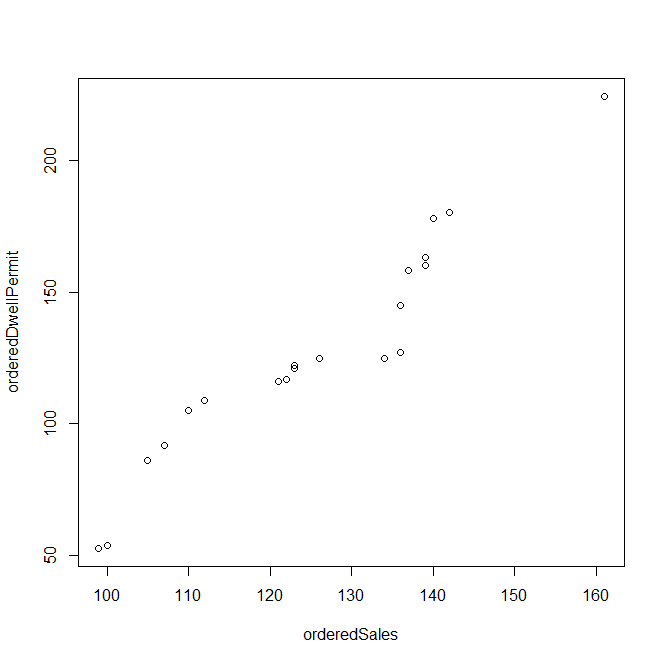
DwellPermit Histogram: 

It does not follow the normal histogram, however it shows a strong correlation between DwellPermit histogram and normal histogram, it is slightly skewed to the right.

DwellPermit QQ Plot: 

It does not follow the normal QQ plot, however the points are following the direction of the QQ line strongly; this QQ plot has the strongest correlation out of the three QQ plots.

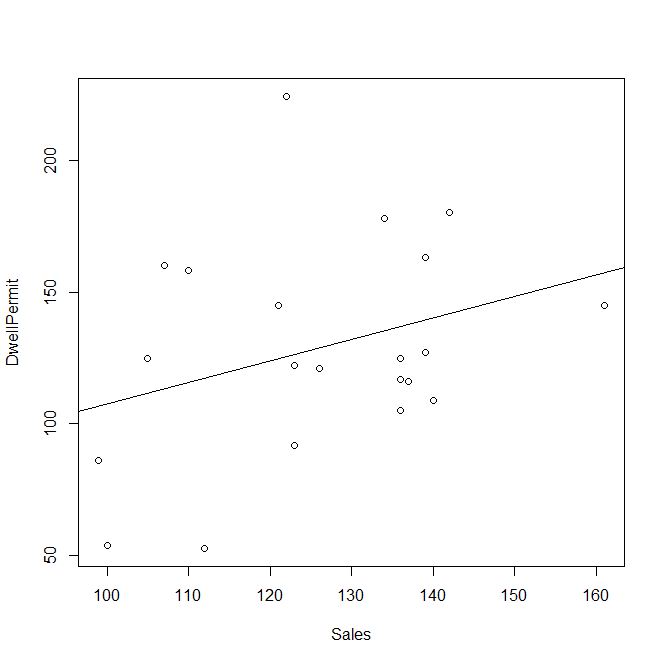
1. Scatter plot Sales vs. DwellPermit

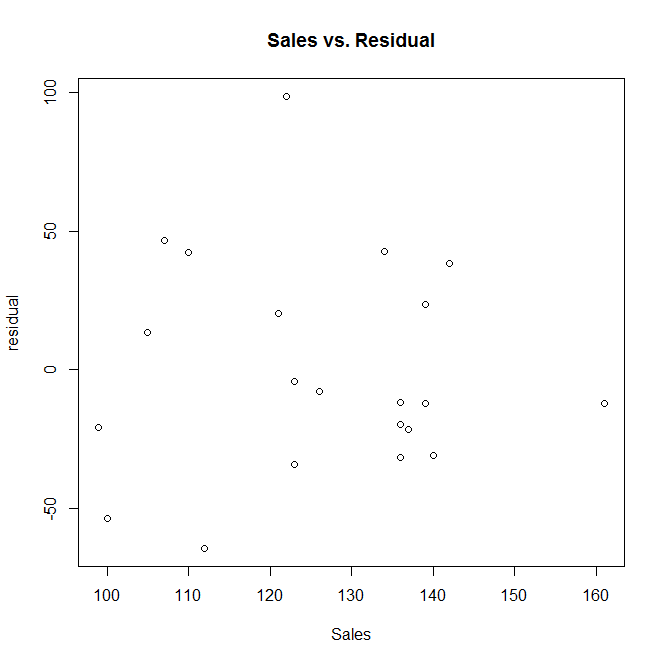


Correlation between Sales and DwellPermit = 0.959

Strong positive progressive relation when sales is response variable, and DwellPermit as explanatory variable.

Sales vs. DwellPermit

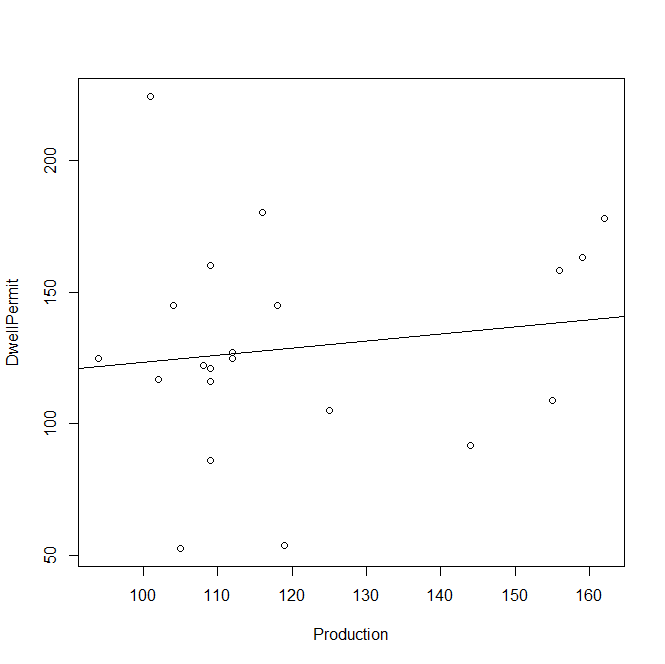
Regression line 

Residual Plot

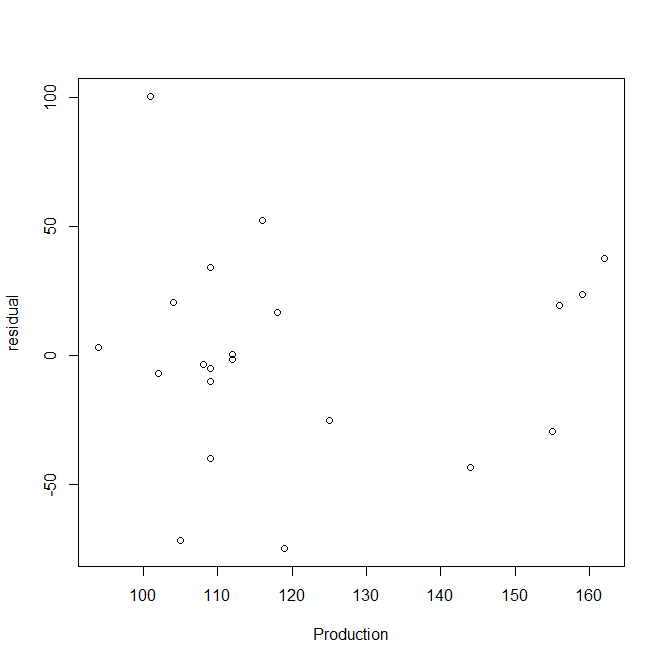
The positive correlation for this graph is extremely weak compared to the other graphs.

1. Production vs. DwellPermit

Regression Line of Production vs DwellPermit



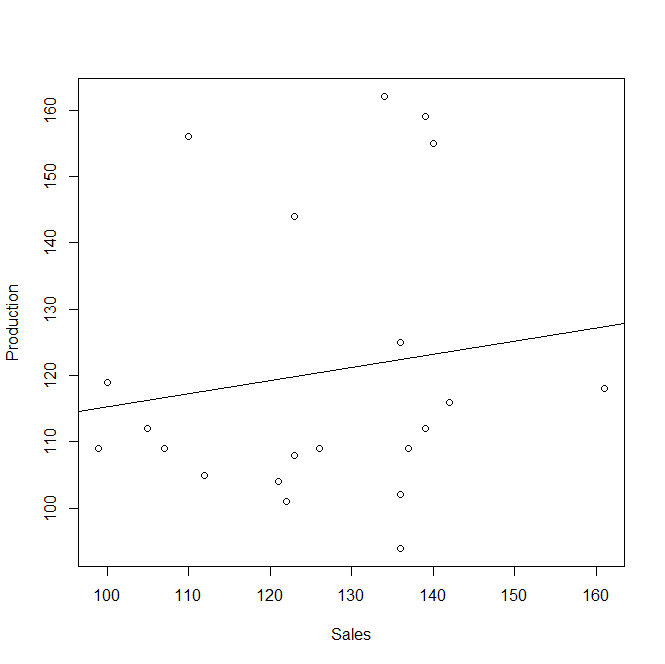
Residual Plot



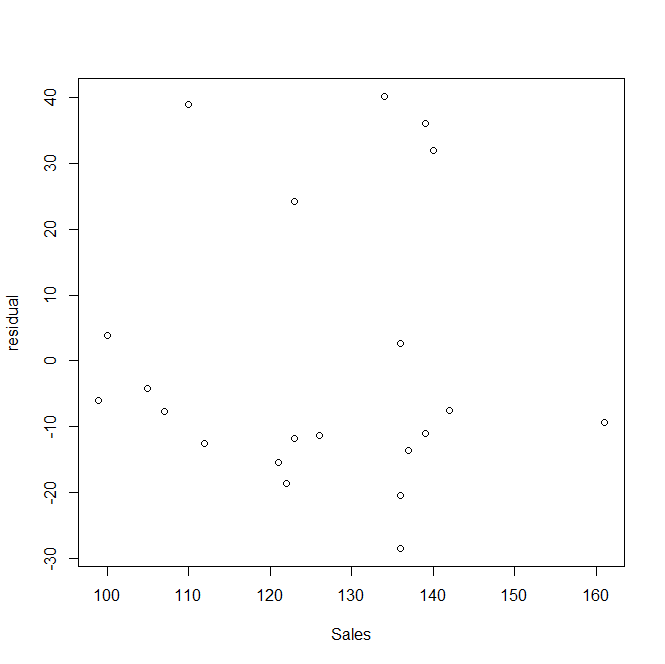
The residual plot and the regression line show a very weak, positive correlation.

1. Sales vs. Production

Regression Line:



Residual Plot



These residual plots and regression line also show a very weak, positive correlation.

1. In relation to the three variables from part d to f, it shows that the sales vs dwelling permit graph for part d has the weakest positive correlation, while the production vs dwelling permit graph for part e has the strongest positive correlation.