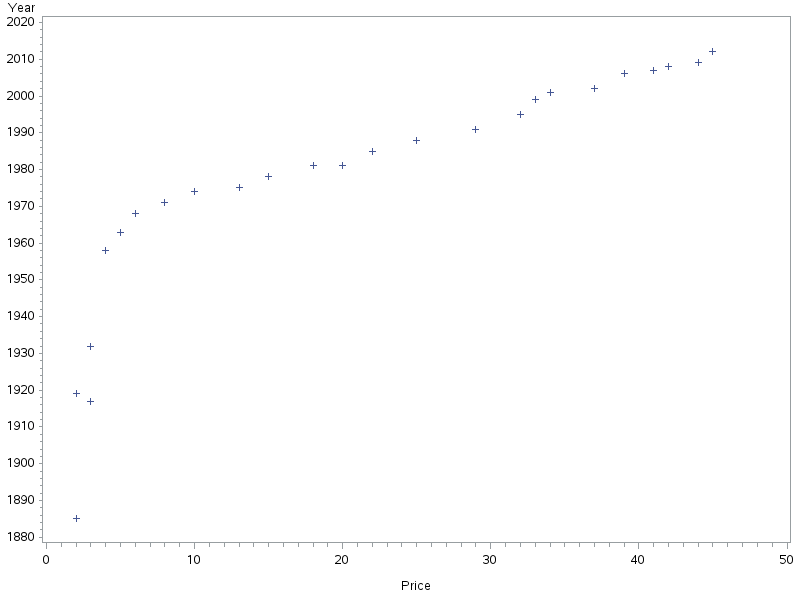
Dylan Smith

Assignment 1

1(a)

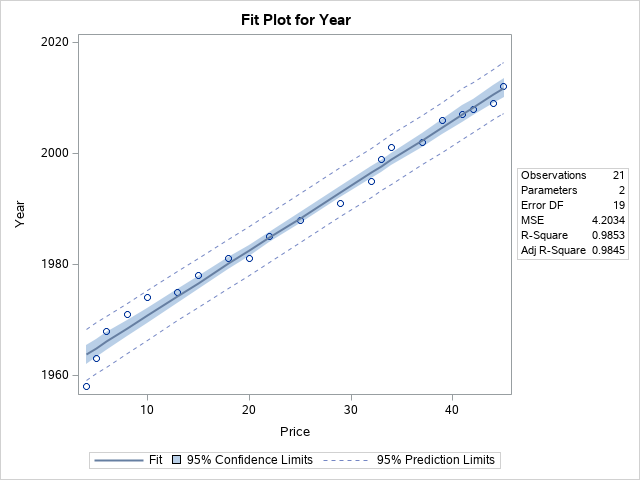


The data seems mostly linear after 1960.

1(b)

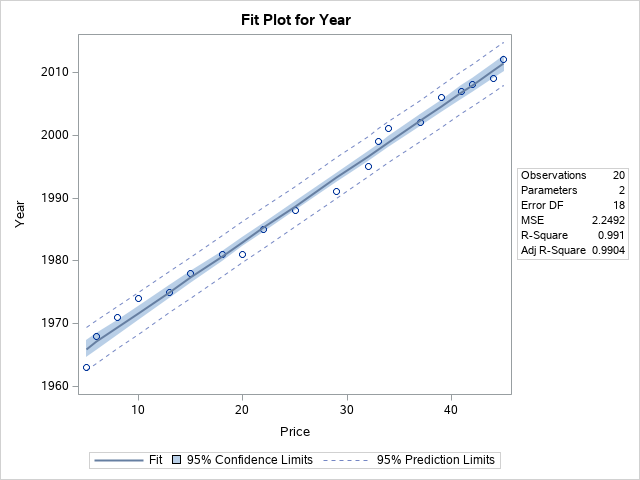
Y(Year)=1959+1.17x(Price)

1(c)



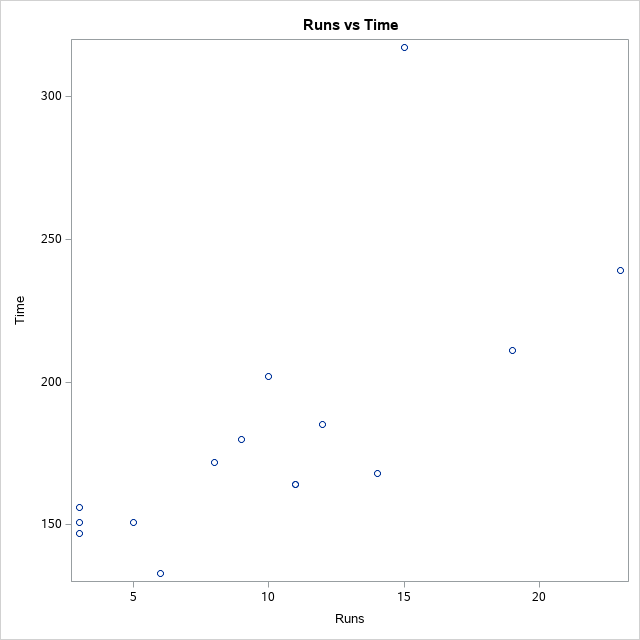
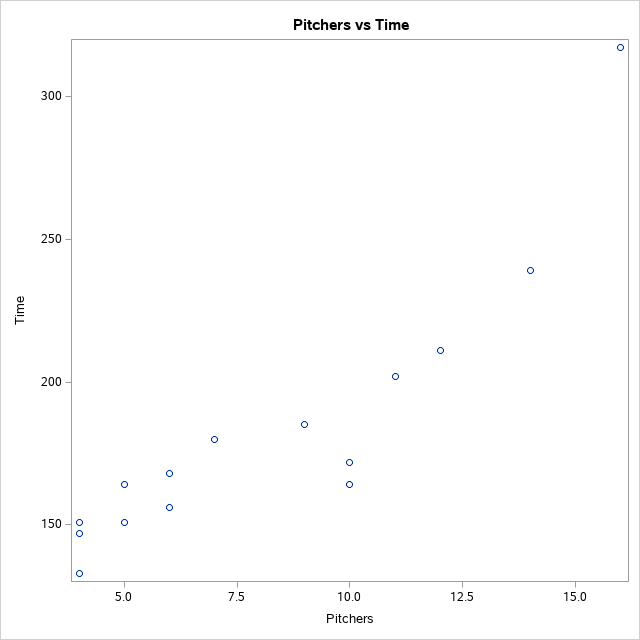
This line is a very good fit with the only concern coming from the first observation.

1(d)



Above is the plot with the problematic point removed. The R-Squared is slightly better. For the best fit the removal of the first 5 observations would be better although the plot in 1(c) is enough.

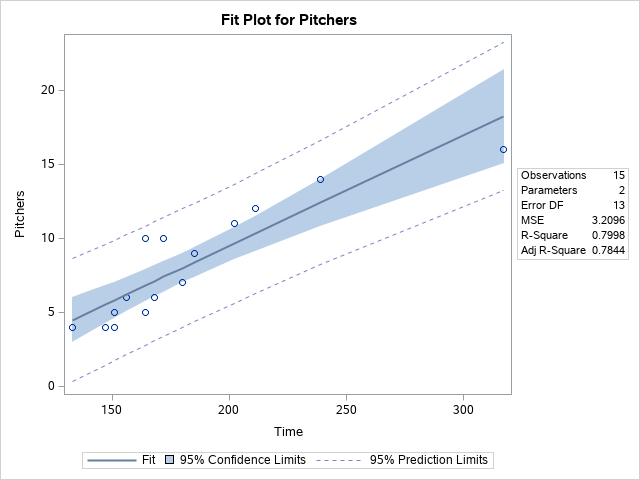
2(a)



2(b)

Pitches would have a stronger fit. The data seems to be more linear with less spread.

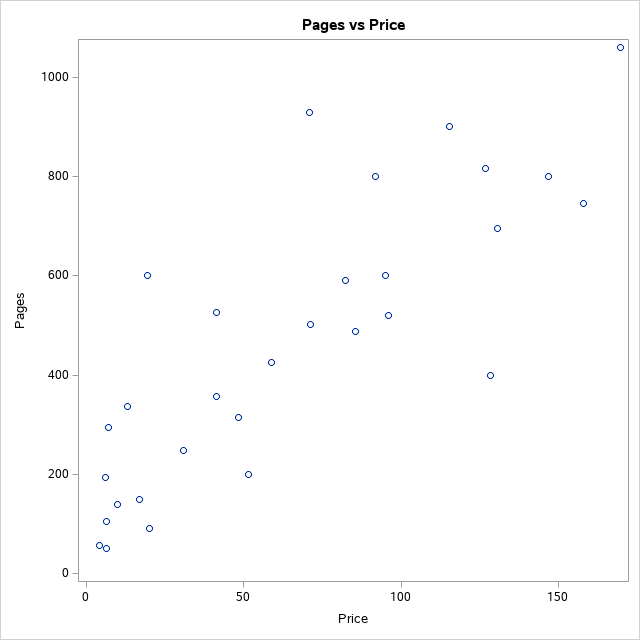
2(c)



2(d)

The residual vs predicted value plot does not show any patterns. The residual vs quantile plot shows conformity and linearity around the prediction line.

3(a)



There seems to be somewhat of a correlation between the number of pages and the price of the book, but it does not seem to be a strong one.

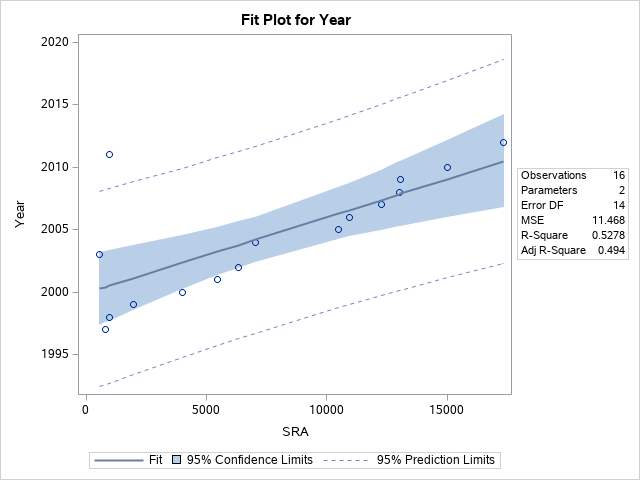
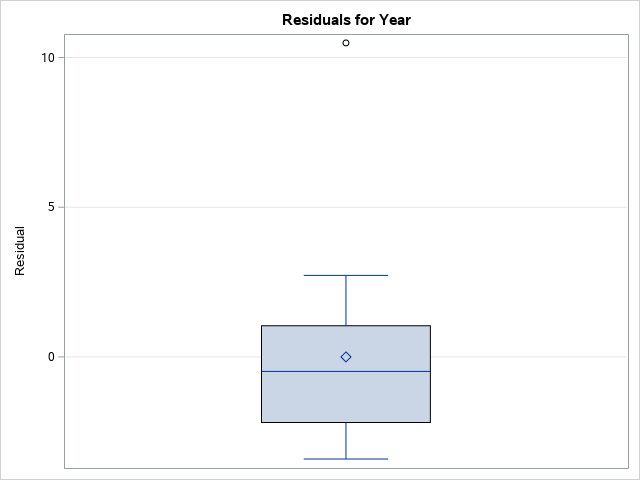
3(b)

Y(Price)=165.96+4.59x(Pages)

3(c)

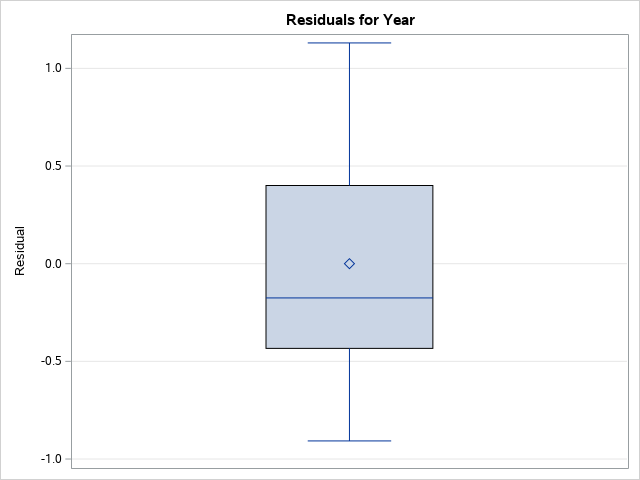
The Residual vs Quantile plot shows linearity around the prediction line. The residual vs predicted value plot shows a seemingly random spread, possibly a slight megaphone affect.

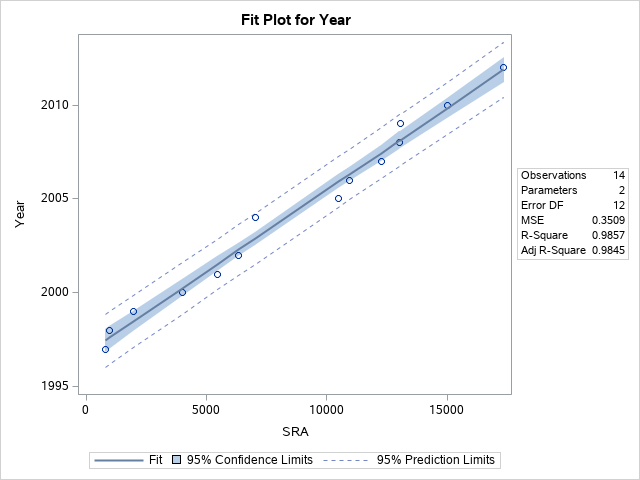
4(a)



There is one definite outlier that appears in both the box plot and they fit plot. The two years with outlying residuals are the year 1997 with a residual of -3.4 and the year 2011 with a residual of 10.49.

4(b)





At first glance, the new box plot and fit plots show no outliers and the fit plot shows a tighter spread. Its R-Square is much higher as well.