

CH4- Model Assessment

Amin Shirazi

2/11/2020

Assessing models

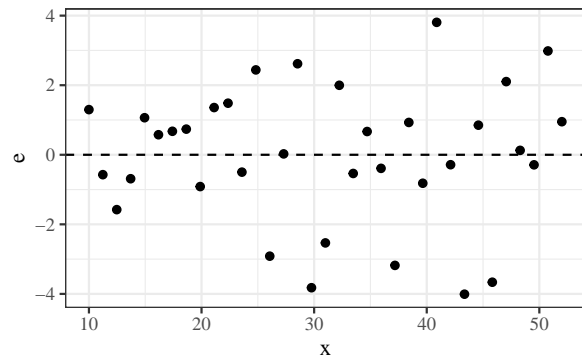
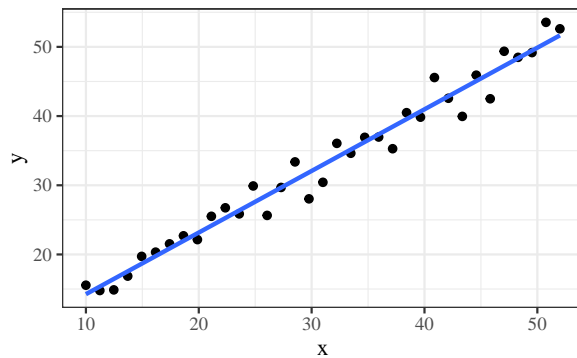
When modeling, it's important to assess the (1) **validity** and (2) **usefulness** of your model. These are model assumptions and we need to check them before making any decisions. If these assumptions are violated, we need to go back and review things and maybe refit the data using other methods to have a valid fit on the data.

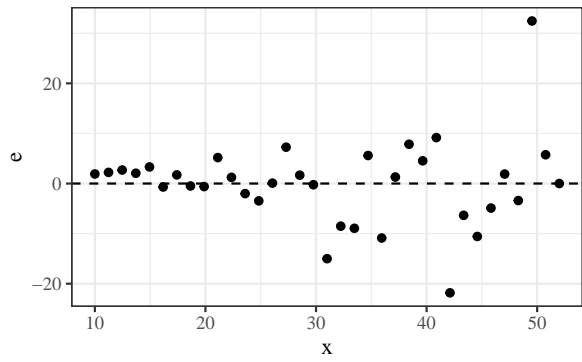
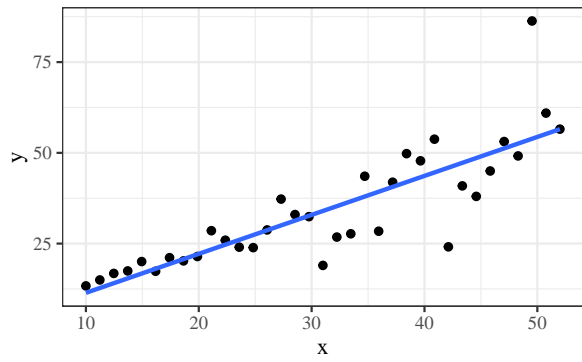
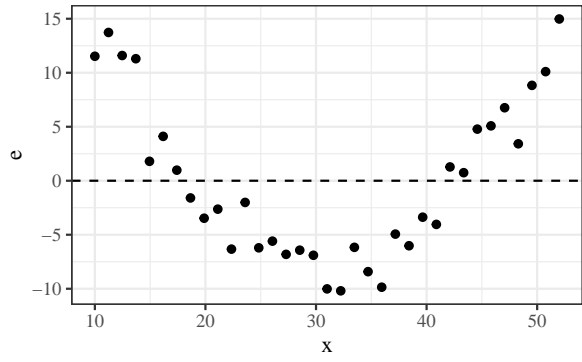
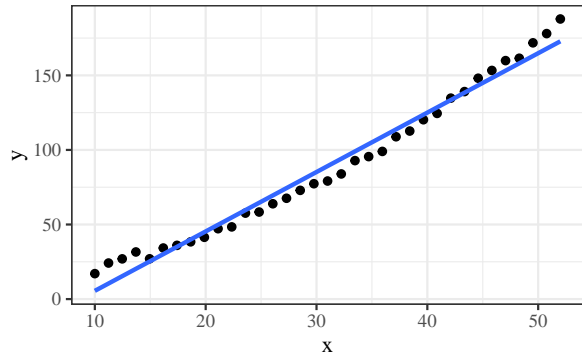
To assess the validity of the model (i.e. to see if model assumptions are met), we will look to the residuals. If the fitted equation is the good one, the residuals will be:

- Ptternless (cloud like, random scatter)
- Centered at zero
- Bell shaped distribution

To check if these three things hold, we will use two plotting methods.

A **residual plot** is a plot of the residuals, $e = y - \hat{y}$ vs. x (or \hat{y} in the case of multiple regression, Section 4.2).



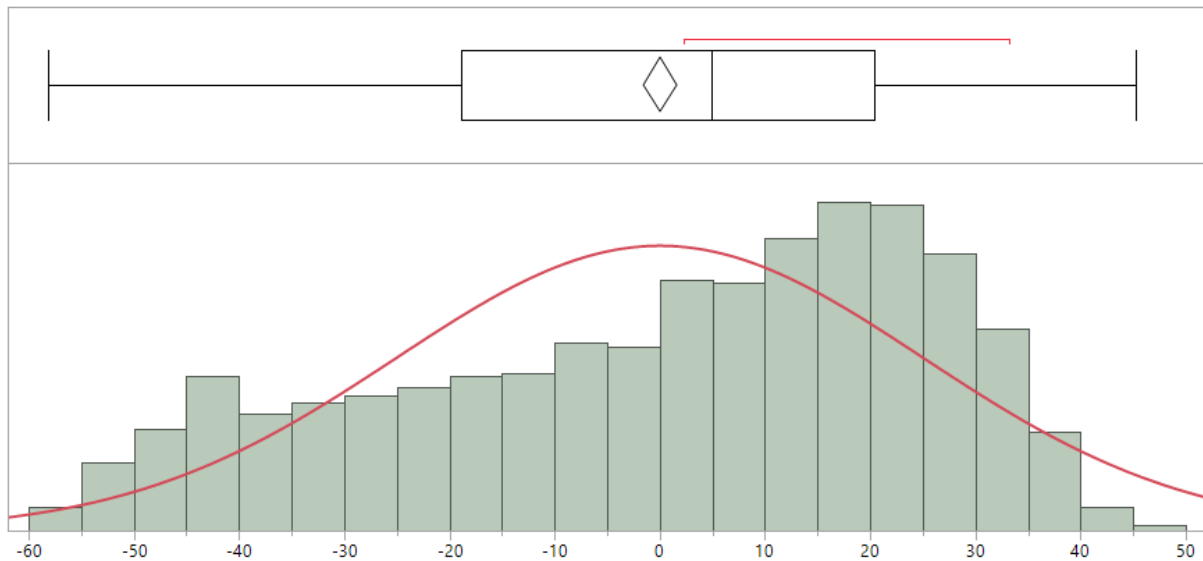


Normality of residuals

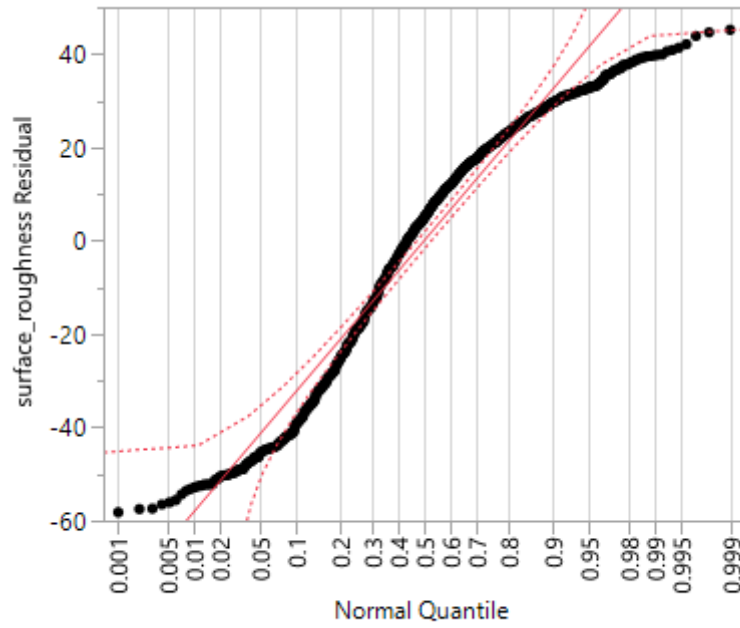
- In addition to the residual versus predicted plot, there are other residual plots we can use to check regression assumptions.
- A **histogram of residuals** and a **normal probability plot (QQ-plot)** of residuals can be used to evaluate whether our residuals are approximately normally distributed.
 - However, unless the residuals are far from normal or have an obvious pattern, we generally don't need to be overly concerned about normality.
- Note that we check the residuals for normality. We don't need to check for normality of the raw data. Our response and predictor variables do not need to be normally distributed in order to fit a linear regression model.

How to check the Normality assumptions of residuals? We are already familiar with theoretical Normal QQ-plots and histograms. So, in order to check normality assumptions of residuals, we need to see if (1) the histogram of the residuals are bell shaped and (2) check their normal QQ-plot to see if the residuals lie around a straight line.

Draw a histogram of the residuals (review the JMP tutorial for histograms)



It seems the residuals are not normally distributed in this example. The residuals have a left skewed distribution.



- Again, the QQ-plot also confirms that the assumption of Normal distribution of residuals is violated to some extent in this example.
- More examination is required to fix the issue or to find the problem.