

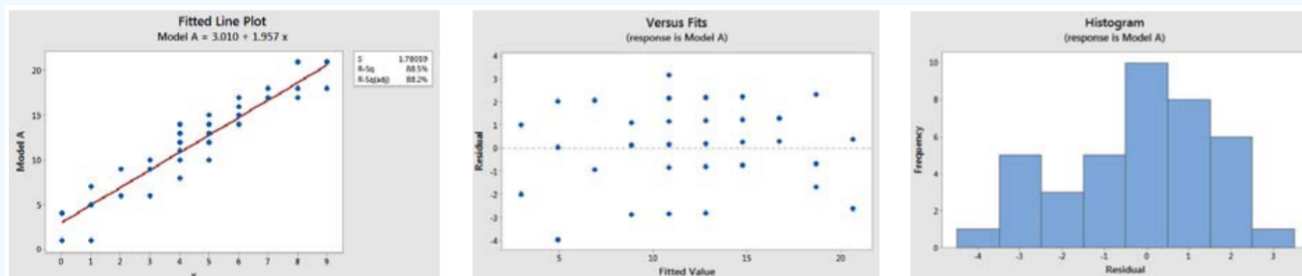
3.3.9: Residual Analysis

In regression, we assume that the model is linear and that the residual errors ($Y - \hat{Y}$ for each pair) are random and normally distributed. We can analyze the residuals to see if these assumptions are valid and if there are any potential outliers. In particular:

- The residuals should represent a linear model.
- The standard error (standard deviation of the residuals) should not change when the value of X changes.
- The residuals should follow a normal distribution.
- Look for any potential extreme values of X .
- Look for any extreme residual errors.

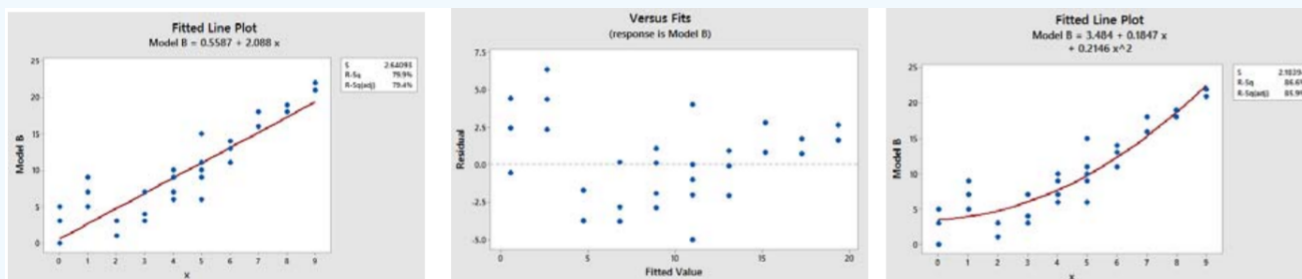
✓ Example: Model A

Model A is an example of an appropriate linear regression model. We will make three graphs to test the residual; a scatterplot with the regression line, a plot of the residuals, and a histogram of the residuals



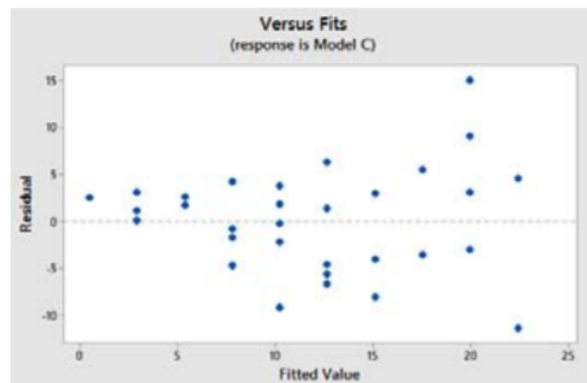
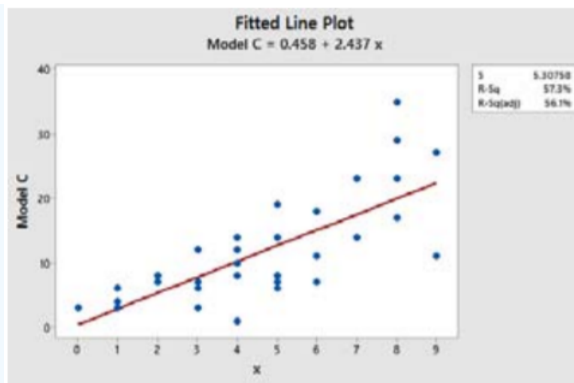
Here we can see that the residuals appear to be random, the fit is linear, and the histogram is approximately bell shaped. In addition, there are no extreme outlier values of X or outlier residuals.

✓ Example: Model B



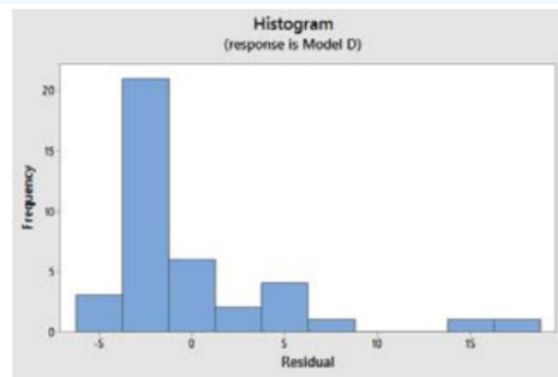
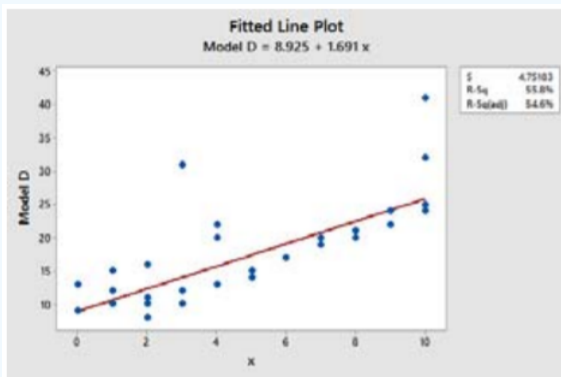
Model B looks like a strong fit, but the residuals are showing a pattern of being positive for low and high values of X and negative for middle values of X . This indicates that the model is not linear and should be fit with a non-linear regression model (for example, the third graph shows a quadratic model).

✓ Example: Model C



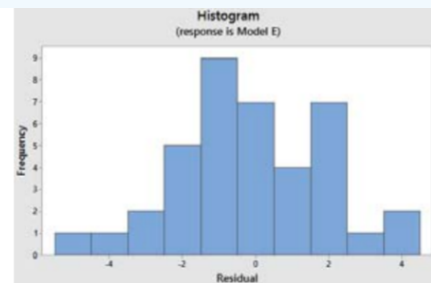
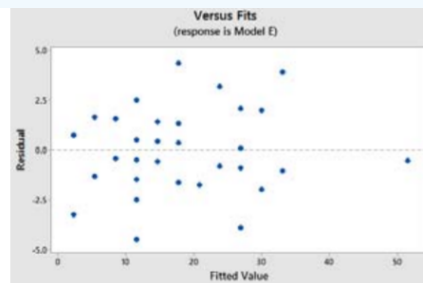
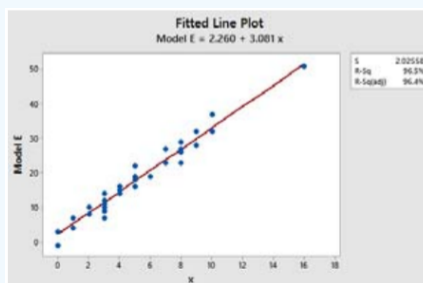
Model C has a linear fit, but the residuals are showing a pattern of being smaller for low values of X and higher for large values of X . This violates the assumption that the standard error should not change when the value of X changes. This phenomena is called **heteroscedasticity** and requires a data transformation to find a more appropriate model.

✓ Example: Model D



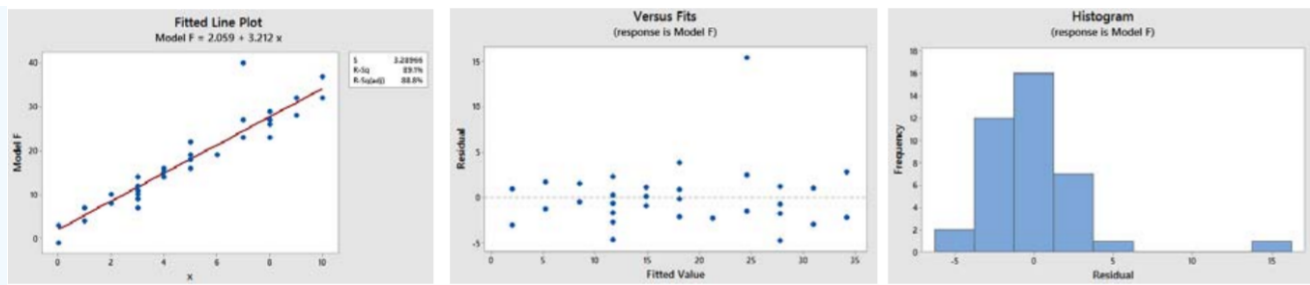
Model D seems to have a linear fit, but the residuals are showing a pattern of being larger when they are positive and smaller when they are negative. This violates the assumption that residuals should follow a normal distribution, as can be seen in the histogram.

✓ Example: Model E



Model E seems to have a linear fit, and the residuals look random and normal. However, the value (16,51) is an extreme outlier value of X and may have an undue influence on the choosing of the regression line.

✓ Example: Model F



Model F seems to have a linear fit, and the residuals look random and normal, except for one outlier at the value (7,40). This outlier is different than the extreme outlier in Model E, but will still have an undue influence on the choosing of the regression line.

3.3.9: Residual Analysis is shared under a CC BY-SA license and was authored, remixed, and/or curated by LibreTexts.

- **14.9: Residual Analysis** by Maurice A. Geraghty is licensed CC BY-SA 4.0. Original source: <http://nebula2.deanza.edu/~mo/holisticInference.html>.