What to do when assumptions aren't met

Assumption 1:
Relationship is linear.

How to detect a problem:
Plot residuals versus fitted values. If you see a pattern, there is a problem with the assumption.

What to do about the problem:

Transform the **X** values, X' = f(X). Then do the regression using X' instead of X:

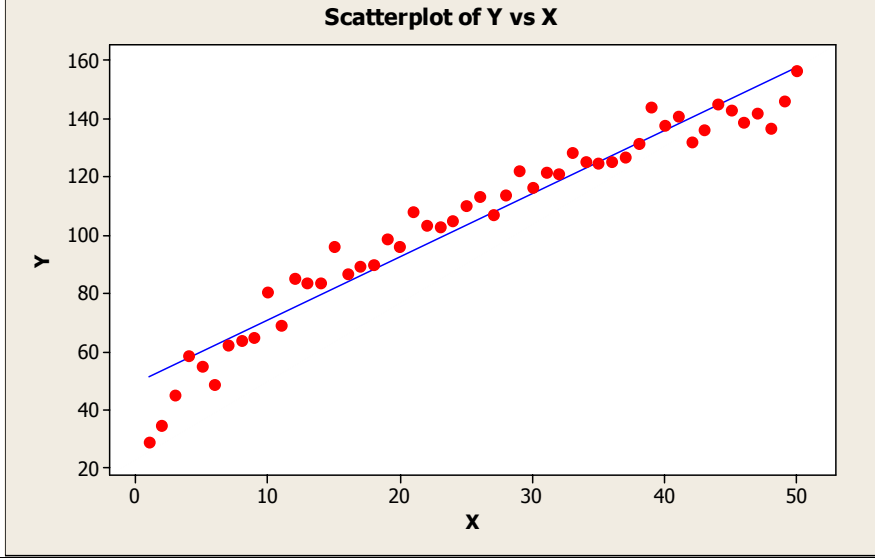
$$Y = \beta_0 + \beta_1 X' + \varepsilon$$

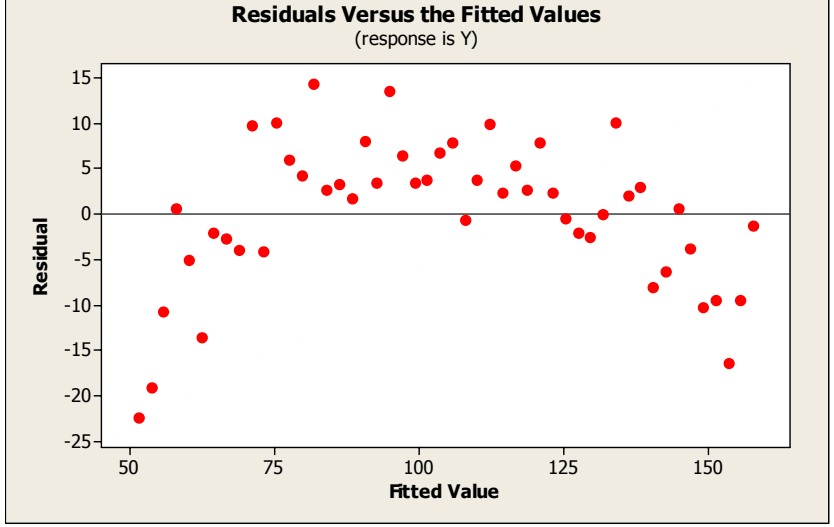
where we still assume the ε are $N(0, \sigma^2)$.

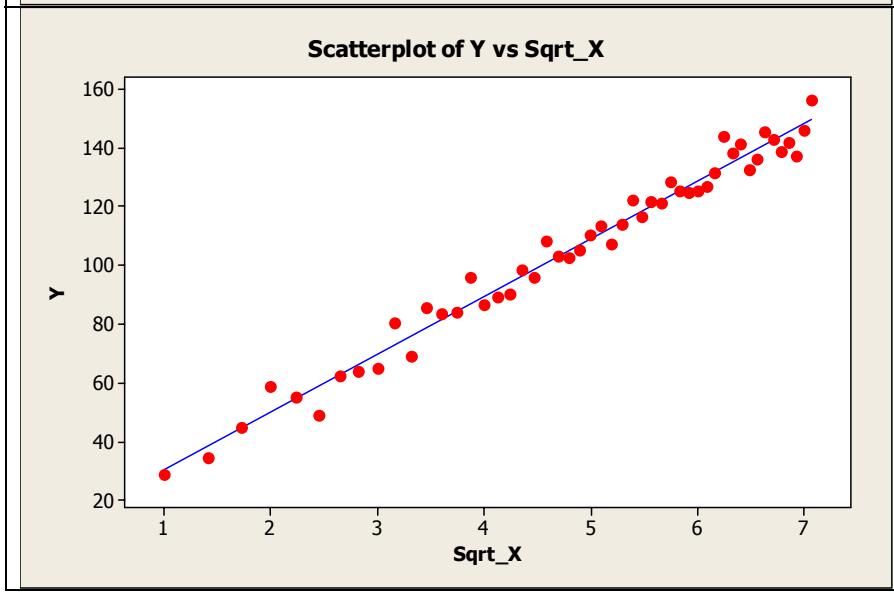
NOTE: Only use this "solution" if non-linearity is the *only* problem, not if it also looks like there is non-constant variance or non-normal errors. For those, we will transform Y.
Reasoning: The errors are in the vertical direction. Stretching or shrinking the X-axis doesn't change those, so if they are normal with constant variance, they will stay that way.

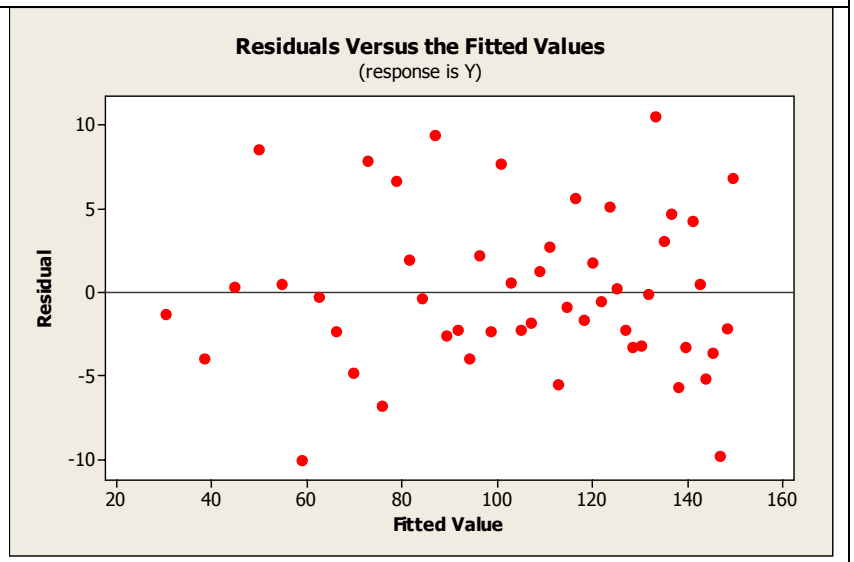
Let's look at what kinds of transformations to use.

Residuals are inverted U, use X' = $\sqrt{X}$ or $\log_{10}$ X

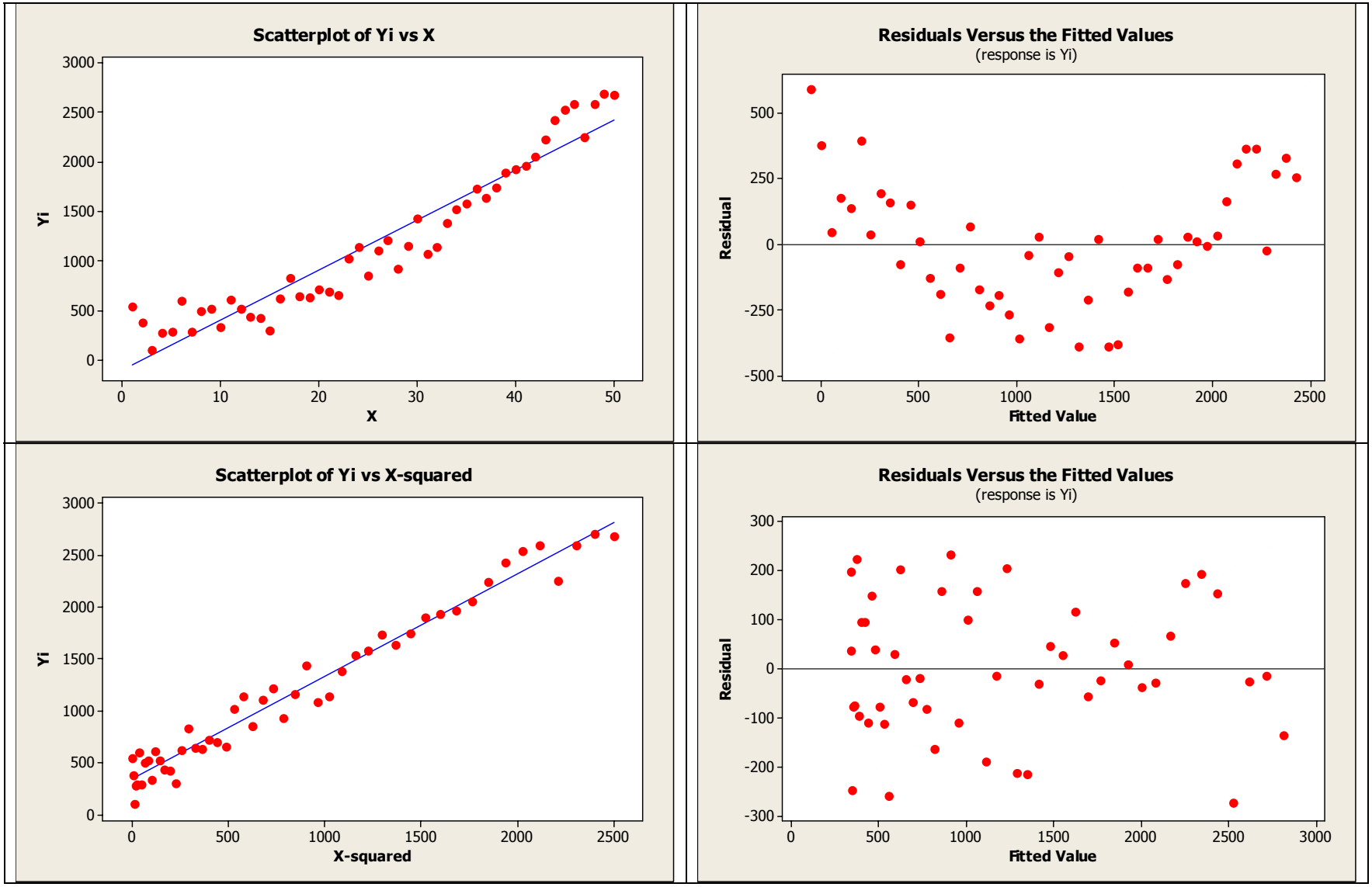








Residuals are U-shaped and association between X and Y is positive: Use $X' = X^2$



Residuals are U-shaped and association between X and Y is negative:
Use X' = 1/X or X' = exp(-X)