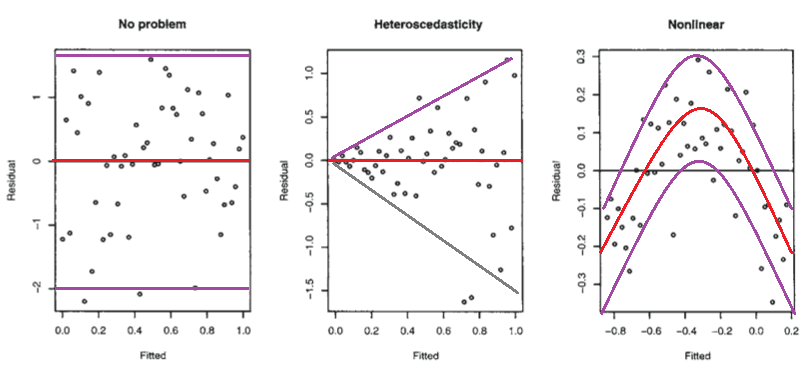
Below are those residual plots with the approximate mean and spread of points (limits that include most of the values) at each value of fitted (and hence of x) marked in - to a rough approximation indicating the conditional mean (red) and conditional mean ± (roughly!) twice the conditional standard deviation (purple):



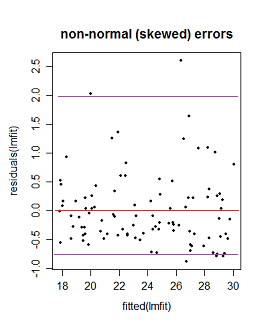
* The second plot shows the mean residual doesn't change with the fitted values (and so is doesn't change with x), but the spread of the residuals (and hence of the y's about the fitted line) is increasing as the fitted values (or x) changes. That is, the spread is not constant. Heteroskedasticity.
* the third plot shows that the residuals are mostly negative when the fitted value is small, positive when the fitted value is in the middle and negative when the fitted value is large. That is, the spread is approximately constant, but the conditional mean is not - the fitted line doesn't describe how y behaves as x changes, since the relationship is curved.

Isn't it possible that it is linear, but that the errors are either not normally distributed, or else that they are normally distributed, but do not center around zero?

No, in those situations the plots look different to the third plot.

(i) If the errors were normal but not centered at zero, but at θ, say, then the intercept would pick up the mean error, and so the estimated intercept would be an estimate of β0+θ (that would be its expected value, but it is estimated with error). Consequently, your residuals would still have conditional mean zero, and so the plot would look like the first plot above.

(ii) If the errors are not normally distributed the pattern of dots might be densest somewhere other than the center line (if the data were skewed), say, but the local mean residual would still be near 0.



Here the purple lines still represent a (very) roughly 95% interval, but it's no longer symmetric. (I'm glossing over a couple of issues to avoid obscuring the basic point here.)