Lecture 18

Regression Diagnostics

Regression diagnostics include using Cook's Distances to find points that exert undue influence over the regression line, looking at your residuals for evidence of lack of normality and heterogeneity of variance, and testing whether the model is really linear or something else. It incorporates topics from both the previous lecture on outliers and leverage points and the next section on transformations.

Is a line the correct model?

Minitab can perform two different tests to determine if a line is a good model for your data. One looks to see if a separate means model fits the data better. To conduct this test, you need repeated values for your predictor variable - something you can guarantee if you choose the predictor values but not if you don't. We don't get to choose the biparietal diameter of fetuses, but, with 1000 data points, we do have many biparietal diameters that are repeated in the data set. The other looks to see if their is curvature in the data. It divides the predictor values in half (or so) and looks to see if fitting a line to the first half and a line to the second half which meets up with the first half appropriately leads to lines with the same slop or different slopes. If the slopes are different, then the data has curvature which is reported in the output.

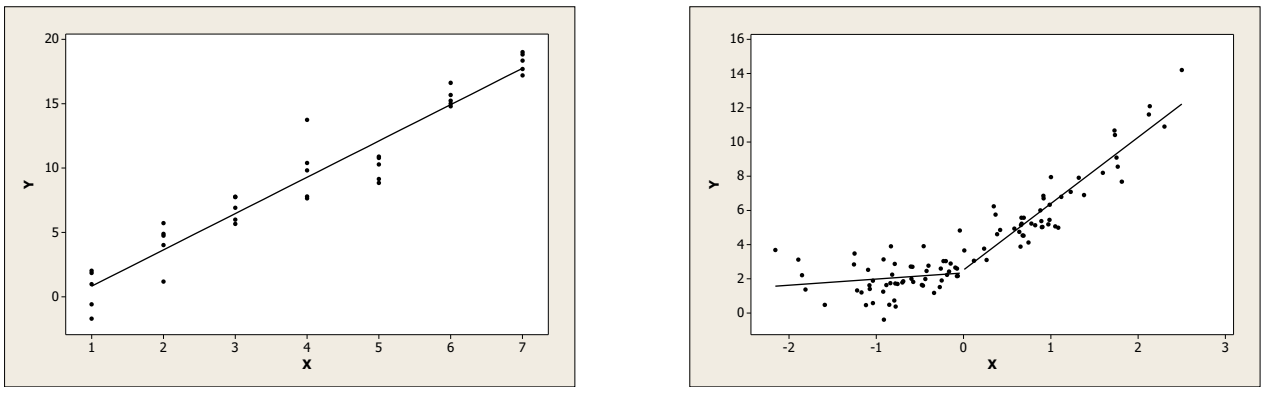

Figure 18.1: Examples when a separate means model may be better than a linear one and when there is curvature in the data that can be determined by the data subsetting procedure.

Exercises for Lecture 18

1. –

2. –