

3.3: Influential Observations and Outliers

Dr. Bean - Stat 5100

Is it possible for a model outlier to not be reflected in a boxplot of Y ? Explain why or why not.

Recall that $h_{i,l}$ represents the element in row i and column l of H
- sometimes called “leverage” (influence of obs. i on its fitted value)

Since $\hat{Y} = HY$, then $\hat{Y}_i = \sum_{l=1}^n h_{i,l} Y_l$

What would a “larger” diagonal element $h_{i,i}$ mean?

What would cause an observation to have high leverage? (Hint: It has nothing to do with the values of Y .)

True or False (and explain): An observation with high leverage will have a large influence on the model fit, and an observation with low leverage will not have a large influence on the model fit.

For the remaining questions, please refer to Figure 1 and assume the theoretical linear model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$.

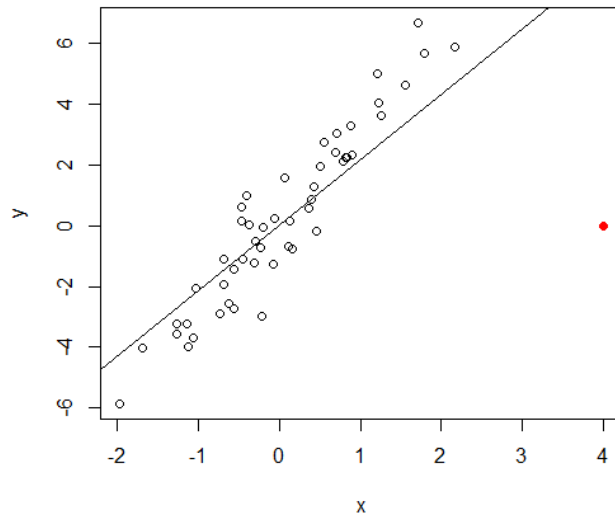


Figure 1: Sample scatterplot with estimated regression model.

Would the DFBETA associated with the filled in point be positive or negative? Explain.

Would the DFFITS associated with the filled in point be positive or negative? Explain.

Without any formal diagnostic checks, do you think that the filled in point is an outlier, influential point, both, or neither? Explain.