Influential Points

High Leverage versus High Influence

"Influence" does not have a formal statistical definition. In SDS 291, we'll consider a point to be highly influential if the final fitted regression model (as summarized by the predicted values, $\hat{y}$) changes substantially when that point is used to fit the regression model versus when it is omitted from the sample.

In a simple linear regression context, sketch a scatterplot that illustrates the following in each of the following scenarios:

a. The sample contains a point that has **high leverage but low influence** over the fitted model.

b. The sample contains a point that has **high leverage and high influence** over the fitted model.

c. The sample contains a point that has a **large studentized residual but low influence** over the fitted model.

d. The sample contains a point that has a **large studentized residual and high influence** over the fitted model.