Outlier detection

- Generally, outliers will have a large externally studentized residual in absolute value.
- **Recall:** $h_{ii} = [\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}']_{ii}$ is a diagonal entry of the hat matrix and measures how much y_i contributes to the estimation of $\hat{y}_i$. We call h_{ii} leverages.
- Typically, observations with

$$|r_i^*| = \left| \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1 - h_{ii}}} \right| > 2$$

  are considered to be possible outliers.
- In SAS, the `RStudentByLeverage` plot can help us to visualize observations exceed this cutoff.
- **Note:** If you decide to make any modifications to the model (e.g., transformations to variables), you should check again for outliers.

High-Leverage Points I

- Observations that are outliers in X-space are considered high-leverage points.
- Outliers in X-space are high leverage values h_{ii}.
- It can be shown that $\sum h_{ii} = p + 1$, therefore the average value for h_{ii} is $(p + 1)/n$.
- Observations with h_{ii} greater than $2(p + 1)/n$ are generally considered to be high-leverage points.
- A plot of h_{ii} against the index i can be useful for this purpose.