Lecture Summary: March 10, 2023

• Identifying influential cases

An outlying case isn't necessarily influential.

To identify an influential case we consider Cook's distance, defined as

$$D_i = \frac{\sum_{j=1}^{n}\{\hat{Y}_j - \hat{Y}_{j(-i)}\}^2}{p\text{MSE}},$$

where $\hat{Y}_j$ is the predicted value of Y_j via regression with the full data, and $\hat{Y}_{j(-i)}$ is the predicted value of Y_j via regression with the data without the ith case. Large value of D_i indicates potential influential case.

Another expression (computationally more convenient):

$$D_i = \frac{h_{ii}\hat{\epsilon}_i^2}{p(1-h_{ii})^2\text{MSE}}.$$

• Logistic regression

Introduction: Post -Challenger risk analysis of the space shuttle [Reference: Dalal *et al.* (1989), Rick analysis of the space shuttle: Pre-Challenger prediction of failure, *J. Amer. Statist. Assoc.* 84, 945–957].