Cook's Distance

- In statistics, Cook's distance or Cook's D is a commonly used estimate of the influence of a data point when performing least squares regression analysis.

- In a practical ordinary least squares analysis, Cook's distance can be used in several ways: to indicate data points that are particularly worth checking for validity; to indicate regions of the design space where it would be good to be able to obtain more data points.

- It is named after the American statistician R. Dennis Cook, who introduced the concept in 1977.

- Cook's distance measures the effect of deleting a given observation. Data points with large residuals (outliers) and/or high leverage may distort the outcome and accuracy of a regression.

- Points with a large Cook's distance are considered to merit closer examination in the analysis.

It is calculated as:
$$Di = nj = 1(Y^j Y^j(i))2pMSE,$$

0.1 Leverage

- In statistics, leverage is a term used in connection with regression analysis and, in particular, in analyses aimed at identifying those observations that are far away from corresponding average predictor values.

- Leverage points do not necessarily have a large effect on the outcome of fitting regression models.

- Leverage points are those observations, if any, made at extreme or outlying values of the independent variables such that the lack of neighboring observations means that the fitted regression model will pass close to that particular observation.[1]

- Modern computer packages for statistical analysis include, as part of their facilities for regression analysis, various quantitative measures for identifying influential observations: among these measures is partial leverage, a measure of how a variable contributes to the leverage of a datum.