

## Seminar - Cook's Distance

### Math 567: Winter 2016

## 1 Little Background

In Regression analysis, Cook's Distance is used to estimate the influence of a data point over the model when performing the ordinary least Square regression. It was introduced by R. Dennis Cook in 1977. [Cook, 1977]

## 2 Definition

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. For the algebraic expression, first define.

## 3 Highly influential data point

A data point is said to be influential if when removed from the calculation change the regression line significantly. Data point with high leverage can be influential if they it is an outlier. A data point can have an high leverage but not influential, goes the same way for an outlier(all outlier are not influential).

## 4 Interpretation of Cook's Distance

## 5 Cook's Distance using R

## 6 Discussion

## References

R.Dennis Cook. Detection of influential observation in linear regression, 1977. URL [http://www.jstor.org/stable/1268249?origin=crossref&seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/1268249?origin=crossref&seq=1#page_scan_tab_contents).