Iliass Tiendrebeogo
February 29, 2016

Seminar 4 - Hat Value
Math 567: Winter 2016

1 Little Background

2 Highly influential data point

A data point is said to be influential if when removed from the calculation change the regression line significantly. How influential a data-point is, is the combination of how much leverage it has and how extreme it is in the y direction. However, A data point can have an high leverage but not influential, and it goes the same way for an outlier(all outlier are not influential).

3 Definition

Hat-matrix

The algebraic expression.

$$\hat{y} = Hy$$

Where:
$\hat{Y}_j$ is the prediction from the full regression model for observation j;
$\hat{Y}_{j(i)}$ is The prediction for observation j from a refitted regression model in which observation i has been omitted;

The **leverage** define how far apart is a given data point from the average(mean/median). Points with high leverage tend to pull the regression line toward themselves and have impact on the slop of the regression line hence **influential**.

4 Interpretation of Hat values

There are several rules when interpreting **cook's distance**. The widely used criterion is that a point is considered to be highly influential if $D_i > 1$ [**?**]

Different rules have been defined such as: $D_i > 8.5$ if $p > 3$ [**?**] where p is the number of regression parameter. [**?**] declares a data-point to be influential when $D_i > \frac{4}{n}$ where n is the number of observation.

5 Hat values using R

Packages use:
`install.packages (QuantPshyc)`
`library(QuantPshyc)`
call Cook's D_i from the library: `linearmodel.cook()`
manually computing cook's distance:

6 Discussion

What to do when a given data-point's $D_i > 1$? has an