# Assignment 4

*Aytijhya Saha*
Roll No. BS2002

Indian Statistical Institute, Kolkata

*October 18, 2022*

## Influential points: Introduction

An influential point is one whose removal from the dataset would cause a large change in the fit. An outlier may or may not be an influential point.

For example, we consider the 10 points in figure 1. They are pretty much on a straight line, that is the least square fitted line.

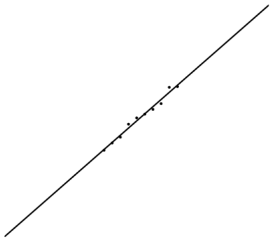

Figure 1

Now suppose we add the the blue point, shown in figure 2. This point is an outlier because it is far from the bulk of the data. We observe that the resulting fit, which is shown by the red line has deviated a lot from this black line. Because of the single point, that line swung from this good fit to a rather bad fit position. We will call such a point an influential point. It has so much influence that it could outweight the initial 10 points.

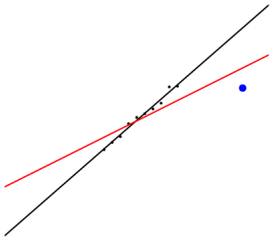

Figure 2

Now suppose, I take the outlier as shown in figure 3, keeping this vertical distance from the original fit same as before. But, here the fitted line is essentially ignoring this point. It is still adhering to its old position. Here we say that this outlier is not an influential point.

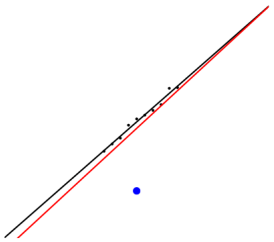

Figure 3

This example shows that the influence is not directly dependent on only the distance of the point from the original fit.

## Delete one out technique

This discussion pertains mainly to the situation where you are in the regression setup and not in an anova setup. So the inputs will be considered to be continuous. Even if there are some discrete inputs, we shall focus on only the continuous inputs. We have a data matrix, which has various columns of the variables including the responses and the predictors and we fit a linear model based on that and we get certain outputs. We primarily focus on the following four components of the output-

1. Estimates of the coefficients
2. Fitted values
3. Residuals
4. Covariance matrix of the coefficients

Suppose we delete the $i$th row of the data matrix and then again re-fit the entire model and see how much has the values of the four components listed above changed. Based on that, we try to measure the influence that particular case has. We say, one case is influential, if omission of that makes a drastic change in the values of any one of these four items. This is the delete one out technique.