# DFBETAS

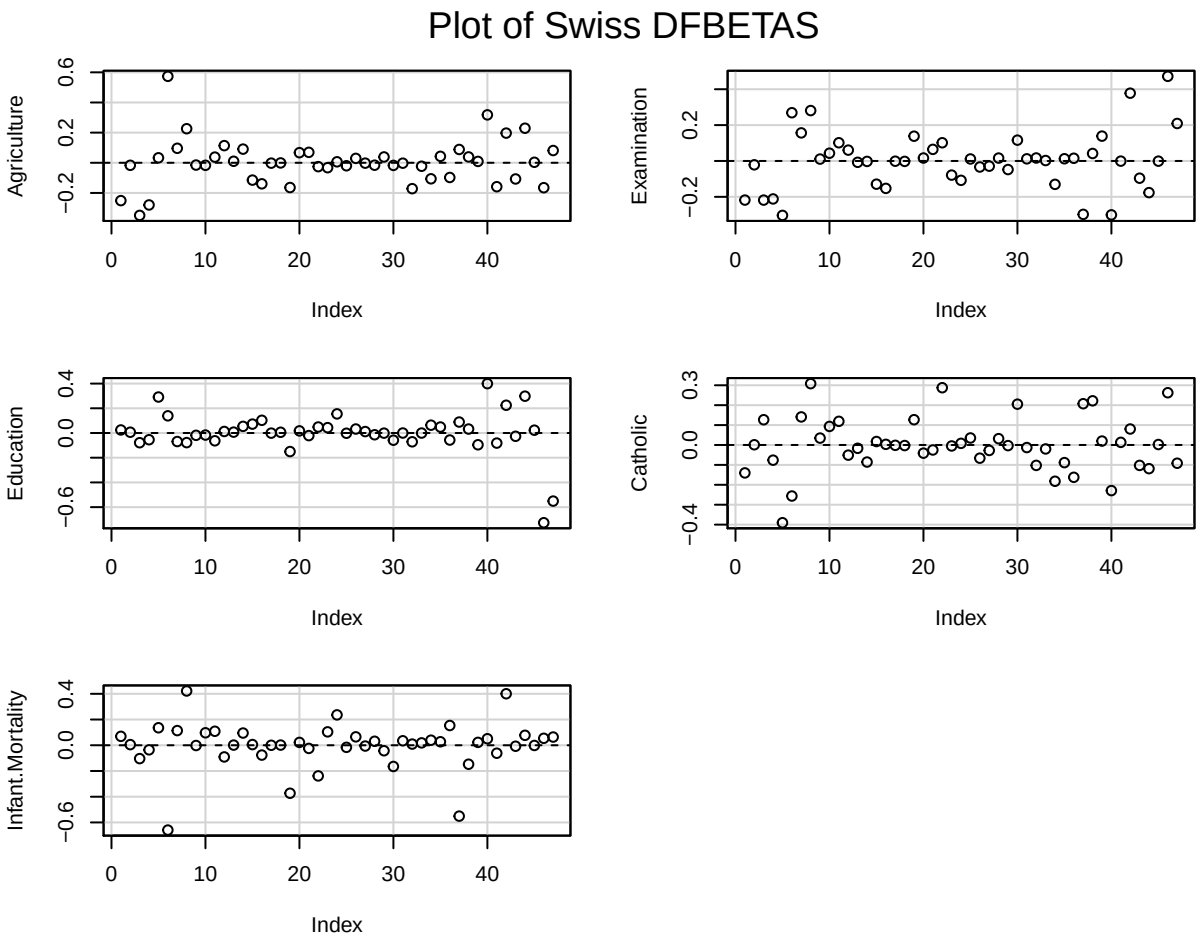
## Definition

**DFBETAS** is a post-regression tool for measuring the influence of outlying data points. It measures the effect on the regression coefficient for a variable of deleting each data observation. This difference is scaled using the standard error when the observation is omitted. When observations exceed the DFBETA cut-off value ($\frac{2}{\sqrt{n}}$), they should be investigated further as the observation could be overly influential on the estimated coefficent.

### Equations

$$DFBETAS_{ij} = \frac{\hat{\gamma}_i - \gamma_{i\hat{(-j)}}}{se(\gamma_{i\hat{(-j)}})}$$

## Example DFBETAS Plots for Swiss Data Set



## Interpretation and Use

DFBETAS is useful, especially in concert with DFFITS and Cook's Distance, to identify and investigate high influence points in a dataset. High influence by itself is not a valid reason to eliminate an observation but indicates that a researcher should evaluate the point for validity in the data.

**Further Avenues**
*DFFITS/Cook's Distance* - In conjunction with DFBETAS, DFFITS and Cook's Distance should corroborate any high influence points in the data.

## R Code

```
# use swiss dataset
data(swiss)
model <- lm(Fertility ~ ., data=swiss)
#dfbetasPlots(model, main="Plot of Swiss DFBETAS")
```