DFFITS

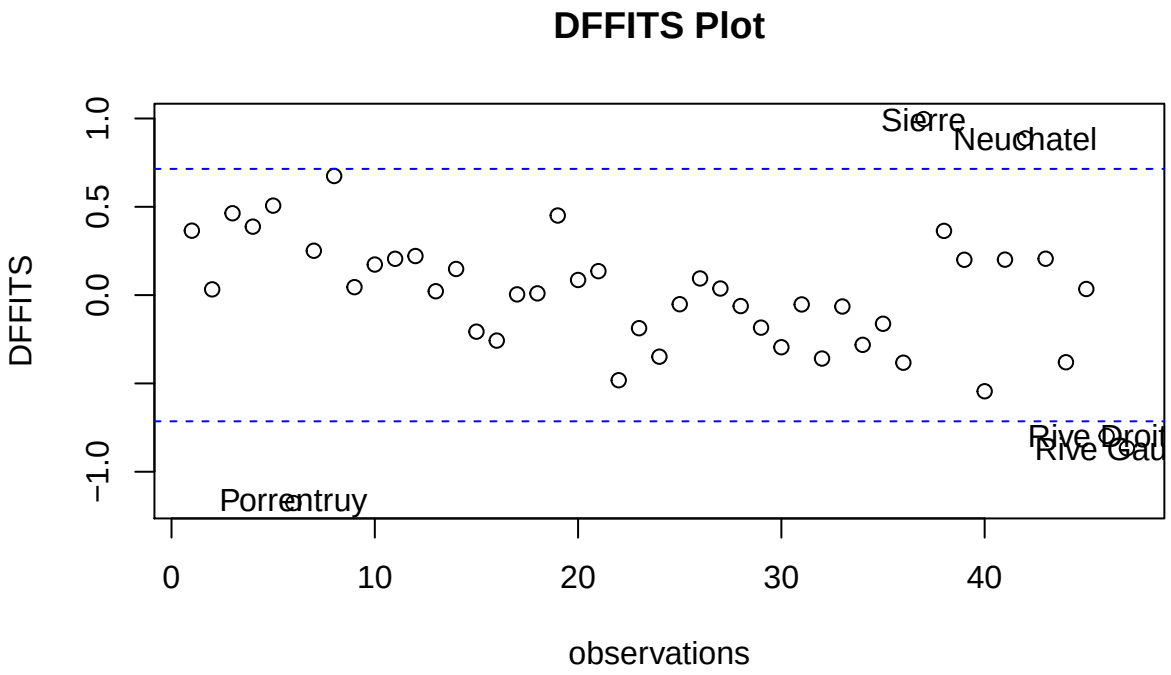
Definition

DFFITS 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 and "studentized" by dividing by the standard deviation. The cut-off for DFFITS is $2\sqrt{\frac{p}{n}}$.

Equations

$$DFFITS = \frac{\hat{y}_i - y_{\hat{i(i)}}}{s_{(i)}(\sqrt{h_{ii}})}$$

Example DFFITS Plots for Swiss Data Set



Interpretation and Use

DFFITS is useful, especially in concert with DFBETAS 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

DFBETAS/Cook's Distance - In conjunction with DFFITS, DFBETAS 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)
# dffitsData <- as.data.frame(dffits(model) )
# names(dffitsData) <- c("observations")
# cutoff <- 2*sqrt(6/47) #2*sqrt(p/n)
# plot(dffitsData$observations, main="DFFITS Plot", xlab="observations", ylab = "DFFIT
# abline(h=cutoff,col=4,lty=2)
# abline(h=-cutoff,col=4,lty=2)
# labels=row.names(swiss)
# text(c(6,37,42,46,47),dffitsData[c(6,37,42,46,47),],labels=labels[c(6,37,42,46,47)])
```