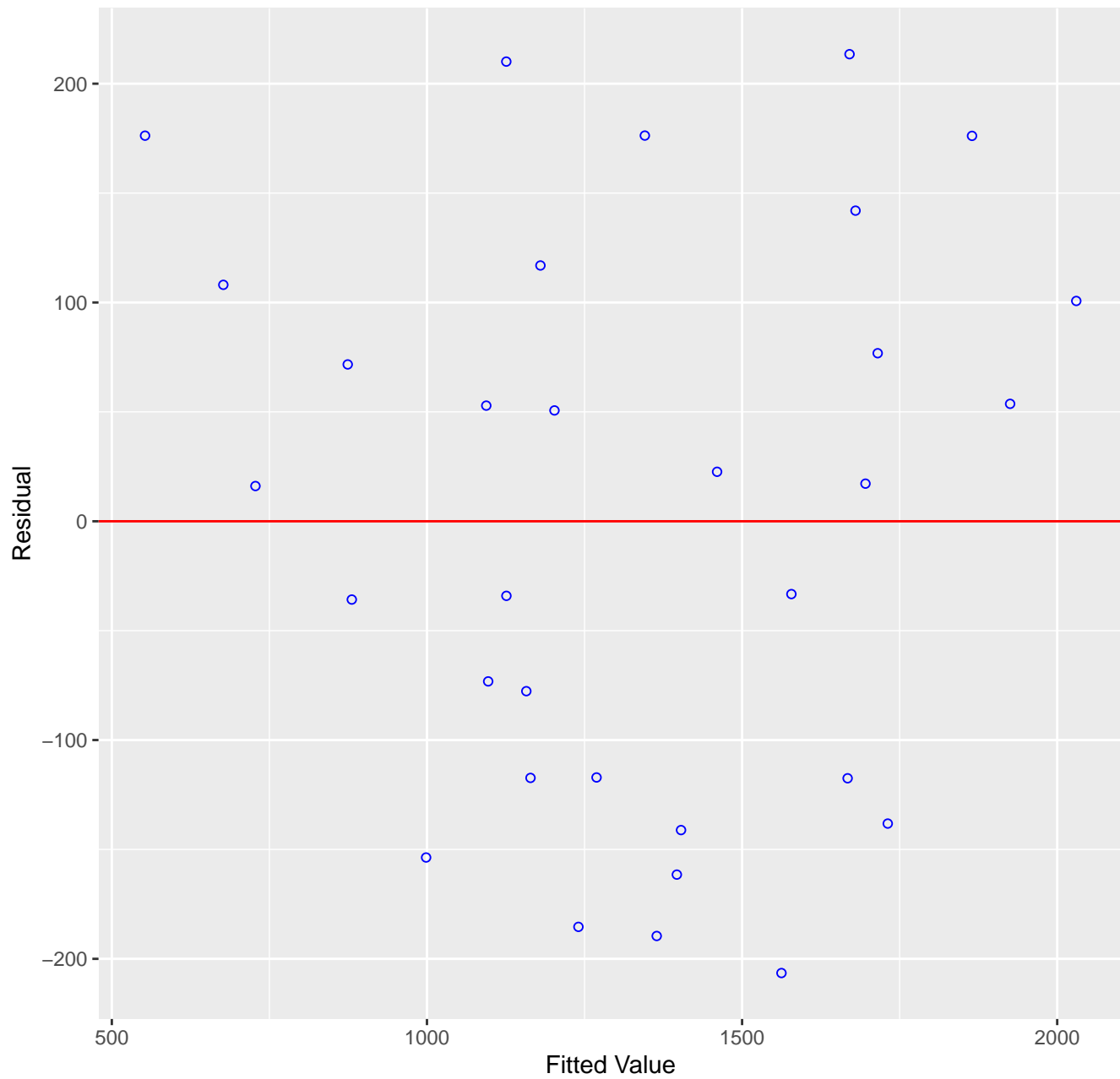
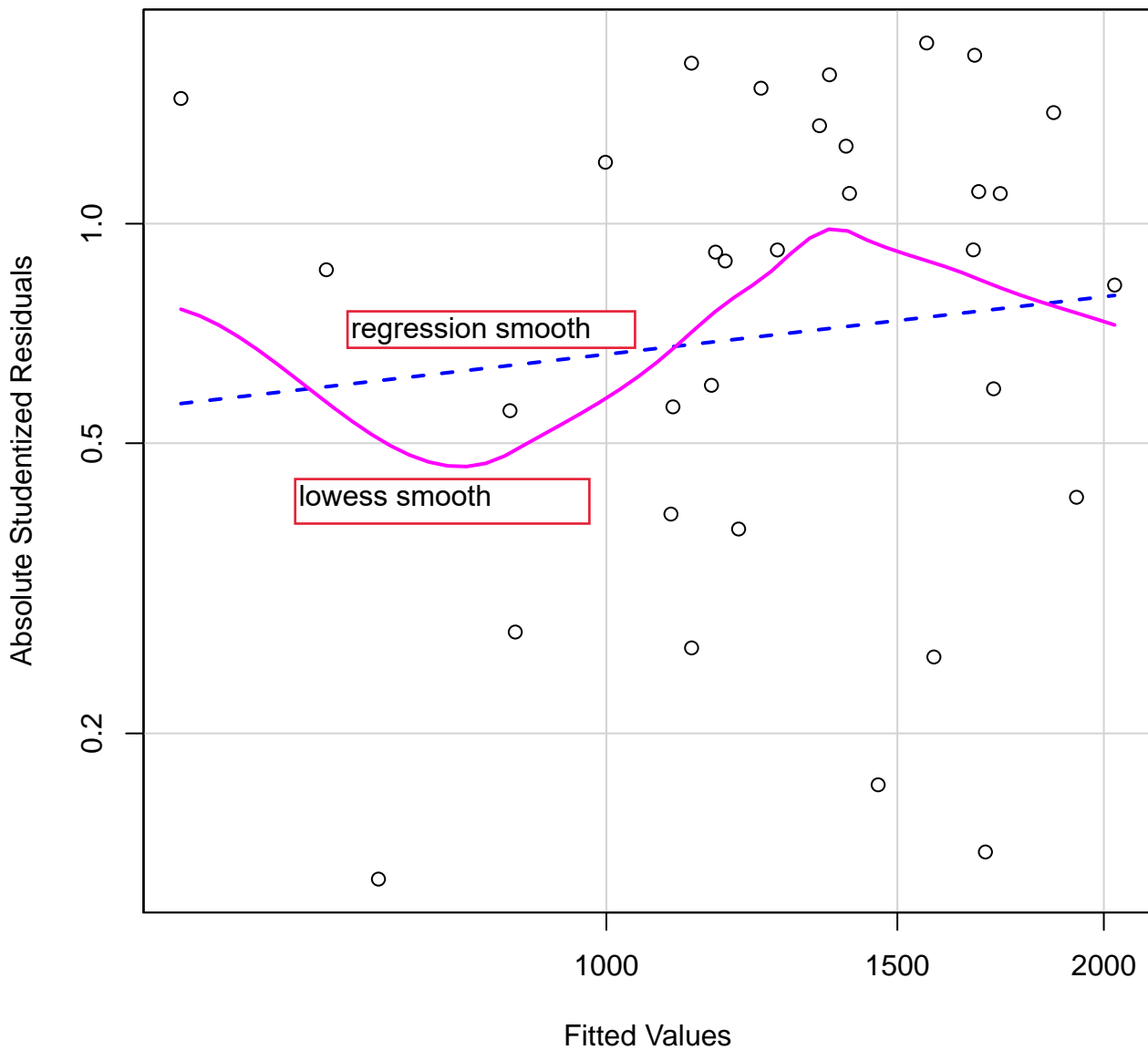


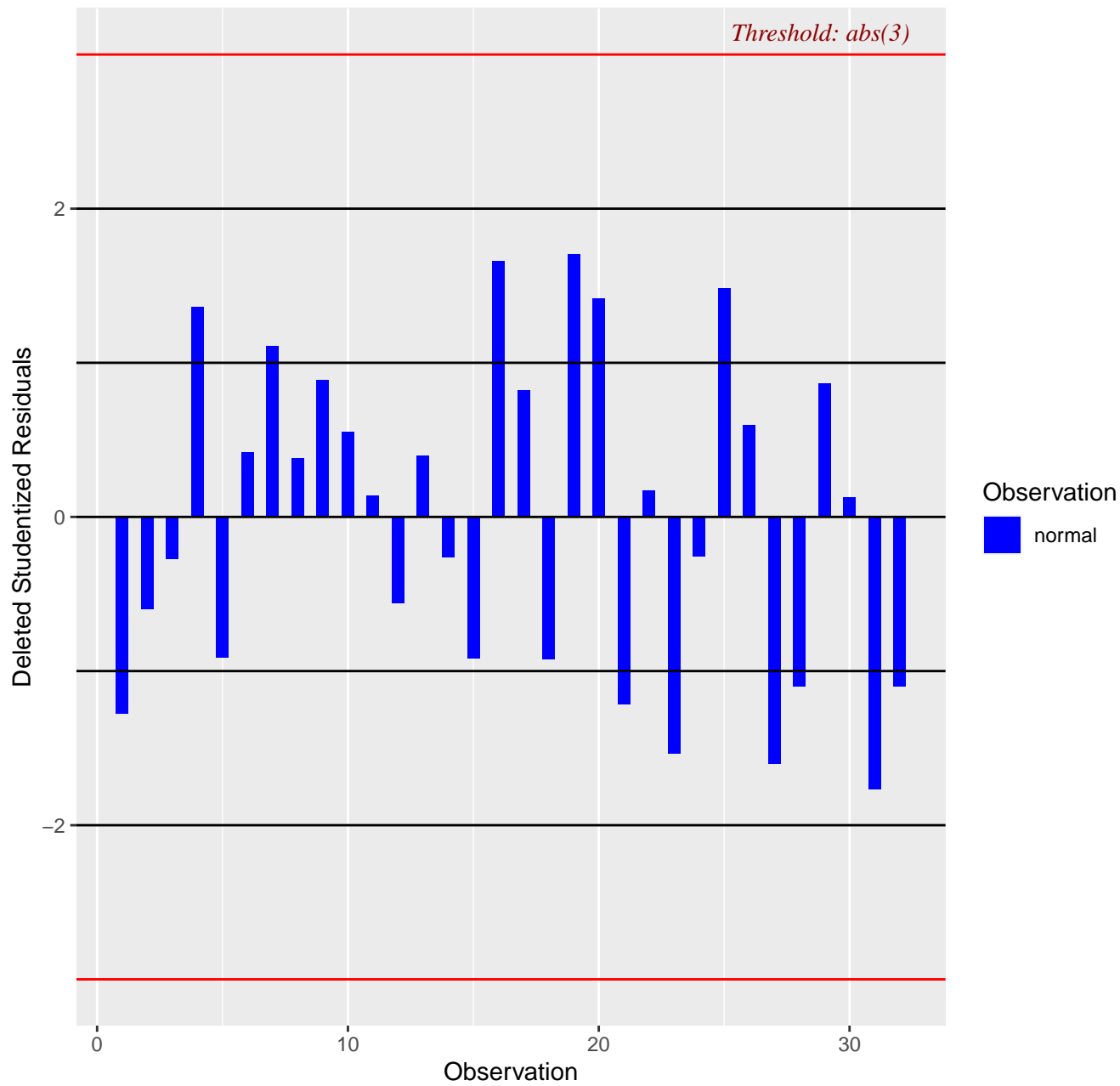
Residual vs Fitted Values



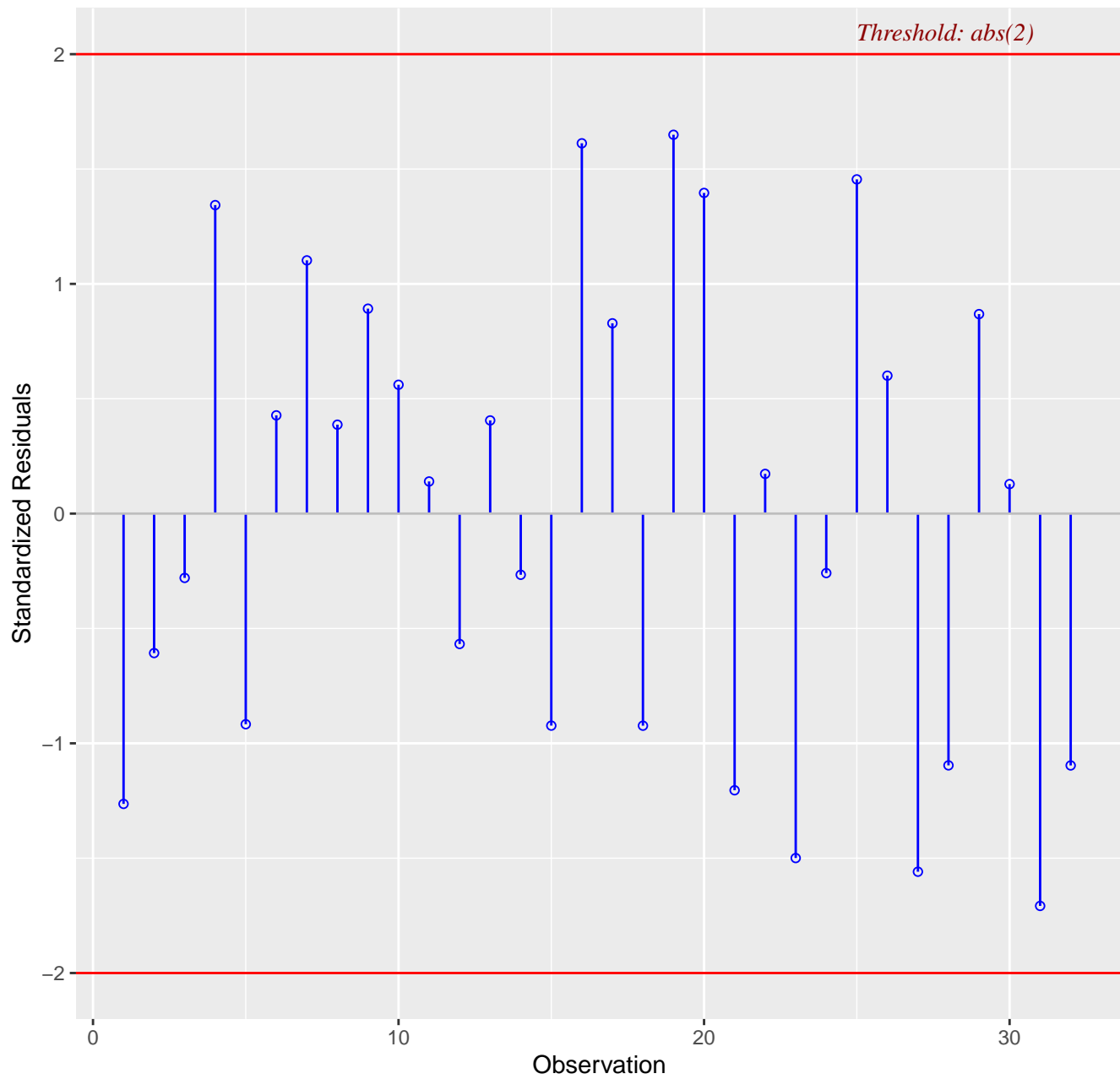
Spread–Level Plot for lmod



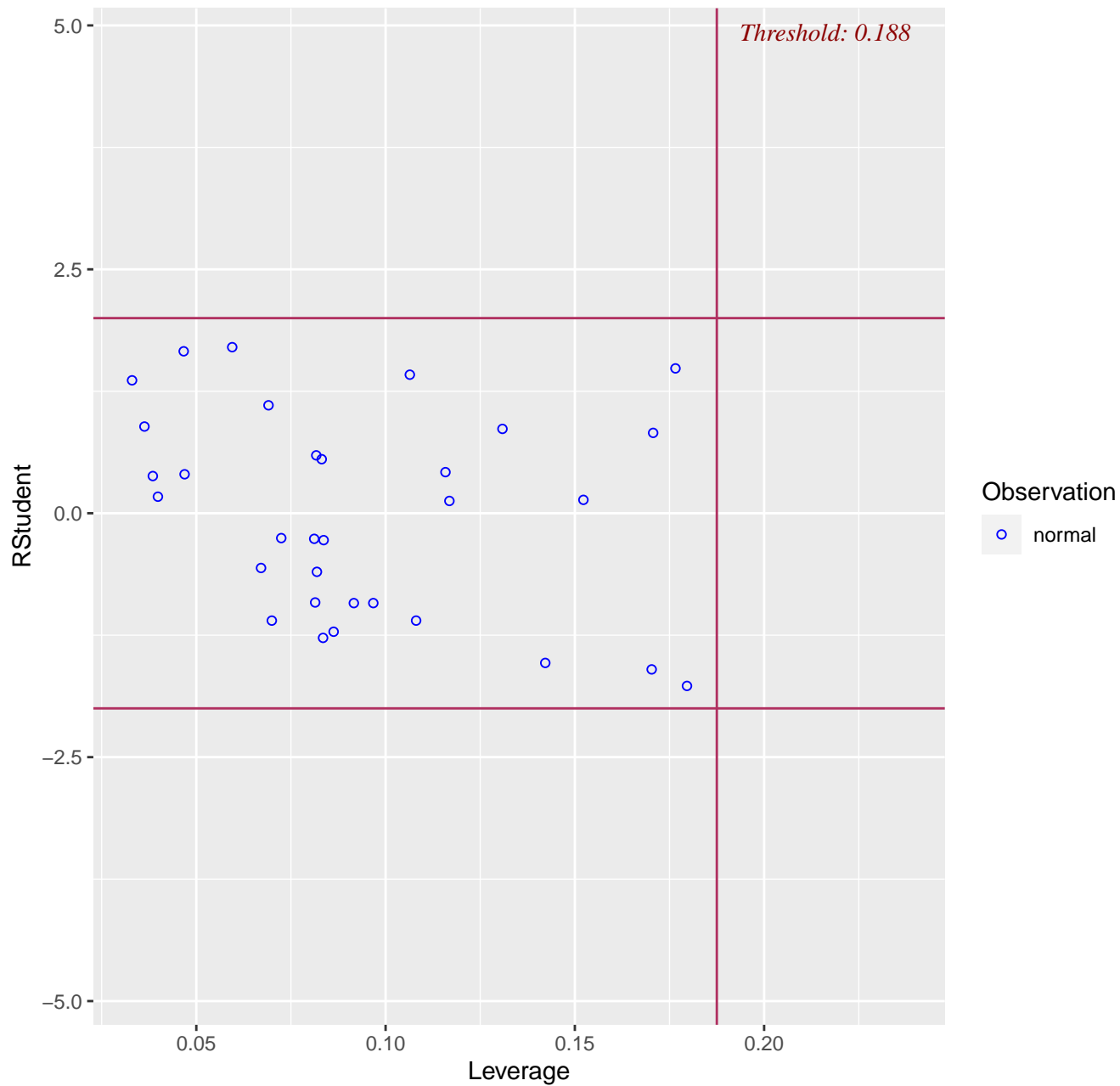
Studentized Residuals Plot



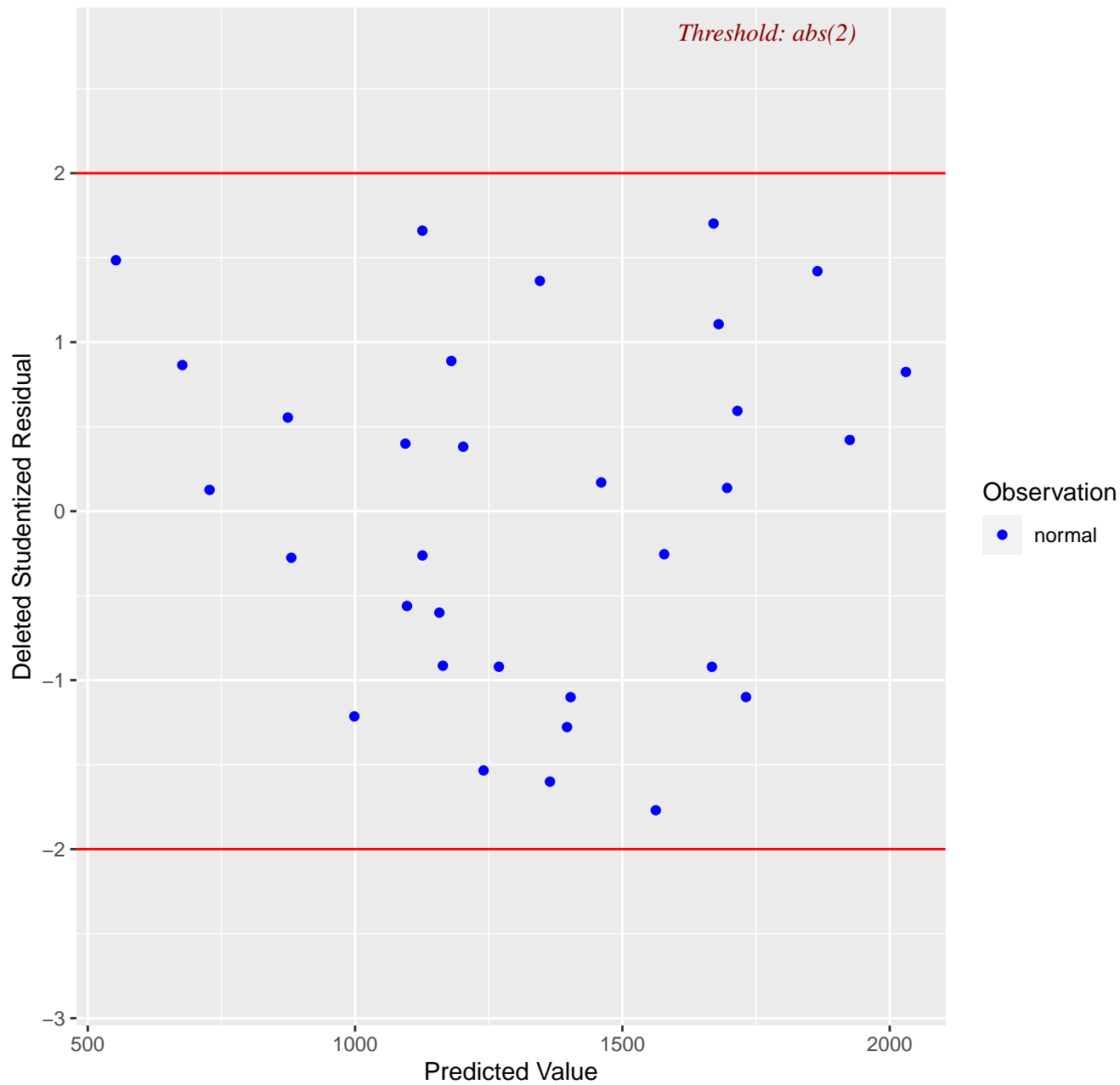
Standardized Residuals Chart



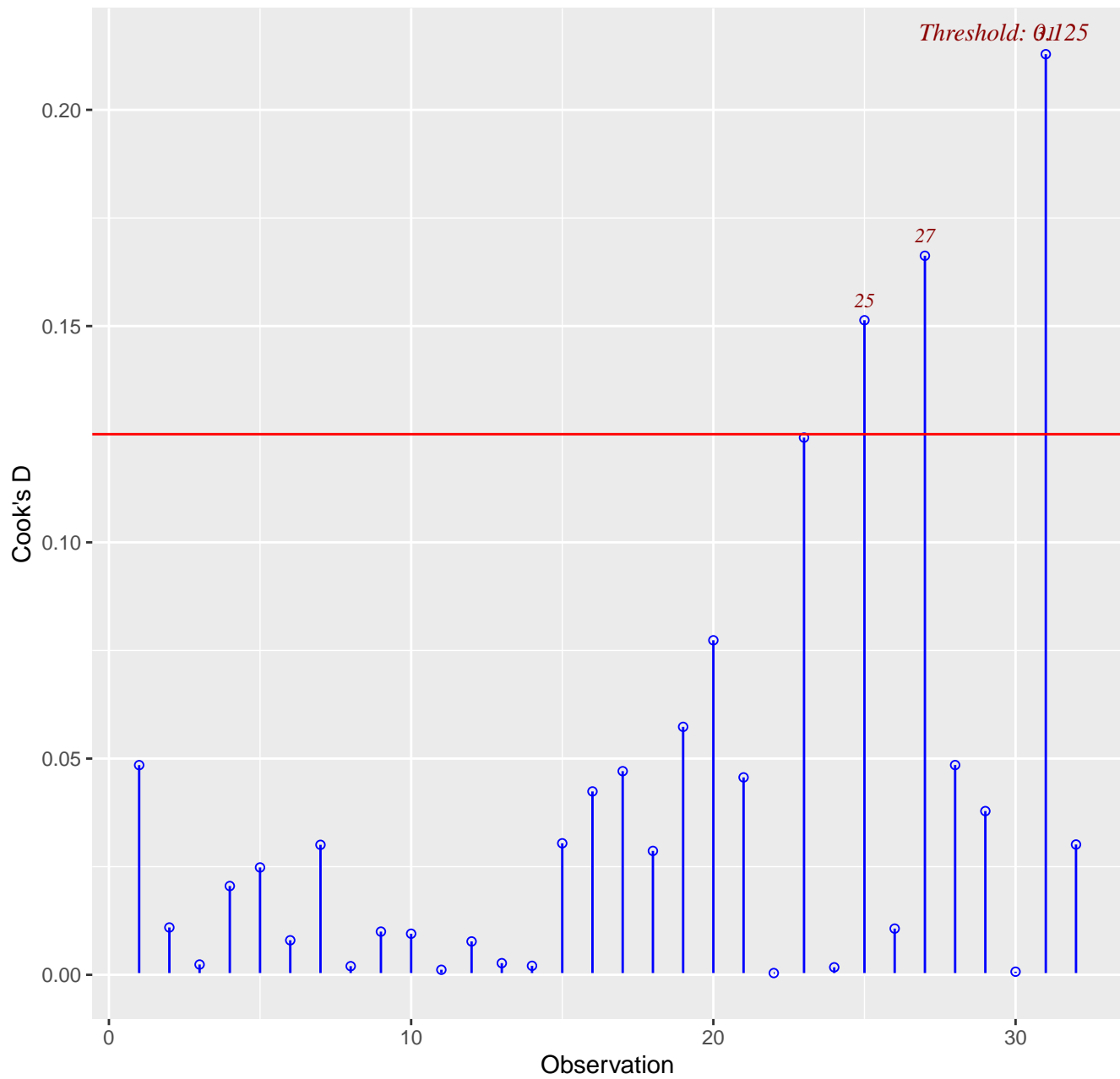
Outlier and Leverage Diagnostics for PRICE



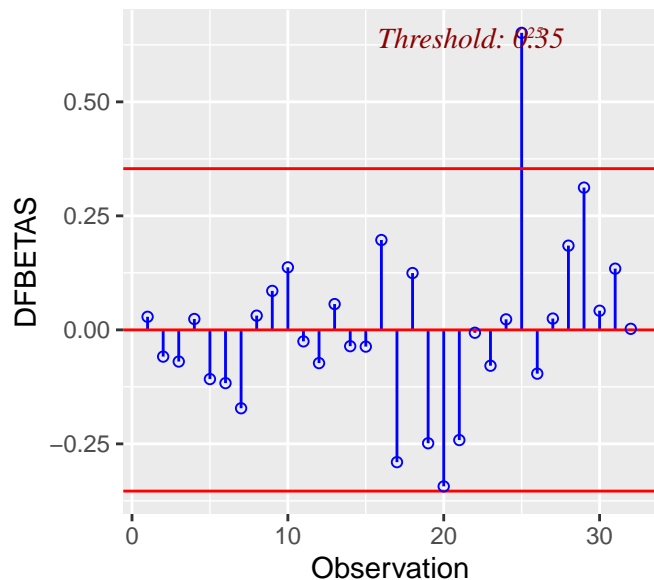
Deleted Studentized Residual vs Predicted Values



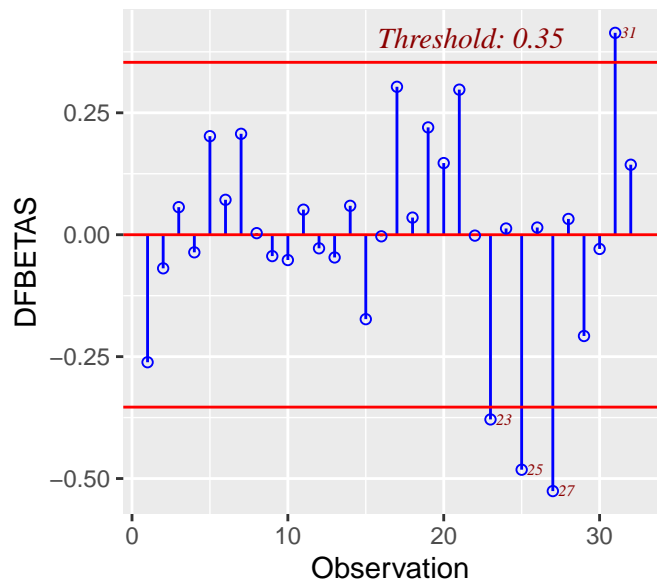
Cook's D Chart



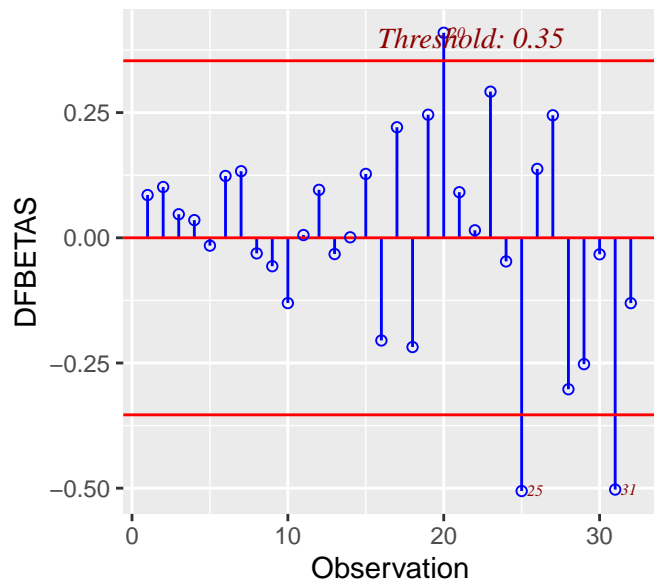
Influence Diagnostics for (Intercept)



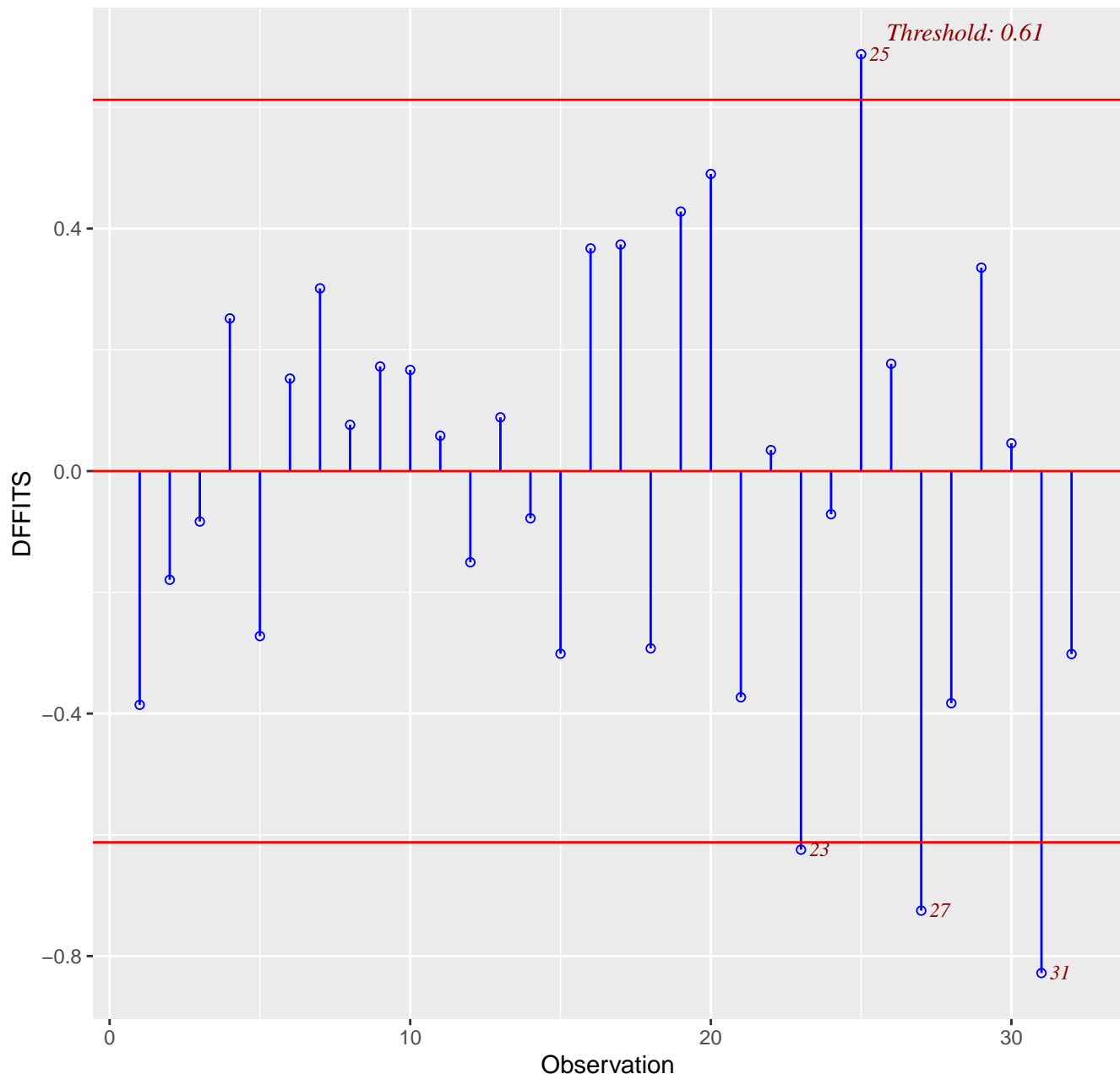
Influence Diagnostics for NUMBIDS



Influence Diagnostics for AGE



Influence Diagnostics for PRICE



ProQuest Ebook Central - Reader

ebookcentral-proquest-com.proxy.libraries.rutgers.edu/lib/rutgers-ebooks/reader.action?docID=918623

ProQuest Ebook Central™

Home Search Bookshelf Settings ? Sign In

Page 113 of 403

MEASURES OF INFLUENCE 113

4.9.3 Hadi's Influence Measure

Hadi (1992) proposed a measure of the influence of the i th observation based on the fact that influential observations are outliers in either the response variable or in the predictors, or both. Accordingly, the influence of the i th observation can be measured by

$$H_i = \frac{p_{ii}}{1 - p_{ii}} + \frac{p + 1}{1 - p_{ii}} \frac{d_i^2}{1 - d_i^2}, \quad i = 1, 2, \dots, n, \quad (4.24)$$

where $d_i = e_i / \sqrt{\text{SSE}}$ is the so-called *normalized residual*. The first term on the right-hand side of (4.24) is the potential function which measures outlyingness in the X -space. The second term is a function of the residual, which measures outlyingness in the response variable. It can be seen that observations will have large values of H_i if they are outliers in the response and/or the predictor variables, that is, if they have large values of r_i , p_{ii} , or both. The measure H_i does not focus on a specific regression result, but it can be thought of as an overall general measure of influence which depicts observations that are influential on at least one regression result.

ProQuest Ebook Central - Reader

ebookcentral-proquest-com.proxy.libraries.rutgers.edu/lib/rutgers-ebooks/reader.action?docID=918623

ProQuest Ebook Central™

Home Search Bookshelf Settings ? Sign In

Page 115 of 403

4.10 THE POTENTIAL-RESIDUAL PLOT

The formula for H_i in (4.24) suggests a simple graph to aid in classifying unusual observations as high-leverage points, outliers, or a combination of both. The graph is called the *potential-residual* (P-R) plot (Hadi, 1992) because it is the scatter plot of

Potential Function $\frac{p_{ii}}{1 - p_{ii}}$ versus Residual Function $\frac{p + 1}{1 - p_{ii}} \frac{d_i^2}{1 - d_i^2}$.

The P-R plot is related to the L-R (*leverage-residual*) plot suggested by Gray (1986) and McCulloch and Meeter (1983). The L-R plot is a scatter plot of p_{ii} versus d_i^2 . For a comparison between the two plots, see Hadi (1992).

As an illustrative example, the P-R plot obtained from fitting model (4.18) is shown in Figure 4.8. Observation 5, which is a high-leverage point, is located by itself in the upper-left corner of the plot. Four outlying observations (3, 7, 4, and 8) are located in the lower-right area of the graph.

It is clear now that some individual data points may be flagged as outliers, leverage points, or influential points. The main usefulness of the leverage and

ProQuest Ebook Central - Reader

ebookcentral-proquest-com.proxy.libraries.rutgers.edu/lib/rutgers-ebooks/reader.action?docID=918623

ProQuest Ebook Central™

Home Search Bookshelf Settings ? Sign In

Page 116 of 403

The P-R plot is related to the L-R (*leverage-residual*) plot suggested by Gray (1986) and McCulloch and Meeter (1983). The L-R plot is a scatter plot of p_{ii} versus d_i^2 . For a comparison between the two plots, see Hadi (1992).

As an illustrative example, the P-R plot obtained from fitting model (4.18) is shown in Figure 4.8. Observation 5, which is a high-leverage point, is located by itself in the upper-left corner of the plot. Four outlying observations (3, 7, 4, and 8) are located in the lower-right area of the graph.

It is clear now that some individual data points may be flagged as outliers, leverage points, or influential points. The main usefulness of the leverage and influence measures is that they give the analyst a complete picture of the role played by different points in the entire fitting process. Any point falling in one of these categories should be carefully examined for accuracy (gross error, transcription error), relevancy (whether it belongs to the data set), and special significance (abnormal condition, unique situation). Outliers should always be scrutinized carefully. Points with high leverage that are not influential do not cause problems. High-leverage points that are influential should be investigated because these points are outlying as far as the predictor variables are concerned and also influence the

ProQuest Ebook Central - Reader

ebookcentral-proquest-com.proxy.libraries.rutgers.edu/lib/rutgers-ebooks/reader.action?docID=918623

ProQuest Ebook Central™

Home Search Bookshelf Settings ? Sign In

Page 116 of 403

116 REGRESSION DIAGNOSTICS: DETECTION OF MODEL VIOLATIONS

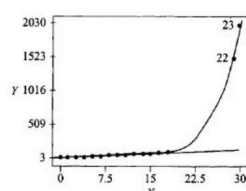
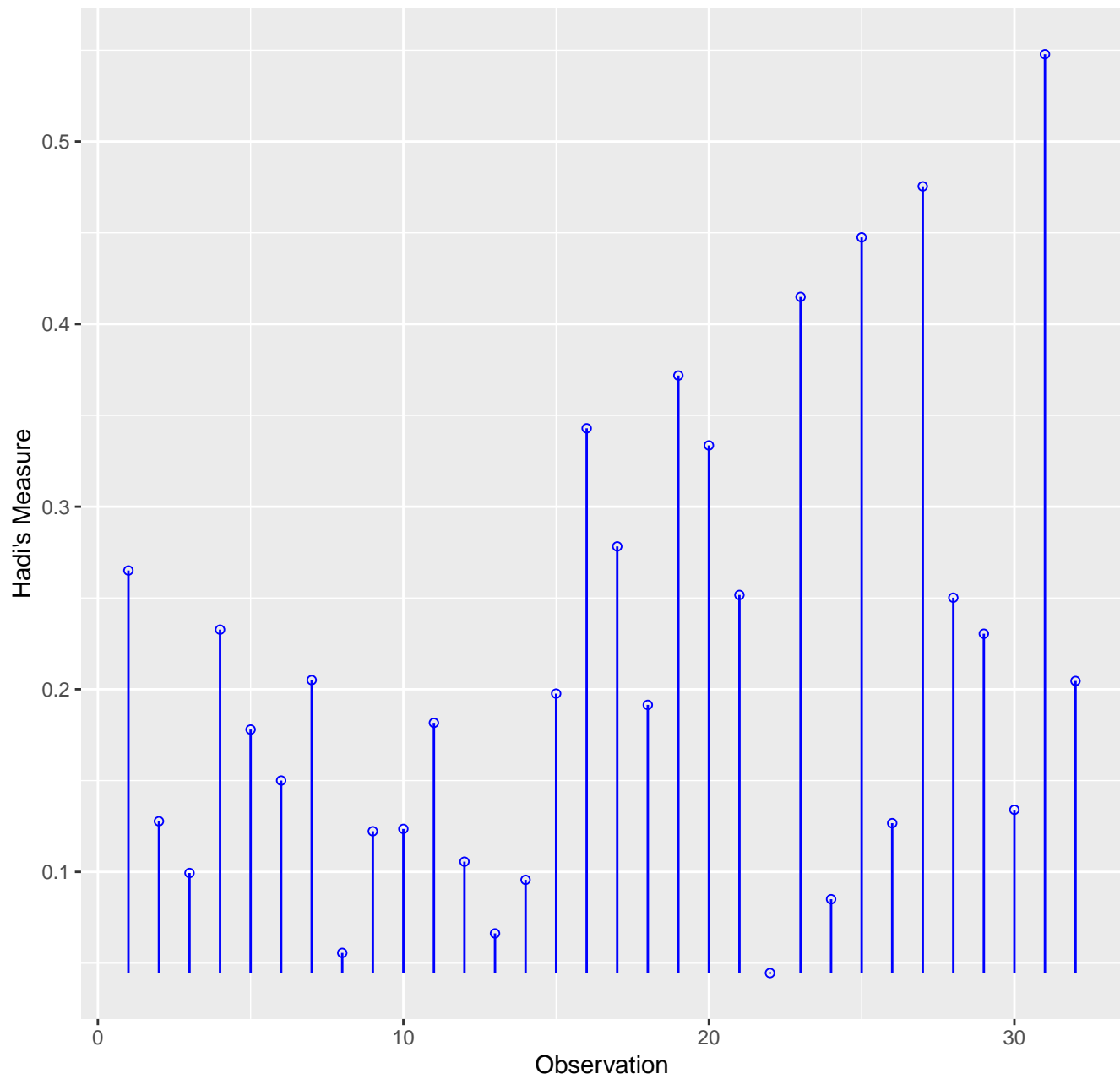


Figure 4.9 Scatter plot of population size, Y , versus time, X . The curve is obtained by fitting an exponential function to the full data. The straight line is the least squares line when observations 22 and 23 are deleted.

fit. To get an idea of the sensitivity of the analysis to these points, the model should be fitted without the offending points and the resulting coefficients examined.

4.11 WHAT TO DO WITH THE OUTLIERS?

Hadi's Influence Measure



Potential-Residual Plot

