Definition 8.1 The **regression residual** is the observed value of the dependent variable minus the predicted value, or

$$\hat{\varepsilon} = y - \hat{y} = y - (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k)$$

Definition 8.3 The **standardized residual**, denoted z_i, for the ith observation is the residual for the observation divided by s, that is,

$$z_i = \hat{\varepsilon}_i / s = (y_i - \hat{y}_i)/s$$

1. Standardized Residuals

$$d_i = \frac{e_i}{\sqrt{MS_{\mathrm{Res}}}}$$

- d_i's have mean zero and variance *approximately* equal to 1
- Large values of d_i ($d_i > 3$) may indicate an outlier

Definition 8.5 The **studentized residual**, denoted z_i^*, for the ith observation is

$$z_i^* = \frac{\hat{\varepsilon}_i}{s\sqrt{1-h_i}} = \frac{(y_i - \hat{y}_i)}{s\sqrt{1-h_i}}$$

s = sqrt(MSE)

2. Studentized Residuals
   - MS_{Res} is only an approximation of the variance of the ith residual.
   - Improve scaling by dividing e_i by the exact standard deviation:

$$Var(e_i) = \sigma^2(1 - h_{ii})$$

Var(e) = Var(Y-Xb) = Var(Y-X(X'X)$^{-1}$X'Y) =Var ((I-X(X'X)$^{-1}$X') Y) =sigma2 (I - X(X'X)$^{-1}$X')

Definition 8.7 A **deleted residual**, denoted d_i, is the difference between the observed response y_i and the predicted value $\hat{y}_{(i)}$ obtained when the data for the ith observation is deleted from the analysis, that is,

$$d_i = y_i - \hat{y}_{(i)}$$

di = e(i)

$$e_{(i)} = \frac{e_i}{1-h_{ii}}$$

PRESS residual

$Var(e_{(i)}) = Var(e_i)/(1\text{-}h_{ii})^2 = \sigma^2(1 - h_{ii})/ (1\text{-}h_{ii})^2 = \sigma^2/ (1\text{-}h_{ii})$

So the standardized PRESS residual is

$e_{(i)} / \sigma/ sqrt(1\text{-}h_{ii}) = e_{(i)}\ sqrt(1\text{-}h_{ii}) / \sigma = e_i/ \sigma\ sqrt(1\text{-}h_{ii})$

If you use s=sqrt(MSE) to estimate σ, then the standardized PRESS residual is just the studentized residual. (internally scaled)

(3) Externally studentized residual also known R-student is given by

$t_i = e_i/sqrt(s^2_{(i)}(1\text{-}h_{ii}))$ where $s^2_{(i)}$ is MSE computed with the ith observation excluded.

$t_i \sim t(n\text{-}p\text{-}1)$ MSE from model with ith observation deleted has (n-1)–p df.

- The **hat matrix** is:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

- The diagonal elements of the hat matrix are given by

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$$

- h_{ii} – standardized measure of the distance of the *i*th observation from the center of the x-space.

Cook’s Distance

$$D_i(X'X, pMS_{\mathrm{Res}}) = D_i = \frac{(\hat{\boldsymbol\beta}_{(i)} - \hat{\boldsymbol\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol\beta}_{(i)} - \hat{\boldsymbol\beta})}{pMS_{\mathrm{Res}}}$$

r_i is studentized residual

$$r_i = e_i/\sqrt{MS_{\mathrm{Res}}(1-h_{ii})} \quad (3)$$

$$= \frac{r_i^2}{p}\frac{Var(\hat{y}_i)}{Var(e_i)} = \frac{r_i^2}{p}\frac{h_{ii}}{(1-h_{ii})}$$

What contributes to D_i:

1. How well the model fits the *i*th observation, y_i
2. How far that point is from the remaining dataset.

Large values of D_i indicate an influential point, usually if $D_i > 1$.

DFBETAS – measures how much the regression coefficient changes in standard deviation units if the ith observation is removed.

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S^2_{(i)} C_{jj}}}$$

C_{jj} is the jth diagonal of $(X'X)^{-1}$

where $\hat{\beta}_{j(i)}$ is an estimate of the jth coefficient when the ith observation is removed.

DFFITS – measures the influence of the ith observation on the fitted value, again in standard deviation units.

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S^2_{(i)} h_{ii}}}$$

COVARIANCE RATIO (measure of model performance)

Information about the overall precision of estimation can be obtained through another statistic, COVRATIO_i

$$COVRATIO_i = \frac{|(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1} S^2_{(i)}|}{|(\mathbf{X}'\mathbf{X})^{-1} MS_{\text{Res}}|} = \frac{(S^2_{(i)})^p}{MS^p_{\text{Res}}}\left(\frac{1}{1-h_{ii}}\right)$$