Deletion Residual

(also called PRESS "Prediction Sum of Squares" residuals)

$$e_{i,-i} = Y_i - \hat{Y}_{i,-i}$$

fitted value for the ith observation based on a model which has been fitted to the data with the i^{th} observation deleted (or excluded)

Surely this involves fitting n models (so we can calculate $e_{i,-i}$ for each $i = 1, 2, \ldots, n$)?

No, as it can be shown that

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

and these deletion residuals have $Var(e_{i,-i}) = \frac{\sigma^2}{1-h_{ii}}$

So, if we standardise the deletion residuals

$$\frac{e_{i,-i} - 0}{\sqrt{\sigma^2/(1-h_{ii})}} \simeq \frac{e_{i,-i}}{\sqrt{s^2/(1-h_{ii})}}$$

$$= \frac{e_i}{(1-h_{ii})\, s\sqrt{\frac{1}{1-h_{ii}}}} = \frac{e_i}{s\sqrt{1-h_{ii}}} = r_i$$

Same standardised (internally Studentised) residuals as before!

But the internally Studentised residuals

$$r_i = \frac{e_i}{s\sqrt{1-h_{ii}}}$$

are called "internally" Studentised as the estimate of σ^2 used is based on a model which uses all the data (including the current or i^{th} observation)

Again, we can derive as estimate of σ^2 that excludes the current observation without to fit the entire model to a new reduced data set — it turns out

$$s_{-i} = \sqrt{\frac{(n-p)s^2 - e_i/(1-h_{ii})}{n-p-1}}$$

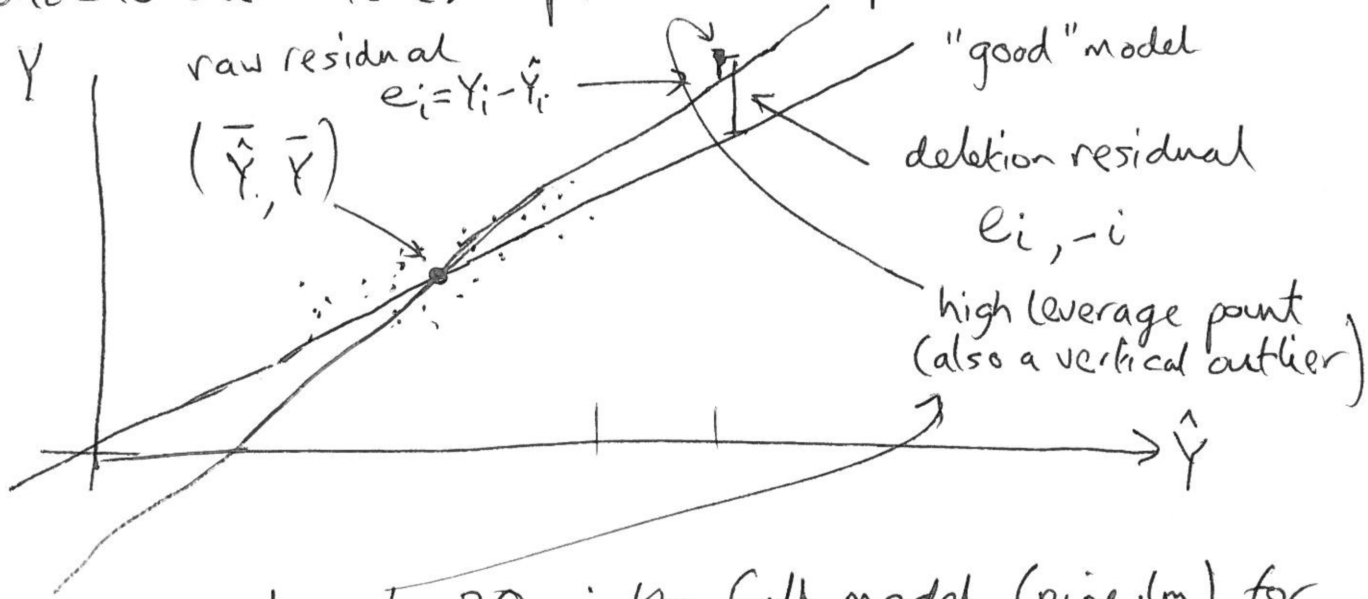
note the new degrees of freedom used here (based on 1 less observation)

This gives an alternative type of standardised residuals; the <u>externally Studentised residuals</u>

$$t_i = \frac{e_i}{s_{-i}\sqrt{1-h_{ii}}} = \frac{e_{i,-i}}{s_{-i}/\sqrt{1-h_{ii}}}$$

this version clearly shows that both the numerator (the deletion residual) and the denominator (a fixed function of the deletion s estimate) come from a model with the i^{th} observation <u>ex</u>cluded, hence the name

Back to the Pine example



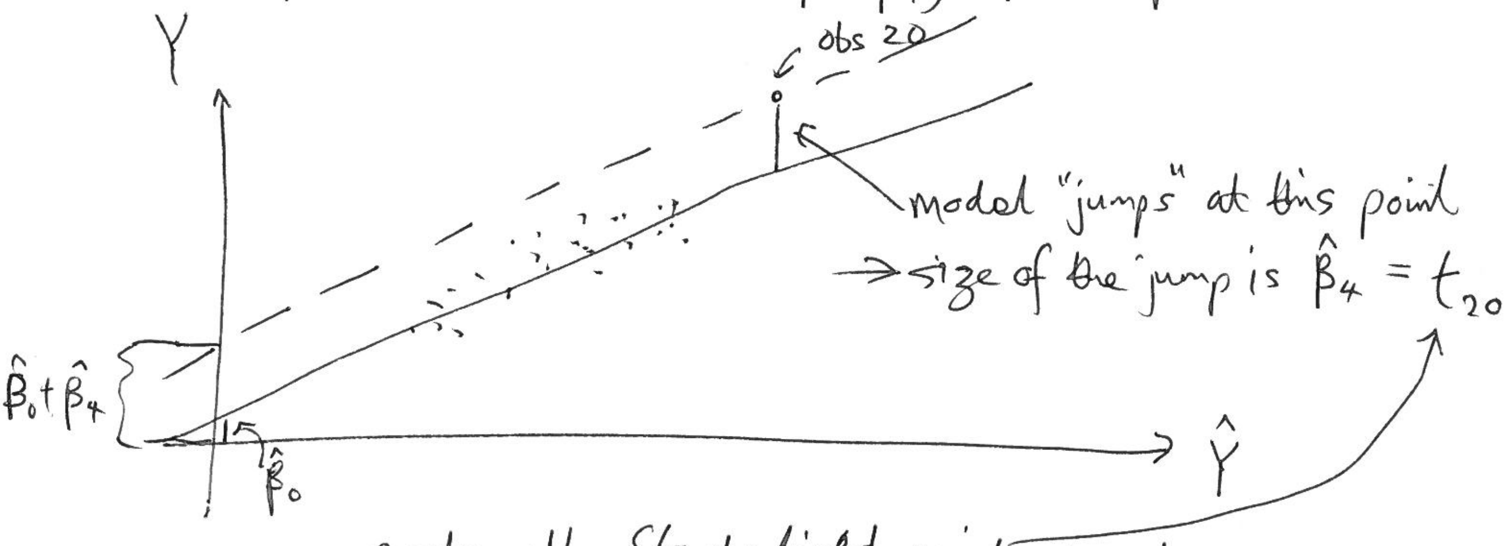
example obs 20 in the full model (pine.lm) for the pine data

indicator variable $I_{20} = \begin{cases} 0 & \text{if } i = 1, 2, \ldots 19 \\ 1 & \text{if } i = 20 \end{cases}$

fitted model $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 I_{20}$

if $I_{20} = 0 \Rightarrow \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$

if $I_{20} = 1 \Rightarrow \hat{Y} = (\hat{\beta}_0 + \hat{\beta}_4) + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$



externally Studentised residual value for observation 20