Residual Diagnostics

Raw residual for the i^{th} observation

$$e_i = Y_i - \hat{Y}_i$$

→ these are estimates of the errors ε_i

i.e. $e_i = \hat{\varepsilon}_i$

Note the assumptions about the errors $Var(\varepsilon) = \sigma^2 I$

but (see earlier) $Var(e) = \sigma^2 (I - H)$

H: hat matrix

So, the standardised (internally Studentised) residuals for the i^{th} observation are:

$$r_i = \frac{e_i - 0}{\sqrt{\sigma^2 (1 - h_{ii})}} \simeq \frac{e_i}{\sqrt{MS_{error}(1 - h_{ii})}}$$

h_{ii}: hat value (leverage) of the i^{th} obs.

$\hat{\sigma}^2 = s^2 = MS_{error}$

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

→ little s or s

As σ^2 is estimated these are approx. distributed as Student's t

Residual Plots (Ian's preferred plots)

(1) Main residual plot

Standardised (internally Studentised) residuals (r_i) against the fitted values ($\hat{Y}_i$)

→ we should check this for every model we fit

→ checks the key assumptions of independence & constant variance

(2) Normal quantile plot (of the standardised residuals)

Default plot (model, which = 2) works fine

→ only bother checking this once plot (1) is okay

→ checks assumption of normality

→ could add 45° line for comparison
abline(0, 1, lty = 2)

(3) Outlier / Influence Plot

My preference is a bar plot of Cook's distances
plot(model, which = 4)

→ only really need if there is some indication that outliers and/or influential points might be a problem on plots (1) and/or (2)

→ can further investigate - check leverage values (bar plot)

→ could also use plot(model, which = 5)
(but ignore the arbitrary cut-offs for Cook's D)

→ also, we could perform a test ....