STAT6038 week 5 lecture 13

Rui Qiu

2017-03-22

blah, blah, blah.

Confidence and Prediction Intervals

For new values of Y given new values of X

$$\hat{Y}|(X = x^*) = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

A 95% **confidence interval** for $E(Y|X = x^*)$ is

$$\hat{Y} \pm t_{\text{error df}}(0.975) \cdot s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}.$$

Note

- when $x^* = 0$, standard error becomes $se(\hat{\beta}_0)$.
- when $x^* = \bar{x}$, standard error becomes $se(\bar{y}) = \frac{s}{\sqrt{n}}$.

A 95% **prediction interval** for $Y|X = x^*$ is

$$\hat{Y} \pm t_{\text{error df}}(0.975) \cdot s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}.$$

Note these are the formulae for SLR, we need to make the usual modifications (switch to matrix notation) for multiple regression.

Modelling process Propose an initial plausible model.

- Is the model appropriate? (Are the underlying assumptions ok?)
  $\rightarrow$ Errors $\overset{iid}{\sim} N(0, \sigma^2)$
  we estimate these Errors using the residuals & produce residual plots.
  (plot(model) in R)

- Is the model adequate? (Does the model have significant explanatory power?)
  Is it a useful model (as per George Box)
  → overall F test from the anova table is a good start here: anova(model) in R.

- If yes to both the above, then we look at the details of the model by try and answer the research question
  → this involves looking at the estimated model coefficients: summary(model) in R.

- Finally, maybe, if everything is good enough, also use predict(model) in R.

- Finally, some overall assessment - is the model just exploratory or can it be sensibly used to make predictions? (i.e., predictive model)