Simple Linear Regression 2

ST552 Lecture 3

Charlotte Wickham

January 8, 2015

Today

- Review SLR model
- Finish's Weds slides
- Review partitioning variation in regression

General Idea: Partitioning the variation

We see variation in the response. We want to attribute that variation to different sources: variation due to the mean varying according to our regression model, and variation due to the random error.

(Sketch)

Partition of variation

$$\text{Total Sum of Squares} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\text{Residual Sum of Squares} = \sum_{i=1}^{n} (y_i - \hat{y})^2$$

Can show

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y})^2 \quad + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$\text{Total SS} = \text{Residual SS} \quad + \text{Regression SS}$$

Many notations:

Total SS = TSS = $SS_{Total} = SS(Total)$

Residual SS = RSS = SSE = $SS_{Res} = SS(Res)$

Regression SS = SSR = $SS_{Reg} = SS(Reg) = SS(Model)$

Degrees of freedom

The degrees of freedom for each sum of squares are also additive

$$\begin{aligned} n - 1 &= n - 2 && +1 \\ \text{Total df} &= \text{Residual d.f.} && +\text{Regression d.f.} \end{aligned}$$

$\text{SS}(.)/\text{d.f.}(.) = \text{Mean sum of squares}(.) = \text{MSS}(.)$

R-squared

R^2 is simply the proportion of variation in the response explained by the model

$$R^2 = \frac{\text{Total SS} - \text{Residual SS}}{\text{Total SS}}$$

In simple linear regression R^2 is the square of the Pearson correlation between x and y.