$$\text{attend}_i = \beta_0 + \beta_1^M \text{monthMAY}_i + \cdots + \beta_6^M \text{monthOCT}_i$$

$\widehat{\text{attend}_i}$ → prediction by the model.

β_0 → 33909.9, β_1^M → −2385, …

$$+ \beta_1^D \text{dayTues}_i + \cdots + \beta_6^D \text{daySun}_i$$

…

$$+ \beta_B \text{BobbleheadYES}_i + \varepsilon_i$$

β_B → 10714.

$\varepsilon_i \sim$ normal $(0, \boxed{6120}^2)$ variance

$\boxed{\text{rse}}$

Least squares.

$$\sum_{i=1}^{81} \Big(\underbrace{\text{attend}_i}_{\text{observed}} - \overline{\text{attend}}\Big)^2$$

TSS

$$= \sum_{i=1}^{81} \Big(\text{attend}_i - \overbrace{\widehat{\text{attend}_i}}^{\text{predicted attend}}\Big)^2$$

RSS ≡ unexplained variance.

$$+ \underbrace{\sum_{i=1}^{81} \Big(\widehat{\text{attend}_i} - \underset{\text{mean}}{\overline{\text{attend}}}\Big)^2}$$

$\overline{\text{attend}} \equiv \overline{\widehat{\text{attend}_i}}$

regression sum of squares ← RegSS = explained variance

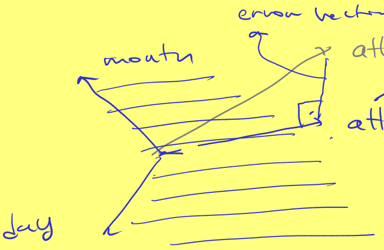
$$\underset{\geq 0}{TSS} = \underset{\geq 0}{RSS} + \underset{\geq 0}{RegSS}$$

$$1 = \frac{RSS}{TSS} + \boxed{\frac{RegSS}{TSS}}$$

$\frac{\text{Explained variance}}{\text{Total variance}} \in [0,1]$

" $R^2 \simeq 54\%$

$$1 - \frac{RegSS}{TSS} = \frac{RSS}{TSS} \to \text{"error"}$$



$$\text{attend} = \begin{bmatrix} \text{attend}_1 \\ \vdots \\ \text{attend}_{81} \end{bmatrix}$$

$$\widehat{\text{attend}} = \begin{bmatrix} \widehat{\text{attend}_1} \\ \widehat{\text{attend}_{81}} \end{bmatrix}$$

orthogonal projection

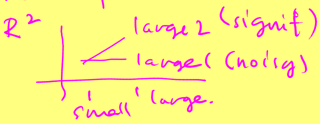
$$R^2 = \frac{\text{Explained variance}}{\text{Total variance}}$$

→ only grows as model gets "complicated" adding more predictor

Increasing R^2 is not enough to justify new predictors.

should increase R^2 by significant amnt. (temp, day-night, sky, opponent)

R^2 does not decrease. noisy/unrelated predictors. mid term scores in GE 461 in 2012



Instead of a graph if you are looking for a single number for comparison of models — then you may use adj R^2

$$= 1 - \frac{RSS/(n-p)}{TSS/(n-1)}$$

Alternative to Adj-R^2,

there are statistically better-motivated model selection criteria

$$\underline{AIC}, \; BIC, \; AIC_c, \ldots$$
$$= -2 \log \text{likelihood} + 2p$$

Out of two models, the better model has

- higher adj R^2
- lower AIC/BIC

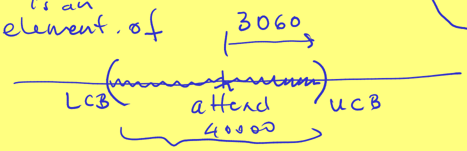
than the poorer model.

Why care about small RSE?

$$\underset{\text{is an element of}}{\overset{\text{unobserved.}}{\text{attend}}} \in \overbrace{\text{attend} \mp \underline{\text{const}}}^{[LCB,\, UCB]} \cdot \underset{6120}{RSE}$$

$\frac{6120}{56000} \simeq 11\%$ error.

bobblehead. $\simeq 11600$



RSE small

narrow conf. int for unobserved true attendance

RSE large.

wide conf. int. for deciding for an investment.

Goodness of fit

- numerical measures: Adj R^2, RSE
- graphical methods.
  - check if regression "line"/model is correct (plot 1)
  - are residuals normally distr. (plot 2)
  - is variance constant (plot 3)
  - are there influential cases (plot "4" or use plot(_, w=4))