Goodness of Fit: R-square

We measure how well the model fits the data

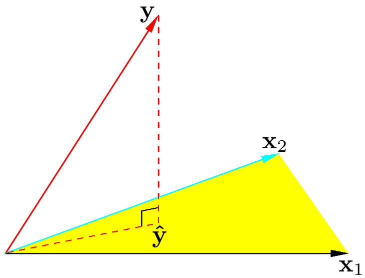
via R^2 (fraction of variance explained)

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{\|\hat{\mathbf{y}} - \bar{y}\|^2}{\|\mathbf{y} - \bar{y}\|^2}$$

$$= \frac{\|\mathbf{y} - \bar{y}\|^2 - \|\mathbf{r}\|^{\mathbf{2}}}{\|\mathbf{y} - \bar{y}\|^2} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

where we use the fact:

$$\|\mathbf{y} - \bar{y}\|^2 = \|\hat{\mathbf{y}} - \bar{y}\|^2 + \|\mathbf{r}\|^{\mathbf{2}}.$$



Goodness of Fit: R-square

We measure how well the model fits the data via R^2 (fraction of variance explained)

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{\|\hat{\mathbf{y}} - \bar{y}\|^2}{\|\mathbf{y} - \bar{y}\|^2}$$

$$= \frac{\|\mathbf{y} - \bar{y}\|^2 - \|\mathbf{r}\|^{\mathbf{2}}}{\|\mathbf{y} - \bar{y}\|^2} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

where we use the fact:

$$\|\mathbf{y} - \bar{y}\|^2 = \|\hat{\mathbf{y}} - \bar{y}\|^2 + \|\mathbf{r}\|^{\mathbf{2}}.$$

$$0 \leq R^2 \leq 1, \quad R^2 = \left[\text{Corr}(\mathbf{y}, \hat{\mathbf{y}})\right]^2.$$

R^2 invariant of any location and/or scale change of Y.

In general, R^2 alone does not tell us much about the effectiveness of the LS model. (Wait till we discuss F-test.)

- A small R^2 does not imply that the LS model is bad.
- Adding a new predictor, even if it is randomly generated and has nothing to do with Y, will decrease RSS and therefore increase R^2.