Statistical Thinking for the 21st Century

Copyright 2020 Russell A. Poldrack

14.1.5 Quantifying goodness of fit of the model

Sometimes it's useful to quantify how well the model fits the data overall, and one way to do this is to ask how much of the variability in the data is accounted for by the model. This is quantified using a value called R^2 (also known as the *coefficient of determination*).

If there is only one x variable, then this is easy to compute by simply squaring the correlation **coefficient**:

$$R^2 = r^2$$

In the case of our study time example, $R^2 = 0.388$, which means that we have accounted for about **39%** of the variance in grades.

More generally we can think of R^2 as **the amount of** variance in the data that is accounted for by the model, which can be computed by breaking the variance into multiple components:

$$SS_{total} = SS_{model} + SS_{error}$$

where SS_{total} is the variance of the data (y) and SS_{model} and SS_{error} are computed as shown earlier in this chapter. Using this, we can then compute the coefficient of determination as:

$$R^2 = \frac{SS_{model}}{SS_{total}} = 1 - \frac{SS_{error}}{SS_{total}}$$

A small value of R^2 tells us that even if the model fit is statistically significant, it may only explain a small amount of information in the data.

practical versus statistical significance!