

Module 12: Statistical inference for multiple linear regression

(The attached PDF file has better formatting.)

Intuition: R^2 VS STANDARD ERROR

Jacob: We have two measures of goodness-of-fit: R^2 and the standard error of the regression. Do they measure the same thing? Is one a function of the other, like the t statistic and the F statistic for the two-variable regression model?

Rachel: The standard error measures the *unexplained variance*. This is the residual sum of squares, or the error sum of squares: what is left over after the regression. The R^2 measures the *percentage* of the total sum of squares that is explained by the regression.

If the total sum of squares is held constant, the R^2 and the standard error measure the same thing. If the total sum of squares differs for two regression equations, the equation with the larger total sum of squares may have the same R^2 but a higher standard error.

Jacob: Why would the total sum of squares be the same for two regression equations? The total sum of squares is the deviation of the Y values from their mean. The Y values are stochastic, so why should the total sum of squares be the same for different regressions?

Rachel: Suppose we seek to explain a phenomenon, such as the scores on an actuarial exam. We consider several explanatory variables, such as the hours the candidate studies, the college grades, the actuarial courses taken, or the years of work experience.

- Given a regression equation, we assume the Y values are stochastic. In theory, the regression is an experiment: we choose values for the independent variable, and we observe the resulting values of the dependent variable.
- In practice, we begin with the observed scores. We hypothesize explanations, and we test each explanatory variable.

If the total sum of squares is fixed before we begin forming regression equations, the R^2 and the standard error measure the same thing. If we start in the opposite direction – first choosing values of the independent variable and then observing the Y values – the total sum of squares is not fixed.

Jacob: If the total sum of squares is not fixed, which is the better measure of the goodness-of-fit, the R^2 or the standard error?

Rachel: If the residuals have a normal distribution with a constant variance, the standard error should not depend on the dispersion of the X values. The standard error is an unbiased estimate of σ^2 , which is assumed to be constant. If the variance is not constant, a wider dispersion of X values may give a higher σ^2 .

