

When we subtract total minus residual error variance ($var(Y) - MSE$), we can call the result “explained error”. It represents the amount of variability in y that is explained away by regressing on x . Then we can compute R^2 as:

$$R^2 = \frac{\text{explained variance}}{\text{total variance}} = \frac{var(Y) - MSE}{var(Y)} \quad (1)$$

When $R^2 \rightarrow 0$ denotes how less information x holds on y , while $R^2 \rightarrow 1$ denotes perfect prediction.

The hypothesis that a parameter equals zero (or any other fixed value) is directly tested by fitting the model that includes the parameter in question and examining its 95% interval. If the interval excludes zero (or the specified fixed value), then the hypothesis is rejected at the 5% level. Testing whether two parameters are equal is equivalent to testing whether their difference equals zero. We do this by including both parameters in the model and then examining the 95% interval for their difference. As with inference for a single parameter, the confidence interval is commonly of more interest than the hypothesis test. For example, if support for the death penalty has decreased by 6% +/- 2% then the magnitude of this estimated difference is probably as important as that the change is statistically significantly different from zero. The hypothesis of whether a parameter is positive is directly assessed via its confidence interval. If both ends of the 95% confidence interval exceed zero, then we are at least 95% sure (under the assumptions of the model) that the parameter is positive. Testing whether one parameter is greater than the other is equivalent to examining the confidence interval for their difference and testing for whether it entirely positive (gelman, reg, 20)

More precisely, one can account for the uncertainty in the standard errors themselves by using the t distribution with degrees of freedom set to the number of data points minus the number of estimated coefficients, but the normal approximation works fine when the degrees of freedom are more than 30 or so (gelman, reg, 40).

As a byproduct of the least squares estimation of β , the residuals r_i will be uncorrelated with all the predictors in the model. If the model includes a constant term, then the residuals must be uncorrelated with a constant, which means they must have mean 0. This is a byproduct of how the model is estimated; it is not a regression assumption.

Standard errors for β are proportional to σ