Print

Adjusted R-Square

Suppose an earlier linear regression analysis identified Lot_Area as an important variable for explaining the values of SalePrice. Remember, R-square measures the proportion of variability in the response variable that is explained by the independent variables in the analysis. When R-square is close to 0, the independent variables don't explain much variability in the data. When it's close to 1, the independent variables explain a relatively large proportion of variability.

In this case, the R-square value, 0.0642, means the Lot_Area explains approximately 6.4% of the variation in SalePrice. This percentage is low, so you might decide to add another variable to the model. In fact, because you know that R-square always increases or at least stays the same as you include more terms in the model, you might be tempted to add as many terms as possible. However, choosing the "best" model isn't as simple as making the R-square as large as possible. To help you decide which model is better, you can compare the adjusted R-square values for the models. The adjusted R square is like R-square, but it considers the number of terms in the model in addition to model fit.

Let's look at the formula for adjusted R-square: $R^2_{ADJ} = 1 - \frac{(n-i)(1-R^2)}{n-p}$

n is the number of observations that are used to fit the model.

i equals 1 if there is an intercept, or zero if there is no intercept.

p is the number of parameters in the model.

You can think of the adjusted R-square as a penalized version of R-square. The penalty increases with each parameter that is added to the model. The adjusted R-square increases only if the additional terms improve the model enough to warrant increasing its complexity.

Close