Assessing the Fit of Regression Models

Karen Grace-Martin

1 Introduction

A well-fitting regression model results in predicted values close to the observed data values. The mean model, which uses the mean for every predicted value, generally would be used if there were no informative predictor variables. The fit of a proposed regression model should therefore be better than the fit of the mean model.

Three statistics are used in Ordinary Least Squares (OLS) regression to evaluate model fit: R-squared, the overall F test, and the Root Mean Square Error (RMSE). All three are based on two sums of squares: Sum of Squares Total (SST) and Sum of Squares Error (SSE). SST measures how far the data are from the mean and SSE measures how far the data are from the model's predicted values. Different combinations of these two values provide different information about how the regression model compares to the mean model.

2 R-squared and Adjusted R-squared

The difference between SST and SSE represents the variation in the response variable that can be "accounted for" by the regression model. Since SSE is the variation in the response left over after fitting the regression model, the difference between SST and SSE is the improvement in prediction given by the regression model compared to the mean model.

Dividing this difference by SST gives R-squared, the proportion of total variance (SST) accounted for by the regression model. R-squared has the useful property that its scale is intuitive: it ranges from zero to one, with zero indicating that the proposed model does not improve prediction over the mean model and one indicating perfect prediction. Improvement in the regression model results in proportional increases in R-squared.

One pitfall of R-squared is that it can only increase as predictors are added to the regression model. This increase is artificial when predictors are not actually improving the model's fit. To remedy this, a related statistic, Adjusted R-squared, incorporates the model's degrees of freedom. Adjusted R- squared will decrease as predictors are added if the increase in model fit does not make up for the loss of degrees of freedom. Likewise, it will increase as predictors are added if the increase in model fit is worthwhile. Adjusted R-squared should always be used with models with more than one predictor variable. It is interpreted as the proportion of total variance that is explained by the model.

There are situations in which a high R-squared is not necessary or relevant. When the interest is in the relationship between variables, not in prediction, the R-square is less important. An example is a study on how religiosity affects health outcomes. A good result is a reliable

relationship between religiosity and health. No one would expect that religion explains a high percentage of the variation in health, as health is affected by many other factors. Even if the model accounts for other variables known to affect health, such as income and age, an R-squared in the range of 0.10 to 0.15 is reasonable.

3 The F test

The F test evaluates the null hypothesis that all regression coefficients are equal to zero versus the alternative that at least one does not. An equivalent null hypothesis is that R-squared equals zero. A significant F test indicates that the observed R-squared is reliable, and is not a spurious result of oddities in the data set. Thus, the F test determines whether the proposed relationship between the response variable and the set of predictors is statistically reliable, and can be useful when the research objective is either prediction or explanation.

4 RMSE

The RMSE is the square root of the variance of the residuals. It indicates the absolute fit of the model to the data—how close the observed data points are to the model's predicted values. Whereas R-squared is a relative measure of fit, RMSE is an absolute measure of fit. As the square root of a variance, RMSE can be interpreted as the standard deviation of the unexplained variance, and has the useful property of being in the same units as the response variable. Lower values of RMSE indicate better fit. RMSE is a good measure of how accurately the model predicts the response, and is the most important criterion for fit if the main purpose of the model is prediction.

The best measure of model fit depends on the researcher's objectives, and more than one are often useful. The above statistics were described for the case of ordinary least squares regression. Other regression models, such as mixed or generalized linear models, have alternative statistics or diagnostics for assessing model fit.

Created May 2005. Last updated April 2022.