Print

Lack-of-Fit Test

The test for lack of fit compares the variation around the model with “pure” variation within replicated observations. In particular, if there are n_i replicated observations $Y_{i1},\ldots,Y_{in_i}$ of the response all at the same values x_i of the regressors, then you can predict the true response at x_i either by using the predicted value Y_i based on the model or by using the mean of the replicated values. The test for lack of fit decomposes the residual error into a component due to the variation of the replications around their mean value (the “pure” error) and a component due to the variation of the mean values around the model prediction (the “bias” error).

$$\sum_i \sum_{j=1}^{n_i} \left(Y_{ij} - \hat{Y}_i\right)^2 = \sum_i \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_i\right)^2 + \sum_i n_i \left(\overline{Y}_i - \hat{Y}_i\right)^2$$

If the model is adequate, then both components estimate the nominal level of error. However, if the bias component of error is much larger than the pure error, then this constitutes evidence that there is significant lack of fit.

Note: A significant result for the lack-of-fit test indicates that the specified model is inadequate, so if this is a problem, you might want to refine the model.

Close