Backward Stepwise Regression

BACKWARD STEPWISE REGRESSION is a stepwise regression approach that begins with a full (saturated) model and at each step gradually eliminates variables from the regression model to find a reduced model that best explains the data. Also known as **Backward Elimination** regression.

The stepwise approach is useful because it reduces the number of predictors, reducing the multicollinearity problem and it is one of the ways to resolve the overfitting.

How To

- ✓ Run: **STATISTICS->REGRESSION -> BACKWARD STEPWISE REGRESSION...**
- ✓ Select the **DEPENDENT** variable (**RESPONSE**) and **INDEPENDENT** variables (**PREDICTORS**).
- ✓ **REMOVE IF ALPHA >** option defines the Alpha-to-Remove value. At each step it is used to select candidate variables for elimination – variables, whose partial F p-value is greater or equal to the alpha-to-remove. The default value is 0.10.
- ✓ Select the **SHOW CORRELATIONS** option to include the correlation coefficients matrix to the report.
- ✓ Select the **SHOW DESCRIPTIVE STATISTICS** option to include the mean, variance and standard deviation of each term to the report.
- ✓ Select the **SHOW RESULTS FOR EACH STEP** option to show the regression model and summary statistics for each step.

Results

The report shows regression statistics for the final regression model. If the **SHOW RESULTS FOR EACH STEP** option is selected, the regression model, fit statistics and partial correlations are displayed at each removal step. Correlation coefficients matrix and descriptive statistics for predictors are displayed if the corresponding options are selected.

The command removes predictors from the model in a stepwise manner. It starts from the full model with all variables added, at each step the predictor with the largest p-value (that is over the alpha-to-remove) is being eliminated. When all remaining variables meet the criterion to stay in the model, the backward elimination process stops.

R^2 (COEFFICIENT OF DETERMINATION, R-SQUARED) - is the square of the sample correlation coefficient between the **PREDICTORS** (independent variables) and **RESPONSE** (dependent variable). In general, R^2 is a percentage of response variable variation that is explained by its relationship with one or more predictor variables. In simple words R^2 indicates the accuracy of the prediction. The larger R^2 is, the more the total variation of **RESPONSE** is reduced by introducing the predictor variable. The definition of the R^2 is

$$R^2 \equiv 1 - \frac{SS_{error}}{SS_{total}}.$$

ADJUSTED R2 (ADJUSTED R-SQUARED) - is a modification of R2 that adjusts for the number of explanatory terms in a model. While R^2 increases when extra explanatory variables are added to the model, the adjusted R2 increases only if the added term is a relevant one. It could be useful for comparing the models with different numbers of predictors. Adjusted R^2 is computed using the formula $Adjusted\ R^2 = 1 - \frac{SS_{error}}{SS_{total}} \frac{df_{total}}{df_{error}} = 1 - (1 - R^2) \frac{N-1}{N-k-1}$, where *k* is the number of predictors.

S – the estimated standard deviation of the error in the model.

MS (MEAN SQUARE) - the estimate of the variation accounted for by this term.

$$MS = SS/DF$$

F - the F-test value for the model.

P-LEVEL - the significance level of the F-test. A value less than α (0.05) shows that the model estimated by the regression procedure is significant.

VIF – variance inflation factor, measures the inflation in the variances of the parameter estimates due to collinearities among the predictors. It is used to detect multicollinearity problems.

TOL - the tolerance value for the parameter estimates, it is defined as *TOL = 1 / VIF*.

Partial Correlations are correlations between each predictor and the outcome variable excluding the effect of other variables.

References

[HRA] Hocking, R. R. (1976) "The Analysis and Selection of Variables in Linear Regression," *Biometrics, 32*

[NWK] Neter, J., Wasserman, W. and Kutner, M. H. (1996). Applied Linear Statistical Models, Irwin, Chicago.