09 December 2017
15:49

BACKWARD ELIMINATION

Step1: Start with Full Model

Step2: Compute partial F-Statistic for each regressor in the presence of other regressors in the model.

Step3: The regressor with smallest partial F_{calc} is removed from the model

If the F_{calc} value is less than F_{Tab} $F(0.05, 1, Erdf)$

Step4: Build the new model by removing the regressor which we found in step2 & step3

Backward elimination procedure will be stopped when we see

$F_{calc} > F_{out}$

Let us work on example

Step1: Let us run all variables and fit the model i,e

$$\hat{y} = 62.4 + 1.55X_1 + 0.510X_2 + 0.102X_3 - 0.144X_4$$

step2: Now we will check each of the variable contribution in the presence of other remaining variables

we need to calculate $F_{1/234}$, $F_{2/134}$, $F_{3/124}$, $F_{4/123}$

(1) $$F_{cal}(1/234) = \frac{SS_{Reg}(Full) - SS_{Reg}(2,3,4)}{MS_{Res}(Full)} = \frac{2667.90 - 2641.9}{5.98} = 4.33$$

(2) $$F_{cal}(2/134) = \frac{SS_{Reg}(Full) - SS_{Reg}(1,3,4)}{MS_{Res}(Full)} = \frac{2667.90 - 2664.9}{5.98} = 0.501$$

(3) $$F_{cal}(3/124) = \frac{SS_{Reg}(Full) - SS_{Reg}(1,2,4)}{MS_{Res}(Full)} = \frac{2667.90 - 2667.7}{5.98} = 0.018$$

(4) $$F_{cal}(4/123) = \frac{SS_{Reg}(Full) - SS_{Reg}(1,2,3)}{MS_{Res}(Full)} = \frac{2667.90 - 2667.65}{5.98} = 0.041$$

Step3: From all of the above, the F_{cal} value which is having should be eliminated from the existing model i.e $F_{cal} = 0.018$

Let us calculate F_{Tab} as well i.e, $F_{Tab}(1, 8) = 5.31$

$\therefore$ $F_{cal} < F_{Tab} \Rightarrow 0.018 < 5.31 \Rightarrow$ Remove X_3 from the Full model.

$\therefore$ we have only X_1, X_2, X_4

Step4: So, we have the full model as $\hat{y} = 71.64 + 1.452X_1 + 0.416X_2 - 0.237X_4$

Now again check each variable F_{cal} in presence of other two variables

i.e. $F_{1/24}$ $F_{2/14}$ $F_{4/12}$

Now again check each variable F_{cal} in presence of other two variables
i.e, $F_{1/24}$, $F_{2/14}$, $F_{4/12}$

$$F_{1/24} = \frac{SS_{Reg}(Full) - SS_{Reg}(2,4)}{MS_{Res}(Full)} = \frac{2667.79 - 1846.88}{5.33} = 154.01$$

$$F_{2/14} = \frac{2667.79 - 2641}{5.33} = 5.02$$

$$F_{4/12} = \frac{2667.79 - 2657.86}{5.33} = 1.86$$

$F_{Tab}(0.05, 1, 9) = 5.12$, among all these values which is every less comparing with F_{Tab} i.e going to be reject. i.e, $F_{4/12}(1.86) < F_{Tab}(5.12)$

$\Rightarrow$ Remove x_4 from the Full model.

Step5: Now we have only $x_1, x_2 \Rightarrow \hat{y} = 52.58 + 1.468x_1 + 0.6623x_2$

Now $F(x_1/x_2)$, $F(x_2/x_1)$

$$F(x_1/x_2) = \frac{SS_{Reg}(Full) - SS_{Reg}(x_2)}{MS_{Res}(Full)}$$

$$= \frac{2657.86 - 1809.4}{5.79} = 146.5$$

$$F(x_2/x_1) = \frac{2657.86 - 1450.1}{5.79} = 208.5$$

$F_{Tab}(0.05, 1, 10) = 4.96$

So, Both variables values are greater than F_{Tab}. So, we cannot remove neither of the variables so the procedure terminates at here. So, $x_1 \& x_2$ are final variables

$\therefore$ $\boxed{\hat{y} = 52.58 + 1.468x_1 + 0.6623x_2}$

NOTE: Forward Selection gives $x_1 \& x_4$

Backward elimination gives $x_1 \& x_2$

So, Both are given different conclusions, so No worries.