

Nested model F test (revisited)

Candidate models \swarrow log(survival)size $p=4$

$$Y = \underbrace{\beta_0 + \beta_1 \text{clot} + \beta_2 \text{prog} + \beta_3 \text{enzyme}}_{\text{base}} + \epsilon \quad (\text{model 13 in big table})$$

size $p=5$

$$Y = \underbrace{\beta_0 + \beta_1 \text{clot} + \beta_2 \text{prog} + \beta_3 \text{enzyme}}_{\text{base}} + \underbrace{\beta_4 \text{prog}^2}_{\text{addition}} + \epsilon \quad (25)$$

size $p=6$

$$Y = \underbrace{\beta_0 + \beta_1 \text{clot} + \beta_2 \text{prog} + \beta_3 \text{enzyme}}_{\text{base}} + \underbrace{\beta_4 \text{prog}^2 + \beta_5 \text{prog:enzyme}}_{\text{addition}} + \epsilon \quad (39)$$

⋮

size $p=8$

$$Y = \text{full model} \\ = \underbrace{\beta_0 + \beta_1 \text{clot} + \beta_2 \text{prog} + \beta_3 \text{enzyme}}_{\text{base}} + \underbrace{\beta_4 \text{liver} + \beta_5 \text{prog}^2 + \beta_6 \text{prog:enzyme} + \beta_7 \text{enzyme}^2}_{\text{addition}} + \epsilon \quad (51)$$

Nested model F-test for (51) vs (13)

$$H_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

$$\equiv \frac{\sigma^2_{\text{addition}}}{\sigma^2_{\text{error}}} = 1$$

vs

$$H_A: \text{not all } \beta_4, \beta_5, \beta_6, \beta_7 = 0 \\ (\text{at least one } \beta_j \neq 0, j=4,5,6,7)$$

$$\equiv \frac{\sigma^2_{\text{addition}}}{\sigma^2_{\text{error}}} > 1$$

Using the cbind approach in R:

$$F = \frac{MS_{\text{addition}}}{MS_{\text{error}}} \sim F_{4, n-p=46}$$

$$= 1.092, \quad p\text{-value} = 0.3717$$

so do not reject H_0 , additions not needed

Even the nested F test is not totally consistent with what happens if we refine models one step (one variable) at a time!

Model Refinement

Sensible systematic approaches to model selection involve changing the model one term at a time.

Forward Selection

1. Start with a base model

This could be a null model or a model that already includes important research & control variables

2. Add the most promising of the optional candidate predictors

This could be the covariate which is most highly correlated with Y or it could be one of a group which is correlated with Y but which is also promising based on the underlying science

3. Look at the last sequential F test for this newly added term and/or at the added variable plot to see if some transformation is required


4. If the test in step 3 is significant, retain this new term in the model; if not, do not retain the term

5. Repeat steps 2, 3, 4 & 5 with the next most promising candidates until all candidates have been examined and included or rejected

Note most of these steps can be automated (by programming in some computer package), however, the underlined bits above require judgement and will probably be omitted in most automated processes.

Backward Elimination

(3)

1. Start with the "full" model which includes all possible candidates
 2. Re-order the model so that the least promising candidates appear towards the end of the model
 3. Select the variable with the smallest sequential F statistic & largest p-value, being careful NOT to select a variable which the research question suggests should be included (same goes for control variables, lower-order terms)
 4. If the variable selected in 3 is not significant then remove (eliminate) it from the model
 5. Repeat steps 2, 3, 4, 5 until all the remaining variables in the model are significant
- 

The underlined points above are the ones that require judgement & therefore are difficult to automate

Stepwise Refinement

We can combine Forward Selection & Backward Elimination in two ways:

Forwards stepwise refinement

1. Start with a base model & use forward selection to select the next candidate to add to the model
2. Each time the model changes in step 1, use backward elimination to check that all the terms in the expanded model are necessary
3. Repeat steps 1 & 2 until there are no more changes to the model

Backwards stepwise refinement

1. ^{Start with} Full model, use Backward elimination to _{remove one variable at a time}
2. Use Forward Selection to check if any of the omitted variables should be added back in
3. Repeat steps 1 & 2 until no more changes