In the following tables columns are:

Complete Model
Reduced Model
SSE_C
SSE_R
k
g
F (test statistic)
$F_{0.05}(k-g, n-k-1)$

We denote the critical value by F_α and check whether $F > F_\alpha$.

Potato Damage Data: ANOVA-F Tests

We compare four models: M_{R_1}, M_{R_2} and M_{R_3} are nested within the complete model M_C.

$$
\begin{aligned}
M_C &: A + B + C + A.B + A.C + B.C + A.B.C \\
M_{R_1} &: A + B + C + A.B \\
M_{R_2} &: A + B + C \\
M_{R_3} &: A + B + A.B
\end{aligned}
$$

COMP.	RED.	SSE_C	SSE_R	k	g	F	F_α
M_C	M_{R_1}	4968.876	5093.746	7	4	0.561	2.76
M_{R_1}	M_{R_2}	5093.746	7183.674	4	3	28.721	3.92
M_{R_1}	M_{R_3}	5093.746	6319.640	4	3	16.846	3.92

Note: The quoted F_α values are approximate as the textbook does not tabulate all Fisher-F distributions. We take $\alpha = 0.05$

Conclusions

Taking the comparisons in order:

1. M_C *vs* M_{R_1} : $F < F_\alpha$. Therefore the result is **not significant**: Model M_{R_1} **is an adequate simplification** of Model M_C, and we choose M_{R_1} over M_C.

   The model M_{R_1} now becomes the complete model.

2. M_{R_1} *vs* M_{R_2} : $F > F_\alpha$. Therefore the result **is significant**: Model M_{R_2} **is not an adequate simplification** of Model M_{R_1}

3. M_{R_1} *vs* M_{R_3} : $F > F_\alpha$. Therefore the result **is significant**: Model M_{R_3} **is not an adequate simplification** of Model M_{R_1}

Thus the final model is

$$A + B + C + A.B$$

i.e. all main effects, plus the interaction between potato variety and acclimatization routine.

We cannot simplify this model further without significant loss in terms of goodness of fit.

Note: $R^2 = 0.631$ and Adjusted $R^2 = 0.610$, so we have a reasonable fit.

Task Distraction Data

Example: Task Distraction Data.

In an experimental study, the number of errors made in performing a specified task was recorded. The experiment investigated the influence of various predictors on the numbers of errors made.

There are two factor predictors (A, B) and one continuous covariate (X).

We have a balanced design with 15 people (replicates) in each factor-level subgroup.

Example: Task Distraction Data.

A	Group	1 : Non-smoker
		2 : Delayed smoker
		3 : Active smoker
B	Task	1 : Pattern Recognition
		2 : Cognitive Task
		3 : Driving Simulation
X	Distraction Level	

We compare four models with the **complete** model

Complete Model : $A * B * X$

$$A + B + X + A.B + A.X + B.X + A.B.X$$

Number of parameters

Term	Parameters		
A	$(a-1)$	$=3-1$	2
B	$(b-1)$	$=3-1$	2
X	(1)		1
$A.B$	$(a-1)(b-1)$	$=2\times 2$	4
$A.X$	$(a-1)(1)$	$=2\times 1$	2
$B.X$	$(b-1)(1)$	$=2\times 1$	2
$A.B.X$	$(a-1)(b-1)(c-1)$	$=2\times 2\times 1$	4
Total			17

For illustration we consider the following sequence of models:

- Reduced Model 1: M_{R_1}

$$A + B + X + A.X + B.X$$

- Reduced Model 2: M_{R_2}

$$A + B + X + B.X$$

- Reduced Model 3: M_{R_3}

$$B + X + B.X$$

- Reduced Model 4: M_{R_4}

$$B + X$$

Task Distraction Data: ANOVA-F Tests

$$
\begin{aligned}
M_C &: A + B + X + A.B + A.X + B.X + A.B.X \\
M_{R_1} &: A + B + X + A.X + B.X \\
M_{R_2} &: A + B + X + B.X \\
M_{R_3} &: B + X + B.X \\
M_{R_4} &: B + X
\end{aligned}
$$

COMP.	RED.	SSE_C	SSE_R	k	g	F	F_α
M_C	M_{R_1}	5660.010	7627.479	17	9	5.084	2.02
M_{R_1}	M_{R_2}	7627.479	7971.274	9	7	2.817	3.07
M_{R_2}	M_{R_3}	7971.274	8404.654	7	5	3.452	3.07
M_{R_3}	M_{R_4}	8404.654	11154.715	5	3	21.105	3.07

Conclusions

Taking the comparisons in order:

1. M_C *vs* M_{R_1} : $F > F_\alpha$. Therefore the result is **significant**: Model M_{R_1} **is not an adequate simplification** of Model M_C

2. M_{R_1} *vs* M_{R_2} : $F < F_\alpha$. Therefore the result **is not significant**: Model M_{R_2} **is an adequate simplification** of Model M_{R_1}

3. M_{R_2} *vs* M_{R_3} : $F > F_\alpha$. Therefore the result **is significant**: Model M_{R_3} **is not an adequate simplification** of Model M_{R_2}

4. M_{R_3} *vs* M_{R_4} : $F > F_\alpha$. Therefore the result **is significant**: Model M_{R_4} **is not an adequate simplification** of Model M_{R_3}

Follow-up Analysis

In a follow up analysis (see Handout), it transpires that the model

$$A + B + X + A.B + A.X + B.X$$

ie selected.

Note: $R^2 = 0.863$ and Adjusted $R^2 = 0.831$, so we have a good fit.

Note: we must take great care with the sequence of models.

Stepwise Selection in SPSS

It is possible to carry out

- Forward
- Backward
- Stepwise

model selection in SPSS using the *Linear Regression* pulldown menu, and the *Method* pulldown list.

SPSS Screen for Stepwise Selection