

10

Model selection I: principles of model choice and designed experiments

10.2 Three principles of model choice

Model choice in the polynomial problem

In the early sections of this chapter, a polynomial model was built up by fitting the three models $Y = X$, $Y = X + X^*X$ (or in Minitab shorthand $Y = X|X$) and $Y = X + X^*X + X^*X^*X$ (or $Y = X|X|X$). Choices can be made between these models on the basis of economy of variables, multiplicity of p -values and considerations of marginality. However, because Minitab uses the adjusted sums of squares by default to construct the ANOVA tables, it was necessary to fit the three separate models, because for each of these models only the final p -value is valid. If fitting models for which such considerations of marginality come into play, it is more efficient and informative to request ANOVA tables based on sequential sums of squares, as is shown below for the *simple polynomial* dataset.

MINITAB COMMANDS FOR BOX 10.5 ANOVA tables using sequential rather than adjusted sums of squares

Commands

```
glm Y1 = X1|X1|X1;
covariates X1;
ssquares 1;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model

Y1 → Response
 X1|X1|X1 → Model

Covariates...

X1 → Covariates

Results...

⊙ In addition, coefficients for all terms

Options...

⊙ Sequential (Type I)

MINITAB OUTPUT FOR BOX 10.5 ANOVA tables using sequential rather than adjusted sums of squares						
General Linear Model: Y1 versus						
Factor	Type	Levels	Values			
Analysis of Variance for Y1, using Sequential SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
X1	1	6663021	2505	6663021	1095.90	0.000
X1*X1	1	256148	4763	256148	42.13	0.000
X1*X1*X1	1	720	720	720	0.12	0.732
Error	76	462078	462078	6080		
Total	79	7381967				
Term	Coef	SE Coef	T	P		
Constant	-15.75	33.92	-0.46	0.644		
X1	6.179	9.625	0.64	0.523		
X1*X1	0.6169	0.6971	0.89	0.379		
X1*X1*X1	0.00500	0.01452	0.34	0.732		
Unusual Observations for YIELD						
Obs	Y1	Fit	SE Fit	Residual	St Resid	
10	319.630	135.467	16.117	184.163	2.41R	
12	405.900	192.144	15.706	213.756	2.80R	
29	904.640	922.875	33.473	-18.235	-0.26 X	
30	929.060	988.514	43.168	-59.454	-0.92 X	
50	973.210	822.857	22.029	150.353	2.01R	
54	323.860	505.919	15.626	-182.059	-2.38R	
57	436.840	639.678	14.448	-202.838	-2.65R	
60	584.290	762.782	17.530	-178.492	-2.35R	
R denotes an observation with a large standardized residual						
X denotes an observation whose X value gives it large influence						

10.4 Orthogonal and near orthogonal designed experiments

Model choice with orthogonal experiments

Three way interactions between categorical variables follow the same shorthand as between continuous variables in the previous example.

MINITAB COMMANDS FOR BOX 10.6 Analysis of a factorial experiment

Commands `glm ROT = BAC|TEMP|OXYGEN;`
 `brief 1.`

Menu route `Stat > ANOVA > General Linear Model`
 `ROT → Response`
 `BAC|TEMP|OXYGEN → Model`

Results...
 Ⓐ Analysis of variance table

MINITAB OUTPUT FOR BOX 10.6 Analysis of a factorial experiment

General Linear Model: ROT versus BAC, TEMP, OXYGEN

Factor	Type	Levels	Values
BAC	fixed	3	1 2 3
TEMP	fixed	2	1 2
OXYGEN	fixed	3	1 2 3

Analysis of Variance for ROT, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
BAC	2	651.81	651.81	325.91	13.91	0.000
TEMP	1	848.07	848.07	848.07	36.20	0.000
OXYGEN	2	97.81	97.81	48.91	2.09	0.139
BAC*TEMP	2	152.93	152.93	76.46	3.26	0.050
BAC*OXYGEN	4	30.07	30.07	7.52	0.32	0.862
TEMP*OXYGEN	2	1.59	1.59	0.80	0.03	0.967
BAC*TEMP*OXYGEN	4	81.41	81.41	20.35	0.87	0.492
Error	36	843.33	843.33	23.43		
Total	53	2707.04				

The same commands are required when orthogonality is lost, but the order of the explanatory variables in the model formulae is varied (see main text).

10.5 Looking for trends across levels of a categorical variable

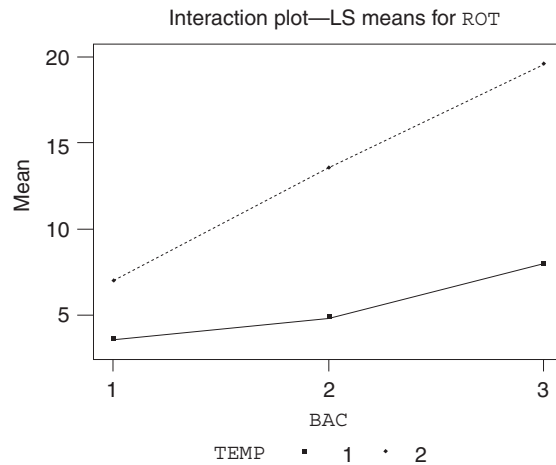
First of all we request two extra pieces of output from the original orthogonal analysis: a table of means for the significant two-way interaction (BAC*TEMP—see Box 10.6), and an interaction diagram. This is done with the following commands. Note that in the command language, plotting the interaction diagram involves calling up a macro, within which Minitab is fussy about having the

variable names within ‘blips’. This is one of the exercises that is more easily executed via the menus!

MINITAB COMMANDS FOR BOX 10.9 AND FIG. 10.4 Means and interaction diagram for ROT by TEMP and BAC	
Commands	<pre> glm ROT = BAC TEMP OXYGEN; brief 1; means BAC*TEMP; smeans c4000; interact TEMP BAC. %gfint 'TEMP' 'BAC'; responses 'ROT'; lsmeans c4000. </pre>
Menu route	<pre> Stat > ANOVA > General Linear Model ROT → Response BAC TEMP OXYGEN → Model Results... ⊙ Analysis of variance table BAC*TEMP → Display least squares means corresponding to the terms Factor Plots... TEMP BAC → Interactions Plot, Factors: </pre>

The following output would then be produced in addition to the ANOVA table:

MINITAB OUTPUT FOR TABLE 10.9			
Least Squares Means for ROT			
BAC* TEMP		Mean	SE Mean
1 1		3.556	1.613
1 2		7.000	1.613
2 1		4.778	1.613
2 2		13.556	1.613
3 1		8.000	1.613
3 2		19.556	1.613



Minitab output for Fig. 10.4 Interaction diagram for the potato rot experiment.

Examination of this interaction diagram is what suggests to us that there is a trend between ROT and BAC, and that this may even be curvilinear at TEMP level 1. This then leads on to the analysis in which BAC is declared as continuous and fitted as a polynomial, interacting with TEMP.

MINITAB COMMANDS FOR BOX 10.10 Reanalysing the potato rot experiment, looking for trends

Commands	<pre>glm ROT = OXYGEN + TEMP BAC BAC; covariate BAC. brief 1;</pre>
Menu route	<pre>Stat > ANOVA > General Linear Model ROT → Response OXYGEN + TEMP BAC BAC → Model Covariates... BAC → Covariates Results... Ⓞ Analysis of variance table</pre>

MINITAB OUTPUT FOR BOX 10.10 Reanalysing the potato rot experiment, looking for trends

General Linear Model: ROT versus OXYGEN, TEMP

Factor	Type	Levels	Values
OXYGEN	fixed	3	1 2 3
TEMP	fixed	2	1 2

Analysis of Variance for ROT, using Adjusted SS for Tests

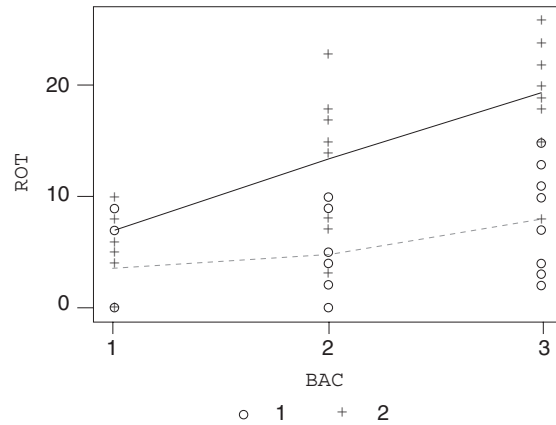
Source	DF	Seq SS	Adj SS	Adj MS	F	P
OXYGEN	2	97.81	97.81	48.91	2.35	0.106
TEMP	1	848.07	4.68	4.68	0.23	0.637
BAC	1	650.25	5.78	5.78	0.28	0.600
TEMP*BAC	1	148.03	15.43	15.43	0.74	0.393
BAC*BAC	1	1.56	1.56	1.56	0.08	0.785
TEMP*BAC*BAC	1	4.90	4.90	4.90	0.24	0.630
Error	46	956.41	956.41	20.79		
Total	53	2707.04				

If you wished to plot an interaction plot containing both the fitted lines and data, you have two (rather convoluted) options in Minitab. Both require you to calculate the least squares mean first (as illustrated in the Minitab output for Box 10.9 above). Using the command language, these values can be entered directly in subcommands to the plot command. The advantage of this method is that you can request that two different symbols are used to represent the two different temperatures, as below:

MINITAB COMMANDS FOR FIG. 10.6

```
Commands      plot ROT*BAC;
              symbol TEMP;
              line 1 3.556 2 4.778 3 8.000;
              type 2;
              line 1 7.000 2 13.556 3 19.556.
```

This would produce the following graph:



Minitab output for Fig. 10.6

Alternatively, the least squares means can be entered into the worksheet. This was done in columns MEANS1 and BAC1 for temperature 1 and MEANS2 and BAC2 for temperature 2. This allows us to use menus to draw the graph, by overlaying three graphs—one containing the data as points, and two others containing the means for the two temperatures separately, joined together with solid and dashed lines. The commands that would produce the equivalent graphs are also included.

MINITAB MENU BASED COMMANDS FOR FIG. 10.6

Commands `plot ROT*BAC MEANS1*BAC1 MEANS2*BAC2;`
 `symbol;`
 `type 6 0 0;`
 `connect;`
 `type 0 1 2;`
 `connect;`
 `type 0 1 2;`
 `overlay.`

Menu route Graph > Plot
 ROT → Graph variables, Y row 1
 BAC → Graph variables, X row 1
 MEANS1 → Graph variables, Y row 2
 BAC1 → Graph variables, X row 2
 MEANS2 → Graph variables, Y row 3
 BAC2 → Graph variables, X row 3

Display ▼

Item 1: Symbol
 Item 2: Connect
 Item 3: Connect

For each ▼

Graph
 Graph
 Graph

Select Symbol in row 1 then

Type ▼

Graph 1: Solid circle

Graph 2: None

Graph 3: None

Select Connect in row 2 then

Line Type ▼

Graph 1: None

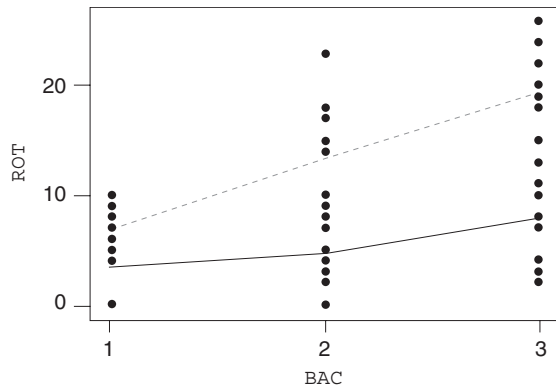
Graph 2: Solid

Graph 3: Dash

Edit attributes...

Edit attributes...

This produces the following graph:



Minitab output from menu based commands for Fig. 10.6

10.7 Exercises

Testing polynomials requires sequential sums of squares

MINITAB COMMANDS FOR BOX 10.11(a) **Analysis based on X**

Commands

```

glm Y = X|X|X;
covariate X;
brief 1;

glm Y = X|X|X;
covariate X;
brief 1;
ssquares 1.
    
```

Menu route

Stat > ANOVA > General Linear Model

Y → Response

X | X | X → Model

Covariates...

X → Covariates

Results...

⊙ Analysis of variance table

Stat > ANOVA > General Linear Model

Y → Response

X | X | X → Model

Covariates...

X → Covariates

Results...

⊙ Analysis of variance table

Options...

⊙ Sequential sums of squares (Type I)

MINITAB OUTPUT FOR BOX 10.11(a)

Analysis based on X: the first analysis bases the p-values on adjusted sums of squares and the second on sequential sums of squares

General Linear Model: Y versus

Factor Type Levels Values

Analysis of Variance for Y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
X	1	58.906	0.719	0.719	2.46	0.121
X*X	1	4.305	0.151	0.151	0.52	0.475
X*X*X	1	0.542	0.542	0.542	1.85	0.178
Error	68	19.876	19.876	0.292		
Total	71	83.629				

General Linear Model: Y versus						
Factor	Type	Levels	Values			
Analysis of Variance for Y, using Sequential SS for Tests						
Source	DF	Seq SS	Adj SS	Seq MS	F	P
X	1	58.906	0.719	58.906	201.53	0.000
X*X	1	4.305	0.151	4.305	14.73	0.000
X*X*X	1	0.542	0.542	0.542	1.85	0.178
Error	68	19.876	19.876	0.292		
Total	71	83.629				

The second analysis requires exactly the same commands, substituting XS for X.

Partitioning a sum of squares into polynomial components

The Minitab output for this exercise may be found in the answers to exercises in Chapter 14.