

3

Models, parameters, and GLMs

In Minitab, the general linear model command (GLM) can be used to analyse models with both continuous and categorical explanatory variables, instead of using ANOVA for categorical and REGRESS for continuous variables. The commands and output are presented in a slightly different format, and this chapter will briefly summarise those conventions. For example, Minitab will always assume that any explanatory variables used in the model formulae are categorical unless you specify otherwise.

3.2 Expressing all models as linear equations

The commands to conduct a general linear model analysis on the *trees* dataset are presented below:

MINITAB COMMANDS FOR BOX 3.1 General Linear models with continuous explanatory variables	
Commands	<code>glm VOLUME = HEIGHT;</code> <code>covariate HEIGHT.</code>
Menu route	Stat > ANOVA > General Linear Model VOLUME → Response HEIGHT → Model <div style="border: 1px solid black; display: inline-block; padding: 2px;">Covariates...</div> HEIGHT → Covariates

Note that we had to declare HEIGHT as continuous. This will produce the following output:

MINITAB OUTPUT FOR BOX 3.1 General Linear Model for the trees dataset						
General Linear Model: VOLUME versus						
Factor	Type	Levels	Values			
Analysis of Variance for VOLUME, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
HEIGHT	1	2901.2	2901.2	2901.2	16.16	0.000
Error	29	5204.9	5204.9	179.5		
Total	30	8106.1				
Term	Coef	SE Coef	T	P		
Constant	-87.12	29.27	-2.98	0.006		
HEIGHT	1.5433	0.3839	4.02	0.000		
Unusual Observations for VOLUME						
Obs	VOLUME	Fit	SE Fit	Residual	St Resid	
31	77.0000	47.1478	4.8600	29.8522	2.39R	
R denotes an observation with a large standardized residual						

It can be seen that exactly the same information is given in the output. All those items that are missing (the fitted equation, s , R -squared) can all be calculated from the ANOVA table and coefficient table as discussed in the main text. The ANOVA table tests *whether* the two variables are related. (Does tree height give us any information about tree volume?—The answer is yes, with a p value of less than 0.0005). The coefficient table tells us *how* the variables are related (Volume = $-87.12 + 1.5433$ Height). Minitab also automatically provides both sequential and adjusted sums of squares in the ANOVA table. (The meanings of these two forms of sum of squares are given in the main text.)

Similarly, the commands for analysing the *fertiliser* dataset are given below:

MINITAB COMMANDS FOR BOX 3.2 General Linear Models with categorical explanatory variables	
Commands	brief 3. glm YIELD = FERTIL.
Menu route	Stat > ANOVA > General Linear Model YIELD → Response FERTIL → Model <div style="border: 1px solid black; padding: 2px; display: inline-block;">Results...</div> ⊙ In addition, coefficients for all terms

This will give the following output:

MINITAB OUTPUT FOR BOX 3.2 General Linear Model for the <i>fertiliser</i> dataset						
General Linear Model: YIELD versus FERTIL						
Factor	Type	Levels	Values			
FERTIL	fixed	3	1	2	3	
Analysis of Variance for YIELD, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
FERTIL	2	10.8227	10.8227	5.4114	5.70	0.009
Error	27	25.6222	25.6222	0.9490		
Total	29	36.4449				
Term	Coef	SE Coef	T	P		
Constant	4.6437	0.1779	26.11	0.000		
FERTIL						
1	0.8013	0.2515	3.19	0.004		
2	-0.6447	0.2515	-2.56	0.016		
Unusual Observations for YIELD						
Obs	YIELD	Fit	SE Fit	Residual	St Resid	
19	6.28000	3.99900	0.30805	2.28100	2.47R	
29	7.00000	4.48700	0.30805	2.51300	2.72R	
R denotes an observation with a large standardized residual						

Note that while we did not need to specify that FERTIL was categorical (as Minitab assumes this by default), we did have to specifically request a coefficients table that gives us parameter estimates for the categorical terms in the model. In the command language, this was by using the 'brief 3' command. This alters the level of output given—brief 3 being the maximum. The simplest solution is to

enter the command `brief 3` at the beginning of any Minitab session, to ensure that the coefficient table is always produced. Using menus, select the ‘Results’ tab in the ‘General Linear Model’ box and request the coefficients table for all terms—unfortunately, this must be done each time.

Also notice that with Minitab, it is the estimate for fertiliser 3 that is aliased (omitted because it can be calculated from the estimates for the first two levels). This follows the same method of aliasing as introduced in the main text, so see Chapter 3 for further details.

With the Minitab general linear modelling facility, it is also possible to request a table of the means, as shown below:

MINITAB COMMANDS FOR BOX 3.3 Obtaining least squares means for a categorical variable	
Commands	<pre>brief 3. glm YIELD = FERTIL; means FERTIL.</pre>
Menu route	Stat > ANOVA > General Linear Model YIELD → Response FERTIL → Model <div style="border: 1px solid black; display: inline-block; padding: 2px;">Results...</div> FERTIL → Display least squares means corresponding to the terms

This will produce the table below in addition to the output shown in Box 3.2. The term ‘least squares means’ is used, as the model is fitted to the data by minimising the residual sum of squares.

MINITAB OUTPUT FOR TABLE 3.3 Obtaining the group means for the fertiliser dataset		
Least Squares Means for YIELD		
FERTIL	Mean	SE Mean
1	5.445	0.3081
2	3.999	0.3081
3	4.487	0.3081

3.3 Turning the tables and creating datasets

These commands will enable you to create the dummy dataset described in Section 3.3 of the main text. Label five blank columns `Y`, `DUM1`, `DUM2`, `F`, and `ERROR`. We now fill the columns as instructed in Box 3.3 (main text) using the Minitab code below. Then follow one of these two routes to create your dataset.

MINITAB COMMANDS FOR BOX 3.3a **Creating your own dataset**

Commands	SET DUM1
Set up the dummy columns (notice no space between the close bracket and the 10) and the treatment subscripts	(1 0 -1) 10 END SET DUM2 (0 1 -1) 10 END SET F (1 2 3) 10 END
Commands	Let K3 = 12.2
Specify parameter values, create error and data	Let K1 = 5.0 Let K2 = -2.5 Let K4 = 1.5 Random 30 error Let Y = K3 + K1*DUM1 + K2*DUM2 + K4*ERROR
Menu route	Calc > Make Patterned Data > Arbitrary Set of Numbers
First, create the variables	DUM1 → Store Patterned Data in 1 0 -1 → Arbitrary Set of Numbers 10 → List each value <i>and repeat with DUM2 (using 0 1 -1) and F (using 1 2 3).</i>
Second, specify your constants	Calc > Calculator K1 (or K2, K3, or K4) → Store result in variable 5.0 (or -2.5, 12.2, or 1.5 respectively) → Expression
Third, create your error variable	Calc > Random data > Normal 30 → Generate – rows of data ERROR → Store in columns 0 → mean 1.0 → standard deviation
Fourth, create Y	Calc > Calculator Y → Store result in variable K3 + K1*DUM1 + K2*DUM2 + K4*ERROR → Expression

Finally, to analyse the dataset you have just created, follow the usual procedure.

MINITAB COMMANDS FOR BOX 3.3b Analysing your own dataset	
Commands	brief 3. glm Y = F; means F.
Menu route	Stat > ANOVA > General Linear Model Y → Response F → Model <div style="border: 1px solid black; padding: 2px; display: inline-block;">Results...</div> ⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 3.4 Recovering the parameter estimates						
General Linear Model: Y versus F						
Factor	Type	Levels	Values			
F	fixed	3	1 2 3			
Analysis of Variance for Y, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
F	2	711.05	711.05	355.53	60.02	0.000
Error	27	159.94	159.94	5.92		
Total	29	870.99				
Term	Coef	SE Coef	T	P		
Constant	12.0655	0.4444	27.15	0.000		
F						
1	5.0358	0.6284	8.01	0.000		
2	-6.5840	0.6284	-10.48	0.000		
Unusual Observations for Y						
Obs	Y	Fit	SE Fit	Residual	St Resid	
23	6.9276	13.6137	0.7697	-6.6861	-2.90R	
R denotes an observation with a large standardized residual						

The means for the three levels of F can be calculated from the coefficients table as described in Chapter 3.

3.5 Exercises

How variability in the population will influence our analysis

See Minitab output for this exercise in the answers to exercises in Chapter 14.