4 Using more than one explanatory variable

Commands and concepts introduced in this chapter

This chapter introduces models with more than one explanatory variable, and so leads to the concepts of statistical elimination and adjusted and sequential sums of squares. Both types of sum of squares are shown in the generic output given in the main text. With Minitab, both types are automatically given in the ANOVA table. Unless you specify otherwise, Minitab will construct the *F* ratio using the adjusted sum of squares. Situations in which it would be more appropriate to use the sequential sum of squares are explored and discussed in later chapters.

4.1 Why use more than one explanatory variable?

Leaping to the wrong conclusion

The first analysis in this example involves just one continuous explanatory variable, as in the previous chapter.

MINITAB COMMANDS FOR BOX 4.1 **GLM with one explanatory variable**

Commands

```
glm AMA = HGHT;
covariate HGHT;
brief 1.
```

Menu route

Stat > ANOVA > General Linear Model
AMA → Response
HGHT → Model
[Covariates...]
HGHT → Covariates
[Results...]
⊙ Analysis of Variance Table

The level of output has been minimised (brief 1 using commands, requesting the analysis of variance table only using menus) as we will be focusing on the ANOVA tables only in the following few examples. These commands give the following output:

MINITAB OUTPUT FOR BOX 4.1 Height explaining mathematical ability

General Linear Model: AMA **versus**

```
Factor       Type Levels Values

Analysis of Variance for AMA, using Adjusted SS for Tests

Source           DF       Seq SS      Adj SS      Adj MS          F         P
HGHT              1       412.77      412.77      412.77     726.87     0.000
Error            30        17.04       17.04        0.57
Total            31       429.81
```

This analysis is then extended to include two explanatory variables. The additional variable is added to the word equation by the + sign (which must be typed in to the model panel, between the two explanatory variables). In this case both are continuous, so both must be included either under the covariate subcommand or in the covariate box (where no + sign is required).

MINITAB COMMANDS FOR BOX 4.2 GLM with two continuous explanatory variables

Commands

```
glm AMA = YEARS + HGHT;
covariate YEARS HGHT;
brief 1.
```

Menu route

Stat > ANOVA > General Linear Model
AMA → Response
YEARS + HGHT → Model
[Covariates…]
YEARS HGHT → Covariate
[Results…]
⊙ Analysis of Variance Table

This gives the following output:

MINITAB OUTPUT FOR BOX 4.2 Years, not height explaining mathematical ability

General Linear Model: AMA versus

```
Factor       Type Levels Values

Analysis of Variance for AMA, using Adjusted SS for Tests

Source         DF       Seq SS      Adj SS      Adj MS          F          P
YEARS           1       422.60        9.84        9.84      39.63      0.000
HGHT            1         0.01        0.01        0.01       0.03      0.860
Error          29         7.20        7.20        0.25
Total          31       429.81
```

This illustrates Minitab's useful convention of automatically providing both the sequential and the adjusted sums of squares side by side in the output. Unless otherwise requested, it will automatically perform the *F* ratio test using the adjusted sum of squares for each factor (calculate the *F* ratio from first principles to satisfy yourself that this is the case). This is the most appropriate test in this example. Later we shall come across examples where the sequential sum of squares proves to be more appropriate.

Missing a significant relationship

This is a second example of the advantages of statistical elimination. The first analysis includes one explanatory variable.

MINITAB COMMANDS FOR BOX 4.3a GLM with one categorical explanatory variable

Commands	`glm FINALHT = WATER;` `brief 1.`
Menu route	Stat > ANOVA > General Linear Model `FINALHT` → Response `WATER` → Model [Results…] ⊙ Analysis of Variance Table.

MINITAB OUTPUT FOR BOX 4.3a Final height alone shows no differences between watering regimes

General Linear Model: FINALHT versus WATER

```
Factor        Type Levels Values
WATER         fixed     4 1 2 3 4

Analysis of Variance for FINALHT, using Adjusted SS for Tests

Source        DF      Seq SS      Adj SS      Adj MS       F        P
WATER          3      12.895      12.895       4.298     1.97     0.136
Error         36      78.461      78.461       2.179
Total         39      91.356
```

The second analysis includes a second explanatory variable.

MINITAB COMMANDS FOR BOX 4.3b GLM with two explanatory variables

Commands

```
glm FINALHT = WATER + INITHT;
covariate INITHT;
brief 1.
```

Menu route

Stat > ANOVA > General Linear Model
FINALHT → Response
WATER + INITHT → Model
[Covariates...]
INITHT → Covariates
[Results...]
⊙ Analysis of Variance Table

MINITAB OUTPUT FOR BOX 4.3b Change in height is significantly different between watering regimes

General Linear Model: FINALHT versus WATER

```
Factor       Type Levels Values
WATER        fixed     4 1 2 3 4

Analysis of Variance for FINALHT, using Adjusted SS for Tests

Source         DF      Seq SS      Adj SS      Adj MS          F         P
WATER           3      12.895       1.052       0.351      64.74     0.000
INITHT          1      78.272      78.272      78.272    1.4E+04     0.000
Error          35       0.190       0.190       0.005
Total          39      91.356
```

4.3 Two types of sum of squares

This section explores two possible consequences of statistical elimination.

Eliminating a third variable makes the second less informative

This is illustrated by comparing two models with one or two explanatory variables.

MINITAB COMMANDS FOR BOX 4.4a Length of right leg used to predict the weight of an individual

Commands

```
glm WGHT = RLEG;
covariate RLEG;
brief 1.
```

Menu route

Stat > ANOVA > General Linear Model
WGHT → Response
RLEG → Model
Covariates…
RLEG → Covariates
Results…
⊙ Analysis of Variance Table

MINITAB OUTPUT FOR BOX 4.4a Length of right leg predicts weight of an individual

General Linear Model: WGHT versus

```
Factor          Type Levels Values

Analysis of Variance for WGHT, using Adjusted SS for Tests

Source          DF       Seq SS       Adj SS       Adj MS            F          P
RLEG             1       3627.7       3627.7       3627.7       125.75      0.000
Error           98       2827.1       2827.1         28.8
Total           99       6454.8
```

MINITAB COMMANDS FOR BOX 4.4b Using both RLEG and LLEG to predict weight

Commands

```
glm WGHT = RLEG + LLEG;
covariate RLEG LLEG;
brief 1.
```

Menu route

Stat > ANOVA > General Linear Model
WGHT → Response
RLEG + LLEG → Model
[Covariates...]
RLEG LLEG → Covariates
[Results...]
⊙ Analysis of Variance Table

MINITAB OUTPUT FOR BOX 4.4b Neither RLEG nor LLEG are significant predictors of weight

General Linear Model: WGHT versus

```
Factor         Type Levels Values
Analysis of Variance for WGHT, using Adjusted SS for Tests

Source          DF       Seq SS       Adj SS       Adj MS         F           P
RLEG             1       3627.7         83.3         83.3      2.93       0.090
LLEG             1         66.0         66.0         66.0      2.32       0.131
Error           97       2761.1       2761.1         28.5
Total           99       6454.8
```

This rather counterintuitive conclusion (that neither RLEG or LLEG are significant predictors of weight) is a consequence of the fact that Minitab will base its *p* values on adjusted sum of squares by default.

Eliminating a third variable makes the second more informative

This example illustrates the converse situation.

MINITAB COMMANDS FOR BOX 4.5a Using year of birth to predict a poet's age

Commands

```
glm POETSAGE = BYEAR;
covariate BYEAR;
brief 1.
```

Menu route

Stat > ANOVA > General Linear Model
POETSAGE → Response
BYEAR → Model
Covariates…
BYEAR → Covariates
Results…
⊙ Analysis of Variance Table

MINITAB OUTPUT FOR BOX 4.5a Age of poets cannot be predicted from birth date alone

General Linear Model: POETSAGE versus

```
Factor      Type Levels Values

Analysis of Variance for POETSAGE, using Adjusted SS for Tests

Source          DF       Seq SS      Adj SS       Adj MS        F         P
BYEAR            1          1.2         1.2          1.2     0.00     0.954
Error           10       3333.5      3333.5        333.4
Total           11       3334.7
```

MINITAB COMMANDS FOR BOX 4.5b Using years of birth and death to predict a poet's age

Commands

```
glm POETSAGE = BYEAR + DYEAR;
covariate BYEAR DYEAR;
brief 1.
```

Menu route

Stat > ANOVA > General Linear Model
POETSAGE → Response
BYEAR + DYEAR → Model
Covariates…
BYEAR DYEAR → Covariates
Results…
⊙ Analysis of Variance Table

MINITAB OUTPUT FOR BOX 4.5b Age of poets can be accurately predicted from birth and death dates

General Linear Model: POETSAGE versus

```
Factor      Type Levels Values

Analysis of Variance for POETSAGE, using Adjusted SS for Tests

Source        DF     Seq SS      Adj SS      Adj MS          F          P
BYEAR          1        1.2      3299.7      3299.7    1.0E+04      0.000
DYEAR          1     3330.6      3330.6      3330.6    1.0E+04      0.000
Error          9        2.9         2.9         0.3
Total         11     3334.7
```

In this case, the p values based on the adjusted sums of squares provide the most useful F ratio test.

No new principles arise in the *urban foxes* example, or in the re-analysis of the *trees* dataset. We therefore move straight on to the exercises.

4.7 Exercises

The cost of reproduction

The analyses involve models with one or two continuous explanatory variables.

MINITAB COMMANDS FOR BOX 4.11 Using reproductive effort to explain survival

Commands

```
glm LLONGVTY = LEGGRATE;
covariate LEGGRATE;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
LLONGVTY → Response
LEGGRATE → Model
[Covariates...]
LEGGRATE → Covariates
[Results...]
⊙ In addition coefficients for all terms

MINITAB OUTPUT FOR BOX 4.11 GLM of survival against reproductive rate for *Drosophila* sp.

General Linear Model: LLONGVTY versus

```
Factor          Type Levels Values

Analysis of Variance for LLONGVTY, using Adjusted SS for Tests

Source              DF      Seq SS      Adj SS      Adj MS       F         P
LEGGRATE             1       7.738       7.738       7.738    5.83     0.024
Error               23      30.507      30.507       1.326
Total               24      38.245

Term                 Coef        SE Coef          T            P
Constant           1.7693         0.2313       7.65        0.000
LEGGRATE           0.2813         0.1165       2.42        0.024

Unusual Observations for LLONGVTY

Obs          LLONGVTY          Fit         SE Fit       Residual        St Resid
  2           4.26300      1.99748        0.25789        2.26552           2.02R

R denotes an observation with a large standardized residual.
```

MINITAB COMMANDS FOR BOX 4.12 Using reproductive effort and size to explain survival

Commands

```
glm LLONGVTY = LSIZE + LEGGRATE;
covariate LSIZE LEGGRATE;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
`LLONGVTY` → Response
`LSIZE + LEGGRATE` → Model
[Covariates...]
`LSIZE LEGGRATE` → Covariates
[Results...]
⦿ In addition coefficients for all terms

MINITAB OUTPUT FOR BOX 4.12 GLM of survival against size and reproductive rate for *Drosophila sp.*

General Linear Model: LLONGVTY versus

```
Factor         Type Levels Values

Analysis of Variance for LLONGVTY, using Adjusted SS for Tests

Source            DF     Seq SS     Adj SS     Adj MS        F        P
LSIZE              1     26.240     21.842     21.842    55.46    0.000
LEGGRATE           1      3.340      3.340      3.340     8.48    0.008
Error             22      8.665      8.665      0.394
Total             24     38.245

Term                  Coef        SE Coef          T           P
Constant            1.6819         0.1266      13.28       0.000
LSIZE               1.4719         0.1976       7.45       0.000
LEGGRATE          -0.28993        0.09956      -2.91       0.008

Unusual Observations for LLONGVTY

Obs          LLONGVTY         Fit        SE Fit       Residual        St Resid
10            4.59500     2.76143       0.21923        1.83357           3.12R

R denotes an observation with a large standardized residual.
```

To plot a graph similar to Fig. 4.10, you need to rank the data, according to size, and then create a new categorical explanatory variable called SIZEGRP. This can then be used to plot a graph in which the different size groups are represented by different symbols.

MINITAB COMMANDS FOR FIGURE 4.10 Sorting the data

Commands

```
sort LSIZE LEGGRATE LLONGVTY LSIZE LEGGRATE & LLONGVTY;
by LSIZE.
```

Menu route

Manip > Sort
LSIZE LEGGRATE LLONGVTY → Sort Columns
LSIZE LEGGRATE LLONGVTY → Store sorted columns
LSIZE → Sort by column, first row

Notice how the columns are sorted and then replaced into the same columns. The next step creates a new column named SIZEGRP containing integers 1–6 as a code for LSIZE. We chose the following coding: LSIZE < −1.5, SIZEGRP = 1; LSIZE −1.5 to −0.5, SIZEGRP = 2; LSIZE −0.5 to 0.25, SIZEGRP = 3; LSIZE 0.25 to 1, SIZEGRP = 4; LSIZE 1 to 1.2, SIZEGRP = 5; LSIZE > 1.2, SIZEGRP = 6.

MINITAB COMMANDS FOR FIGURE 4.10 Coding the data

Commands

```
Name c4 = 'SIZEGRP'.
Code (−3:−1.5) 1 (−1.5:−0.5) 2 (−0.5:0.25) &
3 (0.25:1) 4 (1:1.2) 5 (1.2:3) 6 'LSIZE' &
'SIZEGRP'.
```

Menu route

Manip > Code > Numeric to Numeric
`LSIZE` → Code data from columns
`SIZEGRP` → into columns
`−3: −1.5` → original values, row 1
`1` → new, row 1
`−1.5: −0.5` → original values, row 2
`2` → new, row 2
`−0.5:0.25` → original values, row 3
`3` → new, row 3
`0.25:1` → original values, row 4
`4` → new, row 4
`1:1.2` → original values, row 5
`5` → new, row 5
`1.2:3` → original values, row 6
`6` → new, row 6

These coded data are then used to plot the graph.

MINITAB COMMANDS FOR FIG. 4.10 Plotting the graph

Commands

```
plot LLONGVTY*LEGGRATE;
symbol SIZEGRP.
```

Menu route

Graph > Plot
`LLONGVTY` → Graph variable, Y
`LEGGRATE` → Graph variable, X.

Display ▼

Symbol

For each ▼

Group
`SIZEGRP` → Group variables

This should produce the following graph:

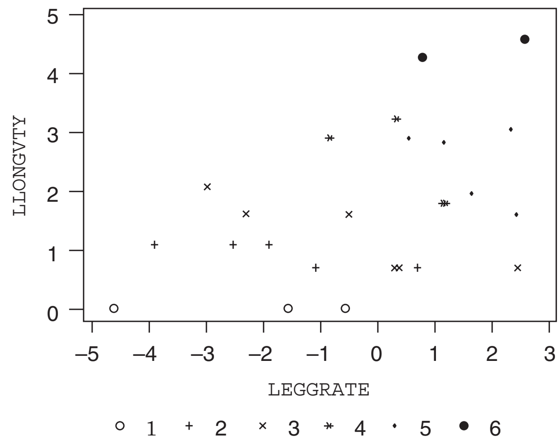

Fig. 4.10 LLONGVTY against LEGGRATE with each point being allocated to one of six size groups.

Investigating Obesity

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