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 SPSS, we have to choose which type of sum of squares is to appear in the ANOVA table. By default, adjusted sums of squares appear, but we can replace them with sequential sums of squares using the METHOD subcommand or via the Model button. We will repeat a number of analyses, once with adjusted and once with sequential sums of squares, in this chapter. The focus for the moment, so far as F ratio tests are concerned, will be on the adjusted sums of squares. Situations in which it would be more appropriate to test 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.

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

Syntax	`glm AMA with HGT` `  /design HGT.`
Menu route	Analyze > General Linear Model > Univariate AMA → Dependent Variable HGT → Covariate(s)

These commands give the following output:

SPSS OUTPUT FOR BOX 4.1 **Height explaining mathematical ability**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: AMA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	412.774[a]	1	412.774	726.866	.000
Intercept	216.887	1	216.887	381.923	.000
HGT	412.774	1	412.774	726.866	.000
Error	17.036	30	.568		
Total	8771.286	32			
Corrected Total	429.811	31			

a. R Squared = .960 (Adjusted R Squared = .959)

This analysis is then extended to include two explanatory variables. The additional variable is added to the syntax in two places. First it must be included in the initial declaration of variables, after the WITH keyword, to show that it is a continuous variable. Second, it is included in the DESIGN statement to add it to the word equation. Using menus, we need only add both variables to the "Covariate(s)" pane.

SPSS COMMANDS FOR BOX 4.2 **GLM with two continuous explanatory variables**

Syntax

```
glm AMA with YEARS HGT
  /design YEARS HGT.
```

Menu route

Analyze > General Linear Model > Univariate

AMA → Dependent Variable

YEARS HGT → Covariate(s)

This gives the following output:

SPSS OUTPUT FOR BOX 4.2 **Years, not height explaining mathematical ability**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: AMA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	422.612[a]	2	211.306	851.229	.000
Intercept	1.057E-02	1	1.057E-02	.043	.838
YEARS	9.838	1	9.838	39.630	.000
HGT	7.877E-03	1	7.877E-03	.032	.860
Error	7.199	29	.248		
Total	8771.286	32			
Corrected Total	429.811	31			

a. R Squared = .983 (Adjusted R Squared = .982)

Missing a significant relationship

This is a second example of the advantages of statistical elimination. The first analysis includes one explanatory variable. Note that if you are doing these examples consecutively via the menu route, you will need to remove variables from the panes in which you last placed them. To do this, just select them and click on the arrow between the source and destination pane — SPSS will then place them back in the original source pane. Alternatively, you can press the 'reset' button at the bottom of the 'Univariate' window. This will remove all variables from destination panes, and set any options back to default.

SPSS COMMANDS FOR BOX 4.3a **GLM with one categorical explanatory variable**

Syntax

```
glm FINALHT by WATER
  /design WATER.
```

Menu route

Analyze > General Linear Model > Univariate

FINALHT → Dependent Variable

WATER → Fixed Factor(s)

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

General linear model

Between-Subjects Factors

		N
WATER	1	10
	2	10
	3	10
	4	10

Tests of Between-Subjects Effects

Dependent Variable: FINALHT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12.895 [a]	3	4.298	1.972	.136
Intercept	1237.211	1	1237.211	567.663	.000
WATER	12.895	3	4.298	1.972	.136
Error	78.461	36	2.179		
Total	1328.567	40			
Corrected Total	91.356	39			

a. R Squared = .141 (Adjusted R Squared = .070)

The second analysis includes a second explanatory variable. Note that you need to use a combination of WITH and BY statements in the syntax route, as you are using both continuous and categorical explanatory variables.

SPSS COMMANDS FOR BOX 4.3b **GLM with two explanatory variables**

Syntax

```
glm FINALHT by WATER with INITHT
  /design WATER INITHT.
```

Menu route Analyze > General Linear Model > Univariate

FINALHT → Dependent Variable

WATER → Fixed Factor(s)

INITHT → Covariate(s)

SPSS OUTPUT FOR BOX 4.3b **Final height is significantly different between watering regimes when initial height is taken into account**

General linear model

Between-Subjects Factors

		N
WATER	1	10
	2	10
	3	10
	4	10

Tests of Between-Subjects Effects

Dependent Variable: FINALHT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	91.166[a]	4	22.792	4207.898	.000
Intercept	3.818E-02	1	3.818E-02	7.048	.012
WATER	1.052	3	.351	64.738	.000
INITHT	78.272	1	78.272	14450.926	.000
Error	.190	35	5.416E-03		
Total	1328.567	40			
Corrected Total	91.356	39			

a. R Squared = .998 (Adjusted R Squared = .998)

4.3 Two types of sum of squares

This section explores two possible consequences of statistical elimination. We need to repeat the analyses so that SPSS will show us both the sequential (Type I) and adjusted (Type III) sums of squares. If you press the 'reset' button in the Univariate pane, when using the menu route between analyses, this will return you to the default option of Type III sums of squares. In the menu route instructions, we assume you use the 'reset' button between analyses.

Eliminating a third variable makes the second less informative

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

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

Syntax

```
glm WGHT with RLEG
  /design RLEG.
```

Menu route Analyze > General Linear Model > Univariate

WGHT → Dependent Variable

RLEG → Covariate(s)

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

General linear model

Tests of Between-Subjects Effects

Dependent Variable: WGHT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3627.670[a]	1	3627.670	125.751	.000
Intercept	7.694	1	7.694	.267	.607
RLEG	3627.670	1	3627.670	125.751	.000
Error	2827.100	98	28.848		
Total	600325.367	100			
Corrected Total	6454.770	99			

a. R Squared = .562 (Adjusted R Squared = .558)

Note the use of the METHOD subcommand to obtain the sequential sums of squares in the next analysis. With the menu route, this option can be obtained using the model button, and selecting Type 1 from the Sum of squares Pop-up-menu.

SPSS COMMANDS FOR BOX 4.4b **Using both RLEG and LLEG to predict weight. The analyses are repeated to produce output with both adjusted and sequential sums of squares**

Syntax

```
glm WGHT with RLEG LLEG
  /design RLEG LLEG.

glm WGHT with RLEG LLEG
  /method sstype(1)
  /design RLEG LLEG.
```

Menu route

Analyze > General Linear Model > Univariate

WGHT → Dependent Variable

RLEG LLEG → Covariate(s)

Analyze > General Linear Model > Univariate

WGHT → Dependent Variable

RLEG LLEG → Covariate(s)

Model

Sum of Squares: Type I ▼

SPSS OUTPUT FOR BOX 4.4b **Showing the results from both analyses: Neither RLEG nor LLEG are significant predictors of WGHT when the adjusted sums of squares are tested**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: WGHT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3693.632[a]	2	1846.816	64.879	.000
Intercept	2.138	1	2.138	.075	.785
RLEG	83.334	1	83.334	2.928	.090
LLEG	65.962	1	65.962	2.317	.131
Error	2761.138	97	28.465		
Total	600325.367	100			
Corrected Total	6454.770	99			

a. R Squared = .572 (Adjusted R Squared = .563)

General linear model

Tests of Between-Subjects Effects

Dependent Variable: WGHT

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3693.632[a]	2	1846.816	64.879	.000
Intercept	593870.597	1	593870.597	20862.935	.000
RLEG	3627.670	1	3627.670	127.442	.000
LLEG	65.962	1	65.962	2.317	.131
Error	2761.138	97	28.465		
Total	600325.367	100			
Corrected Total	6454.770	99			

a. R Squared = .572 (Adjusted R Squared = .563)

This rather counterintuitive conclusion (that neither RLEG nor LLEG is a significant predictor of weight) follows when we base our p-values on adjusted sums of squares.

Notice that with the sequential sums of squares, the three equalities we noted earlier in an SPSS ANOVA table hold true for the sums of squares as well as the degrees of freedom, namely

Total = Intercept + Corrected Total

Corrected Total = Error + Model

Model = Sum of separate terms.

Eliminating a third variable makes the second more informative

This example illustrates the converse situation.

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

Syntax	`glm POETSAGE with BYEAR` `/design BYEAR.`
Menu route	Analyze > General Linear Model > Univariate POETSAGE → Dependent Variable BYEAR → Covariate(s)

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

General linear model

Tests of Between-Subjects Effects

Dependent Variable: POETSAGE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.162[a]	1	1.162	.003	.954
Intercept	361.924	1	361.924	1.086	.322
BYEAR	1.162	1	1.162	.003	.954
Error	3333.505	10	333.350		
Total	30600.000	12			
Corrected Total	3334.667	11			

a. R Squared = .000 (Adjusted R Squared = –.100)

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

Syntax

```
glm POETSAGE with BYEAR DYEAR
  /design BYEAR DYEAR.

glm POETSAGE with BYEAR DYEAR
  /method sstype(1)
  /design BYEAR DYEAR.
```

Menu route

Analyze > General Linear Model > Univariate

POETSAGE → Dependent Variable

BYEAR DYEAR → Covariate(s)

Analyze > General Linear Model > Univariate

POETSAGE → Dependent Variable

BYEAR DYEAR → Covariate(s)

[Model]

Sum of Squares: [Type I ▼]

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

General linear model

Tests of Between-Subjects Effects

Dependent Variable: POETSAGE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3331.734[a]	2	1665.867	5112.883	.000
Intercept	1.557E-02	1	1.557E-02	.048	.832
BYEAR	3299.663	1	3299.663	10127.333	.000
DYEAR	3330.573	1	3330.573	10222.200	.000
Error	2.932	9	.326		
Total	30600.000	12			
Corrected Total	3334.667	11			

a. R Squared = .999 (Adjusted R Squared = .999)

General linear model

Tests of Between-Subjects Effects

Dependent Variable: POETSAGE

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3331.734[a]	2	1665.867	5112.883	.000
Intercept	27265.333	1	27265.333	83682.817	.000
BYEAR	1.162	1	1.162	3.566	.092
DYEAR	3330.573	1	3330.573	10222.200	.000
Error	2.932	9	.326		
Total	30600.000	12			
Corrected Total	3334.667	11			

a. R Squared = .999 (Adjusted R Squared = .999)

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 fox 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

This exercise involves sorting data — so the dataset is stored separately in Drosophila.xls and Drosophila.sav on the website. You will need to read this file in separately for this question. The analyses involve models with one or two continuous explanatory variables. We request the parameter estimates using the PRINT subcommand (or via the options button).

SPSS COMMANDS FOR BOX 4.11 **Using reproductive effort to explain survival**

Syntax

```
glm LLONGVTY with LEGGRATE
  /print parameter
  /design LEGGRATE.
```

Menu route

Analyze > General Linear Model > Univariate

LLONGVTY → Dependent Variable

LEGGRATE → Covariate(s)

Options

☑ Parameter estimates

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

General linear model

Tests of Between-Subjects Effects

Dependent Variable: LLONGVTY

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.738[a]	1	7.738	5.834	.024
Intercept	77.586	1	77.586	58.494	.000
LEGGRATE	7.738	1	7.738	5.834	.024
Error	30.507	23	1.326		
Total	111.974	25			
Corrected Total	38.245	24			

a. R Squared = .202 (Adjusted R Squared = .168)

Parameter Estimates

Dependent Variable: LLONGVTY

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	1.769	.231	7.648	.000	1.291	2.248
LEGGRATE	.281	.116	2.415	.024	4.038E-02	.522

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

Syntax

```
glm LLONGVTY with LSIZE LEGGRATE
  /print parameter
  /design LSIZE LEGGRATE.
```

Menu route

Analyze > General Linear Model > Univariate

LLONGVTY → Dependent Variable

LSIZE LEGGRATE → Covariate(s)

Options

☑ Parameter estimates

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

General linear model

Tests of Between-Subjects Effects

Dependent Variable: LLONGVTY

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	29.580[a]	2	14.790	37.552	.000
Intercept	69.504	1	69.504	176.471	.000
LSIZE	21.842	1	21.842	55.457	.000
LEGGRATE	3.340	1	3.340	8.480	.008
Error	8.665	22	.394		
Total	111.974	25			
Corrected Total	38.245	24			

a. R Squared = .773 (Adjusted R Squared = .753)

Parameter Estimates

Dependent Variable: LLONGVTY

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	1.682	.127	13.284	.000	1.419	1.944
LSIZE	1.472	.198	7.447	.000	1.062	1.882
LEGGRATE	–.290	.100	–2.912	.008	–.496	–8.345E-02

To plot a graph similar to figure 4.10, you need to create a grouping variable called SIZEGRP, based on LSIZE. This can then be used to plot a graph in which the different size groups are represented by different symbols. The first menu action is simply to ensure that the dataset is at the front, so that the Data menu is available.

SPSS COMMANDS FOR FIGURE 4.10 **Creating the grouping variable**

Syntax

```
recode LSIZE
  (lowest thru -1.5=1)  (-1.5 thru -0.5=2)
  (-0.5 thru 0.25=3)  (0.25 thru 1=4)
  (1 thru 1.2=5)  (1.2 thru highest=6)
  into  SIZEGRP .
execute.
```

Menu route

Window > SPSS Data Editor

Transform > Recode > Into Different Variables

`LSIZE` → Input Variable

`SIZEGRP` → Output Variable Name

[Change]

[Old and New Values]

Old Value:
⊙ Range: Lowest through (blank)
`-1.5` → (blank)

New Value:
⊙ Value: (blank)
`1` → (blank)

[Add]

This then adds the condition 'Lowest thru –1.5 → 1' to the Old → New pane. We then need to repeat this for the other values of `SIZEGRP` *as shown below.*

Old Value:

⊙ Range (blank 1) through (blank 2)
`-1.5` → blank 1
`-0.5` → blank 2

New Value:
⊙ Value: (blank)

`2` → (blank)

[Add]

(*Contd.*)

SPSS COMMANDS FOR FIGURE 4.10 (*Contd.*)

`-0.5` → blank 1
`0.25` → blank 2
`3` → New Value blank

[Add]

`0.25` → blank 1
`1` → blank 2
`4` → New Value blank

[Add]

`1` → blank 1
`1.2` → blank 2
`5` → New Value

[Add]

Old value:
⦿Range (blank) through Highest
`1.2` → (blank)
`6` → New Value blank

[Add]

Now we can plot the graph, using SIZEGRP to set markers by.

SPSS COMMANDS FOR FIGURE 4.10 **Plotting the graph**

Syntax

```
graph
  /scatterplot(bivar)=LEGGRATE with LLONGVTY
     by SIZEGRP .
```

Menu route

Graphs > Scatter (then select Simple and click Define)

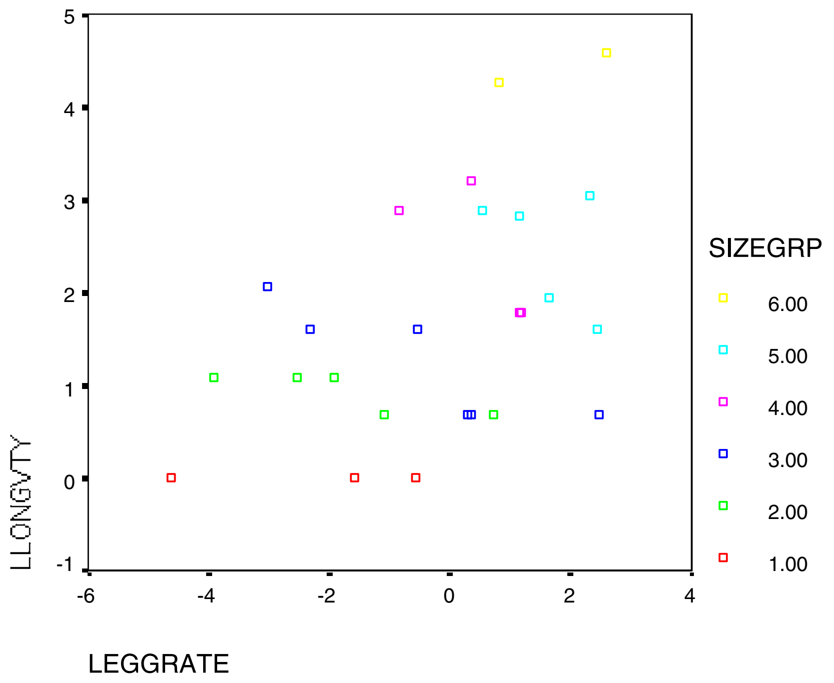
`LLONGVTY` → Y Axis

`LEGGRATE` → X Axis

`SIZEGRP` → Set Markers By

This should produce the following graph:

FIGURE 4.10 **LLONGVTY against LEGGRATE with each point being allocated to one of six size groups**



Investigating obesity

See SPSS output for this exercise in the answers for exercises.