Answers to exercises

Chapter 1

Dioecious trees

The question is whether the number of flowers produced depends upon the sex of the tree. The variable SEX is categorical with two levels. The commands to execute the ANOVA are as before.

SPSS COMMANDS FOR BOX 15.1 **Analysis for dioecious trees**

Syntax	`glm FLOWERS by SEX` `  /design = SEX.`
Menu route	Analyze > General Linear Model > Univariat FLOWERS → Dependent Variable SEX → Fixed Factor(s)

This would produce the following output:

SPSS OUTPUT FOR BOX 15.1 **Analysis for dioecious trees**

General linear model

Between-Subjects Factors

		N
SEX	1	20
	2	30

Tests of Between-Subjects Effects

Dependent Variable: FLOWERS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	171841.333[a]	1	171841.333	1.175	.284
Intercept	6625776.853	1	6625776.853	45.322	.000
SEX	171841.333	1	171841.333	1.175	.284
Error	7017255.167	48	146192.816		
Total	14542709.0	50			
Corrected Total	7189096.500	49			

a. R Squared = .024 (Adjusted R Squared = .004)

There are a number of ways in which these results could be illustrated graphically, one of which is a boxplot. The '`nototal`' subcommand supresses SPSS's urge to produce a plot of total flowers, as well as by sex. Similarly, `statistics=none` prevents a table of descriptive statistics.

Syntax

```
examine FLOWERS by SEX
  /plot=boxplot
  /nototal
  /statistics=none.
```

Menu route

Graph > Boxplot (and choose Simple and click Define)

`FLOWERS` → Variable

`SEX` → Category Axis

This would produce the following graph:

SPSS OUTPUT FOR FIGURE 15.1 **Box plot for dioecious trees**

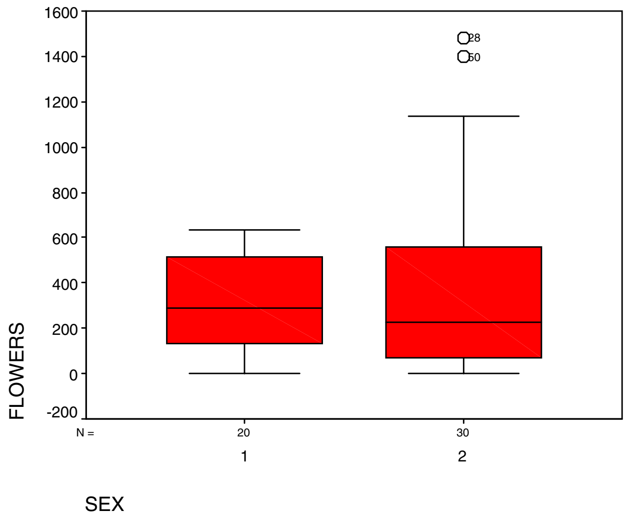
Explore

Sex

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	SEX	N	Percent	N	Percent	N	Percent
FLOWERS	1	20	100.0%	0	.0%	20	100.0%
	2	30	100.0%	0	.0%	30	100.0%

Flowers



Chapter 2

Dioecious trees

The variables FLOWERS and DBH are both continuous, so the best illustration of the relationship between them would be a scatter plot.

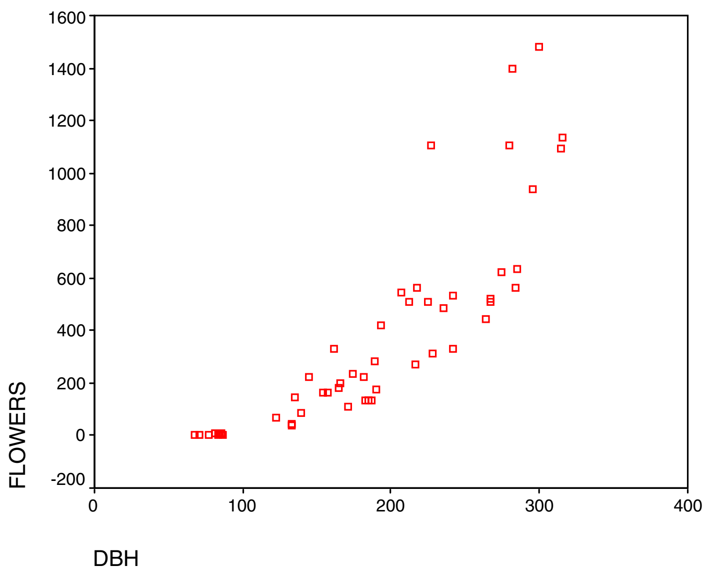
SPSS COMMANDS FOR FIGURE 15.2 **Graph of FLOWERS versus DBH**

Syntax	`graph` `/scatter DBH with FLOWERS`
Menu route	Graphs > Scatter (then select Simple and click Define) FLOWERS → Y Axis DBH → X Axis

This would produce the following graph:

SPSS OUTPUT FOR FIGURE 15.2 **Graph of FLOWERS versus DBH**

Graph



The regression analysis would be executed using the following commands:

SPSS COMMANDS FOR BOX 15.2 **Analysis for dioecious trees**

Syntax

```
glm FLOWERS with DBH
  /print parameters
  /design = DBH.
```

Menu route

Analyze > General Linear Model > Univariate

FLOWERS → Dependent Variable

DBH → Covariate(s)

Options

☑ Parameter estimates

which would give the following output:

SPSS OUTPUT FOR BOX 15.2 **Analysis for dioecious trees**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: FLOWERS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5060722.977[a]	1	5060722.977	114.132	.000
Intercept	1380258.183	1	1380258.183	31.128	.000
DBH	5060722.977	1	5060722.977	114.132	.000
Error	2128373.523	48	44341.115		
Total	14542709.0	50			
Corrected Total	7189096.500	49			

a. R Squared = .704 (Adjusted R Squared = .698)

Parameter Estimates

Dependent Variable: FLOWERS

					95% Confidence Interval	
Parameter	B	Std. Error	t	Sig.	Lower Bound	Upper Bound
Intercept	–481.160	86.241	–5.579	.000	–654.559	–307.762
DBH	4.513	.422	10.683	.000	3.664	5.362

Chapter 3

How variability in the population will influence our analysis

To create data sets for this exercise, just follow the procedure outlined in Box 3.3a 'Creating your own datasets', but alter the standard deviation specified when you define sigma (second group of SPSS commands in Box 3.3a). One advantage of using the syntax method is that you can copy-and-paste the whole set of commands, edit the number you want to change, and then execute the whole lot by selecting all and pressing the run button.

Each of the four created data sets can be analysed using the commands shown below:

SPSS commands to analyse simulated datasets

Syntax	`glm Y by FERTIL`
	`/print parameters`
	`/design = FERTIL.`

Menu route	Analyze > General Linear Model > Univariate `Y` → Dependent Variable `FERTIL` → Fixed Factor(s)

Each analysis will produce one estimate of the intercept, which may then be entered into a new column in the worksheet. We named our new column `SIGMA2` for the estimates obtained when `SIGMA` was set to 2. To add a new variable in SPSS, go to the Variable View sheet and enter a new variable name in the first available row at the end of the list of current variables. On pressing return, SPSS will assume it is numeric by default. This creates a new variable column which will appear on the Data View sheet when you return to it. Now you can enter values into this column as you produce them from your simulations. When ten estimates have been obtained, a histogram may be drawn for the column `SIGMA2` (for example) using the commands:

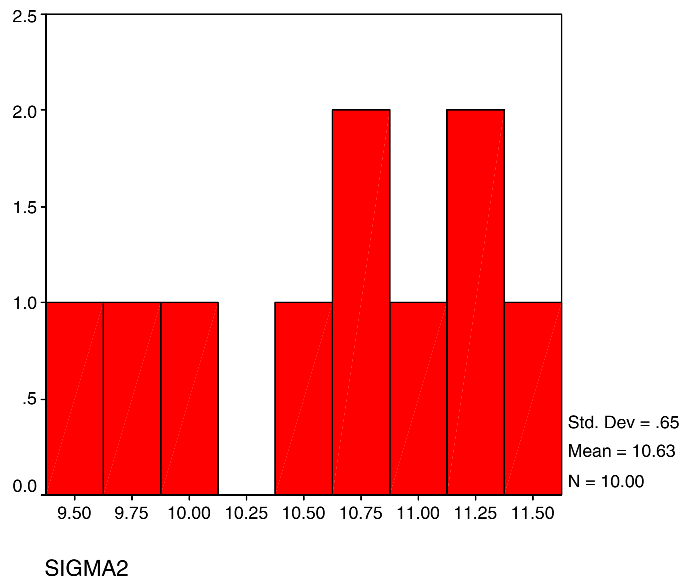
SPSS COMMANDS FOR FIGURE 15.3 **Variation in parameter estimates**

Syntax	`graph` `/histogram SIGMA2.`
Menu route	Graph > Histogram `SIGMA2` → Variable

This would produce a graph similar to the one below:

SPSS OUTPUT FOR FIGURE 15.3 **Variation of parameters estimates when SIGMA = 2**

Graph



Comparison of the four graphs with the error standard deviation set at 2, 4, 8 and 16 will illustrate how the error variance will influence the reliability of the intercept estimate.

Chapter 4

Investigating obesity

First you need to conduct two separate analyses to explain FOREARM using HT or WT.

SPSS COMMANDS FOR BOX 15.4(a) **First analysis of FOREARM**

Syntax

```
glm FOREARM with HT
  /design = HT.
```

Menu route

Analyze > General Linear Model > Univariate

FOREARM → Dependent Variable

HT → Covariate(s)

SPSS OUTPUT FOR BOX 4.13 **First analysis of FOREARM**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: FOREARM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.944[a]	1	.944	.175	.678
Intercept	5.123E-02	1	5.123E-02	.010	.923
HT	.944	1	.944	.175	.678
Error	199.094	37	5.381		
Total	1203.097	39			
Corrected Total	200.038	38			

a. R Squared = .005 (Adjusted R Squared = –.022)

SPSS COMMANDS FOR BOX 15.4(b) **Second analysis of FOREARM**

Syntax

```
glm FOREARM with WT
  /design = WT.
```

Menu route

Analyze > General Linear Model > Univariate
FOREARM → Dependent Variable
WT → Covariate(s)

SPSS OUTPUT FOR BOX 15.4(b) **Second analysis of FOREARM**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: FOREARM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	59.137[a]	1	59.137	15.529	.000
Intercept	13.149	1	13.149	3.453	.071
WT	59.137	1	59.137	15.529	.000
Error	140.901	37	3.808		
Total	1203.097	39			
Corrected Total	200.038	38			

a. R Squared = .296 (Adjusted R Squared = .277)

Then the joint analysis using both explanatory variables. If you are using the menu route, then press the 'reset' button in the Univariate pane between analyses.

SPSS COMMANDS FOR BOX 15.5 **Analysis of FOREARM using both explanatory variables.**
Do once without and once with the bold commands

Syntax

```
glm FOREARM with HT WT
   /print parameters
   /method sstype(1)
   /design = HT WT.
```

(`/method sstype(1)` is printed in bold.)

Menu route

Analyze > General Linear Model > Univariate

FOREARM → Dependent Variable

HT WT → Covariate(s)

[Options]

☑ Parameter estimates

[Model]

Sum of Squares: [Type I ▼]

SPSS OUTPUT FOR BOX 15.5 **Analysis of FOREARM using both explanatory variables**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: FOREARM

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	83.914[a]	2	41.957	13.007	.000
Intercept	1003.059	1	1003.059	310.963	.000
HT	.944	1	.944	.293	.592
WT	82.970	1	82.970	25.722	.000
Error	116.124	36	3.226		
Total	1203.097	39			
Corrected Total	200.038	38			

a. R Squared = .419 (Adjusted R Squared = .387)

Parameter Estimates

Dependent Variable: FOREARM

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	17.452	8.274	2.109	.042	.672	34.232
HT	–.172	.062	–2.772	.009	–.297	–4.606E-02
WT	.233	.046	5.072	.000	.140	.326

and the second analysis would produce the following ANOVA table instead…..

Tests of Between-Subjects Effects

Dependent Variable: FOREARM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	83.914[a]	2	41.957	13.007	.000
Intercept	14.351	1	14.351	4.449	.042
HT	24.777	1	24.777	7.681	.009
WT	82.970	1	82.970	25.722	.000
Error	116.124	36	3.226		
Total	1203.097	39			
Corrected Total	200.038	38			

a. R Squared = .419 (Adjusted R Squared = .387)

Chapter 5

The dorsal crest of the male smooth newt

In the first analysis you are asked to investigate the relationship between LCREST and the body size of the newt. The response variable you are given is logged, therefore the log transformation of the body size variable should be used. The commands and menus are given below:

SPSS COMMANDS FOR BOX 15.6 **Dorsal crest analysis**

Syntax

```
glm LCREST with LSVL
  /print parameters
  /design = LSVL.
```

Menu route

Analyze > General Linear Model > Univariate

LCREST → Dependent Variable

LSVL → Covariate(s)

Options

☑ Parameter estimates

This would result in the following output:

SPSS COMMANDS FOR BOX 15.6 **Dorsal crest analysis**

General linear model

Tests of Between-Subjects Effects

Dependent Variable: LCREST

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.389 [a]	1	2.389	45.808	.000
Intercept	2.036	1	2.036	39.040	.000
LSVL	2.389	1	2.389	45.808	.000
Error	4.434	85	5.216E-02		
Total	59.651	87			
Corrected Total	6.823	86			

a. R Squared = .350 (Adjusted R Squared = .343)

Parameter Estimates

Dependent Variable: LCREST

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	–9.381	1.501	–6.248	.000	–12.366	–6.396
LSVL	5.087	.752	6.768	.000	3.593	6.581

In the second analysis, POND is also included.

SPSS COMMANDS FOR BOX 15.7 **Further dorsal crest analysis**

Syntax

```
glm LCREST by POND with LSVL
  /design = POND LSVL.
```

Menu route

Analyze > General Linear Model > Univariate

LCREST → Dependent Variable

POND → Fixed Factor(s)

LSVL → Covariate(s)

SPSS OUTPUT FOR BOX 15.7 **Further dorsal crest analysis**

General linear model

Between-Subjects Factors

		N
POND	1	3
	2	8
	3	5
	4	8
	5	9
	6	10
	7	20
	8	8
	9	6
	10	10

Tests of Between-Subjects Effects

Dependent Variable: LCREST

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.630[a]	10	.263	4.767	.000
Intercept	1.970	1	1.970	35.701	.000
POND	.241	9	2.674E-02	.485	.881
LSVL	2.305	1	2.305	41.775	.000
Error	4.193	76	5.517E-02		
Total	59.651	87			
Corrected Total	6.823	86			

a. R Squared = .385 (Adjusted R Squared = .305)

Chapter 6

Determinants of the grade point average

This exercise uses the *grades* dataset.

SPSS COMMANDS FOR BOX 15.8 **Analysis of the grades dataset**

Syntax

```
glm GPA by YEAR with VERBAL MATH
  /print parameters
  /design = YEAR VERBAL MATH.
```

Menu route

Analyze > General Linear Model > Univariate

GPA → Dependent Variable

YEAR → Fixed Factor(s)

VERBAL MATH → Covariate(s)

Options

☑ Parameter estimates

SPSS OUTPUT FOR BOX 15.8 **Analysis of the grades dataset**

General linear model

Between-Subjects Factors

		N
YEAR	1	100
	2	100

Tests of Between-Subjects Effects

Dependent Variable: GPA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8.624[a]	3	2.875	9.648	.000
Intercept	.665	1	.665	2.234	.137
YEAR	.846	1	.846	2.839	.094
VERBAL	5.160	1	5.160	17.319	.000
MATH	.709	1	.709	2.380	.124
Error	58.396	196	.298		
Total	1450.400	200			
Corrected Total	67.020	199			

a. R Squared = .129 (Adjusted R Squared = .115)

Parameter Estimates

Dependent Variable: GPA

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	.593	.439	1.350	.179	–.274	1.460
[YEAR=1]	.130	.077	1.685	.094	–2.222E-02	.283
[YEAR=2]	0[a]	.	.	.	.	.
VERBAL	2.288E-03	.001	4.162	.000	1.204E-03	3.372E-03
MATH	9.375E-04	.001	1.543	.124	–2.609E-04	2.136E-03

a. This parameter is set to zero because it is redundant

Chapter 7

Weight, fat and sex

This question requires an analysis with an interaction between a categorical and a continuous explanatory variable.

SPSS COMMANDS FOR BOX 15.9 **Analysis for weight, fat and sex**

Syntax

```
glm FAT by SEX with WEIGHT
  /print parameters
  /design = SEX WEIGHT SEX*WEIGHT.
```

Menu route

Analyze > General Linear Model > Univariate

FAT → Dependent Variable

SEX → Fixed Factor(s)

WEIGHT → Covariate(s)

Options

☑ Parameter estimates

Model

⦿ Custom

Factors & Covariates → Build Terms → Model

SEX	→	SEX
WEIGHT	→	WEIGHT
SEX & WEIGHT	→	SEX*WEIGHT

SPSS OUTPUT FOR BOX 15.9 **Analysis for weight, fat and sex**

General linear model

Between-Subjects Factors

		N
SEX	1	9
	2	10

Tests of Between-Subjects Effects

Dependent Variable: FAT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	188.283[a]	3	62.761	31.237	.000
Intercept	14.859	1	14.859	7.396	.016
SEX	2.108	1	2.108	1.049	.322
WEIGHT	79.542	1	79.542	39.589	.000
SEX * WEIGHT	10.857	1	10.857	5.404	.035
Error	30.138	15	2.009		
Total	15509.000	19			
Corrected Total	218.421	18			

a. R Squared = .862 (Adjusted R Squared = .834)

Parameter Estimates

Dependent Variable: FAT

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	11.571	2.868	4.034	.001	5.458	17.684
[SEX=1]	–6.331	6.181	–1.024	.322	–19.507	6.844
[SEX=2]	0[a]	.	.	.	.	.
WEIGHT	.186	.036	5.199	.000	.109	.262
[SEX=1] * WEIGHT	.217	.094	2.325	.035	1.807E-02	.417
[SEX=2] * WEIGHT	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant

Chapter 8

Combining data from different experiments

It is necessary to combine the columns together to make a single YOUNG column and a single SONGDAY column, and to add the YEAR column. (If you try to do this in the Chapter 8's dataset, you will find the name YEAR already taken — you have a choice of working with a different name, renaming the previous YEAR or working in a new datafile.) We solved this problem by creating a new variable BIRDYEAR. New variables are created in SPSS by going to the 'variable view' sheet, and typing the new name into a free row in the first column. In the 'Data view' sheet, we then cut and paste the SONGDAY and YOUNG data into one column.

SPSS COMMANDS FOR BOX 15.11 **Bird data combined**

Syntax

```
glm YOUNG by BIRDYEAR with SONGDAY
  /print parameters
  /design = BIRDYEAR SONGDAY.
```

Menu route

Analyze > General Linear Model > Univariate

YOUNG → Dependent Variable

BIRDYEAR → Fixed Factor(s)

SONGDAY → Covariate(s)

[Options]

☑ Parameter estimates

SPSS OUTPUT FOR BOX 15.11 **Bird data combined**

General linear model

Between-Subjects Factors

		N
BIRDYEAR	1.00	9
	2.00	12
	3.00	7
	4.00	10
	5.00	6

Tests of Between-Subjects Effects

Dependent Variable: YOUNG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	22.900[a]	5	4.580	3.760	.007
Intercept	49.244	1	49.244	40.432	.000
BIRDYEAR	19.782	4	4.945	4.060	.008
SONGDAY	12.202	1	12.202	10.019	.003
Error	46.282	38	1.218		
Total	430.000	44			
Corrected Total	69.182	43			

a. R Squared = .331 (Adjusted R Squared = .243)

Parameter Estimates

Dependent Variable: YOUNG

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	3.952	.682	5.799	.000	2.572	5.332
[BIRDYEAR=1.00	1.718	.772	2.225	.032	.155	3.280
[BIRDYEAR=2.00	2.949	.788	3.741	.001	1.353	4.545
[BIRDYEAR=3.00	1.049	.615	1.706	.096	–.196	2.293
[BIRDYEAR=4.00	1.985	.809	2.453	.019	.347	3.624
[BIRDYEAR=5.00	0[a]	.	.	.	.	.
SONGDAY	–.109	.034	–3.165	.003	–.179	–3.933E-02

a. This parameter is set to zero because it is redundant

Chapter 9

Stabilising the variance in a blocked experiment

The commands required for the first part of this exercise a) transform the data; b) perform the analysis on the transformed data, storing the residuals and plotting the residuals against the fitted values; c) display descriptive statistics for the residuals by treatment.

SPSS COMMANDS FOR BOXES 15.12 and 15.13, AND FIGURE 15.6

Syntax

```
compute SQRTNCOT = sqrt(NCOT).
execute.
glm SQRTNCOT by BLOCK TRMNT
  /save sresid(SQRTRES) pred(SQRTPRED)
  /design BLOCK TRMNT.
graph
  /scatter SQRTPRED with SQRTRES.
means SQRTRES by TRMNT.
```

Menu route

Transform > Compute
- `SQRTNCOT` → Target Variable
- `sqrt(NCOT)` → Numeric Expression

Analyze > General Linear Model > Univariate
- `SQRTNCOT` → Dependent Variable
- `BLOCK TRMNT` → Fixed Factor(s)
- [Model]
  - ⊙ Custom
  - Factors & Covariates → Build Terms → Model
  - `BLOCK` → `BLOCK`
  - `TRMNT` → `TRMNT`
- [Save]
  - Residuals:
  - ☑ Studentized
  - Predicted Values:
  - ☑ Unstandardized

Graphs > Scatter (then select 'Simple' and click 'Define')
- `Studentized Residuals for SQRTNCOT` → Y Axis
- `Predicted Values for SQRTNCOT` → X Axis

Analyze > Compare Means > Means
- `Studentized Residuals for SQRTNCOT` → Dependent List
- `TRMNT` → Independent List

SPSS OUTPUT FOR BOX 15.12 **Analysis of blocked experiment**

General linear model

Between-Subjects Factors

		N
BLOCK	1	5
	2	5
	3	5
	4	5
	5	5
	6	5
TRMNT	1	6
	2	6
	3	6
	4	6
	5	6

Tests of Between-Subjects Effects

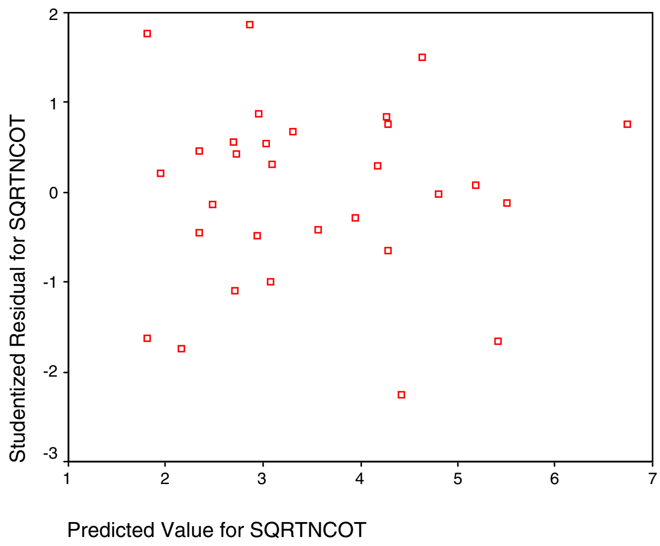
Dependent Variable: SQRTNCOT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	44.444[a]	9	4.938	55.192	.000
Intercept	370.767	1	370.767	4143.868	.000
BLOCK	23.910	5	4.782	53.446	.000
TRMNT	20.534	4	5.134	57.375	.000
Error	1.789	20	8.947E-02		
Total	417.000	30			
Corrected Total	46.233	29			

a. R Squared = .961 (Adjusted R Squared = .944)

SPSS OUTPUT FOR FIGURE 15.6 **Standardised residual plot**

Graph



SPSS OUTPUT FOR BOX 15.13 **Examining residuals of the square root transform analysis**

Means

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Studentized Residual for SQRTNCOT * TRMNT	30	37.5%	50	62.5%	80	100.0%

Report

Studentized Residual for SQRTNCOT

TRMNT	Mean	N	Std. Deviation
1	.0000	6	.93647
2	.0000	6	1.09126
3	.0000	6	.42480
4	.0000	6	1.40246
5	.0000	6	1.33598
Total	.0000	30	1.01710

These commands can then be repeated for the log and inverse transformation.

Checking the 'perfect' model

The commands to produce the histograms of the original and transformed data are reproduced below. We use the command IGRAPH, and the menu route of interactive histograms, and go to some trouble to make the scale the same, to make the histograms easier to compare by eye.

SPSS COMMANDS TO PRODUCE FIGURE 15.6 a AND b **Histograms for the 'perfect' model**

Syntax

```
igraph
  /x1 = var(FEMALE) type = scale
  /y = $count
  /scalerange = var(FEMALE) min=0.25 max=1.15
  /histogram x1interval width = 0.1 .


igraph
  /x1 = var(MALE) type = scale
  /y = $count
  /scalerange = var(MALE) min=0.25 max=1.15
  /histogram x1interval width = 0.1 .
```

Menu route

Graphs >Interactive > Histogram

Assign Variables

FEMALE → (X Axis)

Histogram

– ☑ Set interval size automatically

0.1 → Width of intervals

Options

Scale Range

Variable: FEMALE ▼

– ☑ Auto

0.25 → Minimum

1.15 → Maximum

(Contd.)

SPSS COMMANDS TO PRODUCE FIGURE 15.6 a AND b (*Contd.*)

Graphs >Interactive > Histogram

Assign Variables

MALE → (X Axis)

Histogram

– ☑ Set interval size automatically

0.1 → Width of intervals

Options

Scale Range

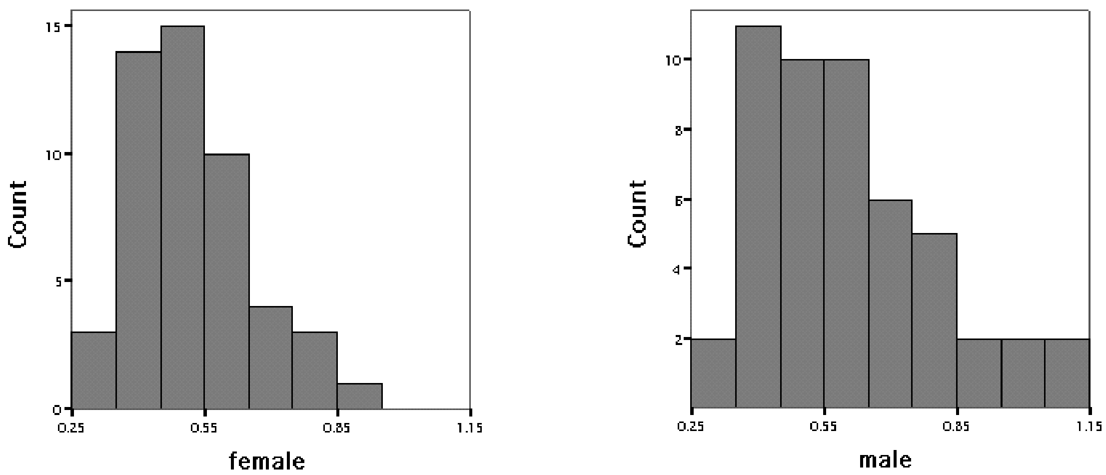
Variable: MALE ▼

– ☑ Auto

0.25 → Minimum

1.15 → Maximum

SPSS OUTPUT FOR FIGURE 15.9 a AND b **Histograms for the 'perfect' model**



and to plot the transformed data…

SPSS COMMANDS TO PRODUCE FIGURE 15.6 c AND d **Histograms for the 'perfect' model**

Syntax

```
compute LFEM = ln(FEMALE).
compute LMALE = ln(MALE).
execute.

igraph
  /x1 = var(LFEM) type = scale
  /y = $count
  /scalerange = var(LFEM) min=-1.25 max=0.15
  /histogram x1interval width = 0.1 .

igraph
  /x1 = var(LMALE) type = scale
  /y = $count
  /scalerange = var(LMALE) min=-1.25 max=0.15
  /histogram x1interval width = 0.1 .
```

Menu route

Transform > Compute

LFEM → Target Variable

ln(FEMALE) → Numeric Expression

Transform > Compute

LMALE → Target Variable

ln(MALE) → Numeric Expression

Graphs >Interactive > Histogram

[Assign Variables]

LFEM → (X Axis)

[Histogram]

– ☑ Set interval size automatically

0.1 → Width of intervals

[Options]

Scale Range

Variable: [LFEM ▼]

– ☑ Auto

–1.25 → Minimum

0.15 → Maximum

(*Contd.*)

SPSS COMMANDS TO PRODUCE FIGURE 15.6 c AND d (*Contd.*)

Graphs >Interactive > Histogram

Assign Variables

LMALE → (X Axis)

Histogram

– ☑ Set interval size automatically

0.1 → Width of intervals

Options

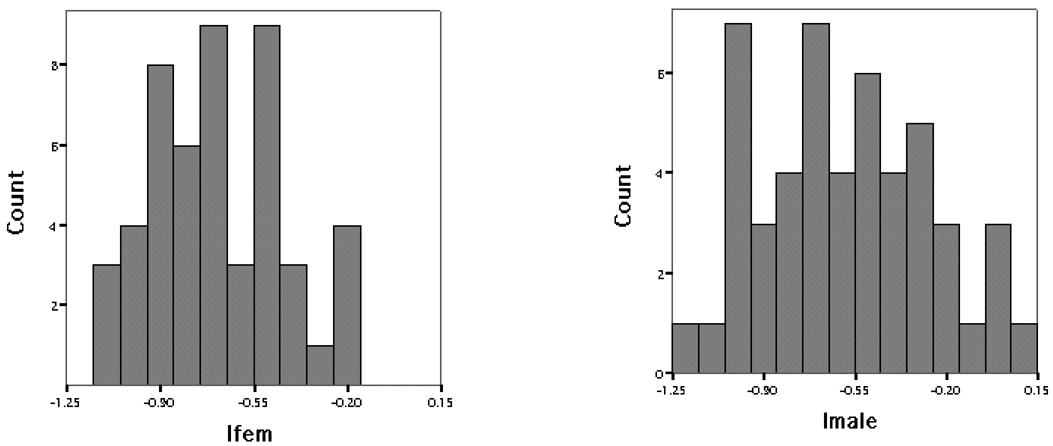
Scale Range

Variable: LMALE ▼

– ☑ Auto

–1.25 → Minimum

SPSS OUTPUT FOR FIGURE 15.9 c AND d **Histograms for 'perfect' model**



We now produce Normal probability plots of the logged and unlogged data, and show the Q-Q plots themselves below (but omit the detrended versions)

SPSS COMMANDS FOR FIGURE 15.10 **Normal probability plots for the *squirrel* dataset**

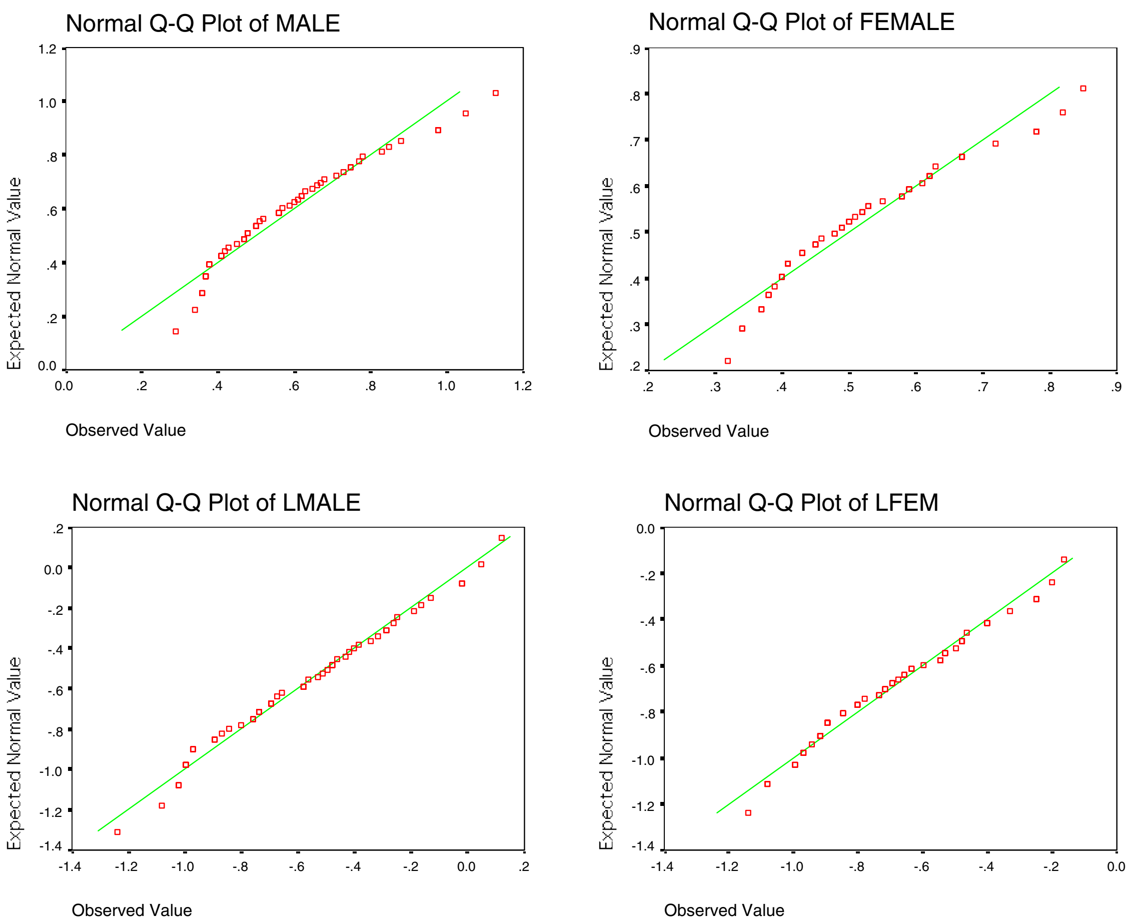
Syntax

```
pplot
  /variables FEMALE MALE LFEM LMALE
  /type q-q.
```

Menu route Graphs > Q-Q

`FEMALE MALE LFEM LMALE` → Variables

SPSS OUTPUT FOR FIGURE 15.10 **Normal probability plots for the *squirrel* dataset**



Tests for Normality can be obtained in SPSS using the EXPLORE command.

The following commands allow you to create the null datasets designed to follow a Normal distribution, with a mean and standard deviation equal to that of the female squirrel weights (which can be obtained using the describe command, as used earlier). The rather peculiar feature of adding and subtracting `MALE` is a trick. It doesn't affect the values, obviously, except that where `MALE` is missing, the result is missing. This makes sure that there are only 50 values in the variable, rather than the 80 of the longest variable in the dataset. (An alternative would be for you to create a new dataset with just the male and female variables). The number of datapoints is important for the comparison of the histograms.

Syntax	`compute SIMFEM1=rv.normal(0.518,0.1315) + MALE - MALE.` `execute.`
Menu route	Transform > Compute `SIMFEM1` → Target Variable `rv.normal(0.518,0.1315) + MALE - MALE` → Numeric Expression

This can be repeated to create as many columns of simulated female (or male) data as you require. This is particularly easy using the written commands in the syntax window, as one can copy-and-paste the relevant commands. These may then be examined using the histogram and Normal probability plot commands used earlier.

Chapter 10

Partitioning a sum of squares into polynomial components

The following commands are required for this exercise:

SPSS COMMANDS FOR BOX 15.18 AND FIGURE 10.8
Perform once without and once with the bold sections. Use reset button in univariate pane of menu route

Syntax

```
glm BYIELD by BBLOCK BSPACE BVARIETY
  /emmeans tables(BSPACE*BVARIETY)
  /plot profile(BSPACE*BVARIETY)
  /method sstype(1)
  /design BBLOCK BSPACE BVARIETY BSPACE*BVARIETY.
```

(**/method sstype(1)** is printed in bold.)

Menu route

Analyze > General Linear Model > Univariate

BYIELD → Dependent Variable

BBLOCK BSPACE BVARIETY → Fixed Factor(s)

[Model]

⦿ Custom

Factors & Covariates → Build Terms → Model

BBLOCK → BBLOCK

BSPACE → BSPACE

BVARIETY → BVARIETY

BSPACE & BVARIETY → BSPACE*BVARIETY

Sum of Squares: [**Type I ▼**]

[Options]

BSPACE* BVARIETY → Display Means for

[Plots]

BSPACE → Horizontal Axis

BVARIETY → Separate Lines

SPSS OUTPUT FOR BOX 15.18 **Analysis for barley yield, with both types of sums of squares**

General linear model

Between-Subjects Factors

		N
BBLOCK	1	9
	2	9
	3	9
	4	9
BSPACE	1	12
	2	12
	3	12
BVARIETY	1	12
	2	12
	3	12

Tests of Between-Subjects Effects

Dependent Variable: BYIELD

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2203.528[a]	11	200.321	11.336	.000
Intercept	120293.361	1	120293.361	6807.274	.000
BBLOCK	255.639	3	85.213	4.822	.009
BSPACE	155.056	2	77.528	4.387	.024
BVARIETY	1027.389	2	513.694	29.069	.000
BSPACE * BVARIETY	765.444	4	191.361	10.829	.000
Error	424.111	24	17.671		
Total	122921.000	36			
Corrected Total	2627.639	35			

a. R Squared = .839 (Adjusted R Squared = .765)

Estimated marginal means

BSPACE * BVARIETY

Dependent Variable: BYIELD

				95% Confidence Interval	
BSPACE	BVARIETY	Mean	Std. Error	Lower Bound	Upper Bound
1	1	47.500	2.102	43.162	51.838
	2	62.250	2.102	57.912	66.588
	3	56.000	2.102	51.662	60.338
2	1	50.750	2.102	46.412	55.088
	2	58.500	2.102	54.162	62.838
	3	64.250	2.102	59.912	68.588
3	1	55.750	2.102	51.412	60.088
	2	52.250	2.102	47.912	56.588
	3	73.000	2.102	68.662	77.338

and the second analysis produces the following ANOVA table instead. . . .

(*Contd.*)

SPSS OUTPUT FOR BOX 15.18 (*Contd.*)

Tests of Between-Subjects Effects

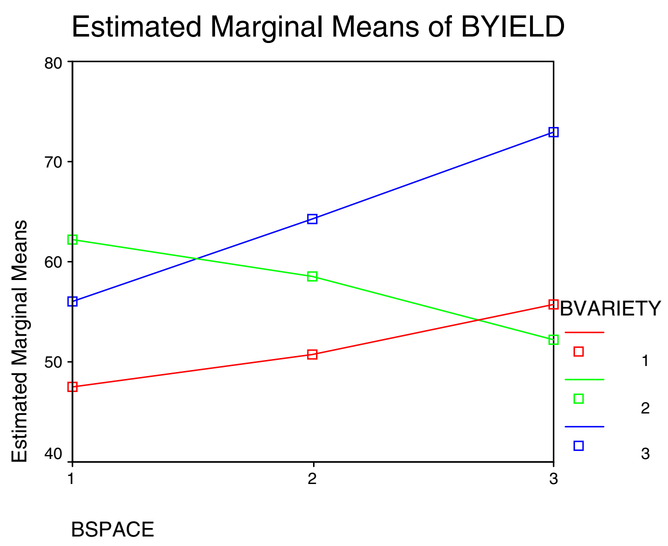
Dependent Variable: BYIELD

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2203.528[a]	11	200.321	11.336	.000
Intercep	120293.361	1	120293.361	6807.274	.000
BBLOCK	255.639	3	85.213	4.822	.009
BSPACE	155.056	2	77.528	4.387	.024
BVARIETY	1027.389	2	513.694	29.069	.000
BSPACE * BVARIETY	765.444	4	191.361	10.829	.000
Error	424.111	24	17.671		
Total	122921.000	36			
Corrected Total	2627.639	35			

a. R Squared = .839 (Adjusted R Squared = .765)

SPSS OUTPUT FOR FIGURE 10.8 **Interaction diagram from the *barley* dataset**

Profile plots



A second analysis is then conducted on the *barley* dataset, involving a polynomial decomposition. We choose to include the interaction of BVARIETY with the quadratic term BSPACE*BSPACE so that the polynomial analysis is a decomposition of the non-polynomial analysis. This means that the error SS, MS and DF will be the same in the two analyses, allowing us to compare them more easily. It also allows us to investigate whether the response of yield to spacing is curvilinear for some varieties but not others. Notice that we need to play the usual trick on SPSS in the menu route to fit polynomial terms.

SPSS COMMANDS FOR BOX 15.19 **Second analysis for barley yield**
Perform once without and once with the bold sections

Syntax

```
glm BYIELD by BBLOCK BVARIETY with BSPACE
  /method sstype(1)
  /design BBLOCK BVARIETY BSPACE BSPACE*BVARIETY
       BSPACE*BSPACE BSPACE*BSPACE*BVARIETY.
```

(The line **/method sstype(1)** is printed in bold.)

Menu route

Transform > Compute

BSPACE2 → Target Variable

BSPACE*BSPACE → Numeric Expression

Analyze > General Linear Model > Univariate

BYIELD → Dependent Variable

BBLOCK BVARIETY → Fixed Factor(s)

BSPACE BSPACE2 → Covariate(s)

Model

⊙ Custom

Factors & Covariates → Build Terms → Model

BBLOCK	→	BBLOCK
BVARIETY	→	BVARIETY
BSPACE	→	BSPACE
BSPACE & BVARIETY	→	BSPACE*BVARIETY
BSPACE2	→	BSPACE2
BSPACE2 & BVARIETY	→	BSPACE2*BVARIETY

Sum of Squares: Type I ▼

SPSS OUTPUT FOR BOX 15.19
Second analysis for barley yield, with both types of sums of squares

General linear model

Between-Subjects Factors

		N
BBLOCK	1	9
	2	9
	3	9
	4	9
BVARIETY	1	12
	2	12
	3	12

Tests of Between-Subjects Effects

Dependent Variable: BYIELD

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2203.528[a]	11	200.321	11.336	.000
Intercept	120293.361	1	120293.361	6807.274	.000
BBLOCK	255.639	3	85.213	4.822	.009
BVARIETY	1027.389	2	513.694	29.069	.000
BSPACE	155.042	1	155.042	8.774	.007
BVARIETY * BSPACE	759.083	2	379.542	21.478	.000
BSPACE * BSPACE	1.389E-02	1	1.389E-02	.001	.978
BVARIETY * BSPACE * BSPACE	6.361	2	3.181	.180	.836
Error	424.111	24	17.671		
Total	122921.000	36			
Corrected Total	2627.639	35			

a. R Squared = .839 (Adjusted R Squared = .765)

and in the second analysis the following ANOVA table is produced inserted….

Tests of Between-Subjects Effects

Dependent Variable: BYIELD

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2203.528[a]	11	200.321	11.336	.000
Intercept	1746.320	1	1746.320	98.822	.000
BBLOCK	255.639	3	85.213	4.822	.009
BVARIETY	38.167	2	19.083	1.080	.356
BSPACE	3.593	1	3.593	.203	.656
BVARIETY * BSPACE	5.655	2	2.827	.160	.853
BSPACE * BSPACE	1.389E-02	1	1.389E-02	.001	.978
BVARIETY * BSPACE * BSPACE	6.361	2	3.181	.180	.836
Error	424.111	24	17.671		
Total	122921.000	36			
Corrected Total	2627.639	35			

a. R Squared = .839 (Adjusted R Squared = .765)

Chapter 11

Finding the best treatment for cat fleas

The final model is:

SPSS COMMANDS FOR BOX 15.20 **Analysis for cat fleas**

Syntax

```
glm LOGFLEAS with NCATS by TRTMT CARPET
   /design TRTMT NCATS CARPET.
```

Menu route: Analyze > General Linear Model > Univariate

LOGFLEAS → Dependent Variable

TRTMT CARPET → Fixed Factor(s)

NCATS → Covariate(s)

SPSS OUTPUT FOR BOX 15.20 **Analysis for cat fleas**

General linear model

Between-Subjects Factors

		N
TRTMT	1	40
	2	49
CARPET	1	54
	2	35

Tests of Between-Subjects Effects

Dependent Variable: LOGFLEAS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	31.445[a]	3	10.482	16.573	.000
Intercept	167.183	1	167.183	264.339	.000
TRTMT	5.750	1	5.750	9.092	.003
NCATS	22.962	1	22.962	36.306	.000
CARPET	6.103	1	6.103	9.650	.003
Error	53.759	85	.632		
Total	1525.926	89			
Corrected Total	85.203	88			

a. R Squared = .369 (Adjusted R Squared = .347)

Multiplicity of p-values

The second exercise involves a simulation. First open the dataset Chapter11sim.sav. This is empty except that it contains one variable with arbitrary values in it, so that SPSS knows the number of datapoints. We then put random values into the variables.

SPSS COMMANDS TO PRODUCE A NULL DATASET

Syntax

```
compute X1= rv.normal(10.5,2.4) .
compute X2= rv.normal(10.5,2.4) .
compute X3= rv.normal(10.5,2.4) .
compute X4= rv.normal(10.5,2.4) .
compute X5= rv.normal(10.5,2.4) .
compute X6= rv.normal(10.5,2.4) .
compute X7= rv.normal(10.5,2.4) .
compute X8= rv.normal(10.5,2.4) .
compute X9= rv.normal(10.5,2.4) .
compute X10= rv.normal(10.5,2.4) .
compute Y= rv.normal(10.5,2.4) .
execute .
```

Menu route

Transform > Compute

X1 → Target Variable

rv.normal(10.5,2.4) → Numeric Expression

and repeat for X2 *to* X10 *and for* Y

These simulated data are then analysed using the following commands:

SPSS COMMANDS TO PRODUCE OUTPUT FOR TABLE 15.6

Syntax

```
regression
  /dependent Y
  /method=enter X1 X2 X3 X4 X5 X6 X7 X8 X9 X10.
```

Menu route

Analyze > Regression > Linear

Y → Dependent

X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 → Independent(s)

EXAMPLE SPSS OUTPUT FOR TABLE 15.6

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X10, X5, X1, X2, X3, X8, X7, X9, X4, X6[a]	.	Enter

a. All requested variables entered

b. Dependent Variable: Y

Model Summary

Mode	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.442[a]	.195	–.252	2.4064

a. Predictors: (Constant), X10, X5, X1, X2, X3, X8, X7, X9, X4, X6

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.318	10	2.532	.437	.909[a]
	Residual	104.237	18	5.791		
	Total	129.555	28			

a. Predictors: (Constant), X10, X5, X1, X2, X3, X8, X7, X9, X4, X6

b. Dependent Variable: Y

Coefficients[a]

Mode		Unstandardize Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.085	6.835		.890	.385
	X1	.179	.254	.176	.706	.489
	X2	-.120	.219	–.147	–.546	.592
	X3	–1.92E-02	.209	–.024	–.092	.928
	X4	–9.61E-02	.239	–.108	–.401	.693
	X5	.171	.286	.181	.599	.557
	X6	.152	.265	.160	.573	.573
	X7	2.978E-02	.190	.039	.157	.877
	X8	.168	.213	.193	.792	.439
	X9	–.164	.269	–.163	–.609	.550
	X10	.152	.233	.159	.651	.523

a. Dependent Variable: Y

The ANOVA table provides a p-value for the regression as a whole, as well as the p-values for each individual X variable in the analysis being found in the table of coefficients. (The p-value resulting from the t-test for the slope against zero for a particular variable is the same as the p-value for the F ratio test for the same explanatory variable in the analysis of variance table. This arises from the fact that $t^2 = F$; see Appendix 3). These results should be used to construct your version of Table 15.6, and then the analysis repeated.

Chapter 12

How a nested analysis can solve problems of non-independence

These commands first calculate the variable `LUPRATE`, and then conduct a nested analysis.

SPSS COMMANDS FOR BOX 15.21 **Nested analysis of *sheep* dataset**

Syntax

```
compute LUPRATE=NLOOKUPS/DURATION.
execute.

glm LUPRATE by SEX SHEEP
  /random SHEEP
  /design SEX SHEEP(SEX).

varcomp LUPRATE by SEX SHEEP
  /random SHEEP
  /method sstype(1)
  /design SEX SHEEP(SEX).
```

Menu route *no nesting permitted in menu route*

SPSS OUTPUT FOR BOX 15.21 **Nested analysis of *sheep* dataset**

General linear model

Between-Subjects Factors

		N
SEX	1	60
	2	60
SHEEP	1	40
	2	40
	3	40

Tests of Between-Subjects Effects

Dependent Variable: LUPRATE

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	5.375	1	5.375	72.935	.001
	Error	.295	4	7.369E-02[a]		
SEX	Hypothesis	.133	1	.133	1.802	.251
	Error	.295	4	7.369E-02[a]		
SHEEP(SEX)	Hypothesis	.295	4	7.369E-02	26.893	.000
	Error	.312	114	2.740E-03[b]		

a. MS(SHEEP(SEX))

b. MS(Error)

Expected Mean Squares[a,b]

Source	Variance Component		
	Var(SHEEP (SEX))	Var(Error)	Quadratic Term
Intercept	20.000	1.000	Intercept, SEX
SEX	20.000	1.000	SEX
SHEEP(SEX)	20.000	1.000	
Error	.000	1.000	

a. For each source, the expected mean square equals the sum of the coefficients in the cells times the variance components, plus a quadratic term involving effects in the Quadratic Term cell.

b. Expected Mean Squares are based on the Type III Sums of Squares.

(Contd.)

SPSS OUTPUT FOR BOX 15.21 (*Contd.*)

Variance components estimation

Factor Level Information

		N
SEX	1	60
	2	60
SHEEP	1	40
	2	40
	3	40

Dependent Variable: LUPRATE

Variance Estimates

Component	Estimate
Var(SHEEP(SEX))	3.548E-03
Var(Error)	2.740E-03

Dependent Variable: LUPRATE
Method: ANOVA (Type I Sum of Squares)

Chapter 13

Fig trees in costa rica

Rather more detail is given by the menu route than the syntax route in this example.

SPSS COMMANDS FOR TABLE 15.7 **Means and standard deviations for fig tree sites**

Syntax	`descriptives SITE1 SITE2 SITE3 NINDIVS.`
Menu route	Analyze > Descriptive Statistics > Explore `SITE1` → Dependent List: `SITE2` → Dependent List: `SITE3` → Dependent List: `NINDIVS` → Dependent List:

SPSS OUTPUT FOR TABLE 15.7 **Means and variances for fig tree sites**

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
SITE1	100	2	17	9.51	3.218
SITE2	100	8	28	18.93	4.409
SITE3	100	27	56	41.43	6.261
NINDIVS	300	2	56	23.29	14.239
Valid N (listwise)	100				

SPSS COMMANDS FOR BOX 15.22

Syntax

```
compute SQRTN = sqrt(NINDIVS).
execute.

glm SQRTN by SITE
  /design SITE.
```

Menu route

Transform > Compute

`SQRTN` → Target Variable

`sqrt(NINDIVS)` → Numeric Expression

Analyze > General Linear Model > Univariate

`SQRTN` → Dependent Variable

`SITE` → Fixed Factor(s)

SPSS OUTPUT FOR BOX 15.22

General linear model

Between-Subjects Factors

		N
SITE	1	100
	2	100
	3	100

Tests of Between-Subjects Effects

Dependent Variable: SQRTN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	582.626[a]	2	291.313	1091.995	.000
Intercept	6325.143	1	6325.143	23709.986	.000
SITE	582.626	2	291.313	1091.995	.000
Error	79.231	297	.267		
Total	6987.000	300			
Corrected Total	661.857	299			

a. R Squared = .880 (Adjusted R Squared = .879)