12 Random effects

Random effects must *not* be analysed with PROC GLM, even though it seems to offer that facility. The tests done for fixed effects ignore the presence of random effects. To perform correct tests with fixed and random effects, we must instead use PROC MIXED. Fortunately, this is very similar to PROC GLM, except that

- Those explanatory variables which are random must be declared as such.
- Additional output may be requested which relates to the new concepts of expected mean squares and variance components introduced in the main text.

It is in this context that you most commonly come across 'nested' variables.

12.3 A one-way ANOVA with a random factor

It is necessary to begin with two remarks about PROC MIXED.

PROC MIXED has a particular eccentricity compared to the specification of models used in the main text, which is that random effects are declared as random using the RANDOM statement, but *are then not included in the MODEL statement*. This makes for some odd looking input, as we will see immediately. Perhaps this allows SAS to choose the order of variables.

Second, PROC MIXED has a vast array of options. We choose among them on the basis of similarity to the output of other packages, and in particular do not wish to claim that these options are statistically preferable to those not taken. Hence some of the options specified may appear rather mysterious to you.

Let us now turn to the first example, which is extremely simple and does not involve nesting, but introduces the extra output that may be produced with a random factor. The model looks odd (just `CACONC=`) for the reasons just given. In the menu route, we uncheck a check box for the first time. We symbolise this with '☑'.

SAS COMMANDS FOR BOX 12.1 **One-way ANOVA with a random factor**

Commands

```
proc mixed data=gandh.Chapter12 method=type1 covtest;
  class LEAF;
  model CACONC = / htype=1 ddfm=satterth;
  random LEAF;
  run;
```

Menu route

Statistics > Anova > Mixed Models…

`CACONC` → Dependent

`LEAF` → Class

Model

☑ Random Effects

`LEAF` → Add

Tests

Fixed Effects Tests:

-☑ Type III

☑ Type I

☑ Tests of variance components

Options

Estimation Method:

⊙ Type 1

SAS OUTPUT FOR BOX 12.1 **One-way ANOVA with a random factor**

The Mixed Procedure

Model Information

Data Set	GANDH.CHAPTER12
Dependent Variable	CACONC
Covariance Structure	Variance Components
Estimation Method	Type 1
Residual Variance Method	Factor
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
LEAF	4	1 2 3 4

Dimensions

Covariance Parameters	2
Columns in X	1
Columns in Z	4
Subjects	1
Max Obs Per Subject	120
Observations Used	16
Observations Not Used	104
Total Observations	120

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF
LEAF	3	0.280862	0.093621	Var(Residual) + 4 Var(LEAF)	MS(Residual)	12
Residual	12	0.129100	0.010758	Var(Residual)	.	.

Type 1 Analysis of Variance

Source	F Value	Pr > F
LEAF	8.70	0.0024
Residual	.	.

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
LEAF	0.02072	0.01914	1.08	0.2792	0.05	-0.01680	0.05823
Residual	0.01076	0.004392	2.45	0.0072	0.05	0.005532	0.02932

Fit Statistics

-2 Res Log Likelihood	-16.1
AIC (smaller is better)	-12.1
AICC (smaller is better)	-11.1
BIC (smaller is better)	-13.4

12.4 A two-level nested ANOVA

This second example introduces the concept of nesting. SAS uses brackets to indicate nesting. The variable `LEAF2(PLANT2)`should be read as 'Leaf2 nested within Plant2'. Notice that the nesting is indicated in the RANDOM statement.

SAS COMMANDS FOR BOX 12.3 **Two-way nested analysis of variance**

Commands

```
proc mixed data=gandh.Chapter12 method=type1 covtest;
  class PLANT2 LEAF2;
  model CACONC2 = / htype=1 ddfm=satterth;
  random PLANT2 LEAF2(PLANT2);
  run;
```

Menu route

Statistics > Anova > Mixed Models…

`CACONC` → Dependent

`PLANT2 LEAF2` → Class

Model

☑ Random Effects

`PLANT2 LEAF2` → Add

Select `LEAF2` in Random Effects box, and `PLANT2` in Class box, and then click 'Nest'.

Tests

Fixed Effects Tests:

-☑ Type III

☑ Type I

☑ Tests of variance components

Options

Estimation Method:

⊙ Type 1

SAS OUTPUT FOR BOX 12.3 **Two-way nested analysis of variance**

The Mixed Procedure

Model Information

Data Set	GANDH.CHAPTER12
Dependent Variable	CACONC2
Covariance Structure	Variance Components
Estimation Method	Type 1
Residual Variance Method	Factor
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
LEAF2	3	1 2 3
PLANT2	4	1 2 3 4

Dimensions

Covariance Parameters	3
Columns in X	1
Columns in Z	16
Subjects	1
Max Obs Per Subject	120
Observations Used	24
Observations Not Used	96
Total Observations	120

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
PLANT2	3	2.714326	0.904775	Var(Residual) + 2 Var(LEAF2(PLANT2)) + 6 Var(PLANT2)
LEAF2(PLANT2)	8	1.261996	0.157749	Var(Residual) + 2 Var(LEAF2(PLANT2))
Residual	12	0.506214	0.042185	Var(Residual)

Type 1 Analysis of Variance

Source	Error Term	Error DF	F Value	Pr > F
PLANT2	MS(LEAF2(PLANT2))	8	5.74	0.0216
LEAF2(PLANT2)	MS(Residual)	12	3.74	0.0200
Residual	.	.	.	.

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
PLANT2	0.1245	0.1238	1.01	0.3147	0.05	-0.1182	0.3672
LEAF2(PLANT2)	0.05778	0.04037	1.43	0.1523	0.05	-0.02133	0.1369
Residual	0.04218	0.01722	2.45	0.0072	0.05	0.02169	0.1149

Fit Statistics

-2 Res Log Likelihood	15.4
AIC (smaller is better)	21.4
AICC (smaller is better)	22.6
BIC (smaller is better)	19.5

12.5 Mixing random and fixed effects

To mix fixed and random factors, only the random factors are declared as such, but both are included in the model formula. The output includes a term Q[1] which stands for:

$$\frac{\Sigma\left(\rho_i-\bar{\rho}\right)^2}{a-1}$$

where ρ_i = the true weight gain for piglets from the ith sire, $\bar{\rho}$ = true mean weight gain for all piglets, and a-1 are the degrees of freedom for SIRE. This follows the same format as the main text.

SAS COMMANDS FOR BOX 12.4 **An analysis mixing fixed and random factors**

Commands

```
proc mixed data=gandh.Chapter12 method=type1 covtest;
  class SIRE DAM;
  model WGAIN = SIRE / htype=1 ddfm=satterth;
  random DAM (SIRE);
  run;
```

Menu route

Statistics > Anova > Mixed Models…

WGAIN → Dependent

SIRE DAM → Class

Model

☑ Fixed Effects

SIRE → Add

☑ Random Effects

DAM → Add

Select DAM in Random Effects box, and SIRE in Class box, and then click 'Nest'.

Tests

-☑ Type III

☑ Type I

☑ Tests of variance components

Options

⊙ Type 1

SAS OUTPUT FOR BOX 12.4 **An analysis mixing fixed and random factors**

The Mixed Procedure

Model Information

Data Set	GANDH.CHAPTER12
Dependent Variable	WGAIN
Covariance Structure	Variance Components
Estimation Method	Type 1
Residual Variance Method	Factor
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
SIRE	5	1 2 3 4 5
DAM	2	1 2

Dimensions

Covariance Parameters	2
Columns in X	6
Columns in Z	10
Subjects	1
Max Obs Per Subject	120
Observations Used	20
Observations Not Used	100
Total Observations	120

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
SIRE	4	1.034387	0.258597	Var(Residual) + 2 Var(DAM(SIRE)) + Q(SIRE)
DAM(SIRE)	5	1.288976	0.257795	Var(Residual) + 2 Var(DAM(SIRE))
Residual	10	0.826581	0.082658	Var(Residual)

Type 1 Analysis of Variance

Source	Error Term	Error DF	F Value	Pr > F
SIRE	MS(DAM(SIRE))	5	1.00	0.4844
DAM(SIRE)	MS(Residual)	10	3.12	0.0593
Residual	.	.	.	.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
DAM(SIRE)	0.08757	0.08359	1.05	0.2948	0.05	-0.07627	0.2514
Residual	0.08266	0.03697	2.24	0.0127	0.05	0.04035	0.2546

Fit Statistics

-2 Res Log Likelihood	17.8
AIC (smaller is better)	21.8
AICC (smaller is better)	22.8
BIC (smaller is better)	22.4

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
SIRE	4	5	1.00	0.4844

These commands will then equip you to conduct the mock analyses of section 12.6. The main text mentions that some tests are approximate, when a denominator needs to be synthesised. SAS does not make a fuss about this, but you can detect it when the denominator degrees of freedom is not an integer, but is given with decimal places. These tests are approximate.

12.8 Exercises

Examining microbial communities on leaf surfaces

SAS COMMANDS FOR BOX 12.10(A) **Analysis 1**

Commands

```
proc glm data=gandh.Chapter12;
  class TREATMNT;
  model DENSITY = TREATMNT;
  run;
```

Menu route Statistics > Anova > Linear Models

DENSITY → Dependent

TREATMNT → Class

SAS OUTPUT FOR BOX 12.10(A) **Analysis 1**

```
                                The GLM Procedure

Dependent Variable: DENSITY

                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F

Model                        1      9543.24817      9543.24817      12.50    0.0008

Error                       58     44265.88433       763.20490

Corrected Total             59     53809.13250

                  R-Square     Coeff Var      Root MSE    DENSITY Mean

                  0.177354      28.18277      27.62616        98.02500

Source                      DF     Type III SS     Mean Square    F Value    Pr > F

TREATMNT                     1     9543.248167     9543.248167      12.50    0.0008
```

SAS COMMANDS FOR BOX 12.10(B) **Analysis 2**

Commands

```
proc mixed data=gandh.Chapter12 method=type1 covtest;
  class PLANT TREATMNT;
  model DENSITY = TREATMNT / htype=1 ddfm=satterth;
  random PLANT(TREATMNT);
  run;
```

Menu route

Statistics > Anova > Mixed Models…

`DENSITY` → Dependent

`PLANT TREATMNT` → Class

Model

☑ Fixed Effects

`TREATMNT` → Add

☑ Random Effects

`PLANT` → Add

Select PLANT in Random Effects box, and TREATMNT in Class box, and then click 'Nest'.

Tests

-☑ Type III

☑ Type I

☑ Tests of variance components

Options

⊙ Type 1

SAS OUTPUT FOR BOX 12.10(B) **Analysis 2**

The Mixed Procedure

Model Information

Data Set	GANDH.CHAPTER12
Dependent Variable	DENSITY
Covariance Structure	Variance Components
Estimation Method	Type 1
Residual Variance Method	Factor
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
TREATMNT	2	1 2
PLANT	6	1 2 3 4 5 6

Dimensions

Covariance Parameters	2
Columns in X	3
Columns in Z	6
Subjects	1
Max Obs Per Subject	120
Observations Used	60
Observations Not Used	60
Total Observations	120

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
TREATMNT	1	9543.248167	9543.248167	Var(Residual) + 10 Var(PLANT(TREATMNT)) + Q(TREATMNT)
PLANT(TREATMNT)	4	14778	3694.453333	Var(Residual) + 10 Var(PLANT(TREATMNT))
Residual	54	29488	546.075389	Var(Residual)

Type 1 Analysis of Variance

Source	Error Term	Error DF	F Value	Pr > F
TREATMNT	MS(PLANT(TREATMNT))	4	2.58	0.1833
PLANT(TREATMNT)	MS(Residual)	54	6.77	0.0002
Residual	.	.	.	.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
PLANT(TREATMNT)	314.84	261.45	1.20	0.2285	0.05	-197.59	827.27
Residual	546.08	105.09	5.20	<.0001	0.05	387.02	828.63

Fit Statistics

-2 Res Log Likelihood	544.6
AIC (smaller is better)	548.6
AICC (smaller is better)	548.8
BIC (smaller is better)	548.2

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
TREATMNT	1	4	2.58	0.1833

How a nested analysis can solve problems of non-independence

The SAS output for this exercise may be found in the answers to exercises.