Repeated Measures Design

Applied Regression and Other Multivariable Methods
Sections 21-1 - 21-6

Introduction

- Previously, each exp unit measured only once
- Often studies where unit measured more than once

Example 1 Subsampling Comparing two fertilizers. A field is divided into 4 sections. Each section randomly assigned one of two fertilizers. After two weeks, three plants in each section are dug up and the number of root tips for each plant is obtained. Are the three plants in each section independent? Perhaps similar environment/soil type?

Example 2 Same subject being measured over time. Consider comparing two blood pressure medicines. In period 1, half of the subjects takes drug 1 and the other half takes drug 2. In period 2, the other drug is taken by each subject. In each period, the subject's blood pressure is measured at times 0, 1 hour, 2 hours, and 4 hours after ingestion. Each subject takes both drugs (subject = block) but some in a different order? Should the four observations in each period be considered independent? They are all from the same subject having taken the medicine only once.

- Subjects/plants usually considered random factor
- Will use EMS to perform appropriate tests
- Requires identification of crossed and nested factors

Definitions

- Factors A and B are considered crossed if

  Every level of B occurs with every level of A

  A two-factor ANOVA involves crossed factors

  Can investigate interactions

	Factor A			
Factor B	1	2	3	4
1	xx	xx	xx	xx
2	xx	xx	xx	xx
3	xx	xx	xx	xx

A	1			2			3			4		
B	1	2	3	1	2	3	1	2	3	1	2	3
	x	x	x	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x	x	x	x

- Factor A a nest factor (B nested within A) if

  Levels of B occur with only one level of A

  One can arbitrarily number levels of B

  Cannot investigate interactions

A	1			2			3			4		
B	1	2	3	4	5	6	7	8	9	10	11	12
	x	x	x	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x	x	x	x

Example 1 : Subsampling Experiment

First generate table to study crossed/nested terms

Fertilizer	1						2					
Section	1			3			2			4		
Plant	1	2	3	4	5	6	7	8	9	10	11	12
	x	x	x	x	x	x	x	x	x	x	x	x

Since we can number the sections 1-4 and plants 1-12, both of these factors are nested. Section is nested within fertilizer. Plant is nested within section which is nested within fertilizer.

```
                    Total Variability


    Within Section                  Between Section
 (plant to plant var)


                         Variation between     Variation of sections
                            Treatments            within  trts

Source               df
-----------------------
Between Sections      3
   Trt              1
   Section(Trt)     2
Within Sections       8
------------------------
Total                11


Sampling more plants within a section does not improve df for
treatment.  One would need more sections to do that.  Sampling plants
within a section simply gives a more precise estimate for that section.
```

The Statistical Model

- Similar to model in Section 21-4-3
- Section = Subject , Plant = Repeats
- Both these terms considered random

$$Y_{ijk} = \mu + \text{Sect}_{i(j)} + \text{Fert}_j + \text{Plant}_{k(ij)} \quad \begin{cases} i = 1,2 \\ j = 1,2 \\ k = 1,2,3 \end{cases}$$

μ - grand mean

Fert_j - jth fertilizer effect

$\text{Sect}_{i(j)} \frown \text{N}(0, \sigma_S^2)$

$\text{Plant}_{k(ij)} \frown \text{N}(0, \sigma^2)$

$\text{Sect}_{i(j)}$'s and $\text{Plant}_{k(ij)}$ are independent

$\text{Plant}_{k(ij)}$ can also be written $E_{k(ij)}$

SAS Commands

```
data example1;
 input fert sect plnt y @@;
 cards;
 1 1 1 18 1 1 2 22 1 1 3 17
 1 3 1 22 1 3 2 28 1 3 3 25
 2 2 1 17 2 2 2 15 2 2 3 14
 2 4 1 12 2 4 2 15 2 4 3 16
 ;

proc glm;
 class fert sect;
 model y = sect(fert) fert;
 random sect(fert) / test;
run;
quit;


** The brackets (  ) contain the nest factor
** The ‘‘/ test’’ option tells SAS to compute EMS and do tests
```


SAS Output

```
                            Sum of
Source             DF      Squares   Mean Square  F Value  Pr > F
Model               3  209.5833333    69.8611111    12.33  0.0023
Error               8   45.3333333     5.6666667
Corrected Total    11  254.9166667

R-Square     Coeff Var      Root MSE        y Mean
0.822164      12.92566      2.380476      18.41667

Source             DF  Type III SS   Mean Square  F Value  Pr > F
sect(fert)          2   55.5000000    27.7500000     4.90  0.0409
fert                1  154.0833333   154.0833333    27.19  0.0008

Source                 Type III Expected Mean Square
sect(fert)             Var(Error) + 3 Var(sect(fert))
fert                   Var(Error) + 3 Var(sect(fert)) + Q(fert)

Tests of Hypotheses for Mixed Model Analysis of Variance

Source             DF  Type III SS   Mean Square  F Value  Pr > F
sect(fert)          2    55.500000     27.750000     4.90  0.0409
Error: MS(Error)    8    45.333333      5.666667

Source             DF  Type III SS   Mean Square  F Value  Pr > F
fert                1   154.083333    154.083333     5.55  0.1426
Error               2    55.500000     27.750000
Error: MS(sect(fert))
```


Example 2 : Crossover Experiment I

Consider an experiment to compare a new blood pressure medicine (Drug 1) with a control (Drug 2). The experiment involves 8 subjects of which four take Drug 1 first followed by Drug 2. The other four will take Drug 2 first followed by Drug 1. First, suppose the systolic blood pressure was measured only at 1 hour after ingestion. The results of this experiment are presented below.

Order	1 followed by 2				2 followed by 1			
Subject	1	2	3	4	5	6	7	8
Period 1	132	130	142	138	126	152	144	144
Period 2	137	138	145	152	161	164	160	166

Subject nested within Order. Period crossed with Subject. Period also crossed with Order. Period*Order combination defines a specific Trt (Drug 1 or Drug 2).

```
 Order                1      Variability Between Subjects
 Subject(Order)       6          8 subjects -> 7 df
----------------------
 Period               1
 Order*Period = Trt   1      Variability Within Subjects
 Period*Subject       6          2 obs/subj -> 8 df
----------------------
Total                15
```

Example 2 : Crossover Experiment I

This example is like the model in 21-4-4. The nest factor is Order with Subjects nested within order. The crossed factors are Order and Period. Because the Order defines which treatment is taken each Period, the Order*Period interaction is the same as the Treatment effect in this example.

$$Y_{ijk} = \mu + \text{Subj}_{i(j)} + \text{Ord}_j + \text{Per}_k + \text{Trt}_{jk} + E_{k(ij)} \quad \begin{cases} i = 1,2,3,4 \\ j = 1,2 \\ k = 1,2 \end{cases}$$

μ - grand mean

Ord_j - jth order effect

Per_k - kth period effect

Trt_{jk} - trt effect

$\text{Subj}_{i(j)} \frown \text{N}(0, \sigma_S^2)$

$E_{k(ij)} \frown \text{N}(0, \sigma^2)$

SAS Commands

```
data example2;
 input order subj period y @@;
 trt = 1;
 if order = 1 & period = 2 then trt = 2;
 if order = 2 & period = 1 then trt = 2;
 cards;
 1 1 1 132 1 1 2 137 1 2 1 130 1 2 2 138
 1 3 1 142 1 3 2 145 1 4 1 138 1 4 2 152
 2 1 1 126 2 1 2 161 2 2 1 152 2 2 2 164
 2 3 1 144 2 3 2 160 2 4 1 144 2 4 2 166
 ;

proc glm;
 class order subj period;
 model y = order subj(order) period period*order;
 random subj(order) / test;

proc glm;
 class order subj period trt;
 model y = order subj(order) period trt;
 random subj(order) / test;
run;
quit;
```


SAS Output

```
                              Sum of
Source             DF        Squares   Mean Square  F Value  Pr > F
Model               9    2115.562500    235.062500     7.59  0.0114
Error               6     185.875000     30.979167
Corrected Total    15    2301.437500

R-Square     Coeff Var      Root MSE        y Mean
0.919235      3.820433      5.565893      145.6875

Source             DF    Type III SS   Mean Square  F Value  Pr > F
order               1    663.0625000   663.0625000    21.40  0.0036
subj(order)         6    436.8750000    72.8125000     2.35  0.1611
period              1    826.5625000   826.5625000    26.68  0.0021
trt                 1    189.0625000   189.0625000     6.10  0.0484

Source                   Type III Expected Mean Square
order                    Var(Error) + 2 Var(subj(order)) + Q(order)
subj(order)              Var(Error) + 2 Var(subj(order))
period                   Var(Error) + Q(period)
trt                      Var(Error) + Q(trt)

Tests of Hypotheses for Mixed Model Analysis of Variance

Source             DF    Type III SS   Mean Square  F Value  Pr > F
order               1     663.062500    663.062500     9.11  0.0235
Error               6     436.875000     72.812500
Error: MS(subj(order))

Source             DF    Type III SS   Mean Square  F Value  Pr > F
subj(order)         6     436.875000     72.812500     2.35  0.1611
period              1     826.562500    826.562500    26.68  0.0021
trt                 1     189.062500    189.062500     6.10  0.0484
Error: MS(Error)    6     185.875000     30.979167
```


Repeated Measures / Longitudinal Study

- In Example 1, 3 plants/section measured at 2 wks
  - Analysis: Based on using section means (4 EUs)
  - Repeat plants per section do not increase $\text{df}_{\text{Section(Trt)}}$
- What if plants measured at diff times (1, 2, 3 wks)
- How do we perform analysis?

1 Compute single summary of information over time
  - Peak response or total concentration in body
  - Response mean (like Example 1)
  - Typically RCBD or one-way ANOVA on summary statistic

2 Bring time into the model as a factor
  - Similar to Period factor in Example 2
  - Interaction of trts with time
  - Look at shape of response curve over time

Repeated Measures / Longitudinal Study

- Common to take similar approach as Example 1
- But also break up Within Section variability

```
Source                  df
-------------------------
Between Sections         3
  Trt             1
  Section(Trt)    2
Within Sections          8
  Time            2
  Time*Trt        2
  Time*Section(Trt)  4
-------------------------
Total                   11
```

- Problems with Assumptions
  - With large changes in response over time, may have problems with constant variance assumption
  - Observations over time within a subject are likely correlated. Adjacent times may be more correlated than times further apart.
- Other approaches (when time a factor)
  - Multivariate analysis (takes correlation into account)
  - Proc Mixed to model correlation structure

The Statistical Model

This example is again like the model in 21-4-4. The nest factor is Fertilizer with Section nested within Fertilizer. The crossed factors are Fertilizer and Time. Here, the Fertilizer*Time interaction tells if the Fertilizers' response curves over time are parallel.

$$Y_{ijk} = \mu + \mathsf{F}_j + \mathsf{S}_{i(j)} + \mathsf{T}_k + (\mathsf{F}*\mathsf{T})_{jk} + E_{ik(j)} \quad \begin{cases} i = 1,2 \\ j = 1,2 \\ k = 1,2,3 \end{cases}$$

μ - grand mean

F_j - jth fertilizer effect

T_k - kth week effect

$\mathsf{F*T}_{jk}$ - interaction effect

$\mathsf{S}_{i(j)} \frown \mathsf{N}(0, \sigma_S^2)$

$E_{ik(j)} \frown \mathsf{N}(0, \sigma^2)$

$$\mathsf{Var}(Y_{ijk}) = \sigma_S^2 + \sigma^2$$
$$\mathsf{Cor}(Y_{ijk}, Y_{ijk'}) = \sigma_S^2/(\sigma_S^2 + \sigma^2)$$

Any two obs in same section/subject has same correlation

Known as assumption of compound symmetry

This model approach appropriate when repeated measures have compound symmetry

SAS Commands

```
** This is the same data set as Example 1 except that plant has been
   replaced with time.


data example1;
 input fert sect time y @@;
 cards;
 1 1 1 18 1 1 2 22 1 1 3 17
 1 3 1 22 1 3 2 28 1 3 3 25
 2 2 1 17 2 2 2 15 2 2 3 14
 2 4 1 12 2 4 2 15 2 4 3 16
 ;

proc glm;
 class fert sect time;
 model y = sect(fert) fert time fert*time;
 random sect(fert) / test;
 lsmeans fert*time / slice=time stderr;
run;
quit;


** The brackets (  ) contain the nest factor
** The ‘‘/ test’’ option tells SAS to compute EMS and do tests
```


SAS Output

```
                                      Sum of
Source                     DF        Squares    Mean Square   F Value   Pr > F
Model                       7    237.9166667     33.9880952      8.00   0.0313
Error                       4     17.0000000      4.2500000
Corrected Total            11    254.9166667

R-Square     Coeff Var      Root MSE        y Mean
0.933312      11.19395      2.061553      18.41667

Source                     DF    Type III SS    Mean Square   F Value   Pr > F
sect(fert)                  2     55.5000000     27.7500000      6.53   0.0550
fert                        1    154.0833333    154.0833333     36.25   0.0038
time                        2     16.1666667      8.0833333      1.90   0.2627
fert*time                   2     12.1666667      6.0833333      1.43   0.3397

Source                     Type III Expected Mean Square
sect(fert)                 Var(Error) + 3 Var(sect(fert))
fert                       Var(Error) + 3 Var(sect(fert)) + Q(fert,fert*time)
time                       Var(Error) + Q(time,fert*time)
fert*time                  Var(Error) + Q(fert*time)

Tests of Hypotheses for Mixed Model Analysis of Variance

   Source                 DF    Type III SS    Mean Square  F Value  Pr > F
   sect(fert)              2      55.500000      27.750000     6.53  0.0550
*  time                    2      16.166667       8.083333     1.90  0.2627
   fert*time               2      12.166667       6.083333     1.43  0.3397
   Error: MS(Error)        4      17.000000       4.250000
* This test assumes one or more other fixed effects are zero.

   Source                 DF    Type III SS    Mean Square  F Value  Pr > F
*  fert                    1     154.083333     154.083333     5.55  0.1426
   Error                   2      55.500000      27.750000
Error: MS(sect(fert))
* This test assumes one or more other fixed effects are zero.
```

Analysis Using Proc Mixed

- Consider covariance of obs within subject
- MIXED allows for different covariance structures
- Consider three time points per subject
- **Compound Symmetry**

$$\begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 \end{bmatrix}$$

- **Unstructured**

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{21} & \sigma_{31} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{32} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 \end{bmatrix}$$

- **First order autoregressive**

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

Using Proc Mixed

```
data example1;
 input fert sect time y @@;
 cards;
 1 1 1 18 1 1 2 22 1 1 3 17
 1 3 1 22 1 3 2 28 1 3 3 25
 2 2 1 17 2 2 2 15 2 2 3 14
 2 4 1 12 2 4 2 15 2 4 3 16
 ;

proc mixed;
 class fert sect time;
 model y = fert time fert*time;
 repeated time / subject=sect type=cs;

proc mixed;
 class fert sect time;
 model y = fert time fert*time;
 repeated time / subject=sect type=ar(1);
run;
quit;
```



```
Compound Symmetry Structure

Cov Parm     Subject    Estimate
CS           sect         7.8333  ** sigma1sq
Residual                  4.2500  ** sigmasq

           Fit Statistics
-2 Res Log Likelihood            33.6
AIC (smaller is better)          37.6
AICC (smaller is better)         41.6
BIC (smaller is better)          36.4

        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
fert            1       2       5.55    0.1426
time            2       4       1.90    0.2627
fert*time       2       4       1.43    0.3397
-----------------------------------------------------
AR(1) Structure

Cov Parm     Subject    Estimate
AR(1)        sect         0.8192  ** rho
Residual                 12.9719  ** sigmasq

           Fit Statistics
-2 Res Log Likelihood            32.1
AIC (smaller is better)          36.1
AICC (smaller is better)         40.1
BIC (smaller is better)          34.9

        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
fert            1       2       4.68    0.1631
time            2       4       4.54    0.0934
fert*time       2       4       3.55    0.1300
```


Example 4 : Crossover Experiment w/ repeated measures

Consider the same experiment as described in Example 2 except that instead of only measuring the blood pressure at one hour, the blood pressure is also measured at 30, 60, 90 and 120 minutes.

This creates two sets of repeated measures. First, measurements within a subject across periods and now measurements within a subject within a period.

The ANOVA table would look like

```
 Order                  1     Variability Between Subjects
 Subject(Order)         6         8 subjects -> 7 df
------------------------
 Period                 1     Variability Within Subjects
 Order*Period = Trt     1          Between Periods
 Period*Subject         6         2 per/subj -> 8 df
------------------------
 Time                   3
 Time*Order             3     Variability Within Subjects
 Time*Subject(Order)   18          Within Periods
------------------------          4 obs/per/subj -> 48 df
 Time*Period            3
 Time*Trt               3
 Time*Period*Subject   18
------------------------
 Total                 63
```