BIOMETRICS INFORMATION

(You're 95% likely to need this information)

SUBJECT: Analysing a split-plot in time with the proper repeated measures ANOVA

Many forestry experiments require repeated measures of the treatment units over time. These have traditionally been analysed using a split-plot ANOVA with time as the split-plot factor even though a repeated measures ANOVA would be more appropriate. Generally, most researchers (and many statisticians!) are familiar and comfortable with split-plot designs but have little experience with the repeated measures variety. This pamphlet will use an example to show how to change the SAS program for a split-plot ANOVA into a repeated measures analysis. Note that in repeated measures parlance, the mainplot factors are called between-subject factors while the split-plot factors are within-subject factors. We might prefer to call these between-plot and within-plot factors, but this is not standard terminology.

The example data is taken from page 332 of Milliken and Johnson. The experiment had three drugs each administered to eight persons (the treatment unit, t.u.) and the response variable was measured at four times for each person. In forestry, the three drugs could be three vegetation management treatments each applied to eight different plots (the t.u.'s) in a completely randomized manner. The percent cover of vegetation could be the response variable measured each year, for four years. The split-plot ANOVA table would be:

Source of Variation		Degrees of Freedom	Error Term
Mainplot:			
Vegetation Treatment	D	d-1 = 2	P(D)
Plots (Veg. Trt)	P(D)	d(p-1) = 21	--
Split-Plot:			
Time	T	t-1 = 3	T x P(D)
T x D		(t-1)(d-1) = 6	T x P(D)
T x P(D)		d(p-1)(t-1) = 63	--

The following SAS program could be used to do the ANOVA analysis including some tests for polynomial trends with time (the `contrast` statements).

```
/* Usual split-plot analysis */

proc glm data=long;
title 'Data taken from Messy Data, page 332';
title2 'Usual split-plot analysis';
  class D T P;
  model y = D P(D) T T*D;
```

```
  test h = D e = P(D);
  contrast 'Linear: time' T -3 -1  1  3;
  contrast 'Quadr:  time' T  1 -1 -1  1;
  contrast 'Cubic:  time' T -1  3 -3  1;
  lsmeans D T T*D P(D);
run;
```

This program requires that each observation or record in the data set contain the variables D, T and P plus the variable y designating the observed response for one plot at one time or year. The repeated measures analysis requires that the data be arranged differently. Each record should contain the response for one plot with a separate variable for the response at each time. For instance, the data set might have 4 variables called t1, t2, t3, and t4, which are the observed responses for each plot at years 1, 2, 3, and 4. In practice, data is often stored this way and must be converted for the split-plot analysis. The SAS program for the repeated measures analysis is:

```
/* Usual Repeated Measures Analysis */

proc glm data=repeat;
title2 'Usual Repeated Measure Analysis';
  class D ;
  model t1 t2 t3 t4 = D / nouni;
  repeated T 4 (1 2 3 4) polynomial / summary nom;
                    /* include printe for sphericity test */;
run;
```

The final ANOVA table for both analyses is:

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F-values Split-plot	F-values Rep. Meas.
D	2	1333.00	666.50	5.99	5.99
P(D)	21	2337.91	111.33	--	--
T	3	289.61	96.54	12.96	12.96
Linear	1	11.10	11.10	1.49	0.91
Quadratic	1	243.84	243.84	32.74	49.64
Cubic	1	34.67	34.67	4.65	6.70
T x D	6	527.42	87.90	11.80	11.80
T x P(D)	63	469.22	7.45	--	--
Linear	21	257.33	12.25	--	--
Quadratic	21	103.16	4.91	--	--
Cubic	21	108.73	5.18	--	--

Except for the partitioning of the error term (T x P(D)) into polynomial contrasts, all the sums of squares (SS's) are easily obtained from the output of the split-plot analysis. Identifying the appropriate numbers from the output of the repeated measures analysis is a less familiar task. The following is the output of the repeated measures analysis.

- -

Data taken from Messy Data, page 332
Usual Repeated Measure Analysis

General Linear Models Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
D	2	1333.000000	666.500000	5.99	0.0088
Error	21	2337.906250	111.328869		

- -

General Linear Models Procedure
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

Source: T

					Adj	Pr > F
DF	Type III SS	Mean Square	F Value	**Pr > F**	**G - G**	**H - F**
3	289.61458333	96.53819444	12.96	0.0001	0.0001	0.0001

Source: T*D

					Adj	Pr > F
DF	Type III SS	Mean Square	F Value	**Pr > F**	**G - G**	**H - F**
6	527.41666667	87.90277778	11.80	0.0001	0.0001	0.0001

Source: Error(T)

DF	Type III SS	Mean Square
63	469.21875000	7.44791667

Greenhouse-Geisser Epsilon = 0.7769
Huynh-Feldt Epsilon = 0.9630

- -

General Linear Models Procedure
Repeated Measures Analysis of Variance
Analysis of Variance of Contrast Variables

T.N represents the nth degree polynomial contrast for T

Contrast Variable: T.1   *Output used to test for a linear trend over time*

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	11.10208333	11.10208333	0.91	0.3520
D	2	83.71666667	41.85833333	3.42	0.0520
Error	21	257.33125000	12.25386905		

```
Contrast Variable: T.2       Output used to test for a quadratic trend over time

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

MEAN                       1     243.84375000    243.84375000      49.64    0.0001
D                          2     397.75000000    198.87500000      40.49    0.0001

Error                     21     103.15625000      4.91220238

Contrast Variable: T.3        Output used to test for a cubic trend over time

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

MEAN                       1      34.66875000     34.66875000       6.70    0.0172
D                          2      45.95000000     22.97500000       4.44    0.0247

Error                     21     108.73125000      5.17767857
```

- -

Note that there are two obvious differences between these two analyses:

1) The split-plot output only provides one set of probability values for the F-tests of the split-plot sources, while the repeated measures provides three (see output above in section titled **Univariate Tests of Hypotheses for Within Subject Effects**). If all three probabilities agree then the appropriate decisions are obvious. If they disagree then the adjusted probabilities (noted by `Adj Pr > F` on the output) should be used for decision-making.

2) The repeated measures analysis uses separate estimates of error for each contrast. The split-plot analysis pools these separate estimates into one since homogeneity of variance is assumed. The occurrence of autocorrelation between successive time measurements invalidates this assumption, hence the repeated measures analysis uses the more appropriate error terms for the time contrast tests.

There are many more points that should be made about performing and interpreting repeated measures analysis (for example, what is autocorrelation? what does a sphericity test test? and what is the difference between the G-G (for Greenhouse-Geisser) probabilities and the H-F (for Huynh-Feldt) probabilities?). The references below (and the SAS/Stat manual) may help you deal with them and later pamphlets may cover some of these topics.

References:

Littel, Ramon C., 1989, *Statistical Analysis of Experiments with Repeated Measurements*, HortScience 24: 36-40.

Milliken, George A. and Johnson, Dallas E., 1984, *Analysis of Messy Data, Volume I*: *Designed Experiments*, Lifetime Learning Publications, Belmont, CA.

CONTACT: Wendy Bergerud
387-5676

─────────────────────────────NEW PROBLEM─────────────────────────────

The following is a typical randomized block repeated measures design. Write the SAS program to do a split-plot analysis and to do a repeated measures analysis.

Source of Variation		Degrees of Freedom	Error Term
Mainplot:			
Blocks	B	b-1 = 7	--
Vegetation Treatment	D	d-1 = 2	B x D
B x D		(b-1)(d-1) = 14	--
Split-Plot:			
Time	T	t-1 = 3	⎰ pooled
T x D		(t-1)(d-1) = 6	⎱ error
T x B		(t-1)(b-1) = 21 ⎤ pooled	--
T x B x D		(t-1)(b-1)(d-1) = 42 ⎦ error	--

─────────────────────────ANSWER FOR PAMPHLET #31─────────────────────────

Model	Number of Parameters	df	SSR Name	SSR Value	Difference	Proportion of SST (R^2)
1. K Separate Lines	2k = 12	19	SSR**KL**	90.57		0.10
					24.00	0.03
2. K Parallel Lines	k+1 = 7	24	SSR**KLP**	114.56		0.13
					630.89	0.73
3. Group Means	k = 6	25	SSR**WG**	745.45		0.56
					not meaningful	--
4. One Line	2	29	SSR**L**	265.01		0.86
					604.86	0.70
5. Grand Mean	1	30	SST	869.87		1.00

   The test for homogeneity of regression is $F = \dfrac{(114.56-90.57)/5}{90.56/19} = 1.01$, df = 5, 19. The resulting p-value is 0.44 so that homogeneity of regression appears to be a reasonable assumption. Hence the usual ANCOVA table can be formed as shown on the next page.

Source of Variation	Sums of Squares Notation	df	Sums of Squares	Mean Square	F-value	p-value
Between Groups	SSR**BG** = SSRL-SSR**KLP**	5	150.45	30.09	6.30	0.0007
Covariate	SS**Cov** = SSR**WG**-SSR**KLP**	1	630.89	630.89	60.58	0.0001
Error	SSR**KLP**	24	114.56	14.2		
Total	SST	30	869.87			

Appendix: The SAS program to read in the data for the analyses described in the text.

```
/* Messy332.sas */

/* Reading in the data and creating two SAS data sets:
        1) Long:   for the split-plot analysis
        2) Repeat: for the repeated measures analysis */

data long    (keep=P D T y)
     repeat (keep=P D t1-t4);
   array tm{4} t1-t4;
   do D = 1 to 3;
      P + 1;
      input t1-t4 @@;
      output repeat;
      do T = 1 to 4;
         y = tm{T};
         output long;
   end; end;
   cards;
72 86 81 77 85 86 83 80 69 73 72 74
78 83 88 81 82 86 80 84 66 62 67 73
71 82 81 75 71 78 70 75 84 90 88 87
72 83 83 69 83 88 79 81 80 81 77 72
66 79 77 66 86 85 76 76 72 72 69 70
74 83 84 77 85 82 83 80 65 62 65 61
62 73 78 70 79 83 80 81 75 69 69 68
69 75 76 70 83 84 78 81 71 70 65 63
; * book has last entry of 65 instead of 63!!;

title 'Data taken from Messy Data, page 332';
run;
```