Topic 12. The Split-plot Design and its Relatives *(continued)*
Repeated Measures

12.9 Repeated measures analysis

Sometimes researchers make **multiple measurements on the same experimental unit**. We have encountered such measurements before and called them **subsamples**; and in the context of the nested designs we have discussed, the primary uses of subsamples are:

1. To obtain a better estimate of the true value of an experimental unit (i.e. to reduce experimental error)
2. To estimate the components of variance in the system.

In these simple nested experiments, "subsample" was not a classification variable but merely an ID. That is, "Subsample 1" from experimental unit 1 was no more similar to "Subsample 1" from experimental unit 2 than it was to "Subsample 2" from experimental unit 2. Even though no two measurements can be made at exactly the same instant in time, we did not refer to these subsamples as repeated measures because we were not interested in characterizing the effect of time between measurements on the response variable.

Now assume that this effect of time is of interest and, again, that the measurements are made repeatedly on the **same experimental unit** (e.g. plant height **over weeks**; yield of perennial crops **over seasons**; animal growth **over months**; population dynamics **over years**, etc.). These observations are not replications because they are made on the same experimental unit (i.e. they are **not independent**). But neither are they subsamples, as defined above, because:

1. We are interested in the effect of time between measurements.
2. "Subsample" is no longer simply an ID: "Subsample 1 from e.u. 1" (e.g. Height of Plant 1 at Week 1) **is** more related to "Subsample 1 from e.u. 2" (Height of Plant 2 at Week 1) than it is to "Subsample 2 from e.u. 2" (Height of Plant 2 at Week 2).

Not replications, not subsamples, these sorts of measurements possess qualities of both and are called **repeated measurements**; and the split-model model offers one means of analyzing such data. It is important to note in this context that an important assumption of the ANOVA is not the independence of *measurements* but the independence of *errors*. Indeed, the split-plot model implicitly assumes a correlation among measurements.

Before outlining the split-plot approach to repeated measures data, know that time series or multivariate methods can also be used to analyze such data. Time series methods are more appropriate when analyzing long series of data, with more than 20 repeated measures per individual. Consequently, such analyses are more frequently applied to stock data or weather data than to agricultural experiments. Multivariate methods are very useful for shorter time series, and researchers who consistently rely on repeated

measures in their experiments should become familiar with them. That being said, the far simpler univariate analysis presented here is also quite useful.

> Note that what distinguishes repeated-measures data from any other multivariate data is not so much the existence of the repeated measurements but the desire to examine *changes* in the measurements taken on each subject.

By simply designating the different points in time as "levels" of a "Time" factor, the split-plot principle can be applied to experiments where successive observations are made on the same experimental unit over a period of time. For example, a fertilizer trial or variety trial with a perennial crop like alfalfa might be harvested several times. Other examples might be repeated picking of fruit from the same trees in an orchard or repeated soil sampling of plots over time for nutrient content. In each case, the experimental units to which the treatment levels are assigned are the "main plots," and the several measurements over time are the "subplots." A subplot in this case, however, differs from the usual subplot in that it consists of data taken from *the entire main plot* rather than from some designated portion of the main plot, as is the case with the usual split-plot.

In the simple, univariate split-plot approach to repeated measures data, the dependency among the observations within a main plot is used to adjust the degree of freedoms in the ANOVA to give approximate tests. The approximate ANOVA for a repeated measures CRD is shown below.

12. 9. 1. Repeated measures ANOVA

Approximate ANOVA of repeated measurement analysis

Source	df	SS	MS	Conservative df
***Among* Experimental Units**				
Treatment (A)	a-1	SSA	SSA/(a-1)	
Rep* Trt (Error A)	a (n-1)	SS(MPE)	SS(MPE)/a(n-1)	
***Within* Experimental Units**				
Response in time (B)	b-1	SSB	SSB /(b-1)	1
Response by Trt. (A*B)	(b-1)(a-1)	SSAB	SSAB/(b-1)(a-1)	a-1
Error B	a(b-1)(n-1)	SS(SPE)	SS(SPE)/a (b-1) (n-1)	a(n-1)

The analysis looks like a split-plot analysis, except that *conservative degrees of freedom* are used in all F-tests for effect of Time (i.e. the repeated measures) and any interactions with Time. No unusual problems arise when analyzing the effects of the main plot (A) because the analysis of the main plot is insensitive to the split, as seen before in the normal split-plot. However, F values generated by testing the effects of Time (subplot B) and the interaction of main plots treatments with Time (A*B) may not follow an F distribution, thereby generating erroneous results.

The normal split-plot model assumes that pairs of observations within the same main plot are equally correlated. With repeated-measures data, however, arbitrary pairs of observations on the same experimental unit are not necessarily equally correlated. Measurements close in time are often more highly correlated than measurements far apart in time. Since this unequal correlation among the repeated measurements is ignored in a simple split-plot analysis, tests derived in this manner may not be valid.

To compensate for this assumption of uniform correlation across repeated measurements, a **conservative approach** is recommended by many statisticians, "conservative" because it requires larger F values to declare significance for B and A*B effects. In this approach, it is suggested that the degrees of freedom of B (response in time) be used to scale the degrees of freedom for B, A*B, and Error B. Finally, critical *F* values should be used that are based upon these conservative degrees of freedom (previous table, right column).

The uncorrected degrees of freedom are appropriate for independent replications within main plots. The corrected ones (right column) are appropriate for totally dependent replications, a situation equivalent to having all responses represented by a single response (this explains why the corrected df is one). Total dependency is the worst theoretically-possible scenario and is therefore a severe condition to impose. The true level of dependency among repeated measurements in a real experiment will probably be somewhere between these two extremes.

12. 9. 2. Example of a repeated measurements experiment

An experiment was carried out to study the differences in yield of four alfalfa cultivars. Five replications of these four varieties were organized according to a CRD, and four cuttings were made of each replication over time. The data represents the repeated measurements of yield (tons/acre) of the four cultivars and is analyzed as a split-plot CRD with repeated measures:

    4 levels of main plot A: **Cultivars** 1 – 4
    4 levels of subplot B: **Cut times** 1 – 4 (9/10/74, 6/25/75, 18/5/75, 9/16/75)

To analyze this data, we begin by carrying out a standard split-plot analysis.

SAS program

```
data rem_mes;
input rep A_var B_time yield @@;
cards;
1 1 1  2.80191  1 1 2  3.73092  1 1 3  3.09856  1 1 4  2.50965
1 2 1  2.76212  1 2 2  5.40530  1 2 3  3.82431  1 2 4  2.72992
1 3 1  2.29151  1 3 2  3.81140  1 3 3  2.92575  1 3 4  2.39863
1 4 1  2.56631  1 4 2  4.96070  1 4 3  2.81734  1 4 4  2.05752
2 1 1  2.96602  2 1 2  4.43545  2 1 3  3.10607  2 1 4  2.57299
2 2 1  3.09636  2 2 2  3.90683  2 2 3  3.26229  2 2 4  2.58614
2 3 1  2.54027  2 3 2  3.82716  2 3 3  2.86727  2 3 4  2.16287
2 4 1  2.31630  2 4 2  3.96629  2 4 3  2.91461  2 4 4  2.15764
```

```
3 1 1  2.43232  3 1 2  4.32311  3 1 3  2.81030  3 1 4  2.07966
3 2 1  3.09917  3 2 2  4.08859  3 2 3  3.13148  3 2 4  2.60316
3 3 1  2.41199  3 3 2  4.08317  3 3 3  3.03906  3 3 4  2.07076
3 4 1  2.65834  3 4 2  3.71856  3 4 3  2.92922  3 4 4  2.15684
4 1 1  2.93509  4 1 2  3.99711  4 1 3  2.77971  4 1 4  2.44033
4 2 1  2.65256  4 2 2  5.42879  4 2 3  2.70891  4 2 4  2.30163
4 3 1  2.30420  4 3 2  3.27852  4 3 3  2.72711  4 3 4  2.04933
4 4 1  2.47877  4 4 2  3.92048  4 4 3  3.06191  4 4 4  2.35822
5 1 1  2.42277  5 1 2  3.85657  5 1 3  3.24914  5 1 4  2.34131
5 2 1  2.63666  5 2 2  3.77458  5 2 3  3.09734  5 2 4  2.30082
5 3 1  2.36941  5 3 2  3.44835  5 3 3  2.50562  5 3 4  2.08980
5 4 1  2.23595  5 4 2  4.02985  5 4 3  2.85279  5 4 4  1.85736
;
proc glm;
  class rep A_var B_cut;
  model yield= A_var rep*A_var B_time A_var*B_cut;
    test h=A_var e=rep*A_var;
    means A_var / lsd e=rep*A_var;
    means B_time/ lsd;
run; quit;
```

Output (split plot CRD)

Class Level Information

```
Class    Levels   Values
REP        5      1 2 3 4 5
A_VAR      4      1 2 3 4
B_TIME     4      1 2 3 4 Number of observations in data set = 80
```

Dependent Variable: **YIELD**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	31	42.884847	1.383382	14.46	0.0001
Error	48	4.592605	0.095679		
Corrected Total	79	47.477452			

R-Square	C.V.	Root MSE	YIELD Mean
0.903268	10.33208	0.3093	2.9938

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A_VAR	3	2.840753	0.946918	~~9.90~~	~~0.0001~~
REP*A_VAR	16	2.048682	0.128043	1.34	0.2141
B_TIME	3	37.447690	12.482563	130.46	0.0001
A_VAR*B_TIME	9	0.547721	0.060858	0.64	0.7606

Tests of Hypotheses using the Type III MS for REP*A_VAR as error term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A_VAR	3	2.8407531	0.9469177	7.40	0.0025

The adjusted degrees of freedom are

Source	DF	Type I SS	Mean Square	F Value	Pr > F
B_TIME	1	37.447690	37.447690	130.46	$F_{1,16}$ **
A_VAR*B_TIME	3	0.547721	0.182574	0.64	$F_{3,16}$ NS
Error	16	4.592605	0.287038		

The *F* test for Time (B) is significant, despite the penalty imposed by conservative df. We can conclude that there are significant differences among cuttings. And since no interactions were detected, subsequent analysis of the main effects is appropriate.

Analysis of the main effect of Cultivar (A) is generated simply by the code above. The important point to remember is to specify the correct error term.

```
T tests (LSD) for variable: YIELD
Alpha= 0.05  df= 16   MSE= 0.128043
Critical Value of T= 2.12, Least Significant Difference= 0.2399

Means with the same letter are not significantly different.
T Grouping          Mean     N   A_VAR
        A         3.2698    20   2
B       A         3.0444    20   1
B       C         2.9008    20   4
        C         2.7601    20   3
```

For the effects of Time (subplot B), however, it is a bit more difficult, requiring some hand calculation in order to stay true to the conservative df approach. A systematic way to proceed here is to request a means separation table *as though the repeated measures were truly independent* from one another. The code above does this, giving the following results:

```
              t Tests (LSD) for Yield
Alpha                         0.05
Error Degrees of Freedom        48 → 16
Error Mean Square         0.095679 → 0.28703782
Critical Value of t        2.01063 → 2.120 (t₀.₀₅ with 16 df )
Least Significant Difference 0.1967 → 0.718349
```

t Grouping	Mean	N	B_Cut
A	4.09959	20	2
B	2.98544	20	3
C	2.59890	20	1
D	2.29123	20	4

This table is nice because it provides a template for all the adjustments that need to be made. The dfError, the MSE, the critical value, and the LSD must all be adjusted, as shown above. Recall that:

$$LSD = t_{\frac{\alpha}{2}, df_{MSE}} \sqrt{MSE \frac{2}{r}} = 2.12\sqrt{0.287\frac{2}{5}} = 0.718$$

where r = 5 in this study. Since we are assuming complete correlation between the data, the number of real replications is only 5, instead of the 20 obtained when all the times
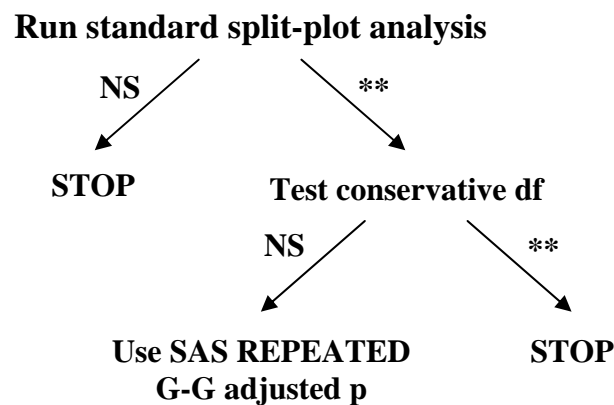
were considered independent replications.
.

This new least significant difference can be used to adjust the significance groupings:

t Grouping	Mean	N	B_Cut
A	4.09959	5	2
B	2.98544	5	3
B	2.59890	5	1
B	2.29123	5	4

Repeated Measures Decision Tree

Below is a schematic illustrating the overall process to follow when dealing with repeated measures data.

Run standard split-plot analysis

NS   **

STOP   **Test conservative df**

NS   **

Use SAS REPEATED   **STOP**
G-G adjusted p

Explanation:

1. Analyze the experiment as a simple split-plot design, treating all repeated measurements as true replications of the subplot effect (Time).

2a. If the results of this split-plot analysis are NS, STOP. Conclusion: NS.

> *You stop here because, if you do not detect differences with this unreasonably liberal analysis, there are no real differences due to Time.*

2b. If the results are significant, proceed to a conservative df analysis.

3a. If the results of conservative df analysis are significant, STOP. Conclusion: Significant.

> *You stop here because, if you detect difference with this unreasonably conservative analysis, there are real differences due to Time.*

4. Finally, if the original split-plot analysis (too liberal) is significant and the conservative df analysis (too conservative) is NS, you should proceed with a more nuanced analysis, like the one available through SAS's REPEATED statement (see next section).

12. 9. 3. Univariate repeated-measures analysis using the REPEATED statement

In the previous example there was no uncertainty because the adjusted and unadjusted df analyses produced the same conclusions. In some cases, however, adjusted and unadjusted df can produce different results. A more precise test, somewhere between the two extremes of complete independence and complete dependence of measurements, can be obtained using SAS's REPEATED statement, designed specifically for this situation.

"Repeated measures" is a special case of multivariate analysis of variance (**MANOVA**). In a MANOVA, the four cutting times in the present example would be treated as *four response variables* (not four levels of a classification variable), with only Variety treated as a classification variable. To use the REPEATED statement, therefore, one must re-organize the dataset. In this case, the dataset must be converted from one in which there are two classification variables (Variety and Time) and one response variable (Yield) to a dataset in which there is one classification variable (Variety) and *four* response variables (Yield of Cut 1, Yield of Cut 2, Yield of Cut 3, and Yield of Cut 4). In SAS, the new data entry, and the appropriate Repeated analysis, would look like this.

```
Data RepMeas;
    Input Rep Var Cut1 Cut2 Cut3 Cut4 @@;
Cards;
1 1 2.80191   3.73092   3.09856   2.50965
1 2 2.76212   5.40530   3.82431   2.72992
1 3 2.29151   3.81140   2.92575   2.39863
1 4 2.56631   4.96070   2.81734   2.05752

2 1 2.96602   4.43545   3.10607   2.57299
2 2 3.09636   3.90683   3.26229   2.58614
2 3 2.54027   3.82716   2.86727   2.16287
2 4 2.31630   3.96629   2.91461   2.15764

3 1 2.43232   4.32311   2.81030   2.07966
3 2 3.09917   4.08859   3.13148   2.60316
3 3 2.41199   4.08317   3.03906   2.07076
3 4 2.65834   3.71856   2.92922   2.15684

4 1 2.93509   3.99711   2.77971   2.44033
4 2 2.65256   5.42879   2.70891   2.30163
4 3 2.30420   3.27852   2.72711   2.04933
4 4 2.47877   3.92048   3.06191   2.35822

5 1 2.42277   3.85657   3.24914   2.34131
5 2 2.63666   3.77458   3.09734   2.30082
5 3 2.36941   3.44835   2.50562   2.08980
5 4 2.23595   4.02985   2.85279   1.85736
;
Proc GLM;
    Class Var;
    Model Cut1 Cut2 Cut3 Cut4 = Var / NoUni;
    Repeated Time / NoM;
Run;
Quit;
```

Take a look at the **Proc GLM** above. As described, we now have a single classification variable (Variety) and four response variables (Cut1 – Cut4, designated as response variables by their position on the left-hand side of the Model equation). The option "/ NoUni" is used to tell SAS to **no**t perform four separate **uni**variate analyses (i.e. four ANOVAs), one for each response variable. Remember, we're entering the world of MANOVA here.

At this point, to perform a repeated measures analysis, one follows the Model statement with the REPEATED statement, the basic syntax of which is:

<div align="center">

REPEATED *factor* / NoM;

</div>

(there are more complex forms of the statement). Here, "*factor*" is the name you give to the *within-subjects* (i.e. subplot) factor. Notice that the name of this factor (e.g. "Time," though you can choose any name you want) has not appeared previously. The option "/ NoM" tells SAS to perform a univariate ANOVA on this newly defined factor and **not** a **M**ANOVA. You are getting a sense of the hybrid nature of this repeated measures analysis, somewhere between univariate and multivariate methods. In general, univariate tests are more capable of detecting existing differences than their multivariate counterparts; hence the motivation for this strategy.

The output, with commentary

1. SAS recognizes that there are only 20 experimental units in the study, not 100.

```
Class Level Information
Class     Levels    Values

VAR           4    1 2 3 4

Number of observations in data set = 20
```


2. And there are four response variables, with correspond to the four levels of the "factor" we have called Time.

```
Repeated Measures Analysis of Variance
Repeated Measures Level Information

Dependent Variable          CUT1     CUT2     CUT3     CUT4
     Level of TIME              1        2        3        4
```


3. Next, SAS produces a table of "Between Subjects Effects" (i.e. the effects of the Main plot factor Variety). Notice that the results we get here **match the results from our earlier split-plot analysis** *exactly*, and we did not even need to specify the correct error term!

```
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects
```

Source	DF	Anova SS	Mean Square	F Value	Pr > F
VAR	3	2.840753	0.946918	7.40	0.0025
Error	16	2.048682	0.128043		

4. Finally comes the results we are looking for. In this table of "**Within Subject Effects**" (i.e. the effects of the **subplot factor Time**) SAS produces the *exact same F values* that we obtained in the split-plot analysis. But these F values are now accompanied by several columns of p-values.

```
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects
```

Source	DF	SS	MS	F Value	Pr > F	Adj Pr > F . G - G	H - F
Time	3	37.45	12.48	130.46	0.0001	0.0001	0.0001
Time*Var	9	0.55	0.06	0.64	0.7606	0.6585	0.6948
Error(Time)	48	4.59	0.096				

```
            Greenhouse-Geisser Epsilon = 0.4984
                  Huynh-Feldt Epsilon = 0.6413
```

The first column, labelled "Pr > F," present the p-values from the split-plot analysis; these are the results one gets when one assumes perfect independence among the repeated measures. Next come two columns of *adjusted* p-values (**G**reenhouse-**G**eisser and **H**uynh-**F**eldt). These p-values have been adjusted based on the *actual* patterns of correlation detected in this particular dataset. As we said before, repeated measures are neither perfectly correlated nor perfectly uncorrelated, but somewhere in between. The two columns represent two different methods for accounting for the patterns of correlation in the data. We recommend using the more conservative of the two: G-G.

This analysis produces the same F values as the previous split-plot analysis but **corrected** p-values for TIME (previous B_cut) and TIME*VAR (previous B_cut*A_var)

Each of the adjustment methods estimate a quantity known as "**Epsilon**" and then multiply the numerator and denominator degrees of freedom by this estimate before determining the significance level for the test, labelled "Adj Pr > F" in the output. This reduces the degrees of freedom less drastically than the conservative df approach, which simply divides the degrees of freedom of B, B*A, and Error by the degrees of freedom of B. At the same time, these adjusted probabilities are more conservative than the original split-plot analysis with unadjusted degrees of freedom.

The F tests from this analysis will be correct *if the ==assumptions of the analysis== are met*. These assumptions can be tested via "sphericity tests" by adding the PRINTE option after "/ NoM" (see <u>SAS for Linear Models</u> 3rd Ed. page 274). Representative output:

<div align="center">

Sphericity Tests

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	5	0.1831747	24.988247	0.4635
Orthogonal Components	5	0.1127544	32.131886	==0.2341== NS

</div>

The test labelled "Orthogonal Components" is the one that is important in determining whether the univariate *F* tests for the within-subjects effects are valid (i.e. that the covariance structure of the dataset is such that the *F* tests are valid). Somewhat simplified, what is being tested is that the measurements conform to a certain pattern. Univariate tests may be too liberal when the data do not conform to the pattern tested by the sphericity test (namely, that the orthogonal components of the data are uncorrelated and have equal variance). In cases where the sphericity test is rejected strongly ($p < 0.0001$), extreme caution should be used in interpreting tests. Better yet, more drastic remedial steps than simple df adjustments should be taken (e.g. use a multivariate test for the Time and Time*MainPlot effects).

12.9.3.1 Mean comparisons   (Cody & Smith 1991, Chapter 8)

In the REPEATED statement, key words can be added after the *factor* name to define single df contrasts for that factor. Since the number of contrasts generated is always one less than the number of levels of the factor, SAS provides ways of controlling which contrasts are included in the analysis.

For example:

```
Repeated Time Contrast(1) / Summary;
```

generates single df contrasts between each level of the Time factor with Level 1 (i.e. Cut1). Any other level can be chosen as the reference level by simply replacing 1 with 2, 3,…,k (where k is the number of level; in this case, k = 4). If you wish to make all possible pairwise contrasts among the levels of the repeated measurement, you will need to write (k − 1) REPEATED statements. In this case, keep in mind the experimental error rate when declaring significance.

The key word HELMERT generates contrasts between each level of the factor and the mean of all subsequent levels:

```
Repeated Time Helmert / Summary;
```

The keyword MEAN(n) generates contrasts between the mean of each level with the group mean of all other levels of the factor. The level specified by (n) is the eliminated contrast. Example:

```
Repeated Time Mean(1) / Summary;
```

Finally, contrasts for each of the (k – 1) polynomial components (i.e. a trend analysis) can also be requested:

```
Repeated Time Polynomial / Summary;
```

Which to use, Split-Plot or REPEATED?

There is no simple answer as to which of these methods is most appropriate. If responses are significant even when the conservative df are used, those conclusions are sound.

For those cases where different conclusions are obtained with the adjusted and unadjusted df, one might think that the REPEATED statement would solve all of the problems. It can solve some problems of correlation, but the interpretation of the results requires a number of assumptions or tests that are beyond the scope of this course. You can use the REPEATED statement, but *only* after making sure you understand and test all the assumptions of the model.

Missing values: In cases where many repeated measurements are missing, the split-plot approach might offer the only way of analyzing a repeated-measures design because the REPEATED statement will simply remove from the analysis any level repeated measures factor where data are missing.