Introduction to Repeated-Measures

William G. LeBlanc, University of Miami, Miami, FL

Introduction

A repeated-measures analysis involves making and analyzing multiple measurements of the same response variable on the same subject. The data for such an analysis can be laid out in two different styles, the univariate and the multivariate, that go hand in hand with two different methods of performing the analysis. Both the univariate method and the multivariate method will be discussed.

Univariate and Multivariate Data

The details of the univariate method will be easier to explain by first presenting an example of the multivariate method of recording data. Table 1 shows a series of heart rate gains measured on 10 subjects after administering each of four drugs, A, B, C, and D. Placing all measurements for one individual on one line is efficient, both in terms of data entry and of visualization. This horizontal orientation presents complete data for one subject in one line.

Table 1. Multivariate Data Layout of HR Gains Measured after Administration of Drugs A, B, C, D

<table>
<thead>
<tr>
<th>ID</th>
<th>GEO</th>
<th>Drug</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>A</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>N</td>
<td>B</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>N</td>
<td>C</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>D</td>
<td>43</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>A</td>
<td>38</td>
</tr>
<tr>
<td>6</td>
<td>E</td>
<td>B</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>E</td>
<td>C</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>E</td>
<td>D</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>E</td>
<td>D</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>D</td>
<td>40</td>
</tr>
</tbody>
</table>

(Means: 34.5 37.5 34.1 40.5)

It will be assumed that the order of the drug administration was properly counterbalanced and there are no carry-over effects from any drug, important considerations in a repeated-measures study.

In Table 1, there are multiple (multivariate) columns of response data, and the thought of comparing or correlating them is obvious. The order and position of the response variables are important—they are related to the levels of the experimental treatments. In Table 1, for example, the first heart rate measurement belongs to the A treatment level, the second to B, etc. There is no variable whose values are "A," "B," "C," or "D." This method of using position to implicitly identify treatment level will be used in the SAS® system programming statements to identify the "within-subject" variables for the repeated-measures analyses. If the programmer is clever, the names of the variables can be designed to help identify treatments and levels, for example, HR1_1AM, HR1_2AM, HR2_1AM, HR2_2AM. (The variable names A, B, C, and D are not clever!)

Table 2. Univariate Data Layout of (abbreviated) HR Data from Table 1

<table>
<thead>
<tr>
<th>ID</th>
<th>GEO</th>
<th>DRUG</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>A</td>
<td>39</td>
</tr>
<tr>
<td>1</td>
<td>N</td>
<td>B</td>
<td>38</td>
</tr>
<tr>
<td>1</td>
<td>N</td>
<td>C</td>
<td>35</td>
</tr>
<tr>
<td>32 lines omitted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>E</td>
<td>D</td>
<td>38</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>A</td>
<td>35</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>B</td>
<td>38</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>C</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>D</td>
<td>40</td>
</tr>
</tbody>
</table>

In contrast, the univariate analysis uses one or more explicit variables to track the levels of the one or more within-subject factors. Table 2 shows an abbreviated version of the same data set written in univariate form. An explicit variable, DRUG, identifies the levels of the treatment factor as they accompany the values of the response variable. This is easily accomplished by pairing the identifying variable with the response variable in adjacent columns. Now the orientation of the data has shifted from horizontal to vertical. One row of data no longer includes all measurements for one complete subject, but rather one measurement at one time or treatment level for one subject. Since every response measurement has identifying information with it, order is no longer essential. For example, the measurement of subject 9 at treatment level D could be placed on any row. The position does not matter, since the value of the DRUG variable identifies the treatment condition of that measurement.

When the data from Table 1 are converted into one (univariate) column in Table 2, the table becomes "taller," and the thought of comparing and correlating...
Contrasts and 2-Group Comparisons

The null hypothesis in the analysis of this data set is that there are no differences in HR gain among the four treatment conditions. An equivalent way of expressing this absence of differences is by saying that the difference between any level, D, for example, and each of the other three levels is zero. The following contrast matrix, similar to the M matrix in the SAS GLM procedure, expresses this pattern of three comparisons (see Table 3). In general, if there are J treatment conditions, then there are J-1 independent or orthogonal comparisons, meaning that any additional comparison(s) will not supply any new information regarding the relationships among any of the variables. The contrast coefficients in Table 3 represent the default contrasts of a four-level within-subject factor. Several other commonly used, built-in contrasts are available. In addition, user-defined contrasts can be specified. Contrasts need not be orthogonal. Not all of the built-in contrasts are orthogonal.

Table 3. Matrix of Comparisons

<table>
<thead>
<tr>
<th>DRUG</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIFF1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DIFF2</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DIFF3</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

A simple way to perform an independent-groups comparison between two means is by a t-test. If the data are in univariate form, as in Table 2, the TTEST procedure would be used. The statements needed to perform the analysis of, for example, level C versus level D would be:

```sas
PROC TTEST;
VAR HR;
CLASS DRUG;
WHERE DRUG='C' or DRUG='D';
```

This test, treating level D measurements as unrelated to level C measurements, compares the level D mean of 40.5 to the level C mean of 34.1. The resulting t value of 2.54 with 18 df is significant at the 0.05 level. Equation 1 gives the formula for the independent groups t-test when the group sizes are equal. With equal group sizes, MSError is the simple average of the variance estimates of the two groups. It is usually referred to in statistics texts as the pooled variance term.

\[
t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\text{MSError} \times \frac{2}{n}}} \quad (1)
\]

If the data are in multivariate form, there is no direct within-subject t-test procedure in SAS. In order to perform a within-subject comparison, a new variable, equal to the difference of the two original variables, would have to be created in the data step. Its mean would then be tested for equality to 0. For example, in order to perform the comparison between level D and level C, a difference variable would be created from the multivariate data of Table 1: DIFF3 = D-C. The name DIFF3 was chosen to correspond to the third comparison from the comparison matrix in Table 3. The mean of this new variable, usually called a transformed variable, is 6.4. This is the same quantity tested in the independent-groups t-test above.

The statements needed to perform the analysis of this difference (D-C) would be:

```sas
PROC MEANS N MEAN VAR T PRT;
VAR DIFF3;
```

The T and PRT options print the Student's t statistic and its associated p value for testing that the mean of the variable(s) defined in the VAR statement is zero. This test, in effect, is a test of a single mean. DIFF3 was designed to be zero if there is no difference between C and D in the population. The error term for such a test of a single mean is the standard error of the mean, computed from the square root of VAR/N, the quantities requested as options in the MEANS procedure statement. The resulting t value of 3.32 with 9 df is significant at the 0.01 level.

The error term of this test is influenced by the degree of correlation (or covariance) between C and D. The greater the correlation, the smaller the variance of the difference scores. The mean of the difference scores (6.4 in the within-subject t-test) will still equal the difference of the means of the two sets of scores (6.4 in the independent groups t-test). However, the variance term in the denominator changes, causing the value of the two t statistics to differ.

This decrease in the error variance due to correlation among response variables is the reason for using a repeated-measures design, in which the response variables are usually highly correlated. In the extreme situation of no correlation between two sets of scores, which "should" occur if the two sets of scores are randomly paired, the within group t-test would give the same results as if the two sets of scores were analyzed as independent groups, i.e., as if there were no correlation between the two sets of scores. This is illustrated by the entries in Table 4A, comparing the six pairwise comparisons among A, B, C, and D performed in two ways—as if the data from each drug treatment came from the same subject (within-subject) or came from two independent sets of subjects (independent groups). Above the diagonal are the t values obtained
by treating drug levels as within-subject data. Below the
diagonal are the $t$ values obtained from the same
comparisons, treating the drug levels as independent
groups. Table 4B shows a modified variance-covariance
matrix.

The largest disparity between the $t$ values occurs
for the B-C comparison, 1.42 (independent groups) vs.
4.02 (within-subject), since that is where the largest
correlation between any two levels is (.89, Table 4B).
Level A correlates minutely with the other three levels,
which is why the $t$ values in row A are not much
different from the $t$ values in column A.

The pooled variance estimate term, $\text{MSerror}$,
from the denominator of the independent groups $t$-test
between B and C is 28.5. If it were reduced by the
covariance term between B and C, 24.9, the
recomputation of Equation 1 would result in a $t$ value of
4.02. In general, the error term is equal to the average
of the variance terms in the variance-covariance matrix
minus the average of the covariance terms, as shown in
Equation 2.

$$\text{MSerror} = \text{MeanVar} - \text{MeanCov} \quad (2)$$

**Table 4A & B. Dependent and Independent Groups t-
tests Influenced by Correlation**

### A. T-test results

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.45</td>
<td>0.16</td>
<td>2.56</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1.40</td>
<td>4.02</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.17</td>
<td>1.42</td>
<td>3.32</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>2.63</td>
<td>1.32</td>
<td>2.54</td>
<td></td>
</tr>
</tbody>
</table>

### B. Correlation, variance, covariance matrix

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23.4</td>
<td>.07</td>
<td>-.03</td>
<td>-.06</td>
</tr>
<tr>
<td>B</td>
<td>1.50</td>
<td>22.5</td>
<td>.89</td>
<td>0.48</td>
</tr>
<tr>
<td>C</td>
<td>-.83</td>
<td>24.9</td>
<td>34.5</td>
<td>0.41</td>
</tr>
<tr>
<td>D</td>
<td>-1.5</td>
<td>12.3</td>
<td>13.1</td>
<td>28.7</td>
</tr>
</tbody>
</table>

Variances can be obtained in SAS by using any
of several procedures, e.g., PROC MEANS or the
UNIVARIATE procedure. Covariances can be obtained
using the CORR procedure or the DISCRIM procedure.
If there is a CLASS variable, PROC DISCRIM gives
separate and pooled variance-covariance matrices, and
a test of homogeneity of the matrices.

### Repeated-Measures the Univariate Way

An ordinary one-way analysis of variance
(ANOVA) on the univariate data from Table 2 will
simultaneously analyze the differences among the four
groups. Such an analysis would ignore the correlations
among the response variables within the same subjects,
and thus would not be the proper analysis for a
repeated-measures design. This ANOVA is the
extension to the independent groups $t$-test above. The
following commands produce the results shown in Table 5:

```
PROC GLM;
CLASS DRUG;
MODEL HR = DRUG;
```

**Table 5. Univariate ANOVA, Treating the Four Drug
Treatment Levels as Independent Groups**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRUG</td>
<td>3</td>
<td>266.70</td>
<td>88.90</td>
<td>3.26</td>
<td>0.0326</td>
</tr>
<tr>
<td>Error</td>
<td>36</td>
<td>982.45</td>
<td>27.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>1249.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The $F$ value is significant at the .05 level and the
conclusion is that at least one of the groups is different
from one of the others. The next step would normally
be to perform some kind of post hoc test.

An additional step will now be added that will
capitalize on the intercorrelations among the response
variables in the four treatment levels that are shown in
Table 4B. This will create a more powerful test through
a reduction of the error term, as was demonstrated in
the discussion of the $t$-tests. There are two ways of
performing the repeated-measures analysis, depending
on how the data are oriented, and on what assumptions
are made about the structure of variance-covariance
matrix. These are the sphericity assumptions, and will
be discussed in detail later.

A univariate repeated-measures ANOVA will
be performed using the univariate arrangement of data
in Table 2. This method is sometimes called the "mixed
model" approach, referring to the fact that subjects are
now used as one of the factors, and subjects are
considered a random factor. It is also sometimes
referred to as a "split-plot" analysis.

The results in Table 6 are produced by adding
subjects (ID) as a new independent variable to the
analysis, using these commands:
PROC GLM;
CLASS DRUG ID;
MODEL HR = DRUG ID;

Table 6. Univariate ANOVA, Treating the Four Drug Treatment Levels as Repeated-Measures

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRUG</td>
<td>3</td>
<td>266.70</td>
<td>88.90</td>
<td>4.67</td>
<td>0.0094</td>
</tr>
<tr>
<td>ID</td>
<td>9</td>
<td>468.10</td>
<td>52.01</td>
<td>2.73</td>
<td>0.0209</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>514.30</td>
<td>19.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>1249.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The CLASS statement in the GLM procedure and the TTEST procedure designates the variables whose values identify the different groups. Thus, the CLASS statement identifies the independent variables used in a univariate data layout.

The $F$ value for DRUG is larger, due to the reduction of the Mean Square Error (MSerror) term, in spite of the fact that the error df also were reduced. Subjects were incorporated as a factor by listing ID as a factor in the CLASS and MODEL statements. The intersubject variability accounts for a substantial amount of systematic variance when the responses of each subject are correlated. Intersubject variability was removed from the error term, making the analysis more powerful, in spite of the loss of 9 df associated with the 10 subjects (compare Tables 5 and 6).

How much systematic variance was removed from the MSerror? Notice that the MSerror in Table 5 is equal to the mean of the variances along the diagonal of the variance-covariance matrix of Table 4B. The MSerror term in Table 6 is equal to the MSerror of Table 5 minus the mean of the covariances (8.25) below the diagonal in Table 4B. See Equation 2. Substantial intercorrelation among the dependent variables, as would most likely occur in repeated-measures or multivariate designs when the measurements were made from the same or matched subjects, works to reduce the error term, and therefore increases the overall power of the analysis.

Assumptions of the Univariate Approach to Repeated-Measures

Certain assumptions are made in the use of the univariate method for performing a repeated-measures analysis regarding the pattern of variances and covariances. These assumptions, however, do not apply to the multivariate analysis, or to analyses with only two levels of the repeated factor. One assumption is "compound symmetry", which requires equality of variances among variables and equality of covariances between all pairs of variables. A second is "sphericity", and it requires equality among the variances of difference scores for all pairs of repeated-measures. Compound symmetry implies sphericity, but the reverse is not true. In other words, compound symmetry is a subset of sphericity. Sphericity is thus less restrictive than compound symmetry, since it requires fewer equalities. There are $(J(J+1)/2)-2$ differences among variances and among covariances (assumed by compound symmetry), but only $(J(J-1)/2)-1$ differences among treatments (considered by sphericity). For our data, the compound symmetry assumption requires $(4x5/2)-2 = 8$ independent variance-covariance comparisons. For the sphericity assumption, there are $((4x3)/2)-1 = 5$ independent covariance comparisons.

If there are only two levels of a within-subject variable, sphericity is assured, i.e., it is not an issue because there is only one set of differences. However, compound symmetry could still be violated if the two variances are not equal.

Violations of the Symmetry Assumptions and Corrective Steps

If the sphericity assumption in the univariate analysis is violated, the critical value of $F$ will be based on inflated df, thereby causing the critical $F$ to be smaller than it should be, thus increasing the risk of incorrectly rejecting the null hypothesis (Type I error). There are two adjustment factors that are commonly used to adjust or scale down the numerator and denominator df, thus correcting for the inflated df in the univariate analysis. They are referred to as the Greenhouse-Geisser epsilon (G-G epsilon) or the Huynh-Feldt epsilon (H-F epsilon) in the SAS GLM procedure. G-G epsilon ranges from a maximum value of 1 when there is no departure from sphericity, to a minimum value of 1/(J-1).

Table 7 shows the adjustment factors to the df ranging from least conservative to most conservative that could be used for selecting the critical F value for the univariate analysis of the data in Table 1. The Greenhouse-Geisser epsilon value of 0.68 indicates a substantial departure from sphericity. To maintain the alpha level at approximately the desired level, the original df for the univariate $F$ test would be adjusted by multiplying both numerator and denominator degrees of freedom by either epsilon.
Table 7. Adjustment Factors for Departures from Sphericity for Univariate Analysis

<table>
<thead>
<tr>
<th>adjustment</th>
<th>num</th>
<th>den</th>
<th>critical F</th>
<th>obtained F</th>
</tr>
</thead>
<tbody>
<tr>
<td>original</td>
<td>3</td>
<td>27</td>
<td>2.96</td>
<td>4.67</td>
</tr>
<tr>
<td>H-F (.89)</td>
<td>2.67</td>
<td>24.03</td>
<td>3.40</td>
<td></td>
</tr>
<tr>
<td>G-G (.68)</td>
<td>1.74</td>
<td>18.36</td>
<td>4.41</td>
<td></td>
</tr>
<tr>
<td>max (.33)</td>
<td>1.00</td>
<td>9.00</td>
<td></td>
<td>5.12</td>
</tr>
</tbody>
</table>

The decimal df values are rounded downward to be conservative in the selection of the critical F value. Choosing the more conservative Greenhouse-Geisser adjustment, the critical F would be F(1,18,.05) = 4.41. Since the obtained F of 4.67 exceeds the G-G adjusted critical F value, it can be assumed that the decision to reject the null hypothesis would still maintain a Type I error rate at or below .05.

There are two quick rules that can be used as guidelines for when it is not necessary to adjust the critical df when using the univariate method for repeated-measures, in the event that the analysis is performed by a computer that cannot generate the G-G or H-F epsilon:

1) If the obtained F is less than the unadjusted critical F, then stop. Any adjustment will make the results even less significant.

2) If the obtained F exceeds the critical F adjusted by the maximum adjustment factor, 1/(1-1), then stop. The results are "significant enough".

The HR data in this paper do not satisfy either of the rules above. The obtained F lies between the critical Fs based on the original df and the maximally reduced df (see Table 7), and thus will require the application of the G-G or H-F adjustments. Proc GLM will print both epsilons and the p-values associated with the adjusted Fs. Either may be chosen. Of the two, the G-G adjustment is more conservative.

Repeated-Measures the Multivariate Way

The second method of performing the repeated-measures analysis is the multivariate method, using the REPEATED statement in PROC GLM. The statements used to perform the multivariate repeated-measures analysis for the data in Table 1 are:

PROC GLM;
  MODEL A B C D = / INT;
  REPEATED DRUG 4
  /SUMMARY PRINTE PRINTM PRINTH;

This procedure creates three new variables from the differences of D with each of A, B, and C, and simultaneously tests the means of these three transformed variables for a difference from zero. If all three differences equal zero, then all four original means are equal. The comparison of the last level, D, with each of the others is the default associated with the REPEATED statement, similar to the M matrix of Table 3. The REPEATED statement has other pre-programmed contrast patterns, and also allows for user-defined contrasts. The results of the multivariate test are shown in Table 8. The actual output has been modified slightly.

Table 8. Multivariate One-Within Group ANOVA Performed on Data from Table 1

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks Lambda</td>
<td>0.2517</td>
<td>6.93</td>
<td>3</td>
</tr>
<tr>
<td>Pillai Trace</td>
<td>0.7482</td>
<td>6.93</td>
<td>3</td>
</tr>
<tr>
<td>H-L Trace</td>
<td>2.9721</td>
<td>6.93</td>
<td>3</td>
</tr>
<tr>
<td>Roy's Root</td>
<td>2.9721</td>
<td>6.93</td>
<td>3</td>
</tr>
</tbody>
</table>

In the MODEL statement, the dependent variables are listed to the left of the equal sign. Any between-group variables (explained later), would be placed to the right of the equal sign, and would be declared in a CLASS statement. For this particular repeated-measures analysis, there are no between-group variables, no CLASS statement, and nothing to the right of the equal sign.

The "INT" option following the slash in the MODEL statement prints the test of departure of the means of the dependent variables from zero. Remember, the dependent variables here are the three differences between D and A, B, and C, defined by the M matrix. Rejection of the null hypothesis that all means are zero would imply that there is at least one difference variable with a non-zero mean, indicating a difference between the mean of level D and the mean of at least one of the other levels. Follow-up tests would be performed to pinpoint the location of the difference(s).

The "DRUG 4" in the REPEATED statement establishes DRUG as the name of the repeated factor. Note that there is no (and must be no) explicit variable named DRUG in this data set, as opposed to the univariate data set which used the variable DRUG in a CLASS statement as a "between-group" factor. The "4" is optional in this one-factor example, since there is only one within-subject factor and it consists of all four dependent variables listed to the left of the equal sign. If there were more than one within-subject factor, each within-factor name would be followed by the number of levels in that factor, and each group would be separated by a comma. The product of the number of levels of
each within-subject factor would equal the number of dependent variables listed to the left of the equal sign. An example will be shown below.

The "SUMMARY" option in the REPEATED statement produces an analysis-of-variance summary table for each of the three contrasts, testing each for a difference from zero. The "PRINTE" option prints the E matrix, the error sum of squares and cross-product matrix. It also prints the partial correlation matrix for the dependent variables, controlling for the CLASS variables, if there are any, and it prints the results of the sphericity test when there are more than two levels of the within-subject factor. The E matrix corresponds to the denominator of the univariate F test. The "PRINTE" option prints the H matrix, the sum of squares and cross product matrix for the transformed variables. The H matrix is also called the hypothesis matrix, and it corresponds to the numerator of the univariate F test. The "PRINTM" option prints the M matrix of contrasts that define the transformed variables for the analysis.

The default output of PROC GLM using the REPEATED statement shows both the multivariate test results (see Table 8) and the univariate test results, including the values for G-G and H-F epsilon and the adjusted p values for the univariate analysis (not shown). The four F values of the multivariate output are sometimes identical. When they differ, there is no commonly accepted best one. The univariate test results are the same as those illustrated earlier in the discussion of doing repeated-measures analyses the univariate way.

For the data in Table 1, the multivariate F = 6.93 (p = 0.0167), and the univariate F = 4.67 (p = 0.0094), unadjusted for departure from sphericity; p = 0.022 (Greenhouse-Geisser; p = 0.013, Huynh-Feldt). Hotelling's T^2, a multivariate analog to the within-subject t-test, is not printed, but it can be computed as:

$$T^2 = (n-1)*((1/\lambda - 1)$$  \hspace{1cm} (3)

where \( \lambda \) is Wilk's lambda from the multivariate test results.

The SUMMARY option produced ANOVA summary tables (not shown here) for each of the three contrasts, indicating that the A-D and C-D contrasts are significant at the .05 level. These are the same results that would have been obtained from three within-subject t-tests performed on the same difference variables from the original data of Table 1. Since these are single-df tests, the square roots of the three Fs in the PROC GLM would equal the three t values from the within-subject t-tests from PROC MEANS. If the default M matrix does not include a desired comparison, a customized M matrix could be specified in the REPEATED statement that would perform the desired test, or a different pre-programmed contrast might provide the desired comparison.

The univariate and multivariate Fs in this example differ. Which analysis is more powerful? The answer is: it depends. The multivariate analysis does not depend on the sphericity assumption. If the sphericity assumption is not violated, the multivariate test is not as powerful as the univariate test. Large departures from sphericity can be corrected in the univariate analysis by the G-G and H-F adjustments, but the resulting test may be less powerful than the multivariate test.

Two Within Factors

The following discussion of more complex designs will focus only on the multivariate method of analysis using the REPEATED statement in PROC GLM.

Suppose drugs A and B were in powder form and drugs C and D were in liquid form, and that A and C were administered in the morning, and B and D were administered in the afternoon. To examine FORM and TIME differences, the following SAS statements would be used:

```sas
PROC GLM;
MODEL A C B D = / INT;
REPEATED TIME 2, FORM 2
/ SUMMARY PRINTE PRINTH;
```

The REPEATED statement now shows two within-subject factors, TIME and FORM. As was cautioned before, these names must not be names of variables already existing in the data set. The product of the number of levels in the within-subject factors must equal the number of dependent variables listed to the left of the equal sign in the MODEL statement. The order of TIME and FORM is determined by the ordering of the dependent variables to the left of the equal sign in the MODEL statement. The variables in the REPEATED statement "nest" under each other, the rightmost variables advancing over levels more rapidly than those to the left. According to the REPEATED statement above, reading from the right, the first two of the four dependent variables will be: FORM levels 1 (=A) and 2 (=C) under TIME at condition 1, and then FORM levels 1 (=B) and 2 (=D) under TIME at condition 2. The PROC GLM output verifies the assignment of response variables to levels as part of its default output (Table 9).

**Table 9. Repeated-Measures Level Information**

<table>
<thead>
<tr>
<th>Dep Var</th>
<th>A</th>
<th>C</th>
<th>B</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of TIME</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Level of FORM</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
According to Table 9, A and C belong to the first level of TIME (morning), and B and D belong to the second level of TIME (afternoon). A and B belong to the first level of FORM (powder), and C and D belong to the second level of FORM (liquid). Within limits, the dependent variables can be listed in the MODEL statement in various orders, as long as the within-subject factors in the REPEATED statement can "pick them up" in a meaningful way. The output from the PRINTM option, the M matrix (not shown), is also helpful in confirming that the proper variables are being associated with the proper factor levels.

For the data from Table I, the multivariate analysis reveals a significant TIME effect, but no significant FORM effect or TIME FORM interaction (see Table 10). The significant TIME effect indicates that the response averaged over drug treatments A and C, the morning administration, was different from the response averaged over drugs B and D, the afternoon administration. The univariate analysis is in agreement. It should be noted that there will not be any G-G or H-F adjustments in the univariate analysis, since there are only two levels for both TIME and FORM. Another consequence when there are two levels of a within-subject (or between-group) variable is that a significant effect requires no followup analysis—the location of the difference is not in doubt. It is between the two levels.

An extract from the multivariate tests is shown in Table 10. Only the Wilk's lambda statistic is shown for the sake of brevity. The F values of all four multivariate test statistics are the same for all three effects in this particular analysis. That is not always the case.

Table 10. Multivariate Two-WithinGroup ANOVA

Manova Test Criteria and Exact F Statistics for the Hypothesis of no TIME Effect

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F</th>
<th>DF</th>
<th>DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilk's Lambda</td>
<td>0.3634</td>
<td>15.766</td>
<td>1</td>
<td>9</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

Manova Test Criteria and Exact F Statistics for the Hypothesis of no FORM Effect

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F</th>
<th>DF</th>
<th>DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilk's Lambda</td>
<td>0.9011</td>
<td>0.987</td>
<td>1</td>
<td>9</td>
<td>0.3464</td>
</tr>
</tbody>
</table>

Manova Test Criteria and Exact F Statistics for the Hypothesis of no TIME FORM Effect

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F</th>
<th>DF</th>
<th>DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilk's Lambda</td>
<td>0.8901</td>
<td>1.111</td>
<td>1</td>
<td>9</td>
<td>0.3193</td>
</tr>
</tbody>
</table>

One Within, One Between Factor

Up until now, the term "between-group variable" has been used to differentiate levels of multivariate data configured in a univariate layout. Such variables are "artificial" in the sense that they can be discarded with no loss of information if the data set is transposed or reconfigured into multivariate form, as happens to the variable DRUG when Table 2 is turned into Table 1. Now, for the discussion of between- and within-subject factors, a between-group factor will refer to a variable that serves to separate subjects into differing groups that usually would not ever be combined by transposing or reconfiguring the data set. Common between-group variables include sex, race, place of birth, experimental treatment level. More often than not, data sets are mixtures of between-group and within-subject variables.

The data in Table I include a GEO variable, whose values represent geographic region of birth. When such variables are present, they are used to divide the sample into groups, and then between-group differences in one or more dependent variables are examined.

In order to test for between-group differences, a CLASS statement is added to PROC GLM and the between group variable, GEO, is listed to the right of the equal sign in the MODEL statement. PROC GLM tests the between-group effect by creating a new, univariate dependent variable from the average of the repeated-measures variables and performing a standard one-way ANOVA using the "CLASS" variable to the right of the equal sign as the between-group variable. Since the four repeated measurements are averaged into one dependent variable for the test of the between-group factor, there are no longer repeated or multivariate dependent variables. Therefore, correlation and sphericity are no longer issues for the analysis of the between-group factor.

The interaction effect between the within-subject factor and the between-groups factor is also of interest. Up until now in our discussion, tests of the transformed variables created according to the pattern of the M matrix have been performed against a reference value of zero, i.e., the question has been "are all of the transformed variables equal to zero?" Now, in the presence of a between-group variable, the reference value will not be zero but the different levels of the between-group variables. The question examined in the test of the DRUG x GEO interaction is "is the pattern of differences among the four HR gains uniform across geographic regions?"

The statements needed to perform a one-within, one-between analysis are:
PROC GLM;  
CLASS GEO;  
MODEL A B C D = GEO;  
REPEATED DRUG 4  
/SUMMARY PRINTE PRINTM PRINTH;  

The test for the DRUG effect is very similar to the multivariate analysis just performed in the one-within ANOVA shown in Table 8. The H matrix, which can be thought of as a multivariate counterpart to the simple contrast sum of squares numerator of the F test procedure, is the same for the one-within analysis as for the one-within, one-between analysis. In both analyses, the contrasts among the repeated-measures (the four heart rate gains) are the same. However, the E matrix, analogous to the denominator sum of squares for the F test, is now affected by the presence of the new between-group variable. In this case, whatever systematic variance that the between-group variable, GEO, contributes to heart rate response is removed from the error term, in the same manner that the intersubject variability was removed from the denominators of the t-tests, causing the disparity between the t values above and below the diagonal in Table 4A. Recall that the t-test is a two-group version of the F test.

Additional output includes the univariate test performed on the within-subject variable DRUG, and the interaction of DRUG with the between-group factor, GEO.

A common application of the one-within, one-between repeated-measures analysis occurs when the response variable is measured before and after some type of intervention, a typical "pre-post" analysis between two groups of subjects. If gain scores were computed between the two repeated measurements in a multivariate data set and analyzed by an independent groups t-test, the results would be the same as those obtained by performing a one-between, one-within repeated-measures analysis using the REPEATED statement in PROC GLM, or by configuring the data set in univariate form with TIME, GEO, and ID as between-group variables and performing an ANOVA, focusing on the TIME*GEO interaction effect. The following statements, written in terms of the data sets and variables in this paper, illustrate all three approaches. The between-group variable will be GEO, with two levels, N and E. The two-level within-subject TIME variable will be created by choosing the measurements from drug treatments C and D. As shown in Table 3, the D-C comparison will be named DIFF3 for the t-test. The same comparison is named TIME in the multivariate repeated-measures procedure. In the univariate repeated-measures procedure, it is referred to as HR. It is important to realize that the same effect is being tested, even though the name of the dependent variable is different in each test. This effect is the difference in HR gains between drug treatments C and D between those born in the north and those born in the east.

*test on gain score;  
PROC TTEST DATA=TABLE1;  
CLASS GEO;  
VAR DIFF3;  

*multivariate repeated-measures;  
PROC GLM DATA=TABLE1;  
CLASS GEO;  
MODEL C D = GEO;  
REPEATED TIME 2 / SUMMARY ;  

*univariate repeated-measures;  
PROC GLM DATA=TABLE2;  
CLASS ID GEO DRUG;  
MODEL HR = GEO DRUG GEO*DRUG ID;  
WHERE DRUG='C' OR DRUG='D';  

The square of the t value in the first example is equal to the F values for the interaction effects in the second and third examples. The t test is testing for an interaction, a change in one group that differs from a change in another group.

Use of a 1-Within Design to Compute Reliability

Two of the mean squares from the univariate procedure used to analyze our J level repeated-measures data set could be used to compute the reliability of a J item scale. For this discussion, imagine that the data from Table 1 are the ratings by four judges of some characteristic over 10 subjects. Differences would naturally exist between individuals, barring the possibility that the N individuals all receive identical ratings. However, if the judges are reliable, ratings within subjects should be consistent. Thus, the within-subject variance should be smaller than the between-subject variance. Consistency or reliability of the test items or judges' ratings would be demonstrated by this pattern of low within-subject variance with respect to between-subject variance.

One method of computing reliability using the within and between subject variance computed in the multivariate repeated-measures analysis is shown by the following formula:

\[ r_{ij} = \frac{1}{J} \cdot \frac{MS\text{error}}{MS\text{between}} \]  

(4)

where the term MSerror is the error term for the analysis, and MSbetween is the between-subject mean square. An increase in reliability results from a decrease in variance within a subject with respect to variance between subjects. For the data from Table 1,
using the mean squares from Table 6:

\[ r = 1 - \frac{19.05}{52.01} = .63 \]

An alternate method of computing reliability would be to use:

\[ r_i = \frac{\text{Mean cov}}{\text{Mean var}} \] \hspace{1cm} (5)

where \( r_i \) is the reliability of 1 judge. The Spearman-Brown Prophecy formula gives the reliability for \( J \) judges:

\[ r_J = \frac{J \times r_1}{1 + (J-1) \times r_1} \] \hspace{1cm} (6)

Note that the MSbetween term from Equation 4 does not appear in the univariate analysis produced by the REPEATED statement. It would be obtained by conducting the univariate analysis on the univariate data, in which the subject variable (ID) is specifically designated as an effect (see Table 6).

Conclusion

The SAS System provides several methods for performing a repeated-measures analysis. The univariate approach provides the user with some advantages: it has increased power when certain assumptions are met; it is able to "salvage" remaining data from a subject who is missing one or more measurements at some level of the repeated factor; and it has more power than the multivariate analysis when the sample size is small compared to the number of repeated levels. The multivariate approach has advantages: it is easy to apply to a multivariate data set; contrasts are easily specified; it performs the univariate analysis complete with "corrected" \( F \) and \( p \) values; and it comes with a host of options that provide information to insure that proper effects are being formulated and tested.

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Suggested Readings


The Author may be contacted at:

University of Miami  
P.O. Box 248011  
Coral Gables, FL 33124  
bleblanc@umiamivm.ir.miami.edu (Internet)  
bleblanc@umiamivm (BITNET)

Acknowledgements

The helpful input from Dr. Brooks Applegate is greatly appreciated.