PLE LINEAR REGRESSION VIEWPOINTS
e 13, Number 2 Winter 1984

Testing Hypotheses in a Repeated Measures Design: An Example

John D. Williams and Jole A. Williams

he University of North Dakota and Grafton (N.D.) State School

Summary - The use of a typical repeated measures design is contrasted with using specific hypotheses which would directly address research questions. A complete example is given.

In an earlier paper Williams and Williams (1984) showed three different methods of using linear models to perform multiple comparisons (contrasts) for within subjects effects on a large sample (N=185) of employees in a test of hypotheses regarding improved facilities on employees attitudes. While large sample sizes yield impractical the use of person vectors (1 if person i, 0 if not), it would be useful to use a small sample so that the two approaches might be compared and the utility of using a single vector (predictor) for the subjects effect can be examined. Accordingly, a data set that has been previously used (Williams, 1974, 1980) will again be used here as an example.

An Example

The following problem is taken from Williams (1974).

A researcher may have an interest in the differential effect of two or more methods of instruction over time; thus, measures can be taken at specified intervals on the several instructional methods. From the point of view of the experiment, a repeated measures design can be conceptualized as a treatments X subjects design repeated for each instructional method.*

*This design is called a Type I design by Lindquist (1953).

To make the example more specific, suppose a research is interested in investigating the differences among three approaches to a human relations experience.**

The three different approaches toward the human relations groups selected are (1) structured sessions in which the group participates toward concrete problem solving, (2) an unstructured group, where the group decides upon its own goals, and (3) a group designed to allow the individual to focus on his personal problems with the interest being to help solve these problems. Five groups with 7-9 individuals in each group are assigned to each of the three human relation group situations; i.e., there are five separate groups for each treatment situation. Each group is to have a two hour session once a week for four weeks.

While there are several things that might be of interest to measure, the researcher is interested specifically in the amount of aggression exhibited in the group setting.

Videotapes are made of all sessions, and a group of five experts independently judge the amount of aggression expressed during the sessions on a continuum from 0 to 10, where 0 represents no aggression and 10 represents an extreme amount of aggression. The measurements are made with the group as the unit of analysis. The score to be used is the mean of the five ratings. Results are as follows:

Table 1

GROUP SCORES FROM THREE HUMAN RELATION GROUP METHODS FOR FIVE SESSIONS
(ARTIFICIAL DATA)

Method 1 (Structured Groups)

Group	Session 1	Session 2	Session 3	Session 4
1	3.2	3.4	3.2	2.8
2	4.6	4.0	3.8	3.4
3	5.0	3.8	5.0	3.2
4	2.0	2.0	2.4	1.6
5	3.6	3.2	3.4	3.0

**By human relations experience is meant the meeting of a group of people that has variously been called the T-group (training group), the encounter group, or some similar name.

Method 2 (Unstructured Groups)

Group	Session 1	Session 2	Session 3	Session 4
6	6.2	5.8	6.8	5.0
7	3.6	3.8	7.2	5.4
8	4.0	6.8	7.8	6.0
9	5.0	5.8	6.0	5.0
10	4.8	5.0	6.4	5.8

Method 3 (Personal Problems)

Group	Session 1	Session 2	Session 3	Session 4
11	7.4	7.6	6.8	5.2
12	6.4	6.4	5.6	4.0
13	7.0	6.6	6.6	6.0
14	5.8	7.4	5.0	4.8
15	6.4	5.2	4.0	3.6

To analyze the data in Table 1, it is first useful to define several variables:

Y = the criterion variable.

P_1 - P_{15} are binary variables that identify each group (the "person" vectors)

X_{16} = 1 if the score is from a group in the structured treatments; 0 otherwise.

X_{17} = 1 if the score is from a group in the unstructured treatment; 0 otherwise.

X_{18} = 1 if the score is from a group in the problems treatment.

X_{19} = 1 if the score is from Session 1; 0 othersise.

X_{20} = 1 if the score is from Session 2; 0 otherwise.

X_{21} = 1 if the score is from Session 3; 0 otherwise.

X_{22} = 1 if the score is from Session 4; 0 otherwise.

$X_{23} = X_{16} \cdot X_{19}$,

$X_{24} = X_{16} \cdot X_{20}$,

$X_{25} = X_{16} \cdot X_{21}$,

$X_{26} = X_{17} \cdot X_{19}$,

$X_{27} = X_{17} \cdot X_{20}$,

$X_{28} = X_{17} \cdot X_{21}$, and

$X_{29} = P =$ a sum of each separate group for the four sessions; for example, for group 1, $X_{29} = 3.2 + 3.4 + 3.2 + 2.8 = 12.6$. ($X_{29}$ will be referred to as P.) Each score (rather than each group) is the unit of analysis; thus, there are 60 scores for the data in Table 1. When preparing the data cards for a computer analysis, 60 data cards would be made. The use of P greatly facilitates a regression solution; this suggestion regarding coding was made earlier by Pedhazur (1977) and by Williams (1977), and more recently by Fraas and McDougall (1983).

To analyze the data in Table 1, it is useful to consider two separate analyses; one analysis can be treatments X subjects design, temporarily disregarding the three different kinds of groups. Then, it is useful to conceptualize the data in a two-way analysis of variance, disregarding for the time being that a given group has been measured several times.

The linear models that are useful for conceptualizing the data in Table 1 as a treatments X subjects design are as follows:

$$Y = b_0 + b_1P_1 + b_2P_2 + \ldots + b_{14}P_{14} + e_1 \text{ (for the subjects (groups) effect)} \quad (1);$$

$$Y = b_0 + b_{19}P_{19} + b_{20}P_{20} + b_{21}P_{21} + e_2 \text{ (for the trend effect)} \quad (2);$$

and

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_{14}X_{14} + b_{19}X_{19} + b_{20}X_{20} + b_{21}X_{21} + e_3. \quad (3).$$

When these linear models are used, the following results can be found:

from equation 1, $SS_S = 104.14$;

from equation 2, $SS_{TREND} = 8.63$; and

from equation 3, $SS_{ERROR_1} = 32.52$; also, $SS_T = 145.29$.

While the preceding information would be sufficient for a treatments X subjects design, it should be recalled that in this formulation, the type of human relation group was disregarded.

Actually, the treatments effect is "nested," i.e., totally contained in the variation among subjects. Before pursuing this "nesting" further at this point, it is first useful to complete the analysis for the two-way formulation.

The following four linear models are sufficient:

$$Y = b_0 + b_{16}X_{16} + b_{17}X_{17} + e_4, \quad \text{(for the treatments effect)} \quad (4)$$

$$Y = b_0 + b_{19}X_{19} + b_{20}X_{20} + b_{21}X_{21} + e_2, \quad \text{(for the trend effect)} \quad (2)$$

$$Y = b_0 + b_{16}X_{16} + b_{17}X_{17} + b_{19}X_{19} + b_{20}X_{20} + b_{21}X_{21} + e_5, \quad (5) \text{ and}$$

$$Y = b_0 + b_{16}X_{16} + b_{17}X_{17} + b_{19}X_{19} + b_{20}X_{20} + b_{21}X_{21} + b_{23}X_{23} + \ldots + b_{28}X_{28} + e_6 \quad \text{(Full Model)} \quad (6)$$

When these linear models are used, the following results can be found:

from equation 4, $SS_{METH} = 78.87$;

from equation 2, $SS_{TREND} = 8.63$;

from equation 6, $SS_{ERROR} = 39.71$.

The sum of squares attributed to regression for the full model (equation 6) is 105.58. The sum of squares attributed to regression for equation 5 is 87.50. The difference between these two values is equal to the interaction. Thus, $SS_{METH\ X\ TREND} = 105.58 - 87.50 = 18.08$. A summary table that would contain the foregoing information would appear as follows:

Table 2

SUMMARY TABLE FOR THE HUMAN RELATION GROUPS DATA IN TABLE 1

Source of Variation	df	SS	MS	F
Among Subjects				
Method	2	78.87	39.44	18.69**
Error (a)	12	25.27	2.11	
Total Among Subjects	14	104.14		
Within Subjects				
trend	3	8.63	2.88	7.20**
meth x trend	6	18.08	3.01	7.52**
error (b)	36	14.44	.40	
Total Within Subjects	45	41.15		
Total	59	145.29		

**Significant at .01 level

The summed vector, X_{29}, could have been used to achieve similar results:

$Y = b_0 + b_pP + e_1$ (for the subjects (groups) effect) [1a];

$Y = b_0 + b_pP + b_{19}X_{19} + b_{20}X_{20} + b_{21}X_{21} + e_3$. [3a]

Equation 1a is identical (in sum of squares) to equation 1; $SS_S = 104.14$; similarly, equation 3 yields $SS_{ERROR_1} = 32.52$. Table 1 could have been accomplished by using results from these last two equations in lieu of the original binary person variables (X_1 to X_{14}).

Multiple Comparisons (Contrasts) Within Groups

It would be helpful to give a diagramatic view, in terms of mean of the data described earlier;

	Session 1	Session 2	Session 3	Session 4
Method 1	$\bar{Y}_1$	$\bar{Y}_2$	$\bar{Y}_3$	$\bar{Y}_4$
Method 2	$\bar{Y}_5$	$\bar{Y}_6$	$\bar{Y}_7$	$\bar{Y}_8$
Method 3	$\bar{Y}_9$	$\bar{Y}_{10}$	$\bar{Y}_{11}$	$\bar{Y}_{12}$

Suppose the interest was in testing the long term change (from session 1 to session 4) between methods 1 and 2; that is, the interest is in testing $\bar{Y}_1 - \bar{Y}_4 = \bar{Y}_5 = \bar{Y}_8$. In our earlier paper (Williams and Williams, 1984) we outlined three different approaches to answering this sort of question.

The first approach, also outlined with this same data set in Williams (1980), was to reconstruct the criterion as $Y = Y^* + Y^{**}$ where $Y^* = \hat{Y}$; the $\hat{Y}$ values are the predicted values from using the equation $\hat{Y} = b_0 + b_pP$. For the present data $Y = 1/4P$ where P is the summed person vector described earlier as X_{29}. (Although it is more cumbersome, P_1 to P_{12} could have been used instead of P.)

It is the Y^{**} criterion that can be used to accomplish tests regarding within group cell differences. The full model can be written as:

$Y^{**} = b_1X_1 + b_2X_2 +..+ b_{12}X_{12} + e_4$, where X_1 to X_{12} correspond respectively to binary coded group variables for each cell. [4]

For some computer programs, a reparameterization of equation 4 that includes the unit vector is more useful:

$Y^{**} = b_0 + b_1X_1 + b_2X_2 +...+ b_{11}X_{11} + e_4$; [4a]

many other reparameterizations could have been chosen. For a more complete description of this reparameterization process, see Williams (1976).

The restriction that tests the hypothesis $\bar{Y}_1 - \bar{Y}_4 = \bar{Y}_5 - \bar{Y}_8$ is $b_1 - b_4 = b_5 - b_8$, or $b_1 = b_5 - b_8 + b_4$. Placing this restriction on equation 4 yields:

$Y^{**} = (b_5 - b_8 + b_4)X_1 + b_2X_2 +...+ b_{12}X_{12} + e_5$ [5]

or $Y^{**} = b_2X_2 + b_3X_3 + b_4(X_4 + X_1) + b_5(X_5 + X_1) + b_6X_6 + b_7X_7 + b_8(X_8 - X_1) + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_5$. [6]

Let $D_4 = X_4 + X_1$,

$D_5 = X_5 + X_1$, and

$D_8 = X_8 - X_1$.

Then equation 6 can also be given as:

$$Y^{**} = b_2X_2 + b_3X_3 + b_4D_4 + b_5D_5 + b_6X_6 + b_7X_7 + b_8D_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_5. \quad [6a]$$

Either equation 6 or a reparameterization of it, done by introducing b_0 an arbitrarily dropping any one predictor, can be used as the restricted mode Letting $b_{12} = 0$, one reparameterization, incorporating D_4, D_5 and D_8 is:

$$Y^{**} = b_0 + b_2X_2 + b_3X_3 + b_4D_4 + b_5D_5 + b_6X_6 + b_7X_7 + b_8D_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + e_5. \quad [6b]$$

The test is given by:

$$t = \sqrt{F} = \sqrt{\frac{R_F^2 - R_R^2/1}{(1 - R_F^2)/36}}.$$

Here $R_F^2 = .64899$; $R_R^2 = .57123$

$t = \sqrt{7.975} = 2.824.$

Using Side Conditions

Another approach to the repeated measures designs is to employ side conditions. Since the group effects are nested in the subjects effects, the full model

$$Y = b_pP + b_1X_1 + b_2X_2 + \ldots + b_{12}X_{12} + e_6 \quad [7]$$

can be turned into a model with the groups effects removed by imposing side conditions.

The group effects restrictions can be given as:

$$\frac{n_1b_1 + n_2b_2 + n_3b_3 + b_4b_4}{n_1 + n_2 + n_3 + n_4} = \frac{n_5b_5 + n_6b_6 + n_7b_7 + n_8b_8}{n_5 + n_6 + n_7 + n_8} = \frac{n_9b_9 + n_{10}b_{10} + n_{11}}{n_9 + n_{10} + n_{11} + n_1}$$

Because of equal n's (proportional n's would also suffice) these restrictio can be greatly simplified:

$b_1 + b_2 + b_3 + b_4 = b_5 + b_6 + b_7 + b_8 = b_9 + b_{10} + b_{11} + b_{12}$. Any two of several restrictions could be made. The following two could be chosen:

$b_3 = b_9 + b_{10} + b_{11} + b_{12} - b_1 - b_2 - b_4$ and

$b_6 = b_9 + b_{10} + b_{11} + b_{12} - b_5 - b_7 - b_8$.

Imposing these two restrictions (side conditions) yields:

$$Y = b_pP + b_1X_1 + b_2X_2 + (b_9 + b_{10} + b_{11} + b_{12} - b_1 - b_2 - b_4)X_3 + b_4X_4 + b_5X_5 + (b_9 + b_{10} + b_{11} + b_{12} - b_5 - b_7 - b_8)X_6 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_7. \quad [8]$$

or

$$Y = b_pP + b_1(X_1 - X_3) + b_2(X_2 - X_3) + b_4(X_4 - X_3) + b_5(X_5-X_6) + b_7(X_7 - X_6) + b_8(X_8 - X_6) + b_9(X_9 + X_3 + X_6) + b_{10}(X_{10} + X_3 + X_6) + b_{11}(X_{11} + X_3 + X_6) + b_{12}(X_{12} + X_3 + X_6) + e_7. \quad [9]$$

Equation 9 (or reparameterization of it, either using different restrictions expressing the side conditions, and/or including a unit vector) then serves as a full model for testing within group hypotheses:

$R^2 = .90057$.

Now, direct hypotheses can be tested by placing appropriate restrictions simultaneously with the side conditions. With the hypothesis $\bar{Y}_1 - \bar{Y}_4 = \bar{Y}_5 - \bar{Y}_8$ or, in terms of the regression coefficients, $b_1 - b_4 = b_5 - b_8$ or $b_1 = b_5 - b_8 + b_4$, as before.

Then, placing all three restrictions simultaneously on equation 7 yields:

$$Y = b_pP + (b_5 - b_8 + b_4)X_1 + b_2X_2 + (b_9 + b_{10} + b_{11} + b_{12} - b_5 + b_8 - b_4 - b_2 - b_4)X_3 + b_4X_4 + b_5X_5 + (b_9 + b_{10} + b_{11} + b_{12} - b_5 - b_7 - b_8)X_6 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_8; \quad [10]$$

or

$$Y = b_pP + b_2(X_2 - X_3) + b_4(X_4 - 2X_3 + X_1) + b_5(X_5 + X_1 - X_3 - X_6) + b_7(X_7 - X_6) + b_8(X_8 + X_3 - X_6 - X_1) + b_9(X_9 + X_3 + X_6) + b_{10}(X_{10} + X_3 + X_6) + b_{11}(X_{11} + X_3 + X_6) + b_{12}(X_{12} + X_3 + X_6) + e_8. \quad [11]$$

Note that the restrictions are made simultaneously with the side conditions on the full model (equation 7). Equation 11 could be reparameterized (necessary with computer programs that automatically introduce a unit vector) by setting equal to zero any of the remaining b_i in equation 11 (excepting b_p. Doing this yields $R_R^2 = .87854$; $F = \frac{(.90057 - .87854)/1}{.09943/36} = 7.976$, $t = \sqrt{F} = 2.84$ this is the same t value found earlier.

This process could be repeated for any other hypothesis, imposing the restriction implied by the hypothesis simultaneously with the side conditions. Care must be taken to be sure that hypotheses tested on this model are appropriate; such hypotheses must be some combination of within group contrasts.

Directly Using the Full Model

Had equation 7 been used directly, it can be seen that the outcome is comparable to using side conditions:

$$Y = b_pP + b_1X_1 + b_2X_2 + \ldots + b_{12}X_{12} + e_6. \quad [7]$$

Testing $\overline{Y}_1 - \overline{Y}_4 + \overline{Y}_5 - \overline{Y}_8$ can be done using the restriction $b_1 - b_4 = b_5 - b_8$, or $b_1 = b_5 - b_8 + b_4$, as before.

Then,

$$Y = b_pP + (b_5 - b_8 + b_4)X_1 + b_2X_2 + \ldots + b_{12}X_{12} + e_9, \text{ or}$$

$$Y = b_pP + b_2X_2 + b_3X_3 + b_4(X_4 + X_1) + b_5(X_5 + X_1) + b_6X_6 + b_7X_7 + b_8(X_8 - X_1) + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_9.$$

Reparameterizing by (arbitrarily) choosing $b_{12} = 0$,

$$Y = b_0 + b_pP + b_2X_2 + b_3X_3 + b_4(X_4 + X_1) + b_5(X_5 + X_1) + b_6X_6 + b_7X_7 + b_8(X_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + e_9. \quad [12]$$

Equation 7 yields $R^2 = .90057$, and equation 12 yields $R^2 = .87854$; $F = \frac{(.90057 - .87854)/1}{.09943/36} = 7.976$; $t = \sqrt{F} = 2.842$, the same result as was found by the first two methods.

It can be seen that several different approaches can be used to test hypotheses in a repeated measures designs. The use of the criterion Y** where Y** = Y - Y* when Y* = 1/4P, as was shown in Williams (1980) allows an appropriate testing procedure. The use of side conditions (which uses a model removing the nesting effect) or a model containing the group membership variables and the person-score vector (directly using the full model) yield identical results. Perhaps the latter approach would be conceptually easier to understand. The direct use of equation 7 can be completed despite the nesting of the group effects.

The present paper, like the earlier one, has shown three different types of solutions for testing hypotheses (contrasts) of interest. All three methods yield accurate results for within group comparisons. While they yield results that are equivalent, they are not conceptually equal in terms of their understandability. The first method has the drawback of using a constructed criterion; method two, using side conditions, is unnecessarily complex; extreme care must be used to achieve intended results. Our preference is clearly on the side of the third approach, directly using the full model and making restrictions of research interest upon it. From the point of view of actual use, the third method is sufficient and clearly preferable. On the other hand, the relationship to the other two solutions is at least interesting.

References

Fraas, J. W. and McDougall, W. R. The use of one full MLR model to conduct multiple comparisons in a repeated measures design: An industrial application. Multiple Linear Regression Viewpoints, 1983, 12, No. 1, 42-55.

Lindquist, E. F. Research design in psychology and education. Boston: Houghton-Mifflin, 1953.

Pedhazur, E. J. Coding subjects in repeated measures designs. Psychological Bulletin, 1977, 84, 298-305.

Williams, J. D. Regression analysis in educational research. New York: MSS Publishing Co., 1974.

Williams, J. D. Multiple comparisons by multiple linear regression. Monograph Series #2, Multiple Linear Regression Viewpoints, 1976, 7, No. 1.

Williams, J. D. A note on coding the subjects effect in treatments x subjects designs. Multiple Linear Regression Viewpoints, 1977, 8, No. 1, 32-35.

Williams, J. D. Multiple comparisons in higher dimension designs, Monograph Series #5, Multiple Linear Regression Viewpoints, 1980, 10, No. 3.

Williams, J. D. and Williams, J. A. Testing hypotheses in a repeated measures design on employee attitudes with large samples. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, April, 1984.