MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 11, NUMBER 3 SUMMER 1982

USING LINEAR MODELS TO SIMULTANEOUSLY ANALYZE A SOLOMON FOUR GROUP DESIGN

John D. Williams
University Of North Dakota

Isadore Newman
University Of Akron

Solomon (1949) devised a design to control threats to design validity (Campbell and Stanley, 1966). Using the notation of Campbell and Stanley, the four groups can be diagrammed as

R	0_1	X	0_2
R	0_3		0_4
R		X	0_5
R			0_6

The four groups are respectively Group One: An experimental groups that has been pretested and posttested; Group Two: A control group that has been pretested and posttested; Group Three: An experimental group that has been posttested only; and Group Four: A control group that has been posttested only.

Campbell and Stanley state, "There is no singular statistical procedure which makes use of all six sets of observations simultaneously." (p. 24)

The Solomon Four Group Design, while very simple conceptually, can be very misleading depending upon the statistical analysis. Campbell and Stanley (1966) have a preferred approach, in which they set up a 2x2 factorial design.

Presented at AERA 1982, MLR Special Interest Group
Not refereed by editorial staff

		Treatment	
		Yes	No
Pre-Test	Yes	Gr1	Gr2
	No	Gr3	Gr4

In this design only the posttest scores are statistically analyzed. This procedure does not allow one to control for the pretest scores in groups 1 and 2, thereby losing some power; it does however estimate the effects of treatments that are independent of individuals having a pretest and treatment pretest interaction. It also tests for the effects of pretesting, independent of treatment and pretest-treatment interaction, on posttest scores. Finally, the approach estimates the effects of pretest-treatment interaction, on posttest scores.

One of the advantages of writing specific regression models which reflect research questions is that one is less likely to have a statistical answer that is unrelated to the researcher's question of interest. The following are a variety of regression models which will reflect potential research questions that can be ascertained from the Solomon Four Group Design. It should be remembered that there is not one correct answer.

Recently, Newman, Benz, and Williams (1980) devised a way to analyze data that, by extension, might be applied to Solomon type designs. A unique property of this technique is that, the statement by Campbell and Stanley not withstanding, a single statistical procedure can be employed which makes use of all six sets of observations simultaneously. On the other hand, the solution(s) may prove to be no more satisfactory than existing possibilities that split the data into two sets. In the end, the Solomon Four Group design may prove to be one of those recalcitrant research situations that leave the would be analysts foundered on the shoal of a simple design whose simplicity is only a deception.

Consider the following research situation. Five people in each group have scores such that one experimental group has been pretested and posttested and one experimental group has been posttested only. Two similarly tested control groups are also included. Data for such a situation are given in Table 1.

Table 1

Data for a Solomon Four Group Design

Experimental: Group One		Control: Group Two		Experimental: Group Three	Control: Group Four
Pretest	Posttest	Pretest	Posttest	Posttest	Posttest
5	15	5	8	13	9
7	12	4	7	10	8
5	10	4	8	12	6
12	17	6	6	11	3
6	11	6	6	14	4

Several different approaches might be tried. One approach would be to divide the data into two sets: Groups One and Two (those who were both pretested and posttested) as one set, and the posttested only groups (Groups Three and Four) as the second set. The latter set can be simply tested by the use of the t test:

$t = 4.24$ ($p<.05$).

The former data set (Groups One and Two) can be conceived either as a repeated measures design or as a problem that can be approached through the analysis of covariance (or related techniques such as residual gain analysis).

To approach the problem first as an analysis of covariance, the following variables can be defined:

Y = the criterion, or posttest score;

X_1 = the pretest score;

X_2 = 1 if the score is from the experimental group, 0 if the score is from the control group;

X_3 = 1 if the score is from the control group, 0 if the score is from the control group.

Then either of two full models can be used:

$$Y = b_1X_1+b_2X_2+b_3X_3+e_1, \qquad (1)$$

or

$$Y = b_0+b_1X_1+b_2X_2+e_1. \qquad (2)$$

Equation 2 utilizes the unit vector in the process of generating a constant whereas equation 1 does not. Either model will yield the same R^2 value.

The restricted model (with equation 2 as the full model) is of the form:

$$Y = b_0+b_1X_1+e_2. \qquad (3)$$

For this data set R_2^2 = .79379, R_3^2 = .42334, $F = \frac{(.79379 - .42334)/1}{(1 - .79379)/7} = 12.58$, $p<.05$.

Using a Repeated Measures Approach

If the problem is visualized as a repeated measures design wherein the pretest is the first measure and the posttest is the second measure, then the design is like the Type I design shown in Lindquist (1953) and can be achieved through a regression approach (Williams, 1974). For a regression formulation, see Table 2.

Table 2
Design Matrix for a Repeated Measures Problem

Y	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	X_1	X_2	X_3	X_4	X_5
5	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1
7	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1
5	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1
12	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1
6	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1
5	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0
12	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0
10	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0
7	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0
11	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0
5	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0
6	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0
6	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0
8	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0
7	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0
8	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0
6	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0
6	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0

Here, Y = the criterion test score;

P_1 thru P_{10} are binary coded person vectors (1 if the person, 0 otherwise);

X_1 = 1 if the score comes from a person in the experimental group, 0 otherwise;

X_2 = 1 if score comes from a person in the control group, 0 otherwise;

X_3 = 1 if the score occurs with a pretest situation, 0 otherwise;

X_4 = 1 if the score occurs with a posttest situation, 0 otherwise; and

$X_5 = X_1 . X_3$.

Several models can be used to generate an analysis. The use of the following is instructive:

$$Y = b_0+b_1P_1+b_2P_2+\ldots+b_9P_9+e_3; \qquad (4)$$

(or alternatively, $Y = b_1P_1+b_2P_2+\ldots+b_{10}P_{10}+e_3$)

$$Y = b_0+b_1X_1+e_4; \qquad (5)$$

$$Y = b_0+b_3X_3+e_5; \qquad (6)$$

$$Y = b_0+b_1X_1+b_3X_3+e_6; \qquad (7)$$

$$Y = b_0+b_1X_1+b_3X_3+b_5X_5+e_7; \quad (8)$$

and

$$Y = b_0+b_1P_1+b_2P_2+\ldots+b_9P_9+b_{10}X_3+b_{11}X_5+e_8. \quad (9)$$

For the preceding, $R_4^2 = .54297$;

$R_5^2 = .31250$;

$R_6^2 = .31250$;

$R_7^2 = .62500$;

$R_8^2 = .70312$ and $R_9^2 = .93359$.

What might have occured if a model of the following form were used?

$$Y = b_0+b_1P_1+b_2P_2+\ldots+b_9P_9+b_{10}X_1+b_{11}X_3+b_{12}X_5+e_9.$$

It would not sensibly yield $R^2 = .54297 + .70312 = 1.24609$. Such a model would fail because the effect for experimental-control is "nested" in the subject (or person) effect. To test for the experimental-control effect,

$$F = \frac{R_5^2/1}{(R_4^2-R_5^2)/(P-1-1)} = \frac{.31250/1}{(.54297-.31250)/(10-2)} = 10.85,$$

$p<.05$.

To test for the test-retest effect,

$$F = \frac{R_6^2/1}{(1-R_9^2)/(N-P-1-1)}$$

$$= \frac{.31250}{.06641/8} = 37.65, \; p<.01.$$

The interaction is tested by

$$F = \frac{(R_8^2-R_7^2)/1}{(1-R_9^2)/(N-P-1-1)} = \frac{(.70312-.62500)/1}{.06641/8} = 9.41, \; p<.05.$$

Note that the interaction effect can be conceptualized as actually being additional evidence for the experimental effect. The higher increases in the experimental group will show up in part as interaction for a repeated measures design.

The usual summary table for the repeated measures design can be constructed. he summary table is shown in Table 3.

Table 3

Summary Table for Repeated Measures Design

	df	SS	MS	F
ubjects	9	139.00		
xperimental-Control	1	80.00	80.00	10.85
rror (a)	8	59.00	7.375	
ithin Subjects	10	117.00		
test-retest	1	80.00	80.00	37.65
interaction	1	20.00	20.00	9.41
rror (b)	8	17.00	2.125	
otal	19	256.00		

d

Using All Six Groups Simultaneously

As the Solomon design is approached, several conceptual issues ensue. Is this to be seen as a six group design with attendant solutions? If the researcher opts for a six group design, person vector information needs to be excluded. Indeed, this was also true in the previous section. At no time were the four groups and person vectors used simultaneously; if it were, the R^2 was theoretically to be 1.24609, obviously an impossibility. If a six group design is to be used, what dimensions would be appropriate? This could be considered to be a one-way lay-out, a two-way lay-out, or a three-way lay-out (but with two missing cells) only the one-way and three-way layouts are discussed here. First hypotheses with a one-way lay-out as addressed.

Consider the following variables:

Y = the criterion score;

X_1 = 1 if the score is a pretest score from a member of the experimental group, 0 otherwise;

X_2 = 1 if the score is a posttest score from a member of the experimental group that has been pretested, 0 otherwise;

X_3 = 1 if the score is a pretest score from a member of the control group, 0 otherwise;

X_4 = 1 if the score is a posttest score from a member of the control group that was pretested, 0 otherwise;

X_5 = 1 if the score is from a member of the experimental group that was <u>not</u> pretested, 0 otherwise; and

X_6 = 1 if the score is from a member of the control group that was <u>not</u> pretested, 0 otherwise.

For the six group situation, the full model is:

$$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_9. \quad (10)$$

At least two different sets of restrictions might make sense in addressing the Solomon design. One such set would be $b_2-b_1=b_4-b_3$, which addresses the hypothesis $\bar{Y}_2-\bar{Y}_1=\bar{Y}_4-\bar{Y}_3$, as the hypothesis that the gains in the twice tested experimental and control groups are equal; also, the second restriction is $b_5=b_6$ as the once tested experimental groups have equal means: $\bar{Y}_5=\bar{Y}_6$.

The first restriction can be rewritten as $b_2=b_4-b_3+b_1$: Placing these two restrictions on the Full Model:

$$Y = b_1X_1+(b_4-b_3+b_1)X_2+b_3X_3+b_4X_4+b_5X_5+b_5X_6+e_{10} \quad (11)$$

$$Y = b_1(X_1+X_2)+b_3(X_3-X_2)+b_4(X_4+X_2)+b_5(X_5+X_6)+e_{10}. \quad (12)$$

Letting $D_1 = X_1+X_2$;

$D_2 = X_3-X_2$;

$D_3 = X_4+X_2$; and

$D_4 = X_5+X_6$, the restricted model is:

$$Y = b_1D_1+b_3D_2+b_4D_3+b_5D_4+e_{10}. \quad (13)$$

Here, $R^2_{10} = .71183$; $R^2_{13} = .42882$.

$$F = \frac{(R^2_{10}-R^2_{13})/2}{(1-R^2_{10})/(N-6)} = \frac{.28301/2}{(1-.71183)/24} = 11.79,\ p<.01.$$

This F test tests simultaneously $\overline{Y}_2-\overline{Y}_1=\overline{Y}_4-\overline{Y}_3$ and $\overline{Y}_5=\overline{Y}_6$; placing both sets of restrictions allows the rejection of the null hypotheses. If these restrictions are equivalent to hypotheses the researcher had in mind, then there is no further problem. Translating the meaning of these two hypotheses into English may leave the researcher somewhat uneasy; however, one attempt at a translation into English is: It is not simultaneously true that there is no differences in the means of the non-pretested group and that there is no differences in the gains of the pre-tested groups.

One approach would be to test each of these hypotheses separately and using Dunn's (1961) test for multiple comparisons. Imposing the first restriction separately ($b_2-b_1=b_4-b_3$) yields $Y = b_1X_1+(b_4-b_3+b_1)X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{11}$;

$Y = b_1(X_1+X_2)+b_3(X_3-X_2)+b_4(X_4+X_2)+b_5X_5+b_6X_6+e_{11}$. (14)

Then using D_1, D_2 and D_3 as previously defined, $Y = b_1D_1+b_3D_2+b_4D_3+b_5X_5+b_6X_6+e_{11}$. (15)

R^2_{15} = .66038 and

$$F = \frac{(R^2_{10}-R^2_{15})/1}{(1-R^2_{10})/(N-6)} = \frac{.71183-.66038}{(1-.71183)/24} = \frac{.05145}{(1-.71183)/24} = 4.29.$$

$t = \sqrt{F} = 2.07$. Since two contrasts are planned, a value of 2.39 is necessary for significance of the .05 level, hence the hypothesis $\mu_2-\mu_1=\mu_4-\mu_3$, corresponding to $\overline{Y}_2-\overline{Y}_1=\overline{Y}_4-\overline{Y}_3$ cannot be rejected. The imposition of the second restriction ($b_5=b_6$) yields:

$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{12}$;

$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5(X_5+X_6)+e_{12}$. (16)

Using D_4,

$Y = b_1X_1+b_2X_2+b_3X_3+b_4X_4+b_5D_4+e_{12}$. (17)

R^2_{17} = .48028 and $F = \frac{(R^2_{10}-R^2_{17})/1}{(1-R^2_{10})/(N-6)} = \frac{.71183-.48028}{(1-.71183)/24} = \frac{.23155}{(1-.71883)/24} = 19.28,$

$t = \sqrt{F} = 4.39$, $t>3.09$ from Dunn's table, so that $p<.01$. Note also that from the numerator of these two tests that $.05145 + .23155 = .28300$, within rounding error of the numerator when both restrictions were applied; this is because these contrasts are independent. From these calculations, it can be seen that the greatest portion of the rejection of the hypotheses tested by the restrictions in equation 13 is due to the differences in the groups that were posttested only rather than due to differential increases.

A second set of restrictions (actually, a single restriction) is given as $(b_2-b_1)-(b_4-b_3)=b_5-b_6$. This restriction tests the hypothesis related to $(\overline{Y}_2-\overline{Y}_1)-(\overline{Y}_4-\overline{Y}_3)=\overline{Y}_6-\overline{Y}_5$; that is, the difference between the mean of the gain scores is equal to the difference in posttest measures of the non-pretested group. The restriction can be stated as $b_2=b_5-b_6+b_1+b_4-b_3$. Imposing this restriction yields:

$$Y = b_1X_1+(b_5-b_6+b_1+b_4-b_3)X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{13}; \qquad (18)$$

$$Y = b_1(X_1+X_2)+b_3(X_3-X_2)+b_4(X_4+X_2)+b_5(X_5+X_2)+b_6(X_6-X_2)+e_{13} \qquad (19)$$

Using D_1, D_2, D_3 and defining $D_5 = X_5+X_2$ and $D_6 = X_6-X_2$, equation 19 can be rewritten as $Y = b_1D_1+b_3D_2+b_4D_3+b_5D_5+b_6D_6+e_{13}$. (20)

$R^2_{20}=.70326$.

Then $F = \dfrac{(.71183-.70326)/1}{(1-.71183)/24} = \dfrac{.01857}{(1-.71183)/24} = 1.55$,

which is non-significant. Thus, while we have previously showed that the differences between the posttested groups is significant ($p<.01$) and the differences in gains in the pretested groups are non-significant ($p>.05$), there are no significant differences between the gain of the mean scores and the posttested only groups differences. This is not to say the outcomes for the Solomon design are uninterpretable; it does say that the interpretations are tricky.

Viewing the Solomon as a Three-Way Design

It is possible to view the Solomon design as a 2x2x2 design with two missing cells. The missing cells are planned, as was the case in missing cell design described by Williams and Wali (1979). In diagramatic form, the three dimensional case can be seen as:

	Pretested Pre	Pretested Post	Non-Pretested Pre	Non-Pretested Post
Experimental	Group 1	Group 2	X	Group 5
Control	Group 3	Group 4	X	Group 6

To test for the experimental-control main effect (A effect), the following restriction can be imposed:

$$b_1+b_2+b_5=b_3+b_4+b_6$$

which yields

$$Y = b_2(X_2-X_1)+b_3(X_3+X_1)+b_4(X_4+X_1)+b_5(X_5-X_1)+b_6(X_6+X_1)+e_{14}. \quad (21)$$

Defining $D_7 = X_2-X_1$;

$D_8 = X_3+X_1$;

$D_9 = X_4+X_1$;

$D_{10} = X_5-X_1$; and

$D_{11} = X_6+X_1$

$$Y = b_2D_7+b_3D_8+b_4D_9+b_5D_{10}+b_6D_{11}+e_{14}. \quad (22)$$

$R^2_{22} = .29159$;

$$F = \frac{(.71183-.29159)/1}{(1-.71183)/24} = \frac{.42024}{.28817/24} = 35.00,\ p<.01.$$

To test the effect of pretesting (the B effect), several rival hypotheses might be used to serve as the main effect.

One such hypothesis is $b_1+b_2+b_3+b_4=b_5+b_6$. This hypothesis does not test the more appropriate hypothesis of interest, since the pretested scores are being compared to the scores which have been posttested only. More interesting is $b_2+b_4=b_5+b_6$ or $b_2=b_5+b_6-b_4$.

Then,

$$Y = b_1X_1+b_3X_3+b_4(X_4-X_2)+b_5(X_5+X_2)+b_6(X_6+X_2)+e_{15}. \quad (23)$$

Defining $D_{12} = X_4-X_2$;

$D_{13} = X_5+X_2$;

$D_{14} = X_6+X_2$,

$$Y = b_1X_1+b_3X_3+b_4D_{12}+b_5D_{13}+b_6D_{14}+e_{15}. \quad (24)$$

$R^2_{24} = .69897$;

$$F = \frac{(.71183-.69897)/1}{(1-.71183)/24} = \frac{.01286}{(1-.71183)/24} = 1.07,\ p>.05.$$

The outcome of this test would suggest that the effect of pretesting per se is minimal for this data set.

To test for pre-post differences (the C main effect), the restriction $b_1+b_3=b_2+b_4$ or $b_1=b_2+b_4-b_3$ can be imposed. Then

$Y = (b_2+b_4-b_3)X_1+b_2X_2+b_3X_3+b_4X_4+b_5X_5+b_6X_6+e_{16}$, or

$$Y = b_2(X_2+X_1)+b_3(X_3-X_1)+b_4(X_4+X_1)+b_5X_5+b_6X_6+e_{16}. \quad (25)$$

Letting $D_{15} = X_3-X_1$, equation 25 can be rewritten

$$Y = b_2D_1+b_3D_{15}+b_4D_9+b_5X_5+b_6X_6+e_{16}. \quad (26)$$

$R^2_{26}=.50600$;

$$F = \frac{(.71183-.50600)/1}{(1-.71183)/24} = \frac{.20583}{(1-.71183)/24} = 17.14,\ p<.01,$$

indicating a pre-test increase in scores.

Interactions in the Three-Way Design

First of all, the two missing cells will cause the non-existence of two interactions. The three way interaction will not exist, since it is impossible to have non-pretested groups who were pretested. For the same reason, the BC interaction will fail to exist. To test for the AB interaction, that is, the interaction between the experimental-control condition

(A) and the effect of pretesting (B), the restriction on the full model would be: $b_2 - b_5 = b_4 - b_6$ or $b_2 = b_4 - b_6 + b_5$.

Then $Y = b_1X_1 + (b_4 - b_6 + b_5)X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + e_{17}$, or

$$Y = b_1X_1 + b_3X_3 + b_4(X_4 + X_2) + b_5(X_5 + X_2) + b_6(X_6 - X_2) + e_{17}. \quad (27)$$

Using previously defined transformations, $Y = b_1X_1 + b_3X_3 + b_4D_3 + b_5D_{13} + b_6D_6 + e_{17}$. (28)

$R^2_{17} = .71183$; R^2_{17} is identical to the R^2 for the full model. This is circumstantially so because $\bar{Y}_2 - \bar{Y}_5 = \bar{Y}_4 - \bar{Y}_6 = 13-12 = 7-6$. Thus, the AB interaction is equal to zero.

To test the AC interaction, that is, the experimental-control condition (A) with pre-post differences (C), the restriction $b_2 - b_1 = b_4 - b_3$ would be imposed on the full model. This in fact was already done in equation 15, yielding $R^2_{15} = .66038$, $F = 4.29$, $p > .05$. The results from the three-way analysis can be placed into a summary table; see Table 4.

Table 4

Summary Table for a Three-Way Solution to the Solomon Design

Effect	Restriction	R^2	df	SS	MS	F
Full Model						
A (experimental-control)	$b_1+b_2+b_5=b_3+b_4+b_6$	.29159	1	163.33	163.33	35.00
B (pretesting)	$b_2+b_4=b_5+b_6$	.69897	1	5.00	5.00	1.07
C (pre-post differences)	$b_1+b_3=b_2+b_4$	.50600	1	80.00	80.00	17.14
AB	$b_2-b_5=b_4-b_6$	.71183	1	0	0	0
AC	$b_2-b_1=b_4-b_3$	.66038	1	20.00	20.00	4.29
Deviation from Full Model		.28817	24	112.00	4.67	

Finding the sum of squares in Table 4 is facilitated by knowing $SS_T = 388.67$. Also, the C effect and the AC effect are identical to the same effects as shown in Table 3.

References

Campbell, D.T. and Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1966.

Dunn, O.J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56, 52-64.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston: Houghton-Mifflin Company, 1953.

Newman, I., Benz, C.R. and Williams, J.D. Repeated measures: Suggested approach for simultaneously independent and dependent measures. Paper presented at the American Psychological Association Annual Meeting, Montreal, Canada, September 1980.

Solomon, R.L. An extension of control group design. Psychological Bulletin, 1949, 46, 137-150.

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

Williams, J.D. and Wali, M.K. Missing cells and a curious case of degrees of freedom. Multiple Linear Regression Viewpoints, 1979, No. 5, 75-87.