

A SAS[®] Program Routine to Derive the Required Data Structure to Select Matched Pairs of Observations for Evaluating Quasi-Experimental Research Designs

Gunapala Edirisooriya
East Tennessee State University, Johnson City, TN

ABSTRACT

A misguided practice among educational researchers is the application of ANCOVA technique to equalize the preexisting differences between control [C] and experimental [E] groups in quasi-experimental research designs. In educational research, this is particularly inappropriate when regression to the mean is more likely to observe. A better approach is to select matched pairs of observations from C and E, through ex-post stratified sampling, and to compare the gains-scores utilizing the dependent t-test. SAS[®] program routines can be used to create an algorithm to select matched pairs of observations effectively and efficiently. Using pretest scores as a covariate I sorted C and E and merged them by a unique identifier variable. This resulted in a data set with matched pairs of observations, which can be used for evaluating project outcomes. The main focus of this paper is to explain a prototype of a SAS[®] program routine, which can be readily adapted for numerous settings, to select matched pairs of observations for evaluating quasi-experimental research designs.

INTRODUCTION

In experimental/quasi-experimental research, researchers generally compare the outcomes of C and E. If E performs better than C, then the experiment/program is

assumed to be effective. For this criterion to hold, at least, C and E must be comparable at the outset of the experiment. If not, E's better performance over C may stem, at least partly, from the preexisting differences between the two groups. In the field of education, researchers have to operate under certain constraints which prevent them conducting true experiments. Specifically, they do not have the freedom to select sample sizes and/or to assign subjects to C and E so that comparable groups can be formed prior to the experiment. Students are selected for special projects on the basis of low scores on a criterion test. In reality, even the evaluation designs are conceptualized only after programs have been fully implemented. This type of project has to be analyzed in the context of quasi-experimental research design. Therefore, in evaluating educational projects we must take special care to examine the data.

At the outset of an experiment, C and E may differ quantitatively and qualitatively. Quantitatively, C and E may seem equivalent, but it does not necessarily imply that C and E are equivalent qualitatively. For example, an equal sample size alone does not necessarily equate the two groups qualitatively. Similarly, C and E may have similar average scores on a pretest, but the two groups can have completely different score distributions. This could well be the case when intact groups are selected as C and E. In educational settings, if the preexisting conditions differ substantially, regression effect poses a major problem.

Therefore, to analyze and interpret the data correctly, every effort should be made to scrutinize them. Application of ANCOVA, a widespread practice among educational researchers, does not necessarily equalize the preexisting differences between C and E. A better approach is to use matched pair comparisons. To use this approach, we need to create an algorithm to select matched pairs of observations from C and E. The following example is used to explain an algorithm derived through the SAS[®] program routines.

AN EXAMPLE

Assume that an LEA introduced a new reading program called README in a number of its elementary schools. README was implemented in grades 1 through 5 in five schools, E1, ..., E5. At the conclusion of the program, this LEA wanted to find out how effective this program was. Based on school characteristics, five comparable schools, C1, ..., C5, were selected as the control group. The total number of students remained in README was 444. The total number of students in the selected C1, ..., C5 in grades 1 through 5 was 1059. The plan was to evaluate the effectiveness of README by comparing the NCE gains between the two groups in Reading Comprehension (RCOM94GN) of the Comprehensive Tests of Basic Skills (CTBS4). The average pretest score (50.0 NCE) of C was higher than the average pretest score (30.0 NCE) of E. Similarly, in every grade, average NCE of C was higher than the average NCE of E. These data are shown in Table 1. This suggests that regression to the mean is a possible threat to internal validity of this experiment. Therefore, a better approach to evaluating this project is to use matched pair comparisons. Details of the algorithm used in selecting matched pairs of observations from C

and E follow.

DATA STRUCTURE

In the data set (READ), we have student identification number (PIFID), the last digit of student identification number (PIFIDL), program type (PROG), school number (SCH94T), grade (GRD94T), pretest scores (RCOM93N), posttest scores (RCOM94N), and score-gains (RCOM94GN). This algorithm involves two procedural steps: merging and concatenation. In each of these procedural steps, a number of data manipulation steps have to be performed. The merging procedure involves the following steps. First, based on the PROG-type the data set, READ has to be divided into two data sets, READC and READE. Second, we have to rename the variables in READE taken from READ to identify them as separate variables since we intend to merge them later with READC. Third, READE is sorted by pretest scores (RCOM93N) within each grade (GRD94T). In all likelihood, in each grade, there would be more than one student with the same score. Therefore, the last digit of the student identification number (PIFIDL) is also used to infuse randomness in sorting the data. The number of variables that should be used in the sorting procedure depends on the number of categorical variables that are critically important for matching subjects. Fourth, in a do loop, a new variable (COUNT) is created by assigning counter values incrementally by 0.01 for each new value of RCOM93N. This would create a unique value for each similar score. Fifth, a unique identification variable (HOOK) is created by combining the pretest scores (RCOM93N) and COUNT. For example, for any given grade, if there are eight students with an RCOM93N score of 32, then COUNT will contain .01, .02, ..., .08 and

HOOK will contain 32.01, 32.02, ..., 32.08. Sixth, the data set, READE, is sorted by GRD94T and HOOK. We have to repeat these data manipulation steps for the data set, READC, except the second step. Then, the two data sets, READE and READC, have to be merged by GRD94T and HOOK. This results in a new data set, BOTH. We have to reorganize some of the variables in BOTH to perform the concatenation procedure.

The data set BOTH includes a number of redundant variables which can be discarded. First, another data set, EXPMENT, is created using BOTH. Second, the variables brought from the data set READC have to be dropped. Third, the renamed variables in READE have to be reversed back to the original names. Similarly, another data set, REGULAR, is created using BOTH. Then, the variables brought from the data set READE have to be dropped. Finally, the data set, READALL, is created by concatenating the two data sets, EXPMENT and REGULAR.

The data set, READALL contains all the matched pairs of observations selected from C and E. READALL includes 570 students; for each student in E, there is a student in C with a matching score. The data set, READALL, can be sorted by PROG and GRT94T and is ready for project evaluation. For comparison purposes, I analyzed these data in three different ways: ANOVA, ANCOVA, and matched-pair comparisons. According to ANOVA results, E performed better than C overall; and in all grades, except in grade 1, E performed better than C. According to ANCOVA, overall, there was no difference between C and E. In all grades, except in grade 1, students in E performed better than students in C. The matched-pair comparison analysis yielded different results.

Overall, students in C did better than students in E. In all other grades, except in grade 5, students in C did better than students in E. This finding supports the warnings on the use of ANCOVA in program evaluation, especially when regression to the mean is a serious threat to internal validity (Edirisooriya, 1996; Loftin & Madison, 1991).

CONCLUSION

Accordingly, in educational research, which are of quasi-experimental nature, the application of ANCOVA to equalize preexisting differences between C and E is not a proper strategy. This is particularly the case when the achievement level of C is higher than the achievement level of E at the beginning of an experiment. On average C tends to regress downward toward the mean while E tends to regress upward toward the mean. More over, when E's scores are adjusted using pretest scores as a covariate, this will create an unfair advantage for E compared to C. Therefore, a better approach would be to select matching pairs of observations from C and E. In this way, we can equalize the distributions of C and E and we can assume that the two groups were of equal standing at the beginning of the experiment. A complete list of SAS[®] programming steps for selecting matched pairs of observations from C and E follows:

```
DATA READE;
  SET READ;
  IF PROG = 'EXPREAD';
  RENAME RCOM93N=RCOM93NS
  RCOM94N=RCOM94NS
  RC94GN=RC94GNS
  PIFID=PIFIDS
  PIFIDL=PIFIDLS
  PROG=PROGS
  SCH94T=SCH94TS;
PROC SORT DATA=READE;
```

```

BY GRD94T RCOM93NS PIFIDLS;
DATA READEN;
SET READE;
BY GRD94T RCOM93NS PIFIDLS;
IF FIRST.RCOM93NS THEN DO;
  COUNT=0;
END;
COUNT=COUNT+.01;
RETAIN COUNT;
DATA READEN;
SET READEN;
HOOK=RCOM93NS+COUNT;
PROC SORT DATA=READEN;
BY GRD94T HOOK;
DATA READC;
SET READ;
IF PROG = 'REGREAD';
PROC SORT DATA=READC;
BY GRD94T RCOM93N PIFIDL;
DATA READCN;
SET READC;
BY GRD94T RCOM93N PIFIDL;
IF FIRST.RCOM93N THEN DO;
  COUNT=0;
END;
COUNT=COUNT+.01;
RETAIN COUNT;
DATA READCN;
SET READCN;
HOOK=RCOM93N+COUNT;
PROC SORT DATA=READCN;
BY GRD94T HOOK;

DATA BOTH BOTHN;
MERGE READEN (IN=A) READCN (IN=B);
BY GRD94T HOOK;
IF A & B THEN OUTPUT BOTH;
IF NOT (A & B) THEN OUTPUT BOTHN;
DATA EXPMENT;
SET BOTH;
DROP PIFID RCOM93N RCOM94N
PROG SCH94T
PIFIDL PIFIDLS RC94GN;
DATA EXPMENT;
SET EXPMENT;
RENAME RCOM93NS=RCOM93N
RCOM94NS=RCOM94N

```

```

RC94GNS=RC94GN
PIFIDS=PIFID
PROGS=PROG
SCH94TS=SCH94T;
DATA REGULAR;
SET BOTH;
DROP PIFIDS RCOM93NS RCOM94NS
PROGS S SCH94TS
PIFIDLS PIFIDL RC94GNS;
DATA READALL;
SET EXPMENT REGULAR;
PROC SORT DATA=READALL;
BY PROG GRD94T;

```

□

BIBLIOGRAPHY

- Edirisooriya, G. (1996). Compared to ANCOVA, Matched Pairs of Observations is a Better Option to Equalize Preexisting Differences Between Control and Experimental Groups. Paper presented at Annual Meeting of the American Educational Research Association in New York.
- Lofin, L. & Madison, S. (1991). The Extreme dangers of covariance corrections. In B. Thompson (Ed.), *Advances in educational research: Substantive findings, methodological developments* (vol. 1, pp. 133-148). Greenwich, CT: JAI Press.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart, and Winston.
- Thompson, B. (1992). Misuse of ANCOVA and related "statistical control" procedures. *Reading Psychology*, 13, iii-xviii.
- Wildt, A. R. & Ahtola, O. T. (1978). *Analysis of covariance*. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 07-012, Beverly Hills: Sage.

ACKNOWLEDGMENT

I highly appreciate Maggie Wyatt's assistance in preparing this paper.

Gunapala Edirisooriya
 Department of Educational Leadership and
 Policy Analysis
 East Tennessee State University
 Johnson City, TN 37614-0550
 Tel: (423) 439 4255
 E-Mail: edirig@etsu.etsu-tn.edu

Table 1

Pretest, Posttest, and Gain-Scores of C and E

Unit	RCOM93N		RCOM94N		RCOM94GN	
	C	E	C	E	C	E
Aggregate	50.0	30.0	44.7	30.1	-5.0	-0.0
Grade 1	60.1	53.1	50.7	39.4	-9.4	-13.7
Grade 2	58.7	31.5	43.1	26.9	-15.7	-4.5
Grade 3	51.7	26.8	48.9	28.6	-2.8	1.9
Grade 4	43.3	23.2	46.0	30.7	2.7	7.5
Grade 5	42.6	29.7	39.1	29.1	-3.5	-0.6