

Quasi-experimentation and Educational Research: A Case Study on the Experimentation of Different Statistics Education Strategies¹

*Quasi-sperimentazione e Ricerca in Educazione:
uno Studio su la Sperimentazione di Diverse Strategie Didattiche
per l'Insegnamento della Statistica*

Maria Gabriella Ottaviani
Dipartimento di Statistica, Probabilità e Statistiche Applicate,
P.le A. Moro 5, 00185 Roma,

Silvia Testa
silviatesta@tiscali.it

mariagabriella.ottaviani@uniroma1.it

Riassunto: L'impossibilità di assegnare casualmente gli studenti a strategie didattiche alternative per l'insegnamento della statistica, rende di fatto la sperimentazione effettuata una quasi-sperimentazione. Per effettuare confronti fra i risultati di strategie alternative, si propone allora di utilizzare una procedura di pair-matching che paragona i risultati degli studenti con la distanza di Mahalanobis. I risultati della sperimentazione tuttavia richiedono, per essere interpretati, informazioni sia quantitative sia qualitative.

Keywords: Quasi-experimentation, nonequivalent group designs, pair matching procedure.

1. Introduction

When designing experimental research in education, particular attention must be paid to the procedures involving the creation of control group. In fact, the way by which experimental and control groups are created implies that participants in any experimental research are placed in different treatment conditions and so they are not treated equally. By this ethical problems are originated. Moreover, the fact that the participants are volunteers creates a strong obstacle to the random assignment of participants to groups. Besides this, the experimental mortality - that is the lost of research participants during the course of the experiments - may cause the breakdown of any random assignment situation initially designed. As participants are not randomly assigned to the experimental and control group, experiments in education are not "real" experiments, but they are quasi-experiments

After summarising some quasi-experimentation problems and some possible ways to cope with them, a matching procedure will be applied to a set of data collected in the occasion of an educational research in order to compare different strategies to teach statistics at upper secondary schools.

¹ This paper is financially supported by the grant: MURST 2000, MM13233892_004.

2. Quasi-Experimentation and Statistical Analysis: an Overview

The most important consequence of not random assigning participants to experimental and control group is that "the comparisons from which treatment-caused change is inferred (...) depend on nonequivalent groups that differ from each other in many ways other than the presence of a treatment whose effects are being tested"(Cook, Campbell, 1979, p.6). So the problem is to separate the effects of the treatment from those due to the initial non-comparability between two groups. Distinctions among different types of quasi-experiments have been done. In particular there are nonequivalent group designs and interrupted time-series designs. The latter imply that many measures are taken at many time intervals, before and after treatment. The former are those in which "responses of a treatment group and a comparison group are measured before and after a treatment" (*idem*, p.6). Nonequivalence of control and experimental groups means that the "expected value of at least one characteristic of the groups are not equal even in the absence of a treatment effect." (*idem*, p.148). This implies that "differences resulting from the nonrandom assignment may produce posttest differences between the groups even in the absence of a treatment effect" (*idem*, p. 149). So what researchers need is to separate the effect of the treatment from the effect of the selection differences.

Among the different type of quasi-experimental design, in this paper a nonequivalent control group design is utilised. This is a design in which each group - the control or comparison group and the treatment group - takes a pretest and a posttest that is a test administered prior to or following an experimental treatment. ANCOVA and matching procedures are suitable techniques to analyse the data from a pretest-posttest control-group experiment. ANCOVA allows in fact to compare the posttest mean of the experimental group with the posttest mean of the control group using the pretest as a covariate. The matching procedures allow to select the research participants for the experimental and control group, so to reduce the bias due to non random assignment. This is done matching variables that are supposed to be linked to the treatment variable. Matching variables make explicit some of those "irrelevant causal forces hidden within the *ceteris paribus*" (*idem*, p.6) condition underlined by random assignment.

3. The Data and the Analysis

The National Research Group (NRG) carried out by the University of Padua, Palermo, Perugia and Rome "La Sapienza", and supported by a grant of the Italian Ministry of University and Scientific Research (ex40% 1998) experimented some new strategies to enhance the teaching of statistics at three different school levels: primary, secondary and upper secondary. The NRG followed the international line of thought suggesting that statistics is better taught and learned if real data are proposed in the classroom (Moore, 1997). Thus, in order to teach statistics in the upper secondary schools, three different strategies were proposed: an instructional strategy based on a data oriented approach (DOA); a strategy that accompanies data oriented approach with a computer laboratory as a tool (LAB); an approach where the pedagogical model of Co-operative Learning is added to real data and computer laboratory (CL). The researchers of the four Universities have proposed the mathematics teachers of upper secondary schools to volunteer in the experimental research. The teachers were offered in-service training according to the strategy they had chosen to apply in their classroom. They received

materials, tests and suggestions so that the activities in the classrooms experimenting the same strategy were as similar as possible. From March to May 2000, the participant teachers taught their students the following content units: Statistical information; Statistical survey; A questionnaire for collecting data; Data collection; Frequency distributions; Reading and explanation of tables; Graphic representations; Central measures, Variability measures.

To assess which strategy was more effective the teachers were requested: i) to record the mathematics mark of each of their students at the end of the first four month school period; ii) to submit their students a pretest in order to assess students' skills in reading graphs and tables before the experiment; iii) to submit their students an "intermediate" test to assess their understanding of statistical concepts after being taught the first seven units; iv) to submit their students a "final" test at the end of the experiment.

The analysis of the data at the disposal of the NRG put in evidence that: the experiment made in Rome was significantly different from the experiences of the other places; different strategies had in general different effects, as well as different types of upper secondary schools (Galmacci, Scrucca, 2000). Consequently this papers concentrates on Roman students in order to examine if significant differences exist among strategies. To enhance *ceteris paribus* conditions only a subset of data are analysed. They refer to students of the Technical Institute for Commerce, in class 3rd and 4th. Two data-sets are examined. The first one, referring to the intermediate test, has 276 observations and 5 variables (strategy; mathematics score; pretest score; pretest score/mathematics score; intermediate test score); the second, concerning the final test has 216 observations and 5 variables (strategy; mathematics score; intermediate test score; intermediate test score/mathematics score; final test score). As mathematics scores depends on teacher's way of scoring, the influence of this factor is eliminated considering the minimum and maximum scores utilised by each teacher and introducing the ratios: pretest score/mathematics score and intermediate score/mathematics score. Pretest, intermediate and final test are scored applying to each correct reply a coefficient decreasing according to the item facility. Table 1 shows: the mean scores according to the test and the strategy, and the number of students in each group. It also enable to see that experimental mortality exists and is different for each strategy (DOA -24%; LAB -19%; CL -17%). As the teachers are volunteers and each class must be treated as an "intact group" the one performed is a quasi-experiment.

Table 1: *Mean scores by test and strategy; number of students in each group*

Test	Strategy					
	DOA		LAB		CL	
Intermediate test	40.25	<u>176</u>	36.55	<u>42</u>	46.78	<u>58</u>
Final test	43.16	<u>134</u>	33.70	<u>34</u>	36.63	<u>48</u>

Two questions are suggested by data in Table 1: i) referring to intermediate test, is CL strategy significantly different from LAB? ii) referring to the final test, is LAB strategy significantly different from DOA?

To reply to these questions a matching procedure is utilised. Following Rubin's paper (Rubin, 1973) the pair matching method is chosen. By this, samples are pair matched individual by individual. As three matching variables exist, we propose to pair individuals so that they have comparable pretest scores, using Mahalanobis' distance. The treatment effect is estimated from the posttest difference between pairs. The

analysis reduces to the pair difference t test. The estimate of the treatment effect is the difference of the mean scores of the treatment group and the mean scores of the comparison group.

Referring to the intermediate test, the group of 58 students receiving CL is the treatment group, while the group of 42 students that received LAB is the comparison group. The effect of the treatment is measured on the intermediate test scores after performing the pair matching procedure using the variables: mathematics score; pretest score; pretest score/mathematics score. To remove the effect of co-operative groups one student has been considered for each of the 17 groups. Hence, the samples have 17 pair matched individuals. Table 2 shows that CL strategy is significantly different from LAB.

With reference to the final test, the group of 34 students receiving LAB is the treatment group, and the group of 134 that received DOA is the comparison group. The effect of the treatment is measured on the final test scores after performing the pair matching procedure using the variables: mathematics score; intermediate test score; intermediate test score/mathematics score. The pair matched individuals in each sample are 34. Table 2 shows that LAB strategy is significantly different from DOA.

All the other possible comparisons are not significant.

Table 2: *Strategies comparisons by pair matched samples*

Test	Mean scores		P value of the t test	Number of students
Intermediate test	CL	48.06	0.043	17
	LAB	38.71		
Final test	LAB	33.70	0.007	34
	DOA	44.21		

4. Conclusion

Pair-matching procedure allows to compare the performance of different strategies in this educational quasi-experimental design. Obviously the inferential analysis does not exempt from trying to understand the reason why CL strategy works significantly well in the intermediate test, and DOA in the final test. This may partly be explained looking at the experimentation registers filled by the teachers. They testify that in the final phase teachers had not at their disposals the number of lessons needed to develop strategies other than DOA. By this the needs of quantitative and qualitative information in educational research is made evident.

References

- Cook T. D., Campbell D. T. (1979) *Quasi-experimentation. Design & Analysis Issues for Field Settings*, Houghton Mifflin Company, Boston.
- Galmacci G., Scrucca L. (2000) Strategie didattiche per l'insegnamento della statistica: alcuni risultati di una sperimentazione, Rapporto tecnico, Stesura provvisoria.
- Moore D. S. (1997) New pedagogy and new Content: the Case of Statistics (with discussion), *International Statistical Review*, 65 (2).
- Rubin D.B. (1973) Matching to remove bias in observational studies, *Biometrics*, 29.