S-601
Prepared by George Bobinski, Jr.

Experimental and Quasi-Experimental Designs for Research
by Donald T. Campbell and Julian C. Stanley, pp.13-34

Abstract

The designs described in this section of the book are true experimental designs which allow for the manipulation of a study's independent variable and the measurement of any impact on the dependent variable. The key is that the experimenter has full control over the independent variable. These are the most strongly recommended designs. They generally have high internal validity and their weakness is external validity.

Summary

Three designs are proposed and explained. The first, (design 4), is shown below:

$R \; O_1 \quad X \quad O_2$

$R \; O_3 \qquad\quad O_4$

This design is the most common of the three. It has good internal validity. It controls for history (not intrasession, solution for this is to run simultaneously). Maturation and selection should be equal in each group. The problem of instrumentation is handled by using a fixed test. If observers are used, the same ones should be used for both experimental and control sessions and they shouldn't be told which is which. Regression should affect both groups equally. Selection effects are generally ruled out by randomization, matching is no substitute for a randomization.

External validity of this design is weak. First there can be interaction effects involving X and another variable. Ex: effect may be specific to groups warmed up by the pretest. This is not really solvable, Interaction of selection and X, and Maturation can both be problems. These are reduced by extending the boundaries in terms of settings and times. The artificiality of the experimental setting (reactive arrangements) are also a confounding factor. A final difficulty in this area is that often you can't randomize who is in a class. The solution is to use the classroom as a unit.

When performing statistical tests Campbell and Stanley say that both assignment to groups and sampling must be random. It is acceptable to do a t-test on gain score or use pretest scores as a covariate.

The second design is the Solomon Four-Group design diagrammed below:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \\ R & & X & O_5 \\ R & & & O_6 \end{array}$$

If $O_2 > O_1$, $O_2 > O_4$, $O_5 > O_6$, $O_5 > O_3$ inference is much stronger.(A 2x2 ANOVA can be used for analyzing this design). The main effect of testing and the interaction of testing and X are both determinable. However other problems remain and this method requires a high expenditure of effort that may not be worth it.

Design 6 is shown below:

$$\begin{array}{ccc} R & X & O_1 \\ R & & O_2 \end{array}$$

This design recognizes that a pretest is not essential. Sometimes it is recommended that this design be used as a pretest would be hard to administer or have an effect of its own. However where pretest scores are already available the first design should be used since the statistical tests for it are more powerful. The availability of pretest scores also makes the examination of the interaction of X and the pretest ability level possible. For statistical analysis the t-test is often used.

If there are two or more treatment variables, each at several levels this complicates the analysis. Interaction effects can be very important especially for generalizations. Sometimes the main effect is nonexistant but interaction is very important. Nesting means that certain interactions cannot be measured. Ex: teachers cannot be at all schools so teacher school interaction cannot be measured.

Some other extensions are possible with the designs discussed in this section. Testing for effect can be extended in time because the experimenter is often interested in long term results. However it is hard to control for history. To be able to generalize to other populations X has to be varied. This is juxtaposed to the demand for experimental control. To find out which aspects are most essential you must either make the control treatment closer to that of the experimental group or try to reduce the treatment of the experimental group to the essential elements. Finally since most measures are imperfect, multiple O's are desirable.

Possible Questions for Discussion:

1. How important is external validity? Is repeatability more important? (pg. 18).

2. If the subjects are in a school setting Campbell and Stanley recommend using the same setting and teachers(to reduce reactive arrangements effect). What problems does this create? Is their recommendation a valid one?

Campbell and Stanley

"Experimental and Quasi-Experimental Designs for Research" (p. 1-12)

Abstract-
The authors are concerned with the threats to validity on various experimental designs. They view experimentation as basic in the evolution of knowledge. Threats to internal and external validity are defined. Three pre-experimental designs are presented and their handling of the various threats to internal validity are examined.

Through the examination of 16 research designs in the light of 12 threats to validity, the authors intend to provide guidance in the selection of designs that may provide "adequate and proper data to which to apply statistical procedure." (p. 1) The first section of their paper introduces the problem and examines three "pre-experimental" designs.

A very high value is placed on experimentation as "...the only means for settling disputed..." (p. 2) The authors give it a very basic position in their perspective on knowledge. "(I)t is a refining process superimposed upon the probably valuable cumulations of wise practice." (p. 4) In other words, experimentation is a means of testing the meaning of experience. Advancement is characterized by "continuous, multiple experimmtation". They do note that we should not expect that the results of testing rival theories to be unambiguous. Given the authors' perspectives, it would seem likely that when opposing views are presented by those competent to state them, that they would all be partially right.

A list of eight threats to internal validity is presented by the authors, along with 4 threats to external validity. Seven of the threats to internal validity are given a detailed discussion in this section of the paper.

-History- the effect of events external to the experiment on subjects between measurements. History becomes a more plausible confounding factor, the longer the time between measurements.

-Maturation- The nature of the subjects changes over time. Obviously, this can occur in most any natural system.

-Test effects- the effects of one measurement on succeeding measurements. The familiar example of improving GMAT scores on the second try would fit here. A subset of concern here is that of reactivity. Since measurement may change what is being measured, one wants non-reactive measures; measures that will not cause a change in the subject.

-Instrumentation- changes in the measuring instrument between measurments. This could include the scoring of measures, as in using different judges to categorize open-ended responses on questionnaires.

-Statistical Regression (aka regression towards the mean)- the movement of post-test measurements of extreme groups towards the pre-test mean of all groups. (Regression towards the mean is a very important concept in evolutionary biology)

-Selection- the differences between groups may be due to recruitment ra rather than treatment.

-Mortality- differences between groups due to differential drop-out. We discussed an example of this under subjects reactions to experimental treatments.

Three research designs are presented and the influences of these various threats on each are discussed. (see Table 1, p. 8)

-The "One-Shot" Case Study

$X \quad O$

The authors denigrate this as the least worthwhile design. The lack of control and lack of comparison except with "implicit common knowledge" lead to the "error of misplaced precision." This design also suffers from all of the threats to validity except those that involve multiple measurements.

-The One-Group Pretest-Posttest Design

$O_1 \quad X \quad O_2$

While this design is deemed better than doing nothing, the authors use it as an example of many of the threats to internal validity. Of all of the threats, only selection and mortality are controlled. As an interesting note, it is the feature of "experimental isolation" that allows the physical sciences to often gain control over the threat of history when using this design.

-The Static-Group Comparison

$$\begin{array}{ll} X & O_1 \\ \hline & O_2 \end{array}$$

Here, one group experiences the treatment, and the other does not. The dotted line shows that no means are used to test whether the differences between groups ~~is~~ are due to treatment. This design controls for all threats, except for selection and mortality.

Questions for discussion

—Aren't the authors somewhat harsh on case studies?

—Aren't there some cases where these designs are the only alternatives ?

—Is the disillusionment that the authors refer to a result of over-optimism or is it due to the difficulties inherent in experimental designs?

Analysis of Variance with Repeated Measures, CH 6, Huck, Cormier, and Bounde

Abstract

The chapter presents two of the most frequently used types of ANOVA with repeated measures: Lindquist Type I and Lindquist Type III. Specifically, attention is given to the interpretation of reported results on the possible research questions, follow-up analyses.

Summary

ANOVA with repeated measures is appropriate for studies where the same subjects are measured more than twice. There are two broad categories for the studies using repeated measures: (1) studies in which there are several dependent variables, with the data from each of these variables being subjected to a separate analysis and (2) studies in which there is only one dependent variable, with subjects measured across all levels of one (or more) of the independent variables. This chapter focuses on the two types of ANOVA belonging to the latter category, namely, Lindquist Type I and Lindquist Type III.

1. Lindquist Type I ANOVA's

This type always involves two factors and repeated measures across the levels of one of the two factors.

Research questions: The three research questions the Lindquist Type I ANOVA allows the researcher to answer are: (1) Is there a significant main effect of the factor that does not have repeated measures across its levels (the between-subjects factor) (2) Is there a significant main effect of the repeated measures (within-subject) factor and (3) Is there a significant interaction between the two factors.

Reporting the results: The summary table contains two major sources of variation, between-subjects and within-subjects. The between-subject part of the summary table shows the main effect of the factor that does not have repeated measures and error-between. The within-subjects part consists of the main effect of the repeated measures factor, the interaction between the two factors, and the error within.

Follow-up analysis: If one (or both) of the main effects is significant and if there are three or more levels associated with that factor, a multiple comparison test is desirable to find out where the significant differences lie among the main effect means. If the interaction is significant tests of simple main effects are applicable to find out the cause of the observed interaction. In conducting tests of simple main effects, some authors compare the various groups in the study at each separate level of the repeated measures variable, while others apply the test for each group, comparing that group's performance across the repeated measures variable. Also, it could be done in both directions.

2. Lindquist Type III ANOVA's

This type involves three factors and repeated measures across the levels of one of the factors.

Research questions: Research questions of Type III deal with main effects of the three factors including the one with repeated measures, three first order interactions, and a second order interaction of the three factors.

Reporting the results: The various sources of interaction are divided into two sections as in the case of the Type I, between-subjects and within-subjects. In the between-subjects part, two main effects and one first order interaction are contained whereas in the latter section, one main effect (of the factor with repeated measures), two first order interactions, and one second-order interaction.

Follow-up analyses: If the results indicate a significant main effect and if there are three or more levels associated with that particular factor a multiple comparison test is appropriate to find out where the significant differences lie among the main effect means. If a first order is significant, researcher will often disregard the results related to the main effects of the two factors that are involved in the interaction. Instead they can average the data across the variable that is not involved in the interaction then apply tests of interaction.

Chankon Kim

Quasi-Experimental Designs, pp. 34-64, Campbell and Stanley

Quasi-experimental designs are applicable to settings where the experimenter has a perfect control over scheduling of data collection procedures (e.g., the when and the whom of measurement) but lacks the full control over the scheduling of experimental stimuli (the when and to whom of exposure and the ability to randomize exposures) which makes a true experiment possible. This lack of control in quasi-experiment dictates that one must in interpreting the results, consider in detail the likelihood of uncontrolled factors accounting for the results. The more implausible this becomes, the more valid the experiment. The designs in general are believed to be sufficiently probing, however, to be well worth employing where more efficient probes are unavailable.

7. Time-Series Experiment

$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$

The essence of the time-series design is the presence of a periodic measurement process on some group or individual and the introduction of an experimental change into this time series of measurements the results of which are indicated by a discontinuity in the measurements recorded in the time series. The test of significance in the effect of X concerns intercept and slope of the time series measurements before and after the treatment - test of linear fit to the data.

8. Equivalent Time-Samples Design

$X_1O \quad X_0O \quad X_1O \quad X_0O$

The design can be seen as a form of the time series experiment with the repeated introduction of the experimental variable, and is most useful where the effect of the experimental variable is anticipated to be transient or reversible character.

Tests of significance deals with two dimensions of generalization: across occasions and across persons. If only one or two repetitions of each experimental condition are involved, sampling errors of occasions may be large or the control of history will be poor. Generalization across persons depends on the size and the representativeness of the sample. T-test or ANOVA may be used to test significance.

9 Equivalent Materials Design

M_aX_1O M_bX_0O M_cX_1O M_dX_0O

This design is required whenever the nature of the experimental variables is such that the effects are enduring and the different treatments must be applied to nonidentical content. The sampling equivalence of the two sets of materials is essential -- M_aM_c in sampling terms, equal to the sample M_bM_d).

Statistical tests deal with establishing the generalization across the sample of lists or items and then computing an experimental effects score for a particular person (group) and employing this as a basis for generalizing across persons.

10. Non-equivalent Control Group Design

0   X   0
0   0

The design involves an experimental group and a control group both given a pretest and a posttest, but there is no preexperimental sampling equivalence between the two groups. The more similar the experimental and the contro groups are in their recuitment, and the more this similarity is confirmed by the scores on the pretest, the more effective the control over the extraneous variables becomes. As a useful adjunct to randomization, matching of subjects in terms of pretest scores is effective. Due to the non-equivalent groups, application of ANCOVA is less plausible.

11. Counterbalanced Designs

	Time 1	Time 2	Time 3	Time 4
Group A	X_1O	X_2O	X_3O	X_4O
B	X_2O	X_4O	X_1O	X_3O
C	X_3O	X_1O	X_4O	X_2O
D	X_4O	X_3O	X_2O	X_1O

These designs involve the case where experimental control is achieved or precision enhanced by entering all respondents (or settings) into all treatments. The Latin-square arrangement is typically employed in which treatments are applied in a restrictively randomized manner in turn to the naturally assembled groups (or individuals).

12. Separate-Sample Pretest-Posttest Design

R   0 (X)
R   X   0

(X) is a presentation of X irrelevant to the argument

The design is appropriate where one cannot randomly segregate subgroups for differential treatments. Designs 12, 12a, 12b, and 12c (see Table 2) are superior to the "true" experiment because they put so little demand upon the respondents for cooperation.

13. Separate-Sample Pretest-Posttest Control Group Design

```
R   O (X)
R       X  O
R   O
R          O
```

In addition to the group used in Design 12, if there are comparable groups from which X can be withheld then this design is appropriate.

14. Multiple Time-Series Design

```
O  O  O  O  X  O  O  O  O
O  O  O  O     O  O  O  O
```

This design is appropriate if there is a similar group (or institution) not undergoing X, from which to collect a similar "control" time series.

15. Recurrent Institutional Cycle Design: A "Patched-Up" Design
(For diagram see Table 3)

This design illustrates a strategy for field research in which one starts out with an inadequate design and then adds a specific features to control for one or another of the recurrent sources of invaidity. The design is appropriate to those situations in which a given aspect of an institutional process is continually being presented to a new group of respondents. This design can combine the longitudinal and cross-sectional approaches in order to more effectively prevent validity problems.

16. Regression-Discontinuity Analysis

When people or groups are given awards or those in special need are given extra help, one would like to discover the consequences of such provisions. This design is appropriate for these situations. A treatment effect is detected if there is a discontinuity between the regression lines derived from the scores of pretest and posttest of the two groups -- one with X and one without X. The ANCOVA is the most efficient test of significance.

Questions

1. Do experimental researches in social science fully evaluate and regard importantly the validity treats in their designs ?
2. How frequently are the quasi-experiments used in business research ? Why?

TABLE 3

Sources of Invalidity for Quasi-Experimental Designs 13 through 16

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and *X*	Interaction of Selection and *X*	Reactive Arrangements	Multiple-*X* Interference
uasi-Experimental Designs Continued:												
13. Separate-Sample Pretest-Posttest Control Group Design R O (X) R X O R O R O	+	+	+	+	+	+	+	−	+	+	+	
13*a*. R O (X) R X O R' R O (X) R X O R O (X) R X O R O R O R' R O R O R O R O	+	+	+	+	+	+	+	+	+	+	+	
14. Multiple Time-Series O O OXO O O O O O O O O	+	+	+	+	+	+	+	+	−	−	?	
15. Institutional Cycle Design ass A X O_1 ass B_1 RO_2 X O_3 ass B_2 R X O_4 ass C O_5 X en. Pop. Con. Cl. B O_6 en. Pop. Con. Cl. C O_7												
$O_2 < O_1$ $O_5 < O_4$	+	−	+	+	?	−	?		+	?	+	
$O_2 < O_3$	−	−	−	?	?	+	+		−	?	+	
$O_2 < O_4$	−	−	+	?	?	+	?		+	?	?	
$O_6 = O_7$ $O_{2y} = O_{2o}$		+						−				
16. Regression Discontinuity	+	+	+	?	+	+	?	+	+	−	+	+

eneral Population Controls for Class B, etc.

TABLE 2

Sources of Invalidity for Quasi-Experimental Designs 7 through 12

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and *X*	Interaction of Selection and *X*	Reactive Arrangements	Multiple-*X* Interference
Quasi-Experimental Designs:												
7. Time Series O O O OXO O O O	−	+	+	?	+	+	+	+	−	?	?	
8. Equivalent Time Samples Design X_1O X_0O X_1O X_0O, etc.	+	+	+	+	+	+	+	+	−	?	−	−
9. Equivalent Materials Samples Design M_aX_1O M_bX_0O M_cX_1O M_dX_0O, etc.	+	+	+	+	+	+	+	+	−	?	?	−
10. Nonequivalent Control Group Design O X O O O	+	+	+	+	?	+	+	−	−	?	?	
11. Counterbalanced Designs X_1O X_2O X_3O X_4O X_2O X_4O X_1O X_3O X_3O X_1O X_4O X_2O X_4O X_3O X_2O X_1O	+	+	+	+	+	+	+	?	?	?	?	−
12. Separate-Sample Pretest-Posttest Design R O (X) R X O	−	−	+	?	+	+	−	−	+	+	+	
12*a*. R O (X) R X O R O (X) R X O	+	−	+	?	+	+	−	+	+	+	+	
12*b*. R O_1 (X) R O_2 (X) R X O_3	−	+	+	?	+	+	−	?	+	+	+	
12*c*. R O_1 X O_2	−	−	+	?	+	+	+	−				