Lecture 6: Experimental Design

References

- Donald T. Campbell, Julian Stanley, *Experimental and Quasi Experimental Designs for Research*, Houghton Mifflin Company, **1963**
- John W. Creswell, *Educational research: planning, conducting, and evaluating quantitative and qualitative research, 4th ed.*, Pearson Education, **2012**
- John W. Creswell, J. David Creswell, *Research design: qualitative, quantitative, and mixed methods approaches*, SAGE Publications, **2018**
- Christopher J. Millera, Shawna N. Smith and Marianne Pugatch, *Experimental and quasi-experimental designs in implementation research*, Psychiatry Research, vol. 283, **2020**

Content

- Steps of experimental research
- Types of experimental design
- 3 pre-experimental designs
- 3 true experimental designs
- Factorial designs

Steps in conducting experimental research

1. Define the overall GQM and the metrics
2. Decide if an Experiment addresses your research problem
3. Formulate Hypotheses H0 and H1
4. Select a Population and a Sample
5. Select an Experimental Treatment and introduce it
6. Choose a type of Experimental Design
7. Conduct the Experiment
8. Organize and analyze the data

Experimental Designs

The experimental design is the general plan for setting up and testing a specific hypothesis or research question. Overall, the design directs the who, what, when, and how of the research project.

Selection of a research design is a compromise between the goal of rigorous scientific integrity versus limited resources and reality.

"Gold standard" trials (prospective, randomized, blinded, controlled) are the most expensive type of study to perform and often are not the most appropriate to answer certain types of research questions.

Sometimes they are not even an option for ethical or legal reasons.

Experimental Design Structure – Schema

- Three Research Variables
  - Independent variable (the intervention)
  - Dependent variable (the outcome)
  - Extraneous variables (other, potentially confounding factors)
- Three Design Elements
  - Manipulation (the ability to influence or direct the independent variable)
  - Control (the ability to direct or influence important extraneous variables and study measurements)
  - Randomization (unbiased [random] subject assignment to each group)

Taken verbatim from: Cheryl Thompson, Edward A Panacek "Research study designs: Experimental and quasi-experimental," November 2006, AirMed 25(6):242-6

Prospective and retrospective designs

Designs can be:

- **prospective**: the events of interest have not yet occurred when the study begins) and
- **retrospective**: the events of interest all occurred before the onset of the study.

Taxonomy of Experimental Designs (1/2)

The most common categorization of experimental designs, which we will use through our lectures is the following:

- Pre-experimental designs (a.k.a., non-experimental)
- True experimental designs
- Quasi-experimental designs

Taxonomy of Experimental Designs (2/2)

- Non-experimental (or pre-experimental)
  - Have one or none of the core design elements
  - Lack manipulation and randomization
  - May also lack control
  - Are generally retrospective
  - Have the lowest scientific validity
- (True) experimental
  - Have all three design elements
  - Are always prospective
  - Have high scientific validity
- Quasi-experimental
  - Only have one or two design elements
  - Have manipulation or control
  - Generally lack randomization
  - Are generally prospective in nature
  - Are moderate in scientific validity

Taken almost verbatim from: Cheryl Thompson, Edward A Panacek "Research study designs: Experimental and quasi-experimental," November 2006, AirMed 25(6):242-6

Proposed exercise – Part 1 (1/2)

- Return to the mix team of 3 people of which at least one DS and one SE
- On the already considered experimentation that happened in your life from anyone in the team determine:
  - Whether it was prospective or retrospective
  - The independent variable(s)
  - The dependent variable(s)
  - The independent variable(s)
  - The manipulation
  - The control
  - The randomization
  - Whether it was pre-, quasi-, or true experimental
- Write the results on the relevant columns of `http://tiny.cc/IU_EM_F20_L5`, tab: `TypesOfExperiments`

Proposed exercise – Part 1 (2/2)

`http://tiny.cc/IU_EM_F20_L5`, tab: `TypesOfExperiments`

Pre-experimental designs

Pre-experimental design

The characteristics of this group:

- Lack manipulation and randomization
- May also lack control
- Are generally retrospective
- Have the lowest scientific validity

The One-Shot Case Study

This design involves an exposure of a group to a treatment followed by a measure.

$$Group\ A \quad X\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_O$$

This design have such a total absence of control as to be of almost no scientific value.

The inferences are based upon general expectations of what the data would have been had the X not occurred

One-Shot Case Study (contd.)

One-Shot Case Study design, if taken in conjunction with the implicit "common-knowledge" comparisons, comprises all of the weaknesses and threads to validity.

Example: We had introduced daily inspections in a team, and after a while collected the statistics on defects rate. Then we compare results against industrial standards.

One-Group Pretest-Posttest

This design includes a pretest measure followed by a treatment and a posttest for a single group.

$$Group\ A \quad O_1 \_\_\_\_ X \_\_\_\_ O_2$$

Example: We measured defects rate before and after introducing of daily inspections in a team, and compare results.

One-Group Pretest-Posttest (contd.)

One-Group Pretest-Posttest Design has a number of potential threads to internal validity:

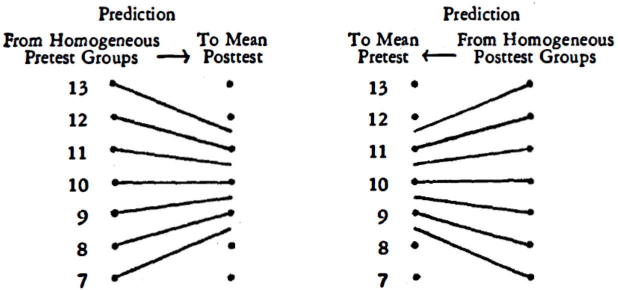
- History. Between O_1 and O_2 many other change-producing events may have occurred in addition to the experimenter's X. History becomes a more plausible rival explanation of change the longer the $O_1 - O_2$ time lapse.
  Experimental isolation, if can be achieved, makes this design acceptable for much of scientific researches.
- Maturation. Could be biological or psychological processes which systematically vary with the passage of time, independent of specific external events.

One-Group Pretest-Posttest (contd.)

- Testing. The effect of the pretest itself on the experiment. The reactive effect can be expected whenever the testing process is in itself a stimulus to change rather than a passive record of behavior (*on therapy for weight control, the initial weigh-in might in itself be a stimulus to weight reduction*).
- Instrumentation. Autonomous changes in the measuring instrument which might account for an $O_1 - O_2$ difference. Where human observers are used to provide O_1 and O_2, processes within the observers will contribute to $O_1 - O_2$ differences.

One-Group Pretest-Posttest (contd.)

- Statistical regression. It is about the tautological aspect of the imperfect correlation between O_1 and O_2.
- The use of time reversed control analyses and the direct examination for changes in population variabilities are useful precautions against such misinterpretation.



from Campbell, Stanley pg 11

Static-Group Comparison

Experimenters use this design after implementing a treatment. After the treatment, the researcher selects a comparison group and provides a posttest to both the experimental group and the comparison group.

Group A	X	____________________	O_1
Group B		____________________	O_2

There is no formal means of certifying that the groups would have been equivalent.

Static-Group Comparison (contd.)

It provides next potential threat to validity - *selection*. The groups might have differed anyway, without the occurrence of X

In our experiment, results (defects rate) of the team 2 could be better even without introducing the treatment (daily inspections) because all team members are senior developers, or they use Haskell and not Java, etc.

Another potential threat here is experimental *mortality*, or the production of $O_1 - O_2$ differences in groups due to the differential drop-out of elements from the groups (or partial data loss).

Obviously, the team composition may changes before the observation

Pre-experimental designs summary

	Sources of invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction with selection	Interaction of testing and X	Interaction of selection and X	Reactive arrangements	Multiple interference
One-Shot Case Study	–	–				–	–			–		
One-Group Pretest-Posttest Design	–	–	–	–	?	+	+	–	–	–	?	
Static-Group Comparison	+	?	+	+	+	–	–	–		–		

"–" indicates a definite weakness,

"+" indicates that the factor is controlled,

"?" indicates a possible source of concern,

a blank space indicates that the factor is not relevant.

True experimental design

True experimental designs

The characteristics of true experimental design:

- Have all three design elements
- Are always prospective
- Have high scientific validity

Pretest-Posttest Control Group Design

The most popular and strongly recommended design. The experimental Group A and the control Group B are selected with random assignment. Both groups take a pretest and posttest. Only the experimental group receives the treatment.

Group A	R	——— O_1 ——— X ——— O_2
Group B	R	——— O_3 ————————— O_4

Equivalence of the groups is achieved by randomization.

Pretest-Posttest Control Group (contd.)

- *History* is controlled as general historical events that might have produced an $O_1 - O_2$ difference would also produce an $O_3 - O_4$ difference.
- *Maturation* and *testing* are controlled in that they should be manifested equally in experimental and control groups. (*Remember, groups are equalized!*)
- *Instrumentation* is controlled where the conditions for the control of intrasession history are met (*scripts for a survey; checklists*). Where observers or interviewers are used and there are few of them to be randomly assigned for observations, then each observer involved both in experimental and control sessions, and the observers should be unaware of who receiving the treatments.

Pretest-Posttest Control Group (contd.)

- *Regression* is controlled as far as mean differences are concerned: both experimental and control groups are randomly assigned from the same extreme pool. In such a case, the control group regresses as much as does the experimental group.
- *Selection* is ruled out o the extent that randomization has assured group equality at time R.
  Obviously, the assurance of equality is greater for large numbers of random assignments than for small.
- *Mortality* The data made available by this design make it possible to tell whether mortality offers a plausible explanation of the $O_1 - O_2$ gain by different ways of comparing control and experimental subgroups who got pretest, posttest or both of them.

It looks like Pretest-Posttest Control Group Design keeps under control
all threats to validity,
so there is no reason for more complex approaches,
isn't it?

Solomon Four-Group

This design deservedly has higher prestige and represents the first explicit consideration of *external validity* factors.

A special case of a 2 × 2 factorial design, this procedure involves the random assignment of participants to four groups. Pretests and treatments are varied for the four groups. All groups receive a posttest.

Group A	R	O_1	X	O_2
Group B	R	O_3		O_4
Group C	R		X	O_5
Group D	R			O_6

Solomon Four-Group (contd.)

Both the main effects of *testing* and the *interaction of testing and X* are determinable.

The effect of X is replicated in four different fashions:
$O_2 > O_1$, $O_2 > O_4$, $O_5 > O_6$, and $O_5 > O_3$

If these comparisons are in agreement, the strength of the inference is greatly increased.
Also, the comparison of O_6 with O_1 and O_3 shows the combined effect of *maturation* and *history*.

Posttest-Only Control Group

This design controls for any confounding effects of a pretest and is a popular experimental design. The participants are randomly assigned to groups, a treatment is given only to the experimental group, and both groups are measured on the posttest.

$$\begin{array}{lll} \text{Group A} & \text{R} & \_\_\_\_\_\text{X}\_\_\_\_\_\_O_1 \\ \text{Group B} & \text{R} & \_\_\_\_\_\_\_\_\_\_\_\_O_2 \end{array}$$

In this design researchers rely on the initial randomization (equalization) of the groups.

Posttest-Only Control Group (contd.)

The design can be considered as the two last groups of the Solomon Four-Group Design, and it is reasonable when extra data from Design 5 may not be worth the more than double effort.

It controls for testing as main effect and interaction as well as the previous design, but without measuring them.

Where anonymity of the participants must be kept, this design is the most convenient option.

True experimental designs summary

	Sources of invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction with selection	Interaction of testing and X	Interaction of selection and X	Reactive arrangements	Multiple interference
Pretest-Posttest Control Group Design	+	+	+	+	+	+	+	+	–	?	?	
Solomon Four-Group Design	+	+	+	+	+	+	+	+	+	?	?	
Posttest-Only Control Group Design	+	+	+	+	+	+	+	+	+	?	?	

"–" indicates a definite weakness,
"+" indicates that the factor is controlled,
"?" indicates a possible source of concern,
a blank space indicates that the factor is not relevant.

Factorial Designs

In some experimental situations, it is not enough to know the effect of a single treatment on an outcome; several treatments may, in fact, provide a better explanation for the outcome.

For example, we want to see not only how the frequency influence the quality of code inspection, but its combined effect with the number of LoC under per reviewer.

Here, "LoC under review" is a blocking or moderating variable and the researcher makes random assignment of each "block" level (0-200 LoC, 200 - 400 LoC, and 400 - 600 LoC) to each experimental group.

Factorial Designs (contd.)

Steps of factorial design:

1. Identify a research question that includes several independent variables (typically at most three due to increasing complexity).
2. Identify the levels of each factor or independent variable (*daily and weekly inspections for the first factor; 3 groups of LoC amount for the second factor*).
3. Assign participants to groups so that all groups receive each level on one independent variable (*2*3 groups in our case*).
4. At the conclusion of the experiment, the posttest will measure defect rates for each group to determine whether they are statistically different.

The null hypothesis would be they are not different, whereas the alternative would be that they are different.

Factorial Designs (contd.)

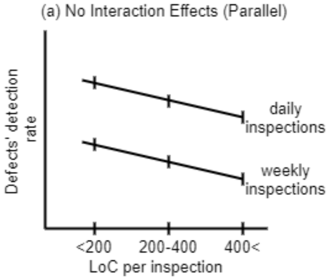
Analysis of variance will produce statistical results for main effects and interaction effects.

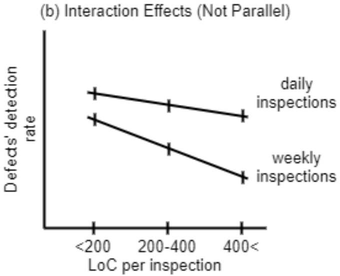
- **Main effects** are the influence of each independent variable on the outcome in an experiment.
- **Interaction effects** exist when the influence on one independent variable depends on (or co-varies with) the other independent variable in an experiment.

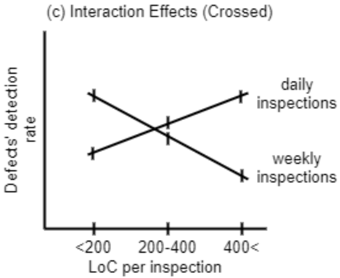
The results of the experiment could be different.

Typically, in factorial designs, the investigator graphs these trends and explains the meaning of the combination of independent variables.

Factorial Designs (contd.)







Proposed exercise – Part 2 (1/2)

- Return to the mix team of 3 people of which at least one DS and one SE
- On the already considered experimentation that happened in your life from anyone in the team determine now also:
  - The overall structure of the experimentation
  - The inherent threats to internal validity
  - The inherent threats to construct validity
  - The inherent threats to external validity
  - How you plan to solve such threats
- Write the results on the relevant columns of `http://tiny.cc/IU_EM_F20_L5`, tab: `TypesOfExperiments`

Proposed exercise – Part 2 (2/2)

`http://tiny.cc/IU_EM_F20_L5`, tab: `TypesOfExperiments`

End of lecture 6