## Chapter 9\_part II: Experiments

## Selecting and Assigning Participants

* The participants selected for the experiment should be representative of the population to which the researcher wishes to generalize the study’s results.
* This may seem self-evident, but we have witnessed several decades of experimentation with college sophomores that contradict that assumption.
* The procedure for random sampling of experimental subjects is similar to the selection of respondents for a survey.
* The researcher prepares a sampling frame and then assigns the subjects for the experiment to groups, using a randomization technique.
* Systematic sampling may be used if the sampling frame is free from any form of periodicity that parallels the sampling ratio.
* Because the sampling frame is often small, experimental subjects are recruited (a self-selecting sample).
* If randomization is used, those assigned to the experimental group are likely to be similar to those assigned to the control group.
* **Random assignment** to the groups is required to make the groups as comparable as possible with the dependent variable.
* Randomization does not guarantee that if a pretest of the groups was conducted, the groups would be pronounced identical.
* However, it is an assurance that those differences remaining are randomly distributed.
* When it is not possible to randomly assign subjects to groups, **matching** may be used.
* Matching employs a nonprobability quota sampling approach.
* The object of matching is to have each experimental and control subject matched on every characteristic used in the research.
* This becomes cumbersome as the number of variables and groups in the study increases.
* Because the characteristics of concern are only those that are correlated with the treatment condition or the dependent variable, they are easier to *identify*, *control*, and *match*.
* In the sales presentation experiment, if a large part of the sample was composed of businesswomen who had recently completed communications training, we would not want the characteristics of gender, business experience, and communication training to be disproportionately assigned to one group.
* Some authorities suggest a **quota matrix** as the most efficient means of visualizing the matching process.
* If matching does not alleviate the assignment problem, a combination of *matching*, *randomization*, and *increasing the sample size* would be used.

## Pilot Testing, Revising, and Testing

* Pilot testing is intended to reveal errors in the design and improper control of extraneous or environmental conditions.
* Pretesting the instruments permits refinement before the final test.
* This is the researcher’s best opportunity to revise scripts, look for control problems with laboratory conditions, and scan the environment for factors that might confound the results.
* In field experiments, researchers are sometimes caught off guard by events that have a dramatic effect on subjects:
* The test marketing of a competitor’s product
* A reduction in force, reorganization, or merger
* The experiment should be timed to that subjects are not sensitized to the independent variable by factors in the environment.

## Analyzing the Data

* Data from experiments are not easy to analyze, they are simply more conveniently arranged because of the:
* Levels of the treatment condition
* Pretests and post-tests
* Group structure
* If adequate planning and pretesting have occurred, the choice of statistical techniques is simplified.
* Researchers have several measurement and instrument options with experiments. Among them are:
* Observational techniques and coding schemes.
* Paper-and-pencil tests.
* Self-report instruments with open-ended or closed questions.
* Scaling techniques (Likert scales, semantic differentials, Q-sort).
* Physiological measures (galvanic skin response, EKG, voice pitch analysis, eye dilation).

# Validity in Experimentation

* Even when an experiment is the ideal research design, there is always a question about whether the results are true.
* Validity means that a measure accomplishes its claims.
* There are several types of validity, here only the two major varieties are considered:
* **Internal validity**—Do the conclusions we draw about a demonstrated experimental relationship truly imply cause?
* **External validity**—Does an observed causal relationship generalize across persons, set-tings, and times?
* Each type of validity has specific threats we must guard against.

## Internal Validity

* Among the many threats to internal validity, we consider seven:
* History
* Maturation
* Testing
* Instrumentation
* Selection
* Statistical regression
* Experimental mortality

## History

* During the time that an experiment is taking place, some events may occur that confuse the relationship being studied.
* In many experimental designs, we take a control measurement (*O***1**) of the dependent variable before introducing the manipulation (*X*).
* After the manipulation, we take an after-measurement (*O***2**) of the dependent variable.
* The difference between *O***1** and *O***2** is the change that the manipulation caused.
* Management may wish to find the best way to educate its workers about the financial condition of the company before labor negotiations.
* To assess the value of such an effort, managers test employees on their knowledge of the company’s finances (*O***1**).
* Then, they present the educational campaign (*X*) to these employees.
* Afterward, they again measure their knowledge level (*O***2**).
* This design, known as a pre-experiment because it is not a very strong design, can be diagrammed as follows:

***O*1 *X O*2**

Pretest Manipulation Posttest

* Between *O*1 and *O*2, many events could occur to confound the effects of the education effort, such as the topic appearing in a newspaper article or being discussed at a union meeting.

### Maturation

* Changes also may occur within the subject that are a function of the passage of time and are not specific to any particular event.
* These are of special concern when the study covers a long time, but they may also be factors in tests that are as short as an hour or two.
* A subject can become hungry, bored, or tired in a short time, and this condition can affect response results.

### Testing

* The process of taking a test can have a learning effect that influences the results of a second test.

### Instrumentation

* This threat to internal validity results from changes between observations in either the measuring instrument or the observer.
* Using different questions at each measurement is an obvious source of potential trouble, but using different observers or interviewers also threatens validity.
* There can even be an instrumentation problem if the same observer is used for all measurements.
* Observer experience, boredom, fatigue, and anticipation of results can all distort the results of separate observations.

### Selection

* An important threat to internal validity is the differential selection of subjects for experimental and control groups.
* Validity considerations require that the groups be equivalent in every respect.
* If subjects are randomly assigned to experimental and control groups, this selection problem can be largely overcome.
* Matching the members of the groups on key factors can also enhance the equivalence of the groups.

### Statistical Regression

* This factor operates when groups have been selected by their extreme scores.
* Example: Measure the output of all workers in a department for a few days before an experiment and then conduct the experiment with only those workers whose productivity scores are in the top 25 percent and bottom 25 percent.
* No matter what is done between *O*1 and *O*2, there is a strong tendency for the average of the high scores at *O*1 to decline at *O*2 and for the low scores at *O*1 to increase.
* This tendency results from imperfect measurement that, in effect, records some persons abnormally high and abnormally low at *O*1.
* In the second measurement, members of both groups score more closely to their long-run mean scores.

### Experiment Mortality

* This occurs when the composition of the study groups changes during the test.
* Attrition is especially likely in the experimental group, and with each dropout the group changes.
* Because members of the control group are not affected by the testing situation, they are less likely to withdraw.
* In a compensation incentive study, some employees might not like the change in compensation method and may withdraw from the test group.
* This could distort the comparison with the control group that has continued working under the established system, perhaps without knowing a test is under way.
* All the threats mentioned to this point are generally dealt with adequately in experiments by random assignment. However, five additional threats to internal validity are independent of whether or not one randomizes.
* The first three have the effect of equalizing experimental and control groups.
* *Diffusion or imitation of treatment*
* *Compensatory equalization*
* *Compensatory rivalry*
* *Resentful demoralization of the disadvantaged*
* *Local history*

## External Validity

* Internal validity factors cause confusion about whether the experimental treatment (X) or extraneous factors are the source of observation differences.
* External validity is concerned with the interaction of the experimental treatment with other factors and the resulting impact on the ability to generalize to (and across) times, settings, or persons.
* Among the major threats to external validity:
* Reactivity of testing on *X*
* Interaction of selection and *X*
* Other reactive factors

### The Reactivity of Testing on *X*

* The reactive effect refers to sensitizing subjects via a pretest so they respond to the experimental stimulus (*X*) in a different way.
* This before-measurement effect can be particularly significant in experiments where the IV is a change in attitude.

### Interaction of Selection and X

* The process by which test subjects are selected for an experiment may be a threat to external validity.
* The population from which one selects subjects may not be the same as the population to which one wishes to generalize results.
* Example: You use a selected group of workers in one department for a test of the piecework incentive system. Can you extrapolate those results to all production workers?
* Example: You ask a cross section of a population to participate in an experiment, but a substantial number refuse. If you conduct the experiment only with those who agree to participate (self-selection), can the results be generalized to the total population?

### Other Reactive Factors

* Experimental settings themselves may have a biasing effect on a subject’s response
to *X.*
* An artificial setting can produce results not representative of larger populations.
* Suppose workers who are given incentive pay are moved to a different work area to separate them from the control group. These new conditions alone could create a strong reactive condition.
* If subjects know they are participating in an experiment, there may be a tendency to role-play in a way that distorts the effects of *X.*
* Another reactive effect is the possible interaction between *X* and subject characteristics.
* An incentive pay proposal may be more effective with persons in one type of job, with a certain skill level, or with a certain personality trait.
* Problems of *internal validity* can be solved by the careful design of experiments, but this is less true for problems of external validity.
* *External validity* is largely a matter of generalization, which is an inductive process of extrapolating beyond the data collected.
* Generalizing means estimating the factors that can be ignored and that will interact with the experimental variable.
* The closer two events are in time, space, and measurement, the more likely they are to follow the same laws.
* Seek internal validity first.
* Try to secure as much external validity as is feasible by making experimental conditions as similar to the conditions under which the results will apply as possible.