Chapter 3: Designing Studies and Obtaining Data

The PICTURE

In Chapters 1 and 2, we learned some basic methods for analyzing the pattern of variation in a set of data. To do scientific work, more is needed. Specifically, **designed studies** should be conducted to confirm the patterns uncovered by data analysis. Designing studies is the topic of Chapter 3.

Preview:

- The role of statistics in producing and analyzing data
- Selecting sampling units
    - Sampling designs
    - Sampling errors
    - Designing sampling plans
- Controlled experiments
    - Principles of experimental design
    - Experimenting with human subjects
    - Steps for planning an experiment
- Observational studies

- Types of observational studies: **cohort** (prospective), **case-referent** (case-control, retrospective) and **sample surveys**.
- Steps for planning an observational study

- Cause and effect in designed studies

What's the IDEA?

Exploratory analysis of available data (data not gathered as part of a designed study, or gathered as part of a study designed for some other purpose) is a vital part of scientific and technological progress, because it can:

- Suggest unexpected connections.
- Identify possibly promising areas for future exploration.

However, exploratory analysis of available data cannot establish scientific validity. For this, designed studies are needed. A good strategy is to use exploratory data analysis to identify specific questions that can then be answered by targeted designed studies.

This means that proper design is crucial to the statistical validity of a study. Light, Singer and Willett put it well:

"You can't fix by analysis what you bungled by design."

Two types of designed studies are

- Controlled Experiments
- Observational Studies

Both use **sampling units**: entities on which measurements or observations can be made.

Selecting Sampling Units

- Target Population: A collection of sampling units about which we want to draw conclusions.
- Frame: A list of all sampling units in the target population.
- Sample: A subset of the target population from which conclusions about the target population will be drawn.
- Sampling Design: A pattern, arrangement or method used for selecting a sample of sampling units from the target population.
- Sampling Plan: The operational plan, including the sampling design, for actually obtaining or accessing the sampling units for the study.

Example 1

Suppose, prior to an election, we want to estimate the outcome in the "population" of WPI students. To do so, we sample 20 WPI students and interview them. For this study:

- Sampling units: Individual students.
- Target population: All WPI students.

- Frame: Campus directory.
- Sample: The 20 selected students.
- Sampling Design: There are a large number of choices, some of which we will see shortly.
- Sampling Plan: Operational plan for deciding whom to interview, how to get them to do the interview, what to do if they can't be found, or won't talk, etc.

Reasons to Sample

- Lower Cost
- Shorter Time
- Little Loss of Precision
- The Only Choice (e.g. Destructive Testing)

What's the IDEA?

If the sampling units are selected by some non-probability method (convenience, for example), the results of the study are, strictly speaking, only applicable to the sampling units in the study. In order to have results of the study apply to the target population, sampling units must be selected from that target population using an appropriate probability sampling design.

Probability Sampling Methods

- Simple Random Sampling

- Stratified Random Sampling
- Cluster Sampling
- Multistage Sampling

Errors in Selecting Sampling Units

- Sampling Error
- Nonsampling Errors (e.g. selection bias)

Some Possible Errors in Selecting Sampling Units for the WPI Election Survey

- Sampling Error The extent to which the sample is unrepresentative of the population, by chance alone.
- Nonsampling Error: Selection Bias Selection bias would occur if the sampling method to at least some extent missed certain segments of the population: for example, if conducted in class, it would miss those who don't come to class.

Types of Designed Studies

- Controlled Experiment
- Observational Study

Controlled Experiments

Before we can discuss controlled experiments, we need some terminology:

- Experimental Unit: A sampling unit selected for use in a controlled experiment.
- Response: A measurement or observation of interest that is made on an experimental unit.
- Factor: A quantity that is thought to influence the response.
    - o Experimental Factor: A factor that is purposely varied by the experimenter.
    - o Nuisance Factor: A factor that is not controlled by the experimenter. Nuisance factors may or may not be known to the experimenter.
- Level: Each value assumed by a factor in an experiment.
- Treatments: The combinations of levels of experimental factors for which the response will be observed.

We are now ready to define **Controlled Experiment**:

A Controlled Experiment is a study in which treatments are imposed on experimental units in order to observe a response.

Example 2

A printing company is having trouble with ink overflow on printed documents After much brainstorming and discussion, they have narrowed the possible causes to two printing machine settings: the pressure plate setting and the ink flow rate setting. They design and run a controlled experiment to evaluate the effect of these settings on the finished product. They decide on three

experimental settings, low, medium and high, for the pressure plate and two experimental settings, low and high, for the ink flow rate. The response is the improperly inked area on a test sheet. The data are (in $(mm)^2$):

Ink flow setting	Pressure plate setting			Row Mean
	Low	Medium	High	
Low	25.300	27.100	19.700	
	21.600	24.200	21.900	
Mean	23.450	25.650	20.800	23.300
High	31.100	25.600	26.600	
	29.500	23.100	23.900	
Mean	30.300	24.350	25.250	26.633
Column Mean	26.875	25.000	23.025	24.967

For this experiment:

- Experimental units: paper sheets.
- Response: improperly inked area.
- Experimental factors: pressure plate, ink flow.
- Nuisance factors: variation in paper characteristics, environmental factors (temperature, humidity, etc.), ink supplier, etc.
- Factor levels: low, medium, high for pressure plate; low, high for ink flow rate.
- Treatments: pressure plate & ink flow rate combinations.

This is a controlled experiment because treatments (pressure plate & ink flow rate combinations) are imposed on experimental units (paper sheets) in order to observe a response (improperly inked area).

Here are some further quantities and concepts of importance in controlled experiments:

o Effect: The change in the average response between two factor levels or between two combinations of factor levels. Here are some effects from example 2:

  o The effect of high ink flow over low ink flow is $26.633 - 23.300 = 3.333$.

  o The effect of high pressure plate setting over medium pressure plate setting is $23.025 - 25.000 = -1.975$.

  o The effect of low ink flow and high pressure plate settings over high ink flow and medium pressure plate settings is $20.800 - 24.350 = -3.55$.

Confounding: Two or more factors are **confounded** if it is impossible to separate their individual effects. Here is an example of confounding in the ink flow experiment:

Suppose that the experimenters ran the trials for the low and high ink flow settings with different kinds of paper, and that the kind of paper used makes a real difference in the response. Then paper and ink flow would be confounded.

Assigning Treatments to Experimental Units: Two Commonly-Used Designs

- Completely Randomized Design (CRD) Treatments assigned to experimental units completely at random. Works well if units and/or experimental conditions are homogeneous.

- Randomized Complete Block Design (RCBD) Experimental units divided into blocks and all treatments are assigned at random within each block. Effective if units and/or experimental conditions within each block are more homogeneous than all units and/or experimental conditions taken as a whole. This is the simplest example of blocking: grouping experimental units by known sources of variation. Blocking is used to reduce the variation involved in comparing the experimental treatments.

Two Different Ways to Design the Ink Overflow Experiment

Suppose there are two printing machines on which the experiment is to be run.

- For a completely randomized design, assign all 12 treatments to the two machines at random. So, for example, machine 1 might get both treatments at high ink flow setting and high pressure plate setting. Of course the order of the runs will also be randomized. This kind of design makes sense if the machines act very much the same.
- For a randomized complete block design, assign one complete set of six treatments to machine 1 and the other to machine 2. Run each set of treatments in random order. This kind of design makes sense if there is substantial machine-to-machine variation.

More on Blocking

Here are two more examples of blocking:

- 25 pleasure boats around the country are available to test two types of marine paint. Make each boat a block by applying both types of paint to each. This (a) reduces boat-to-boat variation and variation due to such things as environment, and (b) makes the results of the study applicable to a wider range of environments and boat types.

- An alloy manufacturer produces aluminum ingots in four furnaces. Each furnace is known to have its own unique operating characteristics, so "furnace" will be a nuisance variable for any experiment run in the foundary that involves more than one furnace. In an experiment to assess the effect of stirring rate on the grain size of the product, the four furnaces are used as blocks: each of four stirring rates (the experimental factor) is assigned in random order to each furnace. This allows comparisons to be made within each furnace and the results extrapolated across furnaces.

Example 3: More on Blocking

Figures 1 and 2 give two different displays of data from a fabricated data set created to illustrate an interesting point about blocking.

Viewing the data as unpaired makes it look as though treatment 1 gives larger responses in general than does treatment 2. However, by looking at the data as paired, we see that except for one pair, treatment 2 gives larger responses. Blocking enables us to compare like with like, which in this case gives a very different analysis.

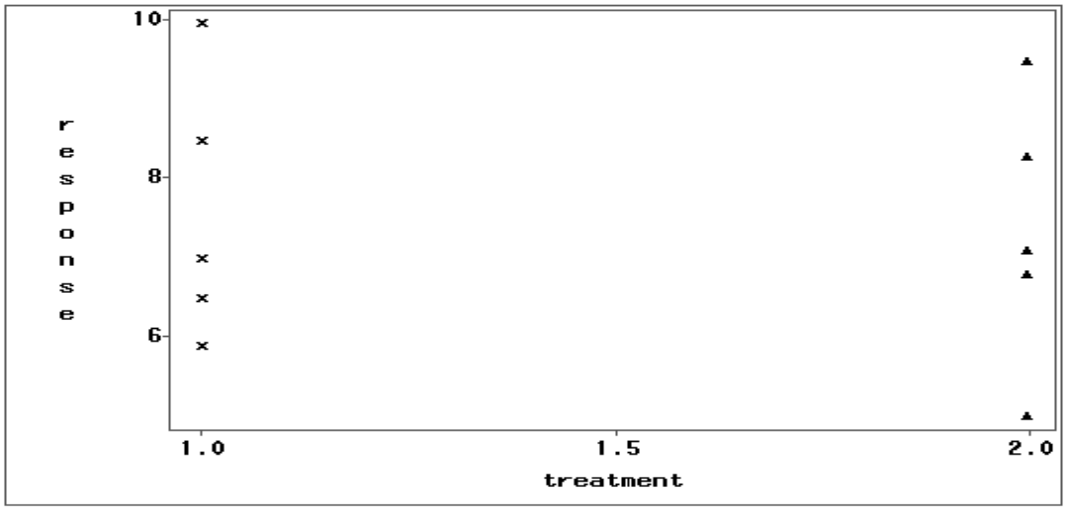

Figure 1: *Scatter plot of fabricated data viewed as unpaired.*

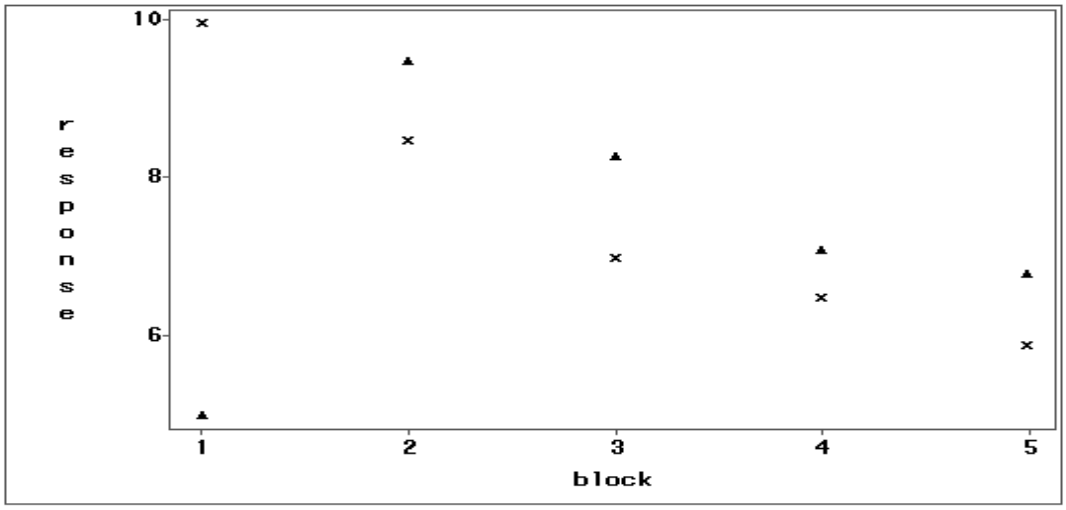

Figure 2: *Scatter plot of fabricated data viewed as paired.*

Principles of Experimental Design

- **Principle 1:** Make Sure the Process Is Stationary.
- **Principle 2:** Block What You Can.

Blocking factor: known nuisance factor used to reduce variance.

- **Principle 3:** Randomize What You Cannot Block.

  Randomization: chance assignment of treatments to experimental units. Used to eliminate bias due to unsuspected nuisance factors.

- **Principle 4:** Replicate as Time and Budget Permit.

  Replication=repetition. Beware of duplication masquerading as replication.

- **Principle 5:** Confirm the Results.

Principles of Experimental Design in the Ink Overflow Experiment

- **Principle 1:** Make Sure the Process Is Stationary.

  The experimenters made sure the process was stationary before taking data. Specifically, the responses were measured over time prior to running the experiment and no instability was observed. To further guarantee stationarity, each time the settings were changed the experimenters ran several sheets to assure that conditions stabilized before taking data.

- **Principle 2:** Block What You Can.

  The design as described could have been run in several ways. As a CRD (see below), treatments could have been assigned to experimental units at random and run in random order. As a RCBD (see below), all treatments could have been run in blocks (say one set in the

morning and one in the afternoon, or one with one operator and one with another, or on two different machines).

- **Principle 3:** Randomize What You Cannot Block.

  Order of run was randomized.

- **Principle 4:** Replicate as Time and Budget Permit..

  Two reps were done.

- **Principle 5:** Confirm the Results.

  To verify the results, confirmatory experiments were run.

Experimenting With Human Subjects

- **Method of Comparison** Compare two (or more) treatments or no treatment.
- **Treatment Group** Group of subjects that receives a treatment.
- **Control Group** Group of subjects that receives no treatment.
- **Placebo** Neutral "treatment" given to subjects in the control group.
- **Double-Blind** Neither subject nor evaluator(s) know which treatment (if any) was given.

Example: Salk Vaccine Field Trial, p. 108 of the text.

Observational Studies

Outside of controlled experiments, observational studies are the main class of designed studies. We will study three types:

- Cohort (Prospective) Study
- Case-referent (Retrospective) Study
- Sample Survey

Example 1, Continued:

The study mentioned earlier in which we selected 20 WPI students to see how they would vote is a sample survey. It is not a controlled experiment since no treatments are assigned to experimental units.

Example 4

Another example is a study published in the **Journal of the American Medical Association**, which assessed the health patterns of 5,000 Canadians, and found that those with the greatest folic acid intake had 68% less fatal coronary disease than those with the lowest intake. It is not a controlled experiment since it merely observed the folic acid intake and coronary death outcomes, rather than assigning a folic acid regimen to individuals.

Cohort Studies

Also known as prospective studies, because they follow established cohorts of subjects to an end result, and as quasi-experiments, because they are controlled experiment "wannabees", cohort studies lack the ability to control the assignment of treatments to experimental units (e.g., we cannot assign a human subject a certain number of cigarettes per day). As a result, they can only

demonstrate association, not cause- effect, between factors and responses.

In a cohort study, "treatment" and "control" groups (or cohorts) are established based on presumed cause (e.g. smokers and non-smokers) and the patterns of the response (e.g. lung cancer or not) are compared.

Example 4, Continued:

If the Canadian folic acid study assigned individuals to groups based on their reported folic acid intake, and compared the incidence of fatal coronary heart disease in the two groups five years later, then it was a cohort study.

Some control can be exerted in observational studies by

- Stratifying by nuisance variables.
- Adjusting responses for values of nuisance variables.

(Note that both these methods can be used to lessen the impact of nuisance variables in controlled experiments as well.)

Example 4, Continued:

In the Canadian folic acid study, researchers might stratify by age (so that comparisons would be done only for subjects of similar ages), or they might adjust the results for age.

Case-referent Studies Case-referent studies are particularly useful when

- The time between the hypothesized cause and observed effect is large, or
- The effect occurs rarely.

In a case-referent study, a group of cases (e.g., subjects having lung cancer), is compared with a group of referents (e.g., subjects not having lung cancer) with respect to differences in the hypothesized causes (e.g. smoking).

Example 4, Continued

If the Canadian folic acid study classified individuals into groups based on whether or not they died of coronary disease, and compared the intake of folic acid for the two groups, then it was a case-referent study.

Caution

In the text, the terminology "prospective study" is used instead of "cohort study", and "retrospective study" instead of "case-referent" study. That terminology gave too many students the mistaken idea that if the study was done in the past, or done using data taken in the past, it is case-referent. Case-referent refers only to the fact that the groups are formed based on the outcome and then differences in potential causal factors are sought. While it is true that outcomes must be observed prior to analyzing a case-referent study, studies based on data taken in the past can also be prospective if groups are established based on characteristics observed prior to the outcomes.

For example, suppose the researchers conducted the Canadian folic acid study by randomly select-

ing 5000 patient records which included folic acid intake, and whether they suffered fatal coronary disease by the end of a 5 year period. Even though the data were taken in the past, if they formed comparison groups based on folic acid intake and compared outcomes, the study is a cohort study.

Cause-Effect

Only a properly-designed and conducted controlled experiment can establish a cause-effect relationship between factors and response.

The biggest difference between controlled experiments and observational studies, such as cohort and case-referent studies, that attempt to show cause-effect, is the idea of **control**: the ability of the experimenter to assign treatments to experimental units. It is the control in controlled experiments that validates cause-effect conclusions and the lack of control in observational studies that invalidates cause-effect conclusions.

Sample Surveys

- Use a sample of sampling units obtained from a population to obtain information about the whole population.
- Have as their primary goals description of various aspects of the population from which the sample is obtained, or comparison of subgroups from that population (not establishment of association).

Non-sampling Errors in Studies of Human Populations

In addition to sampling error and selection bias, studies of human populations in which individuals are asked to respond to questions, verbally or in writing, are subject to other nonsampling errors, such as

- Nonresponse bias: Bias due to failure to obtain responses from some subjects.
- Response bias: Bias due to erroneous responses from some subjects.

Example 4, Continued

Here are some possible non-sampling errors in the Canadian Folic Acid Study:

- Nonresponse bias might occur if certain individuals refuse to supply their medical histories.
- Response bias might occur if the subject feels that giving information about poor health will adversely affect future insurability.

Steps in Designing Observational Studies

- Determine what information is required.
- Design the sampling plan.
- Decide how the data are to be obtained.
- Establish procedures to reduce nonsampling errors.
    - Reduce nonresponse bias by
        * Keeping questionnaires short

    * Building-in incentives

    * Doing follow-ups

  In addition, monitoring differences in known demographic variables for respondents and nonrespondents will help assess the severity of the nonresponse problem.

- o Reduce response bias by careful design, testing and modification, and training.

- Always, always, always Do a pilot study.