



## Chapter 2

### The Data Analysis Process and Collecting Data Sensibly

---

---

---

---

---

---

---

---



### Important Terms

- **Variable** – A variable is any characteristic whose value may change from one individual to another
- Examples:
  - Brand of television
  - Height of a building
  - Number of students in a class

---

---

---

---

---

---

---

---



### Important Terms

- **Data** – Data results from making observations either on a single variable or simultaneously on two or more variables.
- A **univariate data set** consists of observations on a single variable made on individuals in a sample or population.

---

---

---

---

---

---

---

---



## Important Terms

- A **bivariate data set** consists of observations on two variables made on individuals in a sample or population.
- A **multivariate data set** consists of observations on two or more variables made on individuals in a sample or population.

---

---

---

---

---

---

---

---



## Data Sets

- A univariate data set is **categorical** (or **qualitative**) if the individual observations are categorical responses.
- A univariate data set is **numerical** (or **quantitative**) if the individual observations are numerical responses where numerical operations generally have meaning.

---

---

---

---

---

---

---

---



## 2.1: The Data Analysis Process

- Planning and Conducting a Study
  - Understand the Nature of the Problem
  - Decide What to Measure and How to Measure It
  - Collect the Data
  - Summarize the Data & Perform a Preliminary Analysis
  - Do the Formal Data Analysis
  - Interpret the Results

---

---

---

---

---

---

---

---



## 2.2: Types of Bias

- **Selection Bias** is the tendency for samples to differ from the corresponding population as a result of systematic exclusion of some part of the population.
  - Example: Taking a sample of opinion in a community by selecting participants from phone numbers in the local phone book would systematically exclude people who choose to have unlisted numbers, people who do not have phones, and people who have moved into the community since the telephone directory was published.

---

---

---

---

---

---

---

---



## Types of Bias

- **Measurement or Response Bias** is the tendency for samples to differ from the corresponding population because the method of observation tends to produce values that differ from the true value.
  - Example: Taking a sample of weights of a type of apple when the scale consistently gives a weigh that is 0.2 ounces high.

---

---

---

---

---

---

---

---



## Types of Bias

- **Nonresponse Bias** is the tendency for samples to differ from the corresponding population because data is not obtained from all individuals selected for inclusion in the sample..
  - Example: In a study that ask questions of a personal nature, many individuals that are selected might refuse to answer the survey questions. This occurs quite often when the questions are of a highly personal nature or when the individual feels that certain response might prove personally damaging.

---

---

---

---

---

---

---

---



## Important Note on Bias

- Bias is introduced by the way in which a sample is selected so that increasing the size of the sample does nothing to reduce the bias

---

---

---

---

---

---

---

---



## Sampling Methods

- A **Simple Random Sample of size n** is a sample that is selected in a way that ensures that every different possible sample of the desired size has the same chance of being selected.
- A common method of selecting a random sample is to first create a list, called a **sampling frame** of the individuals in the population. Each item on the list can then be identified by a number, and a table random digits or a random number generator can be used to select the sample.

---

---

---

---

---

---

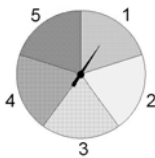
---

---



## Sampling Methods

- **Sampling with replacement** means that after each successive item is selected for the sample, the item is “replaced” back into the population and may therefore be selected again.



Example: Choose a sample of 5 digits by spinning a spinner and choosing the number where the pointer is directed.

---

---

---

---

---

---

---

---



## Sampling Methods

- **Sampling without replacement** means that after an item is selected for the sample it is removed from the population and therefore cannot be selected again.
- Example: A hand of “five card stud” poker is dealt from an ordinary deck of playing cards. Typically, once a card is dealt it is not possible for that card to appear again until the deck is reshuffled and dealt again.

---

---

---

---

---

---

---

---



## Sampling Methods

- A entire population is divided into subpopulations called **strata**.
- **Stratified sampling** entails selecting a separate simple random sample from each of the strata.
- Example: Teachers in a large urban school district are given tenure by subject. The sample is taken by choosing random samples from each of the tenure areas.

---

---

---

---

---

---

---

---



## Sampling Methods

- A entire population is divided into non-overlapping subgroups called **clusters**.
- **Cluster sampling** entails selecting clusters at random and all individuals in the selected clusters are included in the sample.

---

---

---

---

---

---

---

---



## Sampling Methods

- Example: In a large university, a professor wanting to find out about student attitudes randomly selects a number of classes to survey and he includes all the students in those classes.
- Note: The ideal situation occurs when it is reasonable to assume that each cluster reflects the generally population. If that is not the case or when clusters are small, a large number of clusters must be selected to get a sample that reflects the population.

---

---

---

---

---

---

---

---



## Sampling Methods

- **Systematic sampling** is a procedure that can be employed when it is possible to view the population of interest as consisting of a list or some other sequential arrangement. A value  $k$  is specified (a number such as 25, 100, 2500...). The one of the first  $k$  individuals is selected at random, and then ever  $k^{\text{th}}$  individual in the sequence is selected to be included in the sample.

---

---

---

---

---

---

---

---



## Sampling Methods

- Example: In a large university, a professor wanting to select a sample of students to determine the student's age, might take the student directory (an alphabetical list) and randomly choose one of the first 100 students and then take every 100<sup>th</sup> student from that point on.

---

---

---

---

---

---

---

---



## Sampling Methods

- **Convenience sampling** is using and easily available or convenient group to form a sample.
- Example: A “voluntary response sample” is often taken by television news programs. Viewers are encouraged to go to a website and “vote” yes or no on some issue. The commentator then would announce the results of the survey.

---

---

---

---

---

---

---

---



## Sampling Methods

- It is highly unlikely that the responses would be accurately representative of the opinion of the public at large.
- Mainly get the extreme opinions Pro or Con.

---

---

---

---

---

---

---

---



## 2.3: Statistical Studies: Observational & Experimentation

---

---

---

---

---

---

---

---



## Collection of Data

- A **confounding variable** is one that is related both the group membership and to the response variable of interest in the research study. It is generally an explanatory variable of importance that is not included in the study. A carefully designed experiment can do a lot to control confounding variables.

---

---

---

---

---

---

---

---



## Collection of Data

- A study is an **Observational Study** if the values of the variable(s) of a sample from one or more populations is observed.
- Observational studies are usually used to draw conclusions about the population or about differences between two or more populations.

---

---

---

---

---

---

---

---



## Observational Study

- An **observational study** observes individuals and measures variables of interest but does not attempt to influence the responses.
  - Difficult to measure or gauge the effect of an action or procedure
  - Lurking variables are uncontrolled so the study may be confounded
  - + Can use available data

---

---

---

---

---

---

---

---





## Collection of Data

- A study is an **Experiment** if the values of one or more response variables are recorded when the investigator controls (or manipulates) one or more factors.
- Experiments are usually used when attempting to determine the effect of the manipulation of the factors being controlled.

---

---

---

---

---

---

---

---



## 2.4: Simple Comparative Experiments

- An experiment is a planned intervention undertaken to observe the effects of one or more explanatory variables, often called factors, on a response variable.
- Any particular combination of values for the explanatory variables is called an **experimental condition** or **treatment**.

---

---

---

---

---

---

---

---



## More on Experiments

- The **design** of the experiment is the overall plan for conducting the experiment. A good design minimizes ambiguity in the interpretation of the results.

---

---

---

---

---

---

---

---



## Experimental Studies

- An **experiment** deliberately imposes some treatment on individuals in order to observe their responses.
  - + Allows the measurement of effect of a treatment
  - + Can help to control lurking variables
  - + Can give good evidence of causation
  - May not measure realistic effects. Not necessarily workable in real life.

---

---

---

---

---

---

---

---



## Factors

- Any particular combination of values for the explanatory variables is called an **experimental condition** or **treatment**.
- An **extraneous factor** is one that is not of interest in the current study but is thought to affect the response variable.
- Two factors are **confounded** if their effects on the response variable cannot be distinguished from one another.

---

---

---

---

---

---

---

---



## Principles of Experimental Design

- The fundamental principles of statistical design of experiments are:
  - 1) Randomization**  
Random assignment (of subjects to treatments or of treatments to trials) to ensure that the experiment does not systematically favor one experimental condition over another.
  - 2) Blocking**  
Using extraneous factors to create groups (blocks) that are similar. All experimental conditions are then tried in each block.

---

---

---

---

---

---

---

---



# Principles of Experimental Design

- The fundamental principles of statistical design of experiments are:

### 3) Direct Control

Holding extraneous factors constant so that their effects are not confounded with those of the experimental conditions.

### 4) Replication

Ensuring that there is an adequate number of observations in each experimental condition.

---

---

---

---

---

---

---

---

---

---



## 2.5: More on Experimental Design

- A **placebo** is a treatment that resembles the other treatments in an experiment, but which has no active ingredients.
  - In an experiment to test the effectiveness of a new vaccine (by injection), a placebo treatment would consist of injection a neutral substance such as saline solution.
- A **control group** is a group that receives no treatment or a placebo treatment.
  - To see if the "act of living" has any affect...

---

---

---

---

---

---

---

---

---

---



## More on Treatments

- A **Single Blind** study is when the subject does not know if they actually receiving the real treatment
- A **Double Blind** study is when the subject, nor the examiner knows who is actually receiving the real treatment

---

---

---

---

---

---

---

---

---

---



## 2.6: More on Observational Studies (Surveys)

- Fixture in research, especially in the social sciences
- Used routinely in business, government, etc
- Attempt to measure attitudes, opinions, preferences, beliefs, values (Are the tax rates to high? Do you favor a ban on assault rifles?) or to obtain specific facts (What was your GPA in High School? As of the summer of 2006, how many years have you been in school?)

---

---

---

---

---

---

---

---



## More on Surveys

- Survey: voluntary encounter in which an interviewer seeks information from a respondent by engaging in a special conversation (scripted questions).
  - In person
  - Via telephone
  - Written questionnaire (via mail, email, etc)
- The interviewer decides the questions & the respondent decides on which Q's to answer, if any...

---

---

---

---

---

---

---

---



## More on Surveys

- Comprehension is the single most important issue of getting reliable & valid results in a survey.
  - Does the respondent understand what the question is asking. Often the researcher thinks it is obvious what the question is asking, but to the respondent, it's unclear.
    - Appropriate vocabulary
    - Simple sentence structure
    - Little or no ambiguity

---

---

---

---

---

---

---

---



## More on Surveys

- Need to make sure the survey questions have been tested for reliability & validity.
  - Q: Do you think children suffer ill effects from watching violence and/or pornographic material on TV?
  - What is your definition of children?
  - AAP includes 0-18, or until sentioned adult as children
- Actually even more issues, how do you define suffer ill effects, violence, porn?

---

---

---

---

---

---

---

---



## More on Surveys

- Constructing a survey & testing for reliability can be a large task. So when doing so:
  - Make Q's understandable by people in the population being studied.
  - Remember that everyone's is NOT able to remember all the specifics & details of previous events. Make Q's specific for what you really want to know.
  - Try to NOT have questions that make the respondent feel threatened or embarrassed. If you must have sensitive Q's, put them @ the end of the survey.

---

---

---

---

---

---

---

---