DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Chapter 2

The Data Analysis Process and Collecting Data Sensibly

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

2.1 Steps of 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

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Bias in Survey Sampling

- In survey sampling, **bias** refers to the tendency of a **sample statistic** to systematically over- or under-estimate a **population** parameter.
- Increasing the size of the sample does **NOT** reduce the bias.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

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 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 or only have cell phones,
  - people who have moved

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Types of Bias

•**Response Bias**— because of the manner which an interview is conducted, the phrasing of questions, the attitude of the respondent, inaccurate data is collected. **Example**

- **Leading questions. Wording of the question to favor one response over another.**

•**Measurement Bias (also called Response Bias)** described systematic deviations from the true value. **Example**

- **Using faulty measurement instruments (faulty scale)**

•**Voluntary Response Bias** — Voluntary response bias occurs when sample members are self-selected volunteers. **Example**

- **Subjects that respond to a radio station survey.**

•**Nonresponse Bias** — occurs when subjects selected for a sample is unwilling or unable to respond. **Examples**

- **studies asking personal questions such as income or internet downloading of music and videos.**

•**AVOID Undercoverage** — occurs when some members of the population are inadequately represented in the sample.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

2.2 Sampling Methods

A **Simple Random Sample (SRS)**

- a sample in which each member of the sample is chosen by chance and each member of the population has an equal chance to be in the sample.
- Size of Sample is denoted by the variable (n).

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Sampling Methods

Simple Random Sampling (SRS)

- **2 Types**
  1. **SRS WITH replacement** where a member of the population can be selected more than once.
  2. **SRS WITHOUT replacement** where a member of the population can be selected at most once.
- **Simple Random Sampling Method**
  1. first create a **sampling frame** which is a list for individuals in the population from which the sample is selected.
  2. Each item on the list can then be identified by a number, and
  3. Then use a table of random digits created by a **random number generator** to select the sample.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Sampling Methods

Systematic sampling

STEP 1: Divide the population size by the sample size and round the results down to the nearest whole number, ***m***.

STEP 2: Use a random-number table to obtain a number ***k*** between 1 and ***m;*** where ***k*** is the first member chosen from the population.

STEP 3: Then **select every k^{th} individual** to be included in the sample

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^{th} student from that point on.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Sampling Methods

Cluster sampling

STEP 1: Identify **clusters (heterogeneous subgroups of a population).**

STEP 2: Obtain a simple random sample of each **clusters.**

STEP 3: Use all the **clusters** as the sample.

Example: In a large university, a professor wants 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 classes are the clusters.]

9

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Sampling Methods

Stratified Sampling with Proportion Allocation

STEP 1: Divide the population into homogeneous sub grouping called **strata**.

STEP 2: From each **stratum**, obtain a simple random sample size proportional to the size of the **stratum.**

STEP 3: Use all the members in the sample.

Example: Create a sample to survey registered voters for a presidential election based on income levels – 10% upper, 70% middle, and 20% low income families.

10

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Sampling Methods

Convenience sampling

In **Convenience sampling**, individuals are chosen because they are the easiest to reach.

Example: Sampling is done at major public areas such as grocery stores or malls.

- **It is highly unlikely that the responses would be accurately representative of the opinion of the public at large.**

11

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

DAY2

- Review Chapter 1 HW Summer Assignment
- Chapter 1 Test
- Finish Section 2.2
- Review 2A HW #'s 4-13
- Section 2.3 Notes

12

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

2.3 Statistical Studies

See
2.3 Supplemental Notes

for 4 type of Data Collection Methods

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

2.3 Statistical Studies

Observational vs. Experimental Studies

Observational Study in which the individuals are observed and specific variables of interest are measured.

Experimental Study in which the individuals are subject to a particular treatment and the responses to the treatment are measured.

- **Observational studies, although not meaningless, are not as informative as experiments and therefore are not better for obtaining a cause-and-effect relationship.**

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Observational Studies

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

Observational studies are used to draw general conclusions about populations; a direct cause-and-effect relationship can NOT be determined.

15

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Experimental Study

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.

Experiments show a cause-and-effect relationship between 2 variables.

16

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

DAY3

- Finish Chapter 2 Notes
- Review 2A HW #'s 14+

17

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

2.4 to 2.7 Survey Design

Parts of an Experiment

In an experiment, a researcher manipulates one or more variables, while holding all other variables constant. By noting how the manipulated variables affect a response variable, the researcher can test whether a causal relationship exists between the manipulated variables and the response variable

Experimental Units (or Subjects) - the individuals in the experiment.

Response Variable measures the outcome that have been observed.

Treatments - the specific condition that is applied to the **experimental units**

Factors - one or more **explanatory variables** in an experiment. Each **factor** has 1 or more levels (different quantities or categories of the **factor**).

- **Combinations of different levels of factors are the treatments.**

Explanatory Variables - *explains the* ***response variable*** *applied to the* ***experimental units.***.

18

Parts of an Experiment

In a hypothetical experiment, the researcher is studying the possible effects of Vitamin C and Vitamin E on health.

Vitamin E	Vitamin C: 0 mg	Vitamin C: 250 mg	Vitamin C: 500 mg
0 mg	Treatment 1	Treatment 2	Treatment 3
400 mg	Treatment 4	Treatment 5	Treatment 6

There are two factors - dosage of Vitamin C & dosage of Vitamin E.

- The Vitamin C factor has three levels - 0, 250 and 500 mg per day.
- The Vitamin E factor has 2 levels - 0 and 400 mg per day.

The experiment has six treatments

•**The independent variable (the factor)** is an explanatory variable manipulated by the experimenter.

The recipients of experimental treatments are the **experimental units**. And selected based on sampling methods previously discussed.

•The researcher is looking at the effect of vitamins on health. **The dependent variable** in this experiment would be some measure of health (annual doctor bills, number of colds caught in a year, number of days hospitalized, etc.).

19

Principles of Experimental Design

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

The fundamental principles of statistical design of experiments are:

1) **Direct Control**
2) **Randomization**
3) **Replication**
4) **Blocking**

A study that does NOT include Control, Randomization, Replication, Blocking is subject to bias and confounding.

- **A Confounding Variable is one whose effect on the response variable cannot be untangled from the effects of the treatment.**

20

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Principles of Experimental Design ▪ ▪ ▪

1. Direct Control

Direct Control refers to steps taken to reduce the effects of extraneous variables (i.e., variables other than the independent variable and the dependent variable). These extraneous variables are called lurking variables.

Extraneous factor is one that is not of interest in the current study but is thought to affect the response variable.

Lurking variable is one that has an effect on the outcome of a study but was not part of the investigation.

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.

Control group is a group that receives no treatment or a placebo treatment.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Principles of Experimental Design ▪ ▪ ▪

1. Direct Control

Another type of control is blinding.

- In **single blind studies**, the experimenter (or observer) is aware of who or what belongs to the control group and the experimental group.
- In **double blind studies**, the experimenter is **not** aware of who/what belongs to which group. This is to eliminate the subjective bias an experimenter may have.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Principles of Experimental Design ▪ ▪ ▪

2. Randomization

Randomization is the process by which treatments are assigned by a chance mechanism to the experimental units to ensure that the experiment does not systematically favor one experimental condition over another.

- It "averages out" variation due to variables that cannot be controlled.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Principles of Experimental Design ▪ ▪ ▪

3. Replication

Replication is the practice of reducing chance variation by assigning each treatment to many experimental units.

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

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

4. Blocking

Blocking using extraneous factors to create groups (blocks) that are similar. All experimental conditions are then tried in each block.

- Do not confusing **blocking** and **stratification**. They perform similar functions, but **blocks** are part of the experimental design and **strata** are part of a sampling process.

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Types of Experimental Designs

- Completely Randomized Design
  - probably the simplest experimental design
  - All experimental units are assigned treatments solely by chance
  - A completely randomized design relies on randomization to control for the effects of extraneous variables

Treatment	
Placebo	Vaccine
500	500

- Randomized Block Design
  - the experimenter divides participants into subgroups called blocks
  - Then, participants within each block are randomly assigned to treatment conditions
  - Because this design reduces variability and potential confounding, it produces a better estimate of treatment effects.

	Treatment	
Gender	Placebo	Vaccine
Male	250	250
Female	250	250

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

Types of Experimental Designs

•**Matched Pairs Design**

- **is a special case of the randomized block design**
- **It is used when the experiment has only two treatment conditions;**
- **and participants can be grouped into pairs, based on some blocking variable**
- **Then, within each pair, participants are randomly assigned to different treatments.**

	Treatment	
	Placebo	Vaccine
1	1	1
2	1	1
…	…	…
499	1	1
500	1	1

*each pair matched on gender & age

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

DAY4

- Review HW 2B
- CW 2C – Chapter 2 Concept Quiz
- **Start Chapter 2 Review HW**
  - It will be graded
  - Collected next class
  - Will NOT be reviewed prior to Test so see me before class with questions

- **Chapter 2 Test - next class**
  - **27 Multiple Choice AP Questions**
  - **2 points each**
  - **104 total points**
  - **1 hour time limit**

DUXBURY
THOMSON LEARNING
Professor David Crystal
Rochester Institute of Technology
Dept of Mathematics & Statistics

DAY5

- **Chapter 2 Test – 1 hour**
- **Start Chapter 3**