How to succeed in Mr. L's class

- What is an **advocate**?

  Someone who supports you and tries to help you succeed

- What is an *adversary*?

  A rival; Someone who works against you and gets in the way

How to succeed in Mr. L's class

- I am here to be your **ADVOCATE**, please don't treat me as an *adversary*!
- When you take the time, and make the effort, **MATH** can help you succeed, so try to avoid thinking of math as your *adversary* as well.
- Be your own **ADVOCATE!**

Chapter 2: Collecting Data Sensibly & Designing Studies

adapted from The Practice of Statistics, 4th edition – For AP*

Chapter 2 - Collecting Data & Experimental Designs

- **Chapter 2.1 - Observational Studies vs. Experimentation (a controlled experiment)**
- **Chapter 2.2 - Sampling**
- **Chapter 2.3** - Simple Comparative Experiments
- **Chapter 2.4** - More Experimental Design
- **Chapter 2.5** - Interpreting Results of Statistical Analysis

Section 2.1 Statistical Studies: Observation & Experimentation

Learning Objectives

After this section, you should be able to…

- ✓ DISTINGUISH observational studies from experiments
- ✓ DESCRIBE the language of experiments
- ✓ APPLY the three principles of experimental design
- ✓ DESIGN comparative experiments utilizing completely randomized designs and randomized block designs, including matched pairs design

Warm-Up

1) The two major branches of statistics are _____ and _____

2) A sample that consists of people who choose for themselves to participate by responding to a general invitation is called a _____

3) What is this is the formula for: $\bar{x} = \sum \frac{x_i}{n}$

Warm-Up ANSWERS

1) The two major branches of statistics are **descriptive statistics** and **inferential statistics**

2) A sample that consists of people who choose for themselves to participate by responding to a general invitation is called a **voluntary response sample** .

Warm-Up ANSWERS

Sample mean = $\bar{x}$

Greek symbol for Sigma (upper case) = $\sum$

$\sum$ means “summation”

What does this ratio signify: $\frac{x_i}{n}$

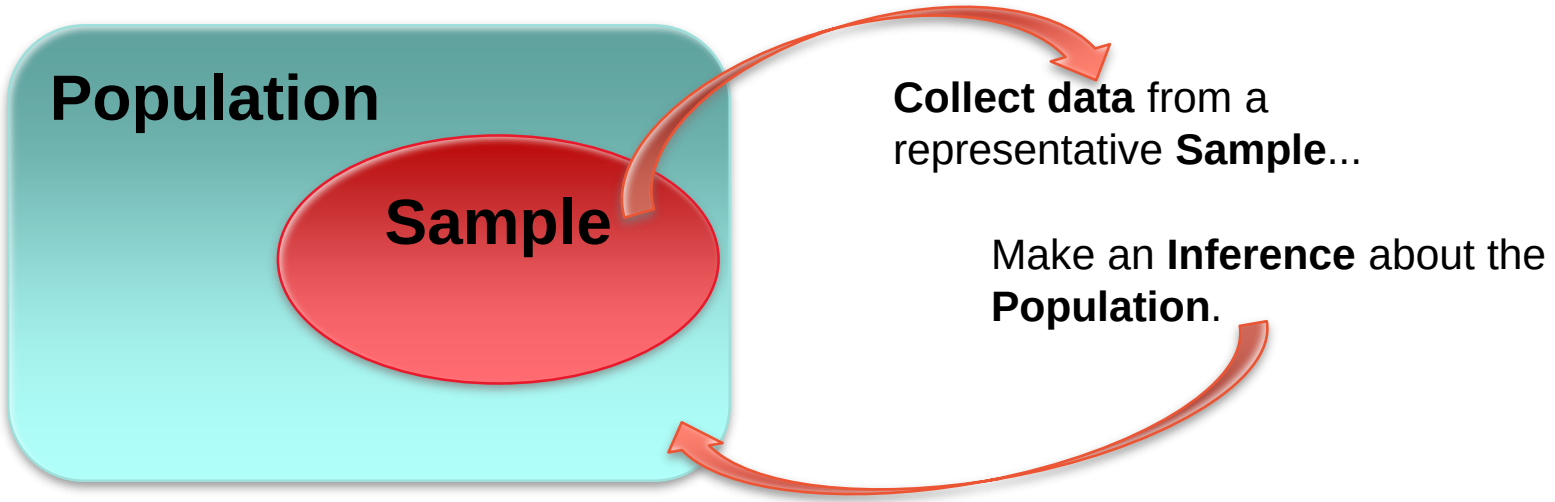
REVIEW: Population and Sample

The distinction between population and sample is basic to statistics. To make sense of any sample result, you must know what population the sample represents

Definition:

The **population** in a statistical study is the entire group of individuals about which we want information.

A **sample** is the part of the population from which we actually collect information. We use information from a sample to draw conclusions about the entire population.



Observational Study versus Experiment

In contrast to observational studies, experiments don't just observe individuals or ask them questions. They actively impose some treatment in order to measure the response.

Definition:

An **observational study** observes individuals and measures variables of interest but does not attempt to influence the responses.

An **experiment** deliberately imposes some treatment on individuals to measure their responses.

When our goal is to understand cause and effect, experiments are the *only* source of fully convincing data.

The distinction between observational study and experiment is one of the most important in statistics.

Section 2.2
Samples and Surveys

Learning Objectives – We will be able to…

After this section, you should be able to…

- ✓ IDENTIFY the population and sample in a sample survey
- ✓ IDENTIFY voluntary response samples and convenience samples
- ✓ DESCRIBE how to use a table of random digits to select a simple random sample (SRS)
- ✓ DESCRIBE simple random samples, stratified random samples, and cluster samples
- ✓ EXPLAIN how undercoverage, nonresponse, and question wording can lead to **bias** in a sample survey

Activity: *See no evil, hear no evil?*

CLASS ACTIVITY! ☺

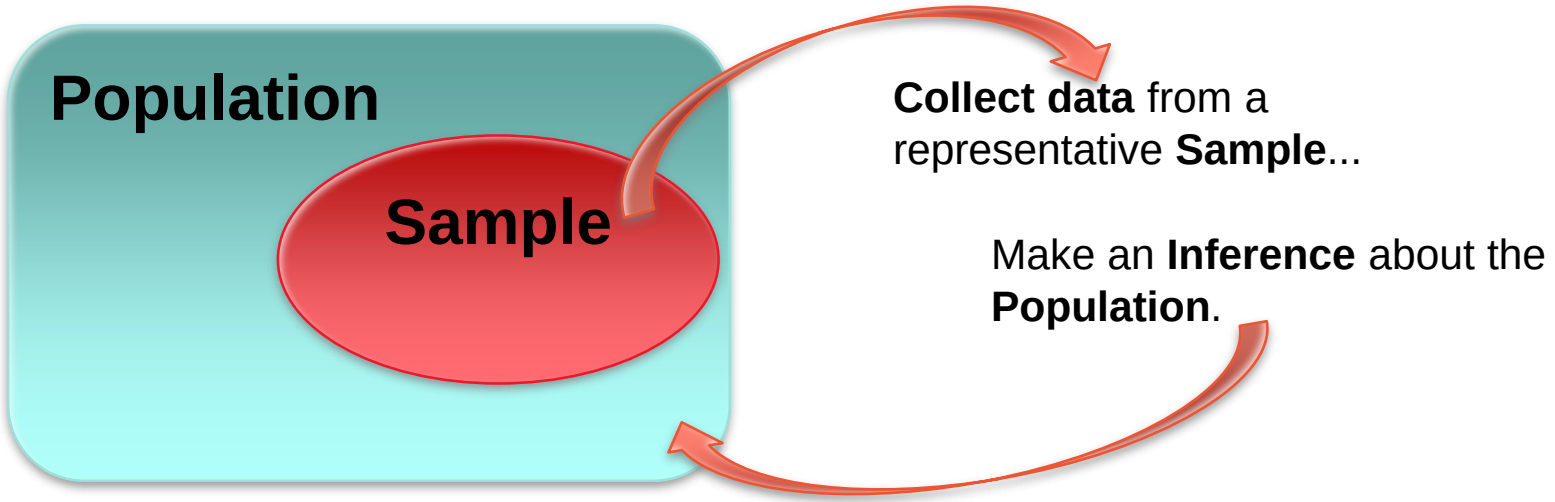
REVIEW: Population and Sample

The distinction between population and sample is basic to statistics. To make sense of any sample result, you must know what population the sample represents

Definition:

The **population** in a statistical study is the entire group of individuals about which we want information.

A **sample** is the part of the population from which we actually collect information. We use information from a sample to draw conclusions about the entire population.



The Idea of a Sample Survey

We often draw conclusions about a whole population on the basis of a sample.

Choosing a representative sample from a large, varied population is not that easy.

Step 1: Define the *population* we want to describe.

Step 2: Decide exactly *what we want to measure.*

A "sample survey" is a study that uses an organized plan to choose a sample that represents some specific, and clearly defined population.

Step 3: Decide how to choose a sample from the population (various sampling methods).

How to AVOID Sampling Badly?

How can we choose a sample that we can trust to represent the population? There are a number of different methods to select samples.

Definition:
Choosing individuals who are easiest to reach results in a **convenience sample.**

Convenience samples often produce unrepresentative data…why? There is a high likelihood of ***bias.***

Definition:
The design of a statistical study shows **bias** if it systematically favors certain outcomes.

Types of Bias in Sampling

Bias often occurs when the survey *sample* does not accurately represent the *population*. Some common examples of selection bias are described below.

- **Selection Bias**: The bias that results from an unrepresentative sample is called **selection bias**. Common examples of selection bias include *undercoverage*, or bias from a voluntary-response survey or a convenience sample.

- **Non-response Bias**: Sometimes, individuals chosen for the sample are unwilling or unable to participate in the survey. Non-response bias is the bias that results when respondents differ in meaningful ways from non-respondents.

- **Measurement or Response Bias**: A poor measurement process can also lead to bias. In survey research, the measurement process includes the environment in which the survey is conducted, the way that questions are asked, and the state of the survey respondent.

How to AVOID Sampling Badly

- Convenience samples are almost guaranteed to show bias. So are **voluntary response samples**, in which people decide whether to join the sample in response to an open invitation.

Definition:

A **voluntary response sample** consists of people who choose themselves by responding to a general appeal. Voluntary response samples show bias because people with strong opinions (often in the same direction) are most likely to respond.

How to Sample Well: Random Sampling

- The statistician's remedy is to allow impersonal chance to choose the sample. A sample chosen by chance rules out both favoritism by the sampler and self-selection by respondents.
- **Random sampling**, the use of chance to select a sample, is the central principle of statistical sampling.

Definition:

A **simple random sample (SRS)** of size *n* consists of N individuals from the population chosen in such a way that every set of *n* individuals has an equal chance to be the sample actually selected.

In practice, people use random numbers generated by a computer or calculator to choose samples. If you don't have technology handy, you can use a **table of random digits.**

How to Choose an SRS

Definition:

A **table of random digits** is a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these properties:

- Each entry in the table is equally likely to be any of the 10 digits 0 - 9.
- The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

How to Choose an SRS Using Table D

Step 1: Label. Give each member of the population a numerical label of the *same length.*

Step 2: Table. Read consecutive groups of digits of the appropriate length from Table D.

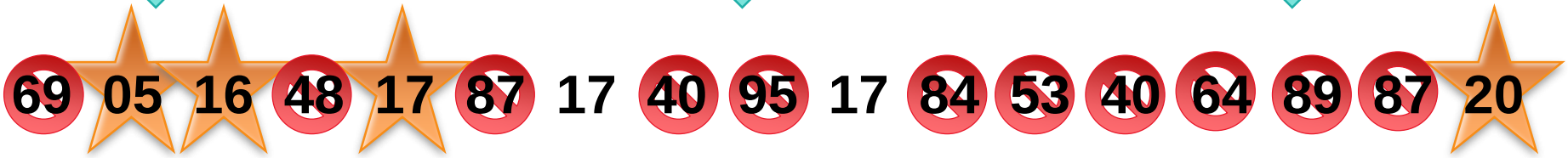
Your sample contains the individuals whose labels you find.

Example: How to Choose an SRS

- Problem: Use Table D at line 130 to choose an SRS of 4 hotels.

01 Aloha Kai	08 Captiva	15 Palm Tree	22 Sea Shell
02 Anchor Down	09 Casa del Mar	16 Radisson	23 Silver Beach
03 Banana Bay	10 Coconuts	17 Ramada	24 Sunset Beach
04 Banyan Tree	11 Diplomat	18 Sandpiper	25 Tradewinds
05 Beach Castle	12 Holiday Inn	19 Sea Castle	26 Tropical Breeze
06 Best Western	13 Lime Tree	20 Sea Club	27 Tropical Shores
07 Cabana	14 Outrigger	21 Sea Grape	28 Veranda

69051 64817 87174 09517 84534 06489 87201 97245



Our SRS of 4 hotels for the editors to contact is: 05 Beach Castle, 16 Radisson, 17 Ramada, and 20 Sea Club.

Sampling Methods

Statistical or Random Sampling Methods

- Simple Random Sample (SRS)
- Stratified Random Sample
- Cluster Sample
- Systematic Sample

Preferred methods

Non-Random Sampling *(often creates bias)*

- Convenience Samples
- Voluntary Response Samples

Other Random Sampling Methods

- The basic idea of sampling is straightforward: take an SRS from the population and use your sample results to gain information about the population. Sometimes there are statistical advantages to using more complex sampling methods.

- One common alternative to an SRS involves sampling important groups (called *strata*) within the population separately. These "sub-samples" are combined to form one stratified random sample.

Definition:

To select a **stratified random sample,** first classify the population into groups of similar individuals, called **strata**. Then choose a separate SRS in each stratum and combine these SRSs to form the full sample.

Other Random Sampling Methods

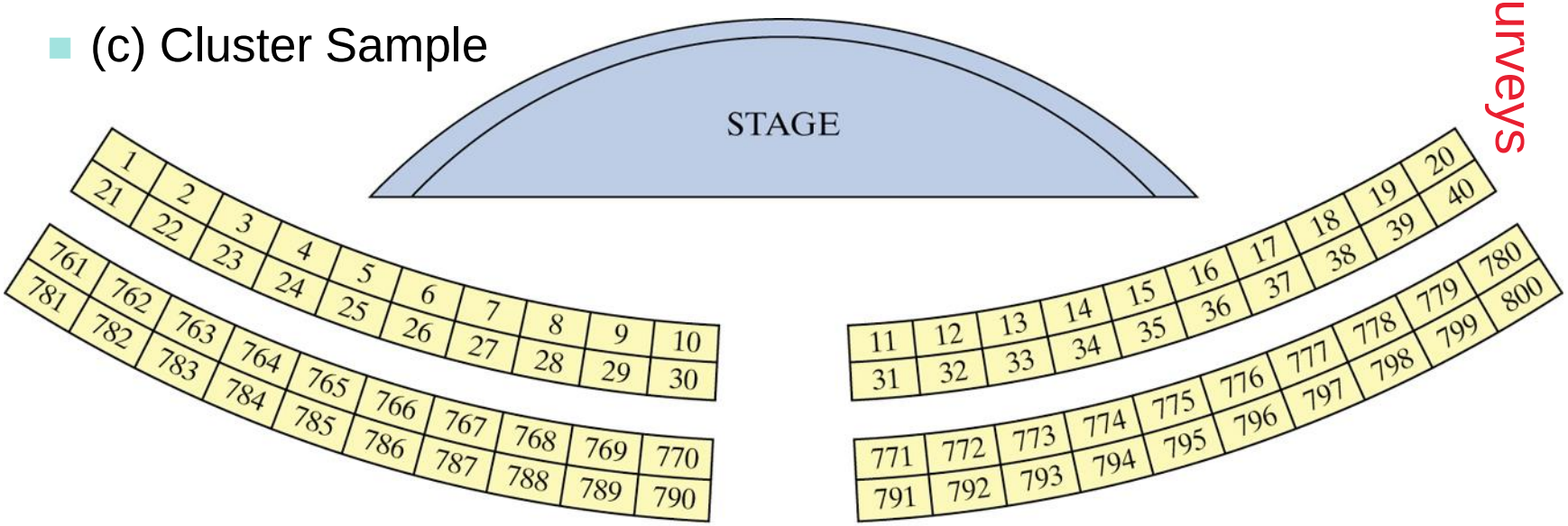
- Although a stratified random sample can sometimes give more precise information about a population than an SRS, both sampling methods are hard to use when populations are large and spread out over a wide area.
- In that situation, we'd prefer a method that selects groups of individuals that are "near" one another.

Definition:

To take a **cluster sample**, first divide the population into smaller groups. Ideally, these clusters should mirror the characteristics of the population. Then choose an SRS of the clusters. All individuals in the chosen clusters are included in the sample.

Example: Sampling at a School Assembly

- Describe how you would use the following sampling methods to select 80 students to complete a survey.
- (a) Simple Random Sample
- (b) Stratified Random Sample
- (c) Cluster Sample



9th grade: Seats 601–800	10th grade: Seats 401–600
11th grade: Seats 201–400	12th grade: Seats 1–200

Inference for Sampling

- The purpose of a sample is to give us information about a larger population.
- The process of drawing conclusions about a population on the basis of sample data is called **inference**.

Why should we rely on random sampling?

1)To eliminate bias in selecting samples from the list of available individuals.

2)The laws of probability allow trustworthy inference about the population

- Results from random samples come with a **margin of error** that sets bounds on the size of the likely error.
- Larger random samples give better information about the population than smaller samples.

Sample Surveys: What Can Go Wrong?

- Most sample surveys are affected by errors in addition to sampling variability.
- Good sampling techniques include the art of reducing all sources of error.

Definition

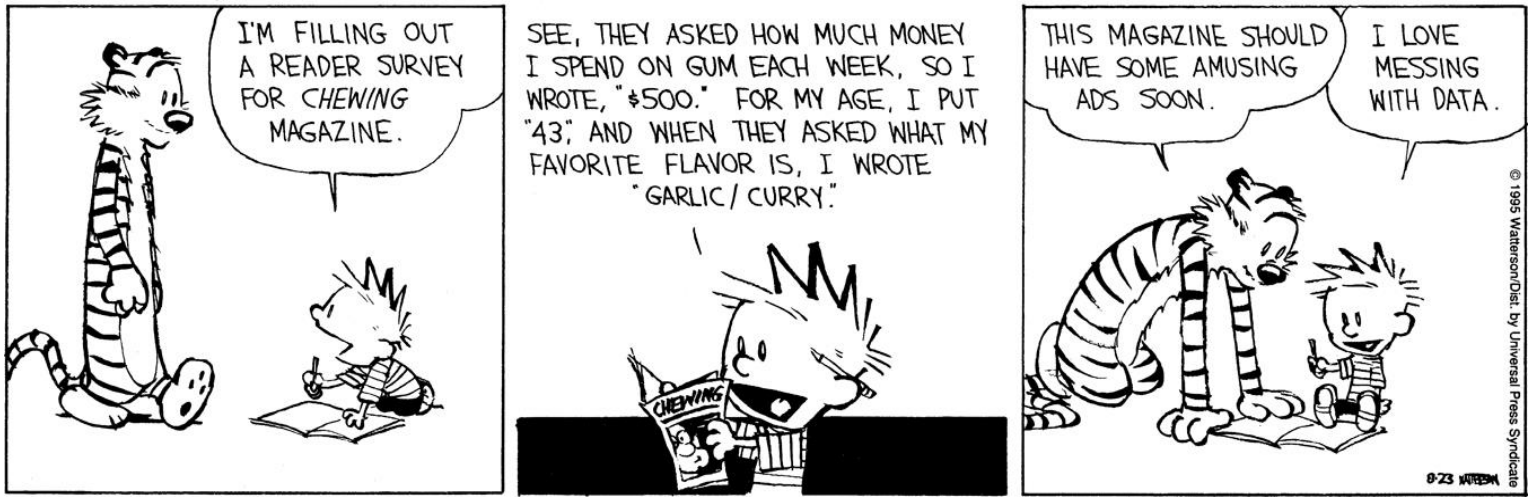
Undercoverage occurs when some groups in the population are left out of the process of choosing the sample.

Nonresponse occurs when an individual chosen for the sample can't be contacted or refuses to participate.

A systematic pattern of incorrect responses in a sample survey leads to **response bias**.

The **wording of questions** is the most important influence on the answers given to a sample survey.

Sample Surveys: What Can Go Wrong?



Definition

Undercoverage occurs when some groups in the population are left out of the process of choosing the sample.

Nonresponse occurs when an individual chosen for the sample can't be contacted or refuses to participate.

A systematic pattern of incorrect responses in a sample survey leads to **response bias**.

The **wording of questions** is the most important influence on the answers given to a sample survey.

Sections 2.1 & 2.2
Types of Samples and Surveys

Summary

In this section, we learned that…

- ✓ A **sample survey** selects a **sample** from the **population** of all individuals about which we desire information.
- ✓ **Random sampling** uses chance to select a sample.
- ✓ The basic random sampling method is a **simple random sample (SRS)**.
- ✓ To choose a **stratified random sample**, divide the population into **strata**, then choose a separate SRS from each stratum.
- ✓ To choose a **cluster sample**, divide the population into groups, or **clusters**. Randomly select some of the clusters for your sample.

Section 2.1 & 2.2
Samples and Surveys

Summary, con't

In this section, we learned that…

- ✓ Failure to use random sampling often results in **bias**, or systematic errors in the way the sample represents the population.
- ✓ **Voluntary response samples** and **convenience samples** are particularly prone to large bias.
- ✓ **Sampling errors** come from the act of choosing a sample. Random sampling error and **undercoverage** are common types of error.
- ✓ The most serious errors are **nonsampling errors.** Common types of sampling error include **nonresponse, response bias,** and **wording of questions.**

Looking Ahead in Chapter 2…

In the next Sections…

We'll learn how to produce data by designing experiments.

We'll learn about

- ✓ **Principles of Experimental Designs**
- ✓ **Various Types of Experiments**
- ✓ **Types of Variables**
- ✓ **Interpreting Results**