Chapter 1

Introduction to Statistics

§ 1.1

An Overview of Statistics

Data and Statistics

Data consists of information coming from observations, counts, measurements, or responses.

Statistics is the science of collecting, organizing, analyzing, and interpreting data in order to make decisions.

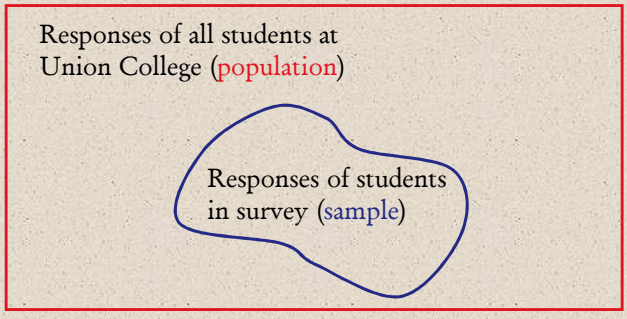
A **population** is the collection of *all* outcomes, responses, measurement, or counts that are of interest.

A **sample** is a subset of a population.

Populations & Samples

Example:

In a recent survey, 250 college students at Union College were asked if they smoked cigarettes regularly. 35 of the students said yes. Identify the population and the sample.



Parameters & Statistics

A **parameter** is a numerical description of a *population* characteristic.

A **statistic** is a numerical description of a *sample* characteristic.

Parameter → Population

Statistic → Sample

Parameters & Statistics

Example:

Decide whether the numerical value describes a population parameter or a sample statistic.

a.) A recent survey of a sample of 450 college students reported that the average weekly income for students is $325.

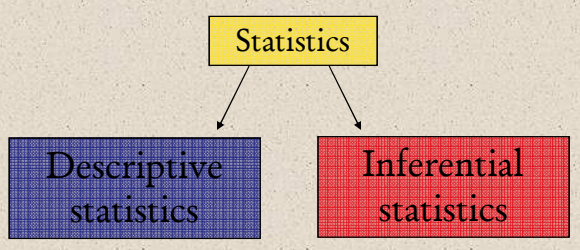
Because the average of $325 is based on a sample, this is a sample statistic.

b.) The average weekly income for all students is $405.

Because the average of $405 is based on a population, this is a population parameter.

Branches of Statistics

The study of statistics has two major branches: **descriptive statistics** and **inferential statistics**.



Involves the organization, summarization, and display of data.

Involves using a sample to draw conclusions about a population.

Descriptive and Inferential Statistics

Example:

In a recent study, volunteers who had less than 6 hours of sleep were four times more likely to answer incorrectly on a science test than were participants who had at least 8 hours of sleep. Decide which part is the descriptive statistic and what conclusion might be drawn using inferential statistics.

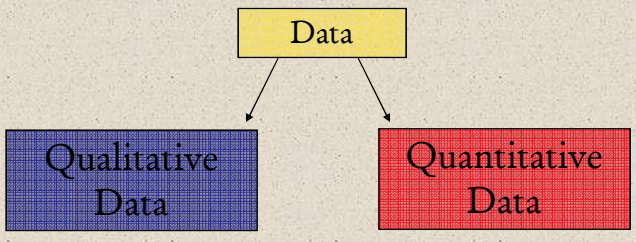
The statement "four times more likely to answer incorrectly" is a descriptive statistic. An inference drawn from the sample is that all individuals sleeping less than 6 hours are more likely to answer science question incorrectly than individuals who sleep at least 8 hours.

§ 1.2

Data Classification

Types of Data

Data sets can consist of two types of data: **qualitative data** and **quantitative data**.



Consists of attributes, labels, or nonnumerical entries.

Consists of numerical measurements or counts.

Qualitative and Quantitative Data

Example:

The grade point averages of five students are listed in the table. Which data are qualitative data and which are quantitative data?

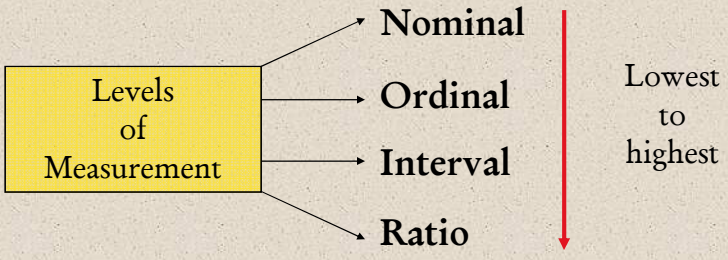
Student	GPA
Sally	3.22
Bob	3.98
Cindy	2.75
Mark	2.24
Kathy	3.84

Qualitative data

Quantitative data

Levels of Measurement

The level of measurement determines which statistical calculations are meaningful. The four levels of measurement are: **nominal**, **ordinal**, **interval**, and **ratio**.



Nominal Level of Measurement

Data at the **nominal level of measurement** are qualitative only.

Levels of Measurement → **Nominal**

Calculated using names, labels, or qualities. No mathematical computations can be made at this level.

- Colors in the US flag
- Names of students in your class
- Textbooks you are using this semester

Ordinal Level of Measurement

Data at the **ordinal level of measurement** are qualitative or quantitative.

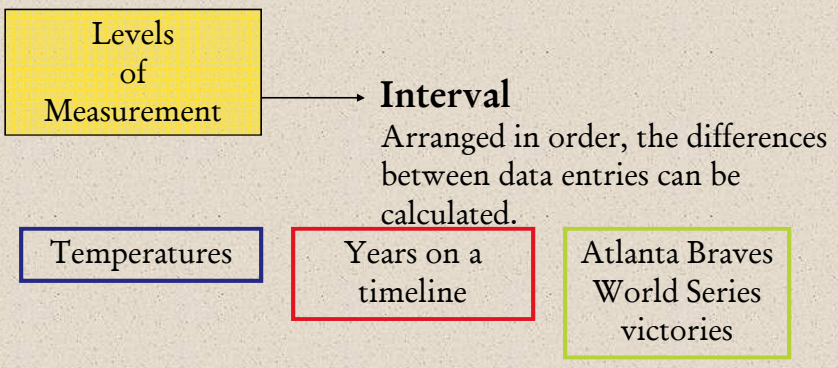
Levels of Measurement → **Ordinal**

Arranged in order, but differences between data entries are not meaningful.

- Class standings: freshman, sophomore, junior, senior
- Numbers on the back of each player's shirt
- Top 50 songs played on the radio

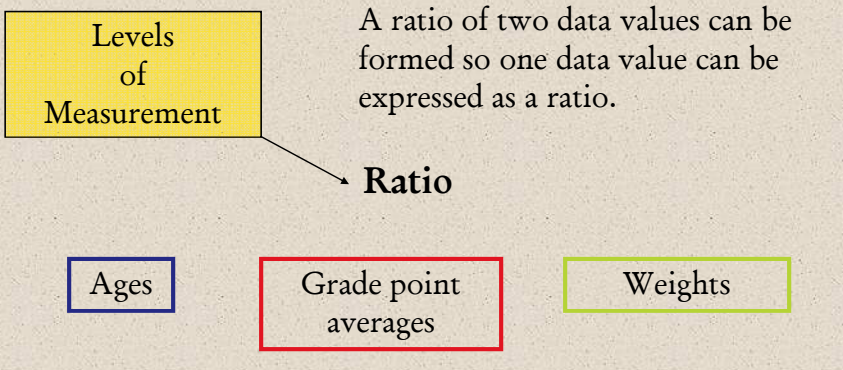
Interval Level of Measurement

Data at the **interval level of measurement** are quantitative. A zero entry simply represents a position on a scale; the entry is not an inherent zero.



Ratio Level of Measurement

Data at the **ratio level of measurement** are similar to the interval level, but a zero entry is meaningful.



Summary of Levels of Measurement

Level of measurement	Put data in categories	Arrange data in order	Subtract data values	Determine if one data value is a multiple of another
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

§ 1.3

Experimental Design

Designing a Statistical Study

GUIDELINES

1. Identify the variable(s) of interest (the focus) and the population of the study.
2. Develop a detailed plan for collecting data. If you use a sample, make sure the sample is representative of the population.
3. Collect the data.
4. Describe the data.
5. Interpret the data and make decisions about the population using inferential statistics.
6. Identify any possible errors.

Random Samples

Random samples are selected by using chance methods or random numbers.

One such method is to number each subject in the population. Then place numbered cards in a bowl, mix them thoroughly, and select as many cards as needed. The subjects whose numbers are selected constitute the sample. Since it is difficult to mix the cards thoroughly, there is a chance of obtaining a biased sample. For this reason, statisticians use another method of obtaining numbers. They generate random numbers with a computer or calculator.

Methods of Data Collection

In an **observational study**, a researcher observes and measures characteristics of interest of part of a population.

In an **experiment**, a treatment is applied to part of a population, and responses are observed.

A **simulation** is the use of a mathematical or physical model to reproduce the conditions of a situation or process.

A **survey** is an investigation of one or more characteristics of a population.

- A **census** is a measurement of an *entire* population.
- A **sampling** is a measurement of *part* of a population.

Stratified Samples

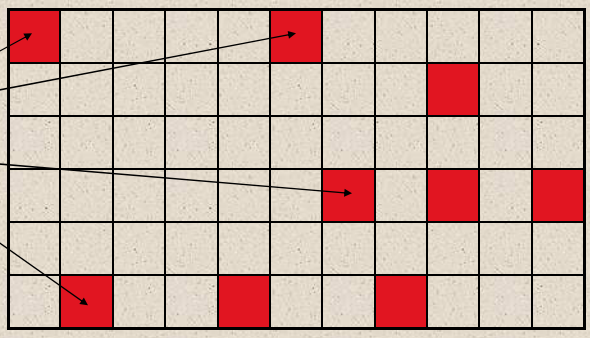
A **stratified sample** has members from each segment of a population. This ensures that each segment from the population is represented.

Freshmen Sophomores Juniors Seniors

Cluster Samples

A **cluster sample** has all members from randomly selected segments of a population. This is used when the population falls into naturally occurring subgroups.

All members in each selected group are used.



The city of Clarksville divided into city blocks.

Systematic Samples

A **systematic sample** is a sample in which each member of the population is assigned a number. A starting number is randomly selected and sample members are selected at regular intervals.

Every fourth member is chosen.

Summary of Sampling Methods

Random: Subjects are selected by random numbers.

Systematic: Subjects are selected by using every *kth number after the first subject is randomly selected from 1 through k.*

Stratified: Subjects are selected by dividing up the population into groups (strata), and subjects within groups are randomly selected.

Cluster: Subjects are selected by using an intact group that is representative of the population.

Identifying the Sampling Technique

Example continued:

You are doing a study to determine the number of years of education each teacher at your college has. Identify the sampling technique used if you select the samples listed.

1.) This is a cluster sample because each department is a naturally occurring subdivision.

2.) This is a stratified sample because the teachers are divided by department and some from each department are randomly selected.