Chapter 1

Introduction to Statistics

What is Statistics?

- The main objective is to make inferences about certain characteristics of a population
- Statistics is the science of 'good' decision making
- It is used in many disciplines, such as financial analysis, econometrics, auditing, production and operations, and marketing research.
- It provides knowledge and skills to interpret
- A typical Statistics course is intended for
  - descriptive statistics (collection, description, analysis, and summary of data),
  - probability, and the binomial and normal distributions,
  - test of hypotheses and confidence intervals,
  - linear regression, and correlation.

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.

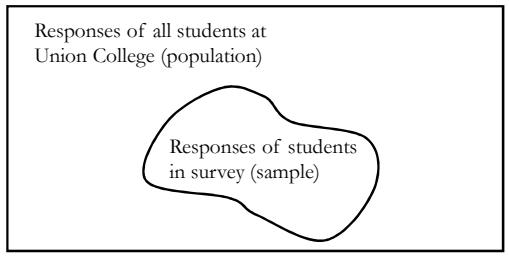
Common Statistical Terminology

An Experiment: An experiment is a process whose outcome is not known in advance with certainty.

Statistical Experiment: An experiment in general is an operation in which one chooses the values of some variables and measures the values of other variables, as in physics.

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.



6

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**

7

Parameters & Statistics

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

a) A recent survey of approximately 400,000 employers reported that the average starting salary for marketing majors is $53,400.

Because the average of $53,400 is based on a subset of the population, it is a sample statistic.

b.) The sophomore class at a university has an average Statistics score of 55.

Because the average Statistics score of 55 is based on the entire sophomore class, it is a population parameter

8

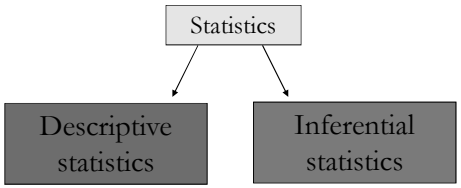
Statistical Notations

Greek Letters Commonly Used as Statistical Notations										
alpha	beta	ki-sqre	delta	mu	nu	pi	rho	sigma	tau	theta
α	β	χ^2	δ	μ	ν	π	ρ	σ	τ	θ

9

Branches of Statistics

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



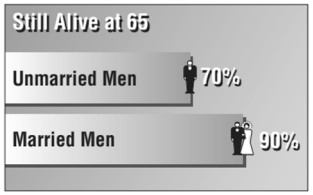
Involves the organization, summarization, and display of data.

Involves using a sample to draw conclusions about a population.

Descriptive and Inferential Statistics

Example:

A large sample of men, aged 48, was studied for 18 years. For unmarried men, approximately 70% were alive at age 65. For married men, 90% were alive at age 65.

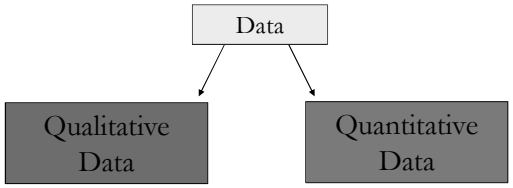


§ 1.2

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

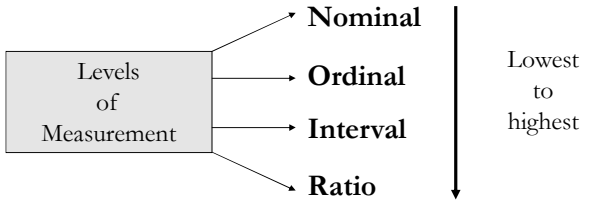
Example:
The grade point averages of five students are listed in the table. Which data are qualitative data and which are quantitative data?

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

Qualitative data ← → Quantitative data

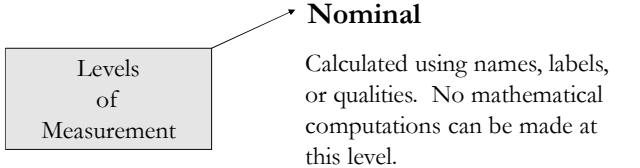
Levels of Measurement

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.



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

- Colors in the Turkish 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.

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.

Levels of Measurement → **Interval**

Arranged in order, the differences between data entries can be calculated.

- Temperatures
- Years on a timeline
- Atlanta Braves World Series victories

Ratio Level of Measurement

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

Levels of Measurement → **Ratio**

A ratio of two data values can be formed so one data value can be expressed as a ratio.

- Ages
- Grade point averages
- Weights

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.

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.

EXPERIMENTAL DESIGN

Three key elements of a well-designed experiment are *control*, *randomization*, and *replication*.

DEFINITION

A **confounding variable** occurs when an experimenter cannot tell the difference between the effects of different factors on the variable.

DEFINITION

Blinding is a technique where the subjects do not know whether they are receiving a treatment or a placebo. In a **double-blind experiment,** neither the experimenter nor the subjects know if the subjects are receiving a treatment or a placebo. The experimenter is informed after all the data have been collected. This type of experimental design is preferred by researchers.

DEFINITION

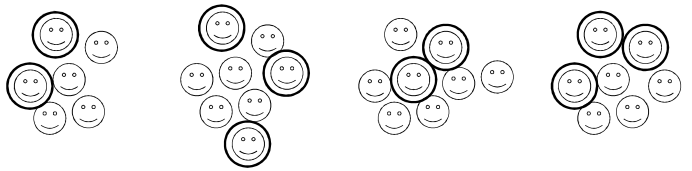
Randomization is a process of randomly assigning subjects to different treatment groups.

DEFINITION

Replication is the repetition of an experiment under the same or similar conditions.

Stratified Samples

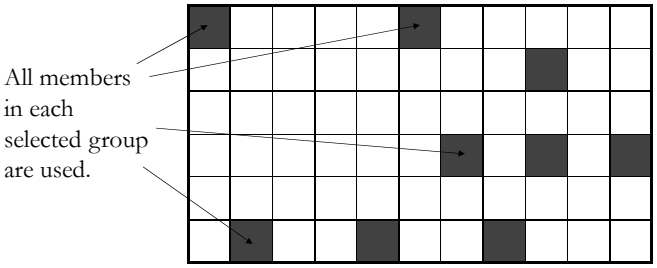
A **stratified sample** has members from each segment of a population. This ensures that each segment from the population is represented.



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.



The city of Konya 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.

Convenience Samples

A **convenience sample** consists only of available members of the population.

Example:

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.) You randomly select two different departments and survey each teacher in those departments.

2.) You select only the teachers you currently have this semester.

3.) You divide the teachers up according to their department and then choose and survey some teachers in each department.

Continued.

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 convenience sample because you are using the teachers that are readily available to you.

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