Eighth Edition

Elementary Statistics

A Step by Step Approach

Bluman

Mc Graw Hill

1

The Nature of Probability and Statistics

Objectives

After completing this chapter, you should be able to

1. Demonstrate knowledge of statistical terms.

2. Differentiate between the two branches of statistics.

3. Identify types of data.

4. Identify the measurement level for each variable.

5. Identify the four basic sampling techniques.

6. Explain the difference between an observational and an experimental study.

7. Explain how statistics can be used and misused.

8. Explain the importance of computers and calculators in statistics.

Outline

Introduction

- **<u>Statistics</u>** is the science of conducting studies to

    collect,

    organize,

    summarize,

    analyze, and

    draw conclusions from data.
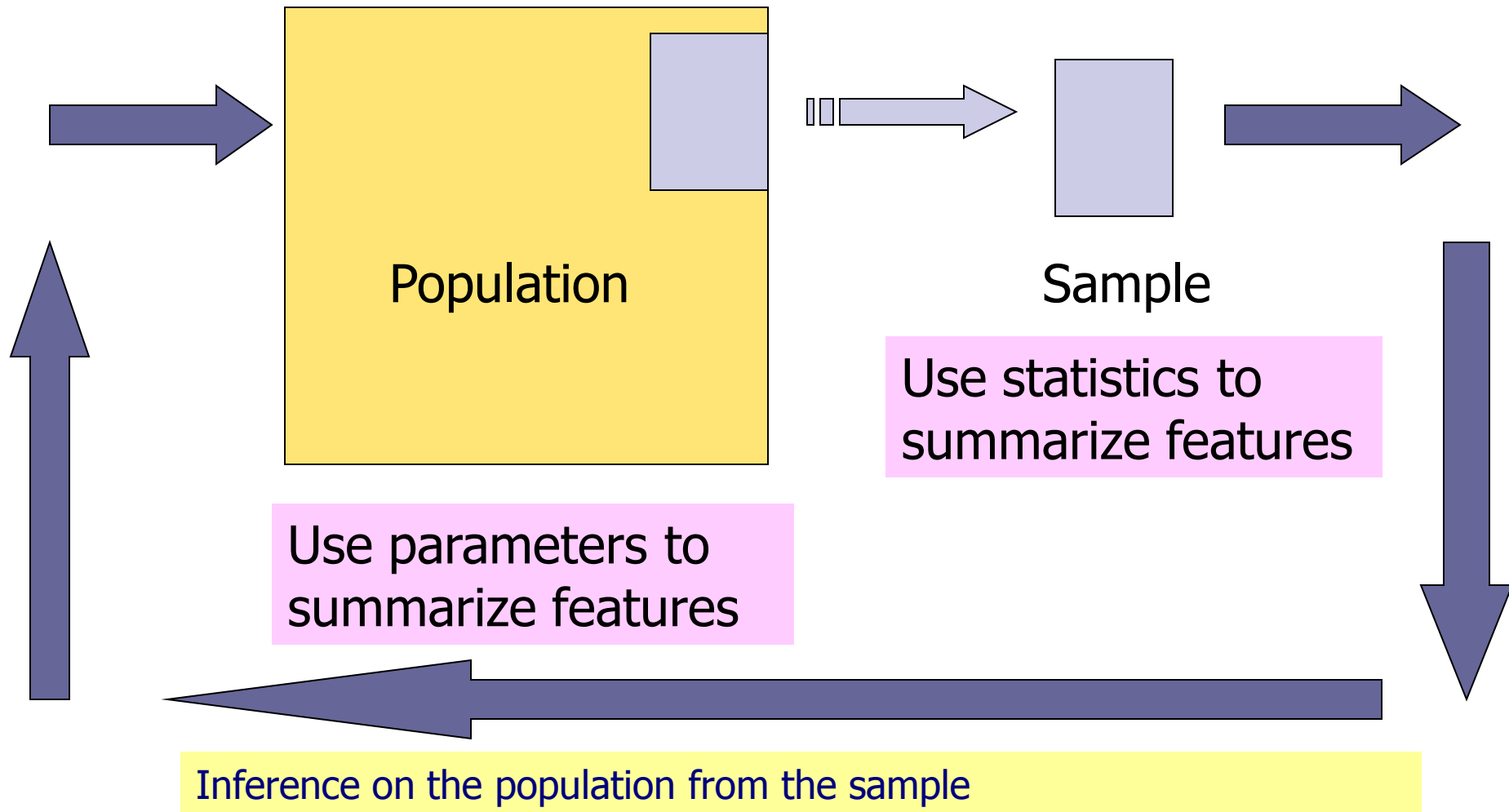
1-1 Descriptive and Inferential Statistics

- A **variable** is a characteristic or attribute that can assume different values.

- Variables whose values are determined by chance are called **random variables**.

- The values that a variable can assume are called **data**.

- A **population** consists of all subjects (human or otherwise) that are studied.

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

E.g. A data set in table form (columns and rows):

id	gender	age
1	male	20
2	male	25
3	female	30
4	male	23
5	female	30

Population and Sample

Population

Sample

Use statistics to summarize features

Use parameters to summarize features

Inference on the population from the sample

Census versus Sample

- ❖ **Census**

  Collection of data from *every* member of a population
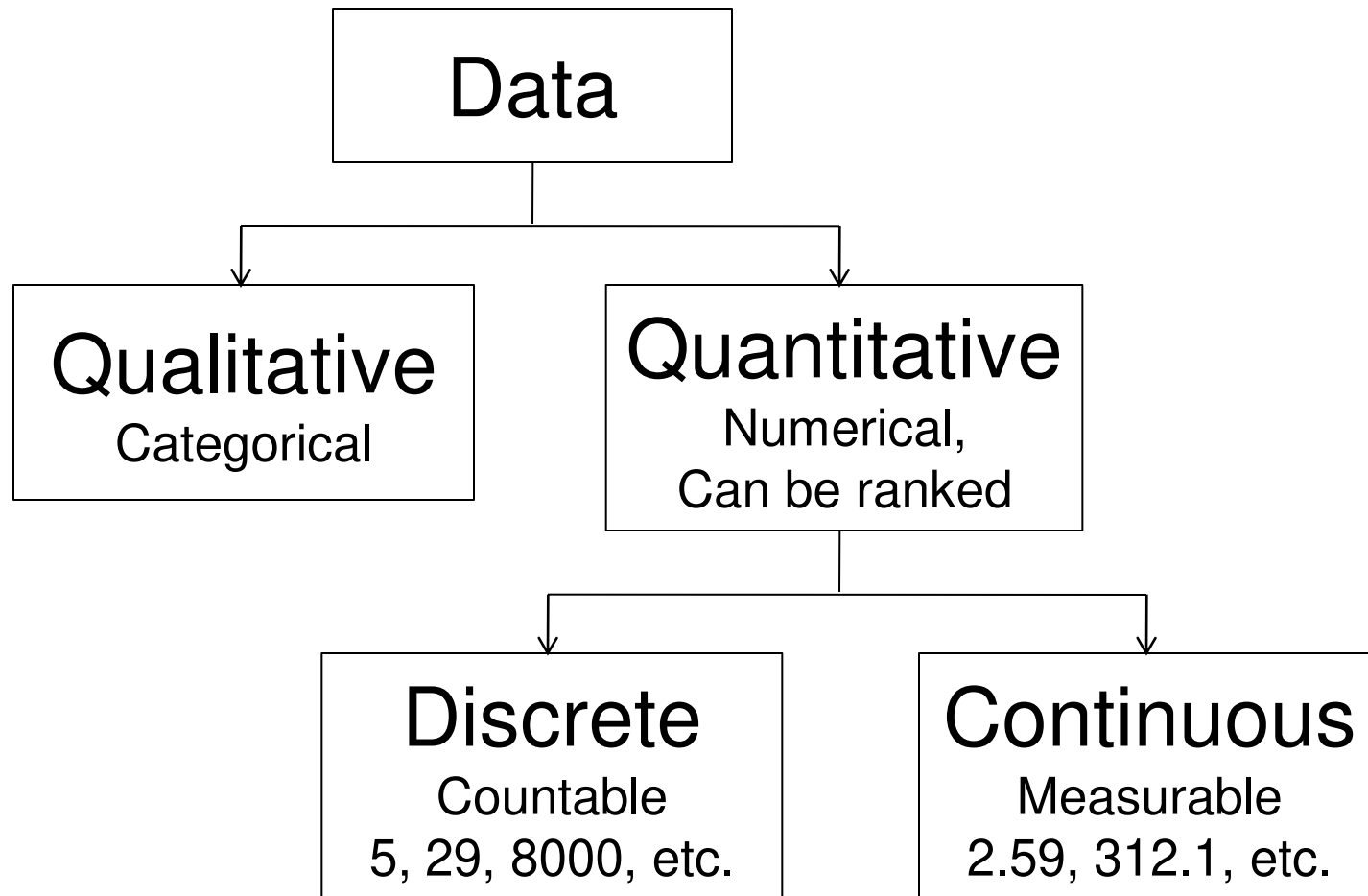
- ❖ **Sample**

  Subcollection of members selected from a population

1-1 Descriptive and Inferential Statistics

- **Descriptive statistics** consists of the collection, organization, summarization, and presentation of data.

- **Inferential statistics** consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships among variables, and making predictions.

1-2 Variables and Types of Data

```
                    ┌─────────────┐
                    │    Data     │
                    └─────────────┘
                      │         │
          ┌───────────┘         └───────────┐
          ▼                                 ▼
┌──────────────────┐            ┌──────────────────────┐
│   Qualitative    │            │    Quantitative      │
│   Categorical    │            │    Numerical,        │
│                  │            │    Can be ranked     │
└──────────────────┘            └──────────────────────┘
                                      │          │
                          ┌───────────┘          └──────────┐
                          ▼                                 ▼
                ┌──────────────────┐            ┌──────────────────────┐
                │    Discrete      │            │    Continuous        │
                │    Countable     │            │    Measurable        │
                │   5, 29, 8000,   │            │   2.59, 312.1, etc.  │
                │      etc.        │            │                      │
                └──────────────────┘            └──────────────────────┘
```

1-2 Variables and Types of Data

Another way to classify data is to use levels of measurement. Four of these levels are discussed in the following slides.

1-2 Variables and Types of Data

Levels of Measurement

1. **Nominal** – categorical (names)

2. **Ordinal** – nominal, plus can be ranked (order)

3. **Interval** – ordinal, plus intervals are consistent (differences but no natural zero starting point)

4. **Ratio** – interval, plus ratios are consistent, true zero (difference <u>and</u> a natural zero starting point)

NOTE: To remember these four terms, think of the French word NOIR. Noir in French is the name of the color black. But the letters of the word represent the first letter of the above four terms. It will help you remember them.

1-2 Variables and Types of Data

Determine the measurement level.

Variable	Nominal	Ordinal	Interval	Ratio	Level
Hair Color	Yes	No			Nominal
Zip Code	Yes	No			Nominal
Letter Grade	Yes	Yes	No		Ordinal
ACT Score	Yes	Yes	Yes	No	Interval
Height	Yes	Yes	Yes	Yes	Ratio
Age	Yes	Yes	Yes	Yes	Ratio
Temperature (F)	Yes	Yes	Yes	No	Interval

1-3 Data Collection and Sampling Techniques

Some Sampling Techniques

-Probability sampling (random sampling):

- **Random** – random number generator
- **Systematic** – every k^{th} subject
- **Stratified** – divide population into group called "strata"
- **Cluster** – use intact groups

-Non-probability sampling: *cannot* be used to infer from the sample to the general population.

- **Convenient** – mall surveys

Table 1-3		Random Numbers										
79	41	71	93	60	35	04	67	96	04	79	10	86
26	52	53	13	43	50	92	09	87	21	83	75	17
18	13	41	30	56	20	37	74	49	56	45	46	83
19	82	02	69	34	27	77	34	24	93	16	77	00
14	57	44	30	93	76	32	13	55	29	49	30	77
29	12	18	50	06	33	15	79	50	28	50	45	45
01	27	92	67	93	31	97	55	29	21	64	27	29
55	75	65	68	65	73	07	95	66	43	43	92	16
84	95	95	96	62	30	91	64	74	83	47	89	71
62	62	21	37	82	62	19	44	08	64	34	50	11
66	57	28	69	13	99	74	31	58	19	47	66	89
48	13	69	97	29	01	75	58	05	40	40	18	29
94	31	73	19	75	76	33	18	05	53	04	51	41
00	06	53	98	01	55	08	38	49	42	10	44	38
46	16	44	27	80	15	28	01	64	27	89	03	27
77	49	85	95	62	93	25	39	63	74	54	82	85
81	96	43	27	39	53	85	61	12	90	67	96	02
40	46	15	73	23	75	96	68	13	99	49	64	11

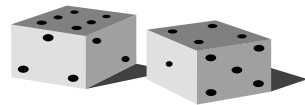Q: Select random sample of 15 subjects out of 85 subjects:
   Skip the number > 85 or already been selected.
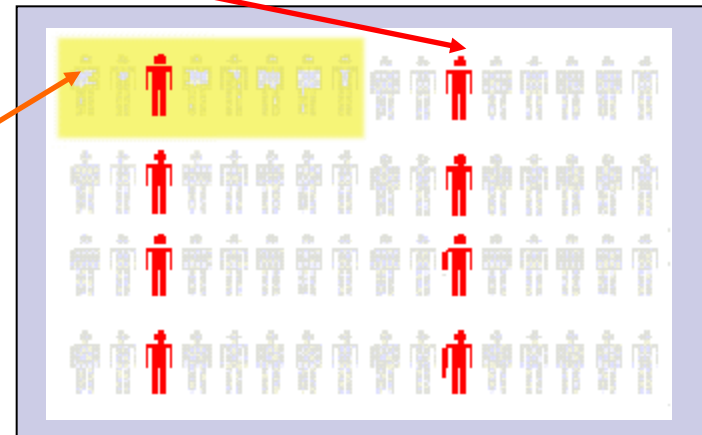A: 12, 27, 75, 62, 57, 13, 31, 06, 16, 49, 46, 71, 53, 41, 02

Systematic Samples

- Decide on Sample Size: n

- Divide Frame of N individuals into Groups of k Individuals: k=N/n

- Randomly Select One Individual from the 1st Group

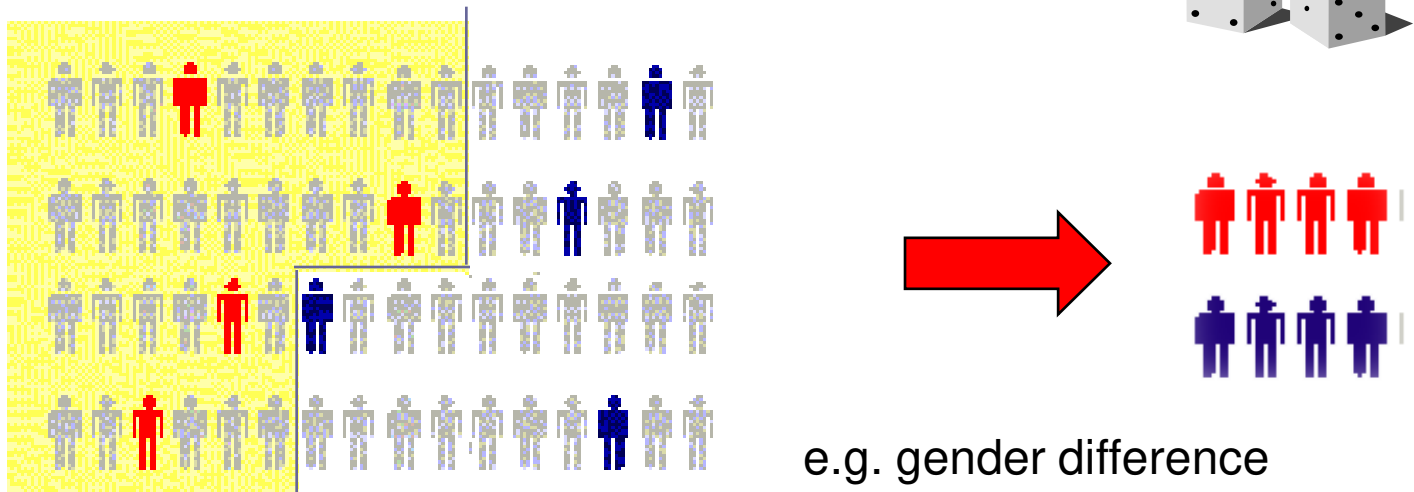- Select Every k-th Individual Thereafter

N = 64

n = 8

k = 8

First Group

Stratified Samples

- Population Divided into 2 or More Groups According to Some Common Characteristic

- Simple Random Sample Selected from Each Group

- The Two or More Samples are Combined into One

e.g. gender difference

Cluster Samples

- Population Divided into Several "Clusters," Each Representative of the Population

- A Random Sampling of Clusters is Taken

- All Items in the Selected Clusters are Studied

Randomly selected 2 clusters

Population divided into 4 clusters
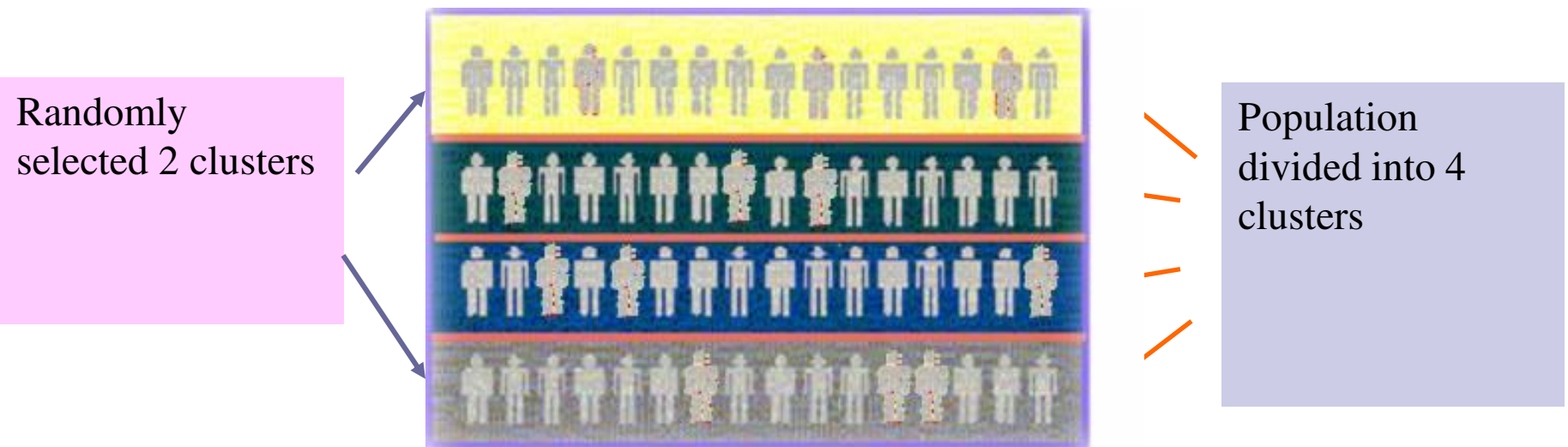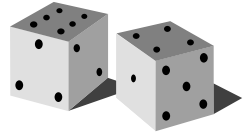
1-4 Observational and Experimental Studies

- In an **observational study**, the researcher merely observes and tries to draw conclusions based on the observations.

- The researcher manipulates the **independent (explanatory) variable** and tries to determine how the manipulation influences the **dependent (outcome) variable** in an **experimental study**.

- A **confounding variable** influences the dependent variable but cannot be separated from the independent variable.

1-5 Uses and Misuses of Statistics

- **Suspect Samples**

  - ☐ **Is the sample large enough?**

  - ☐ **How was the sample selected?**

  - ☐ **Is the sample representative of the population?**

- **Ambiguous Averages**

  - ☐ **What particular measure of average was used and why?**

1-5 Uses and Misuses of Statistics

- **Changing the Subject**
  - **Are different values used to represent the same data?**

- **Detached Statistics**
  - **One third fewer calories.......than what?**

- **Implied Connections**
  - **Studies *suggest* that *some* people *may* understand what this statement means.**

1-5 Uses and Misuses of Statistics

- **Misleading Graphs**

  - ☐ **Are the scales for the x-axis and y-axis appropriate for the data?**

- **Faulty Survey Questions**

  - ☐ **Do you feel that statistics teachers should be paid higher salaries?**

  - ☐ **Do you favor increasing tuition so that colleges can pay statistics teachers higher salaries?**

1-6 Computers and Calculators

- **TI-83/84**
- **Microsoft Excel**
- **Microsoft Excel with MegaStat**
- **Minitab**
- **SAS (Statistical Analysis Software)**
- **SPSS (Statistical Package for the Social Sciences)**

Highlights

- Statistical studies use <u>random variables</u> and values of the variable are called data.

- Statistics can be divided into two main branches, <u>descriptive and inferential</u> (based on probability theory) statistics.

- A very important requirement of a statistical study is to define the <u>population</u> and select a random <u>sample.</u>

- Various type of data and measurement levels.