Chapter

Introduction to Statistics

Chapter Outline

- 1. An Overview of Statistics
- 2. Data Classification
- 3. Experimental Design

Section 1

An Overview of Statistics

Section 1. Objectives

- Define statistics
- Distinguish between a population and a sample
- Distinguish between a parameter and a statistic
- Distinguish between descriptive statistics and inferential statistics

What is Data?

Data

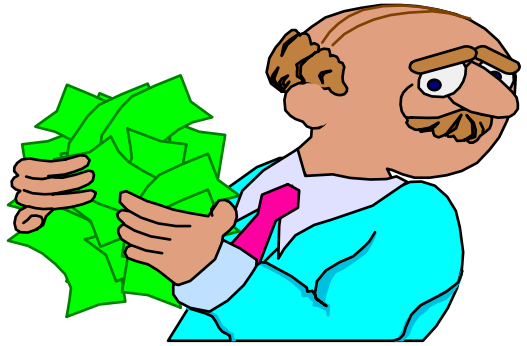
Consist of information coming from observations, counts, measurements, or responses.

- "People who eat three daily servings of whole grains have been shown to reduce their risk of…stroke by 37%."
- "Over 50 percent of the people killed on roads were pedestrians".

What is Statistics?

Statistics

The science of collecting, organizing, analyzing, and interpreting data in order to make decisions.



Data Sets

Population

The collection of ***all*** outcomes, responses, measurements, or counts that are of interest.

Sample

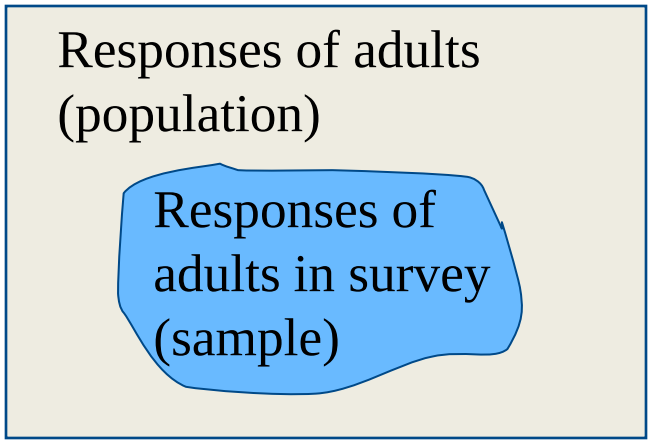
A subset of the population.

Example: Identifying Data Sets

In a recent survey, 1708 adults were asked of a country if they think global warming is a problem that requires immediate government action. Nine hundred thirty-nine of the adults said yes. Identify the population and the sample. Describe the data set.

Solution: Identifying Data Sets

- The population consists of the responses of all adults of a country?
- The sample consists of the responses of the 1708 adults of the country in the survey.
- The sample is a subset of the responses of all adults of the country.
- The data set consists of 939 yes's and 769 no's.

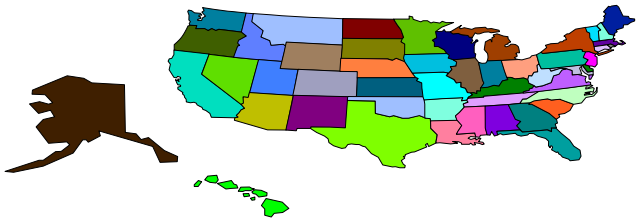


Parameter and Statistic

Parameter

A number that describes a population characteristic.

Average age of all people in the Bangladesh



Statistic

A number that describes a sample characteristic.

Average age of people from a sample of three district .

Example: Distinguish Parameter and Statistic

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

1. A recent survey of a sample of MBAs reported that the average salary for an MBA is more than 40,000 Tk.

Solution:

Sample statistic (the average 40,000 Tk. is based on a subset of the population)

Example: Distinguish Parameter and Statistic

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

2. Starting salaries for the 667 MBA graduates from the Dhaka University Business faculty increased 8.5% from the previous year.

Solution:

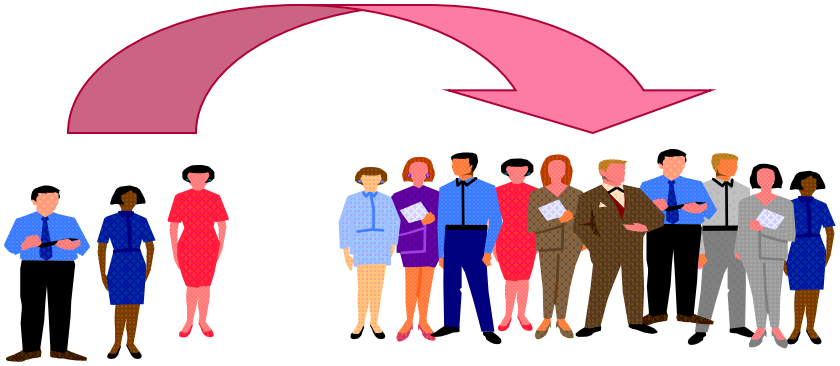
Population parameter (the percent increase of 8.5% is based on all 667 graduates' starting salaries)

Branches of Statistics

Descriptive Statistics
Involves organizing, summarizing, and displaying data.

e.g. Tables, charts, averages

Inferential Statistics
Involves using ***sample data*** to draw conclusions about a ***population.***



Section 1 Summary

- Defined statistics
- Distinguished between a population and a sample
- Distinguished between a parameter and a statistic
- Distinguished between descriptive statistics and inferential statistics

Section 2

Data Classification

Section 2 Objectives

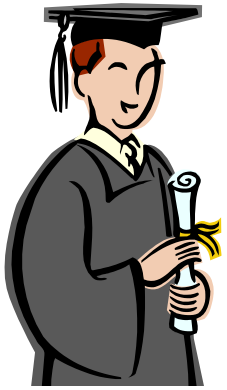
- Distinguish between qualitative data and quantitative data
- Classify data with respect to the four levels of measurement

Types of Data

Qualitative Data

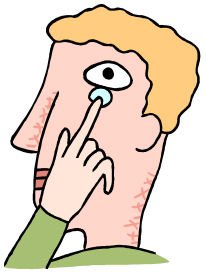
Consists of attributes, labels, or nonnumerical entries.

Major



Place of birth

Eye color



Types of Data

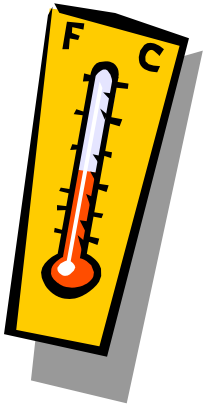
Quantitative data

Numerical measurements or counts.

Age

Weight of a letter

Temperature



Example: Classifying Data by Type

The base prices of several vehicles are shown in the table. Which data are qualitative data and which are quantitative data?

Best HP Laptop List in December, 2020	Latest Price
HP EliteBook 840 G5 Core i5 8th Gen 8GB RAM 256GB SSD	৳ 56,000
HP EliteBook 840 G2 i5 5th Gen 4GB RAM **Laptop**	৳ 26,500
HP Probook 430 G2 Core i5 5th Gen 4GB RAM 13" **Laptop**	৳ 27,500
HP ProBook 440 G3 Core i5 6th Gen **Laptop**	৳ 30,000

Solution: Classifying Data by Type

Best HP Laptop List in December, 2020	Latest Price
HP EliteBook 840 G5 Core i5 8th Gen 8GB RAM 256GB SSD	৳ 56,000
HP EliteBook 840 G2 i5 5th Gen 4GB RAM **Laptop**	৳ 26,500
HP Probook 430 G2 Core i5 5th Gen 4GB RAM 13" **Laptop**	৳ 27,500
HP ProBook 440 G3 Core i5 6th Gen **Laptop**	৳ 30,000

Qualitative Data (Names of models are nonnumerical entries)

Quantitative Data (prices of laptops models are numerical entries)

Levels of Measurement

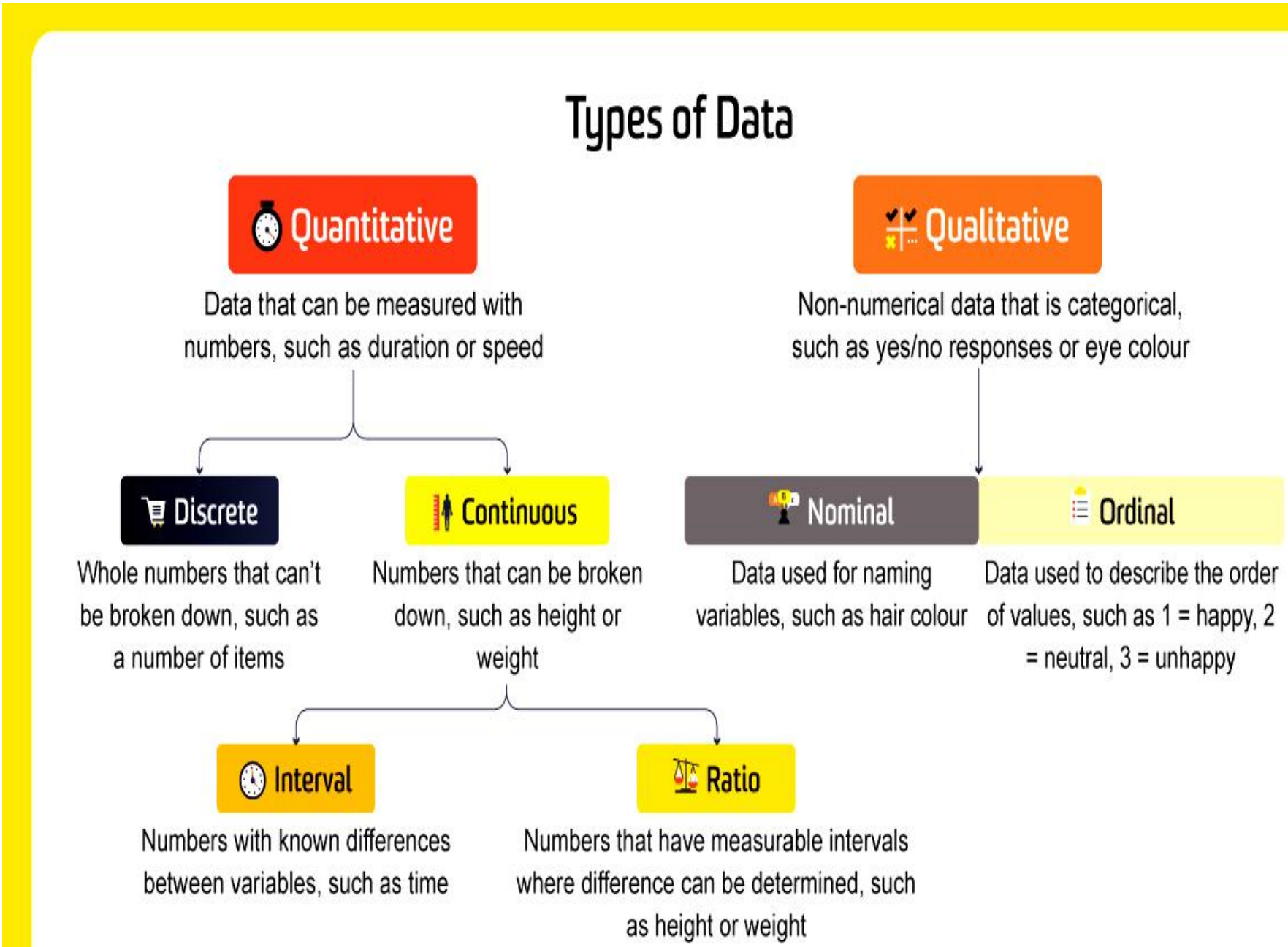
Nominal level of measurement

- Qualitative data only
- Categorized using names, labels, or qualities
- No mathematical computations can be made

Ordinal level of measurement

- Qualitative or quantitative data
- Data can be arranged in order
- Differences between data entries is not meaningful

Types of Data



Levels of Measurement

Interval level of measurement

- Quantitative data
- Data can ordered
- Differences between data entries is meaningful
- Zero represents a position on a scale (not an inherent zero – zero does not imply “none”)

Levels of Measurement

Ratio level of measurement

- Similar to interval level
- Zero entry is an inherent zero (implies “none”)
- A ratio of two data values can be formed
- One data value can be expressed as a multiple of another

Example: Classifying Data by Level

Two data sets are shown. Which data set consists of data at the interval level? Which data set consists of data at the ratio level?

New York Yankees' World Series Victories (Years)
1923, 1927, 1928, 1932, 1936, 1937, 1938, 1939, 1941, 1943, 1947, 1949, 1950, 1951, 1952, 1953, 1956, 1958, 1961, 1962, 1977, 1978, 1996, 1998, 1999, 2000

2003 National League Home Run Totals (by Team)	
Baltimore	164
Boston	192
Chicago	236
Cleveland	196
Detroit	203
Kansas City	124
Los Angeles	159
Minnesota	143
New York	210
Oakland	175
Seattle	172
Tampa Bay	190
Texas	183
Toronto	199

Solution: Classifying Data by Level

New York Yankees' World Series Victories (Years)
1923, 1927, 1928, 1932, 1936, 1937, 1938, 1939, 1941, 1943, 1947, 1949, 1950, 1951, 1952, 1953, 1956, 1958, 1961, 1962, 1977, 1978, 1996, 1998, 1999, 2000

Interval level (Quantitative data. Can find a difference between two dates, but a ratio does not make sense.)

2003 National League Home Run Totals (by Team)	
Baltimore	164
Boston	192
Chicago	236
Cleveland	196
Detroit	203
Kansas City	124
Los Angeles	159
Minnesota	143
New York	210
Oakland	175
Seattle	172
Tampa Bay	190
Texas	183
Toronto	199

Ratio level (Can find differences and write ratios.)

Summary of Four 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

Section 2 Summary

- Distinguished between qualitative data and quantitative data
- Classified data with respect to the four levels of measurement

Section 3

Experimental Design

Section 3 Objectives

- Discuss how to design a statistical study
- Discuss data collection techniques
- Discuss how to design an experiment
- Discuss sampling techniques

Designing a Statistical Study

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 using descriptive statistics techniques.
5. Interpret the data and make decisions about the population using inferential statistics.
6. Identify any possible errors.

Data Collection

Observational study

- A researcher observes and measures characteristics of interest of part of a population.

- Researchers observed and recorded the mouthing behavior on nonfood objects of children up to three years old.

Data Collection

Experiment

- A treatment is applied to part of a population and responses are observed.

- An experiment was performed in which diabetics took cinnamon extract daily while a control group took none. After 40 days, the diabetics who had the cinnamon reduced their risk of heart disease while the control group experienced no change.

Data Collection

Simulation

- Uses a mathematical or physical model to reproduce the conditions of a situation or process.
- Often involves the use of computers.

- Automobile manufacturers use simulations with dummies to study the effects of crashes on humans.

Data Collection

Survey

- An investigation of one or more characteristics of a population.
- Commonly done by interview, mail, or telephone.

- A survey is conducted on a sample of female physicians to determine whether the primary reason for their career choice is financial stability.

Example: Methods of Data Collection

Consider the following statistical studies. Which method of data collection would you use to collect data for each study?

1. A study of the effect of changing flight patterns on the number of airplane accidents.

Solution:

Simulation (It is impractical to create this situation)

Example: Methods of Data Collection

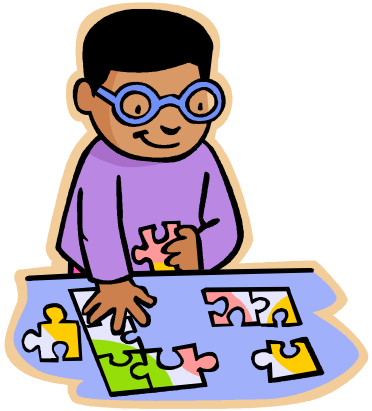
2. A study of the effect of eating oatmeal on lowering blood pressure.

Solution:

Experiment (Measure the effect of a treatment – eating oatmeal)

Example: Methods of Data Collection

3. A study of how fourth grade students solve a puzzle.



Solution:

Observational study (observe and measure certain characteristics of part of a population)

Example: Methods of Data Collection

4. A study of U.S. residents' approval rating of the U.S. president.

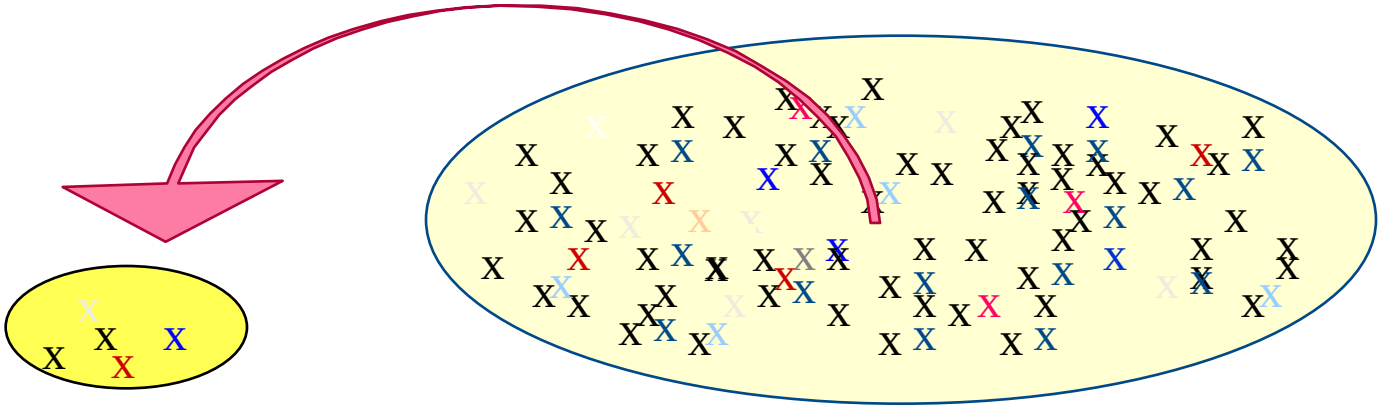
Solution:

Survey (Ask "Do you approve of the way the president is handling his job?")

Sampling Techniques

Simple Random Sample

Every possible sample of the same size has the same chance of being selected.



Simple Random Sample

- Random numbers can be generated by a random number table, a software program or a calculator.
- Assign a number to each member of the population.
- Members of the population that correspond to these numbers become members of the sample.

Example: Simple Random Sample

There are 731 students currently enrolled in statistics at your school. You wish to form a sample of eight students to answer some survey questions. Select the students who will belong to the simple random sample.

- Assign numbers 1 to 731 to each student taking statistics.
- On the table of random numbers, choose a starting place at random (suppose you start in the third row, second column.)

Solution: Simple Random Sample

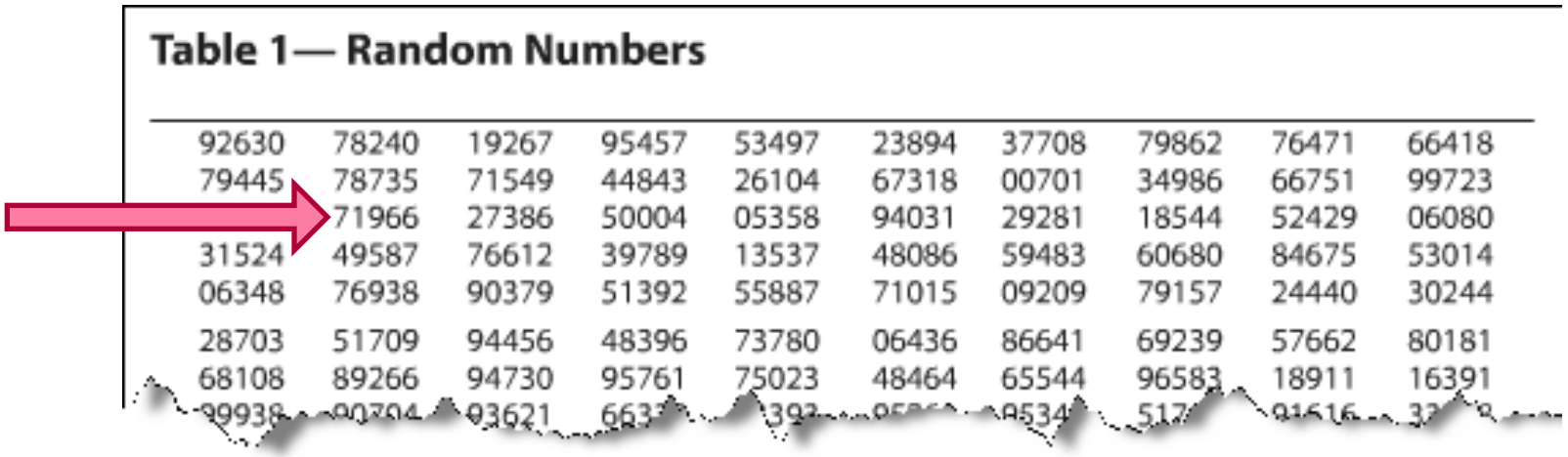


Table 1— Random Numbers

92630	78240	19267	95457	53497	23894	37708	79862	76471	66418
79445	78735	71549	44843	26104	67318	00701	34986	66751	99723
71966	27386	50004	05358	94031	29281	18544	52429	06080	
31524	49587	76612	39789	13537	48086	59483	60680	84675	53014
06348	76938	90379	51392	55887	71015	09209	79157	24440	30244
28703	51709	94456	48396	73780	06436	86641	69239	57662	80181
68108	89266	94730	95761	75023	48464	65544	96583	18911	16391
99938	90704	93621	6633	393		0534	517	91616	32

- Read the digits in groups of three
- Ignore numbers greater than 731

719 | 66 2 | 738 | 6 50 | 004 | 053 | 58 9 | 403 | 1 29 | 281 | 185 | 44

The students assigned numbers 719, 662, 650, 4, 53, 589, 403, and 129 would make up the sample.

Other Sampling Techniques

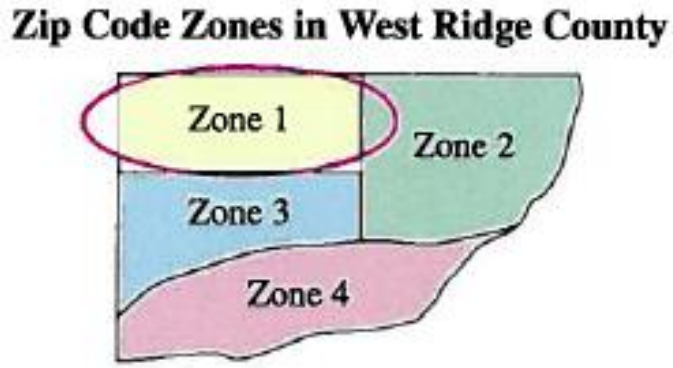
Stratified Sample

- Divide a population into groups (strata) and select a random sample from each group.
- To collect a stratified sample of the number of people who live in West Ridge County households, you could divide the households into socioeconomic levels and then randomly select households from each level.

Other Sampling Techniques

Cluster Sample

- Divide the population into groups (clusters) and select all of the members in one or more, but not all, of the clusters.
- In the West Ridge County example you could divide the households into clusters according to zip codes, then select all the households in one or more, but not all, zip codes.

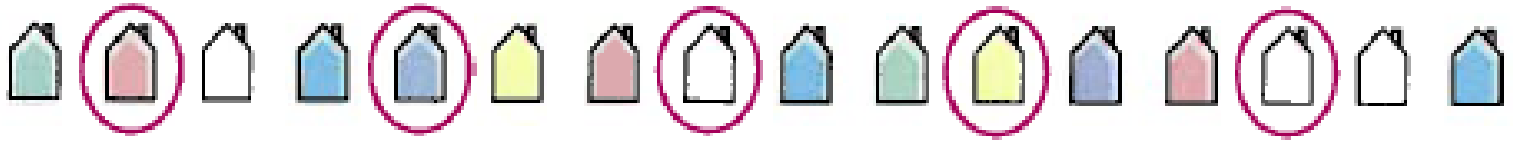


Other Sampling Techniques

Systematic Sample

- Choose a starting value at random. Then choose every k^{th} member of the population.

- In the West Ridge County example you could assign a different number to each household, randomly choose a starting number, then select every 100^{th} household.



Example: Identifying Sampling Techniques

You are doing a study to determine the opinion of students at your school regarding stem cell research. Identify the sampling technique used.

1. You divide the student population with respect to majors and randomly select and question some students in each major.

Solution:

Stratified sampling (the students are divided into strata (majors) and a sample is selected from each major)

Example: Identifying Sampling Techniques

2. You assign each student a number and generate random numbers. You then question each student whose number is randomly selected.

Solution:

Simple random sample (each sample of the same size has an equal chance of being selected and each student has an equal chance of being selected.)

Section 3 Summary

- Discussed how to design a statistical study
- Discussed data collection techniques
- Discussed how to design an experiment
- Discussed sampling techniques