Introduction to Statistics

Objectives:

After studying this chapter, the student will:

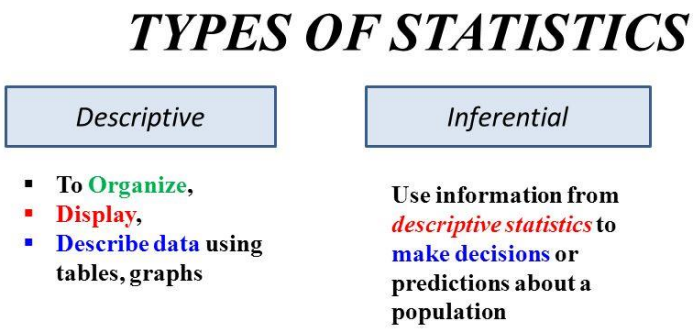
1. Demonstrate knowledge of all statistical terms.
2. Differentiate between the two branches of statistics.
3. Identify types of data.
4. Identify data level of measurement.
5. Differentiate between population and sample data.
6. Identify the four basic sampling techniques.

Some Basic Concepts:

Statistics

Statistics is a field of study concerned with

- collection, organization, summarization and analysis of data (Descriptive Statistics)
- drawing inferences about a body of data when only a part of the data is observed (Inferential Statistics)



Dr. Monjed H. Samuh

Data

Data: the raw material of statistics.

Sources of data: Data are obtained from

- Routinely kept records (Hospital and medical records)
- Surveys (Questionnaires)
- Experiments
- Reports (Published records)

Biostatistics

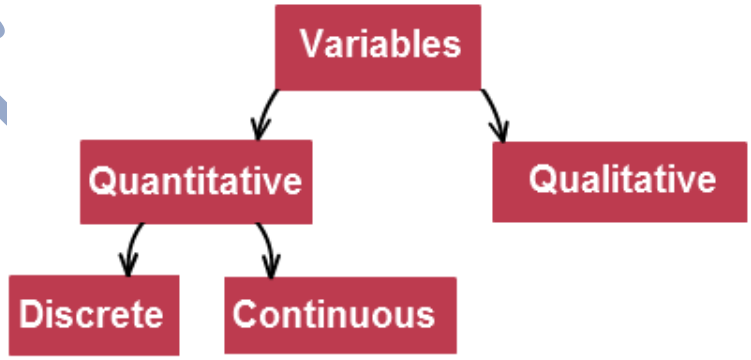
Biostatistics is statistics applied to biological sciences and medicine.

Examples: Blood Type, Heart Rate, Hemoglobin Level, Weight of Infant.

Variable

A variable is an object, characteristic or property that can have different values.

- A quantitative variable can be measured in some way.
- A qualitative variable is characterized by its inability to be measured but it can be sorted into categories.



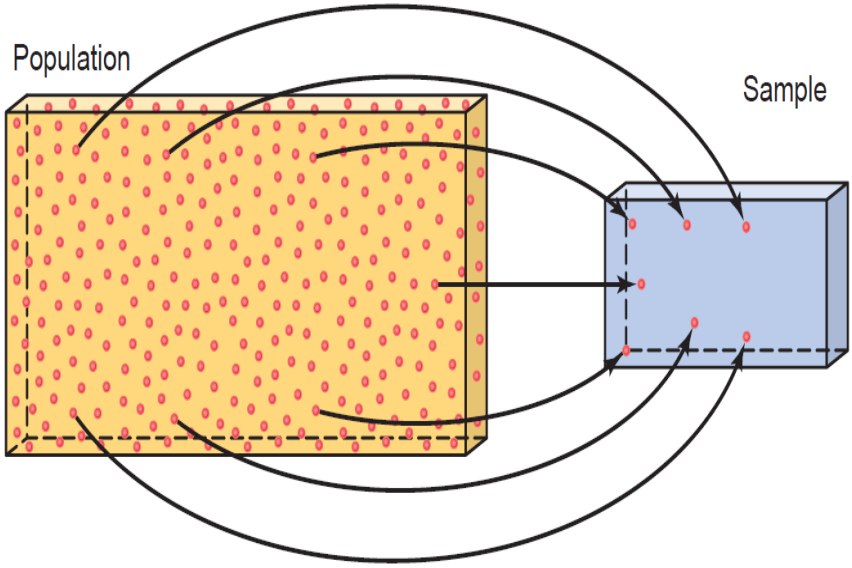
Dr Monjed H. Samuh

Question: **State some examples of quantitative and qualitative variables.**

Population and sample

A **population** is a complete set of elements (persons or objects) for which we have interest.

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



Statistical inference

Statistical inference is the procedure used to reach a conclusion about a population based on the information derived from a sample that has been drawn from that population.

Statistic and Parameter

A **statistic** is a descriptive measure computed from the data of the sample.

A **parameter** is a descriptive measure computed from the data of the population.

Measurement Scales:

Measurement: Assignment of numbers to objects.

	< >	−	c× + ÷	
Scale	*Order*	*Distance*	*True Zero*	*Examples*
Nominal	no	no	no	Color, Gender, Ethnicity, Country
Ordinal	yes	no	no	Rating scales, Rank orders
Interval	yes	yes	no	Time of day, Year, IQ, Likert scales
Ratio	yes	yes	yes	Age, Height, Weight, Rates

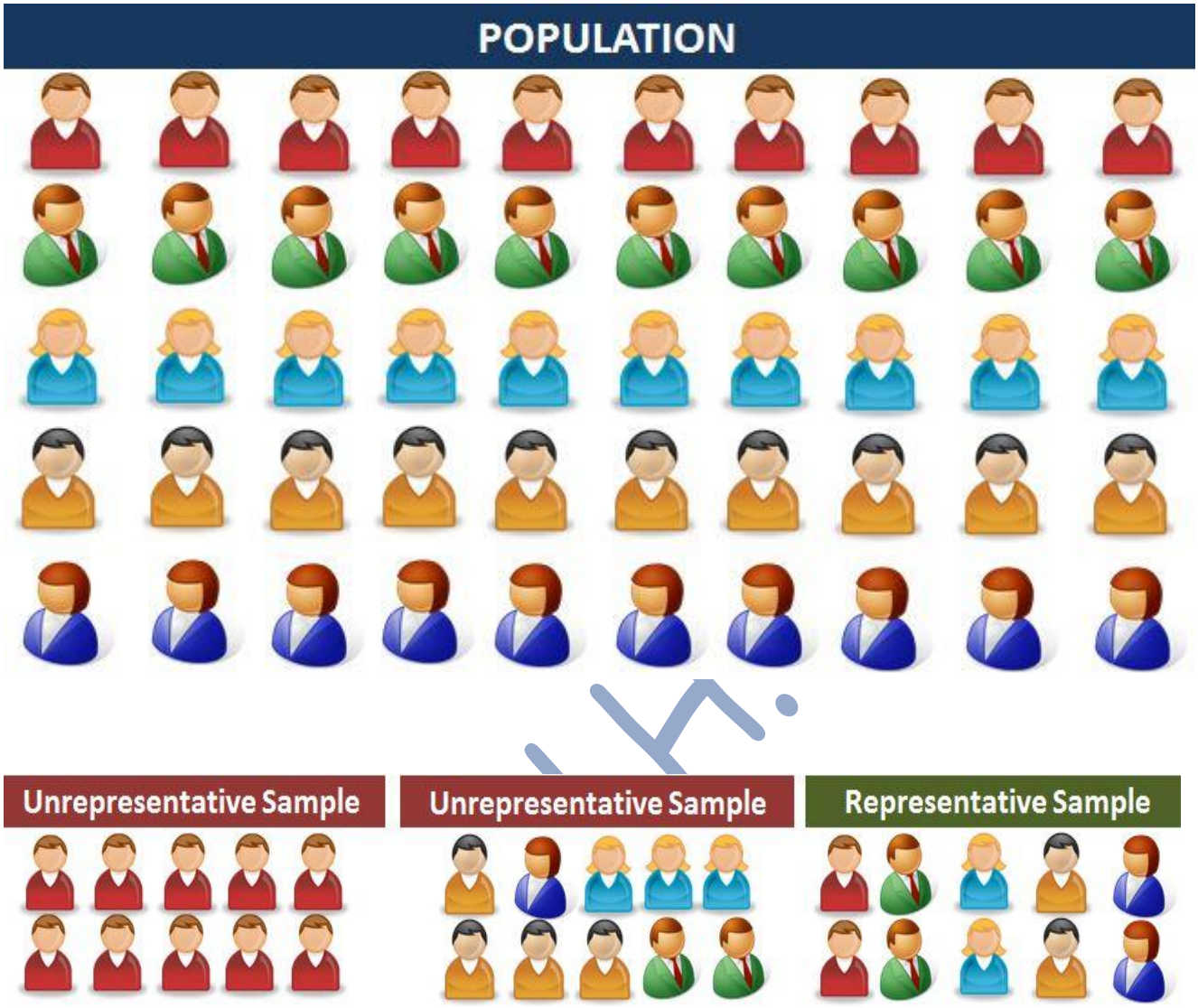
Sampling Techniques

Two advantages of sampling:

1. the cost is lower, and
2. data collection is faster than measuring the entire population.

Because the sample will be used to draw conclusions about the entire population, it should be a ***representative sample,*** that is, it should reflect as closely as possible the relevant characteristics of the population under consideration

Dr. Monied H. S...uh



Sampling Methods (Probability Sampling Methods):

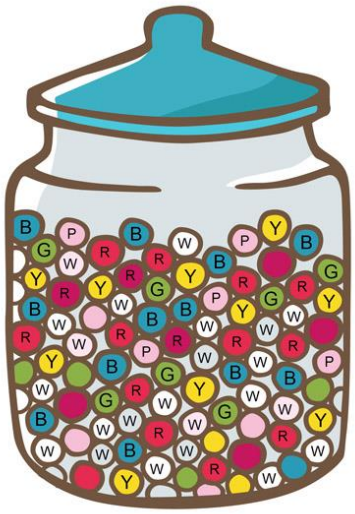
1. Simple Random Sampling
2. Systematic Sampling
3. Stratified Sampling
4. Cluster Sampling

Dr. Mo'ayad H.

Simple Random Sampling

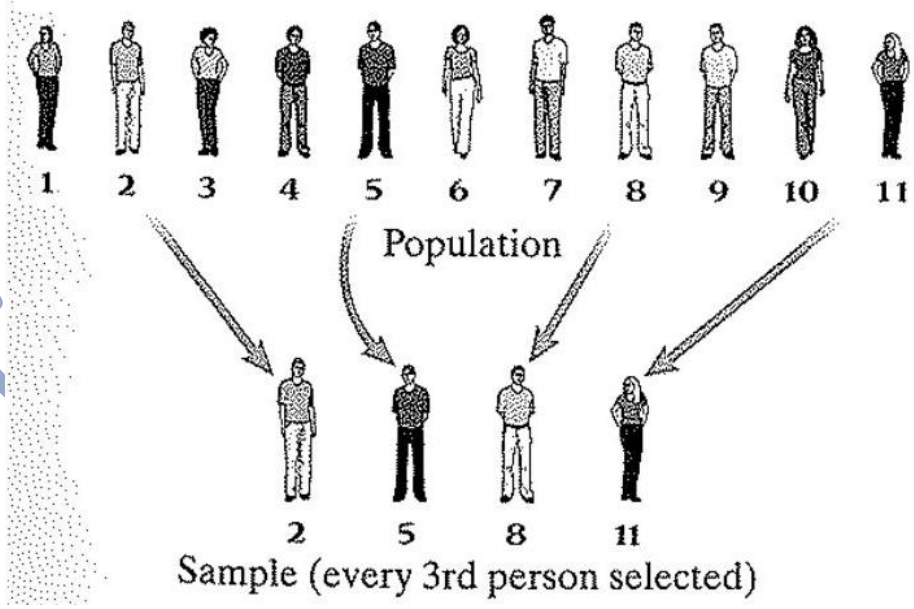
DEFINITION

If a sample of size *n* is drawn from a population of size *N* in such a way that every possible sample of size *n* has the same chance of being selected, the sample is called a simple random sample.



Systematic Sampling

Systematic samples are obtained by numbering each value in the population and then selecting the kth value.

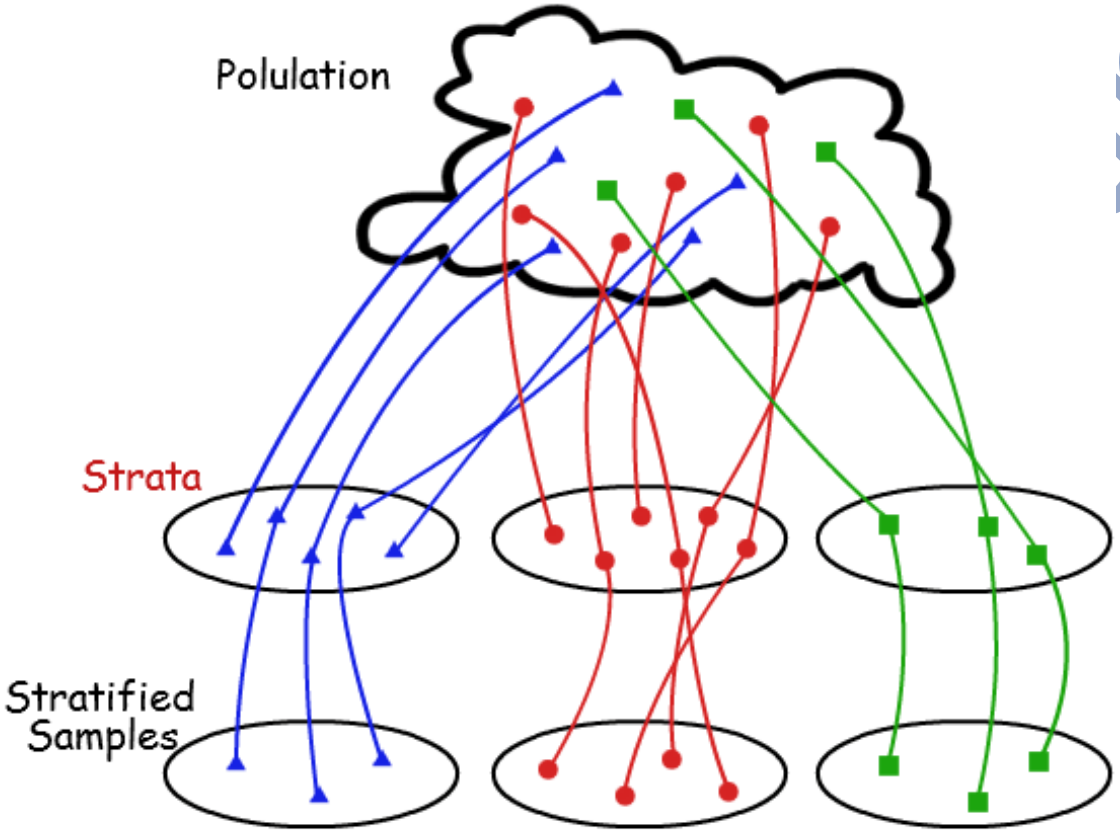


Dr. Nida H. Samur

Stratified Sampling

A stratified sample has members from each segment of a population.

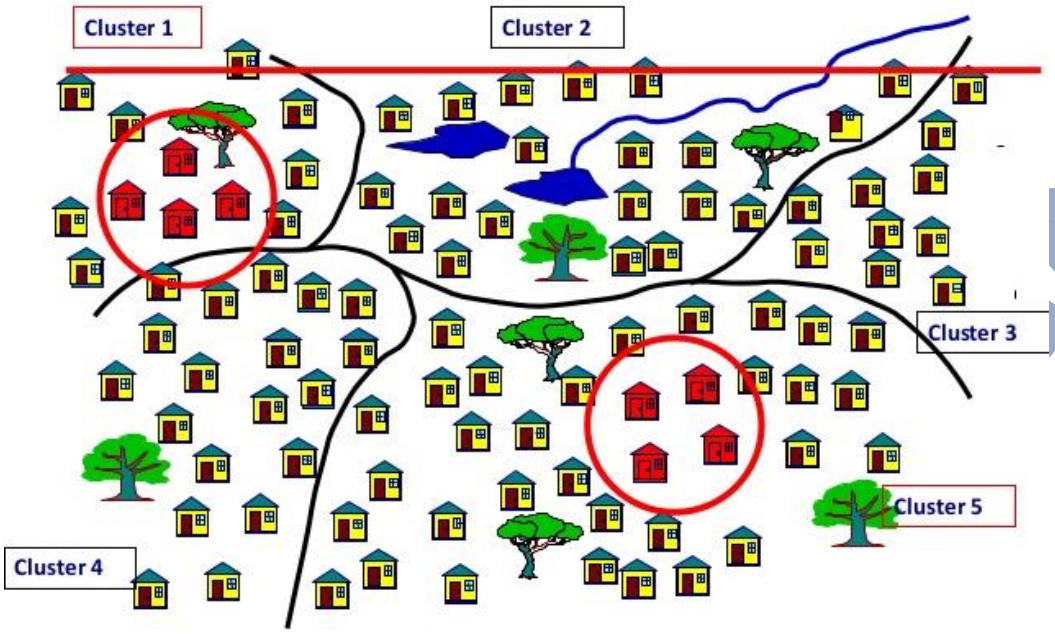
This ensures that each segment from the population is represented.



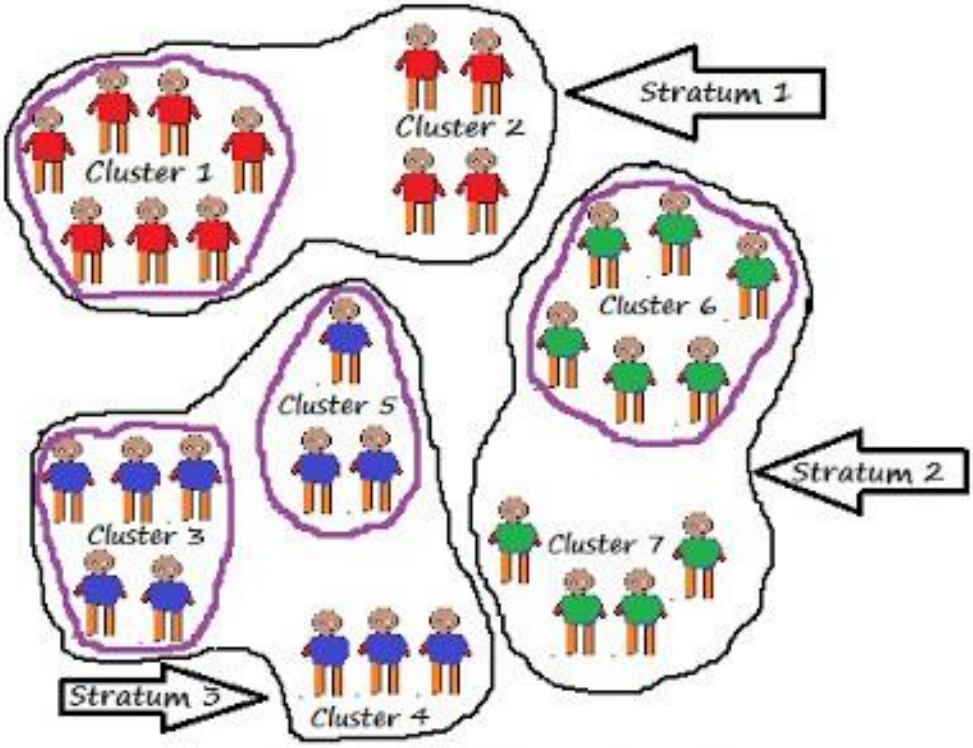
Dr. Mor

Cluster Sampling

Cluster samples are selected by dividing the population into groups and then taking samples of the groups.



Difference between Cluster Sampling and Stratified Sampling



Exercises:

Q1. For each of the following, identify the population and the sample.

1. In a recent survey, 3002 adults in the United States were asked if they read news on the Internet at least once a week. Six hundred of the adults said yes.

2. A study of 254 patient with sleep disorders was conducted to find a link between obesity and sleep disorders.

Q2. For each of the following, determine whether the data are qualitative or quantitative and specify the measurement scale that is employed when taking measurements on each:

1. The telephone numbers in a telephone directory.
2. The scores of a class on a statistics exam.
3. The player numbers for a soccer team.
4. The monthly salaries of the employees at an accounting firm.
5. The ages of a sample of 350 residents of nursing homes.
6. Weights of babies born in a hospital during a year.
7. Gender of babies born in a hospital during a year.
8. Blood type of a person.
9. Color of a flower.
10. Shape of a seed: wrinkled, smooth.
11. Cholesterol concentration in a blood specimen.
12. Optical density of a solution.
13. Number of bacteria colonies in a petri dish.
14. Number of cancerous lymph nodes detected in a patient.
15. Length of a DNA segment in base pairs.

Dr. Monjed H. Samuh

Q3. For each of the following, determine whether the numerical value is a parameter or a statistic.

1. The average annual salary of 35 of a company's 1200 accountants is $68000.
2. In a recent year, the average math scores for all graduates on the ACT was 20.7 .
3. In a survey of a sample of computer users, 8% said their computer had a malfunction that needed to be repaired by a service technician.
4. In a recent survey of 1000 adults from the United States, 47% said using a cell phone while driving should be illegal.

Q4. For each of the following situations, answer questions (a) through (e):

(a) What is the sample in the study?

(b) What is the population?

(c) What is the variable of interest?

(d) How many measurements were used in calculating the reported results?

(e) What measurement scale was used?

Situation A.

A study of 300 households in a small southern town revealed that 20 percent had at least one school-age child present.

Situation B.

A study of 250 patients admitted to a hospital during the past year revealed that, on the average, the patients lived 15 miles from the hospital.

Q5. For each of the following, identify which sampling method is used:

1. You sample every sixth person at a grocery store to see what type of sports events they watch.
2. Your teacher puts the names of all the students in your class into a box. He selects a class representative by drawing a name out of the box.

Dr. Monjed H. Samuh

3. Every 10th bottle of super-Duper cola is selected, and the amount of liquid in the bottle is measured. The purpose is to see if the machines that fill the bottles are working properly.

4. Out of 10 hospitals in the municipality, a researcher selects one and collects records for 24 hour period on the type of emergencies that were treated there.

5. A youth center has 150 male members and 250 female members. The manager of the youth center wants to study the average weights of its members. He randomly chooses 30 male members and 50 female members and measures their weights.

Dr. Monjed H. Samuh