

## Section A

### Basic Introduction to Statistics

What do you think of when you see/hear the word *statistics*? The majority of people immediately think of numerical facts, data, graphs and tables. But not only do statisticians collect, classify and tabulate data, they also analyze data in order to make generalizations and decisions.

Why study statistics?

- 1) Everyone comes in contact with statistics in everyday life.
- 2) People should understand reports in newspapers, magazine and journals.
- 3) People should be able to question the statistics they read, and not blindly accept these as proven fact.
- 4) Many areas of study use statistics, such as; psychology, sociology, business, biology, government, engineering, science education and even areas such as history, language and the arts.

**Statistics** is the science of collecting, organizing, summarizing, and analyzing data to draw conclusions or answer questions. It also provides a measure of confidence in any conclusions.

#### **Two types of statistics:**

- 1) **Descriptive statistics:** the use of numbers to summarize information which is known about some population. [collecting, organizing and summarizing the data]
- 2) **Inferential statistics:** the use of numbers related to a random sample from a population to give numerical information about the population itself. [analyzing the data to draw conclusions or answer questions about the population]

**Probability** is very important in inferential statistics; it's related to the risk of making an error.

**Variables** are characteristics of the individuals or things being studied.

**Raw scores or data:** Numbers obtained in a particular situation. A collection of raw scores is usually called a distribution of scores.

Examples of a distribution:

- a) **Test scores on an exam in a particular class.**
- b) **Ages of students at MCCC.**
- c) **IQ's of a random sample of 6<sup>th</sup> grade students in the Trenton school district.**

**Population** – All people or things being considered in a particular situation

EX: **All students in the 6<sup>th</sup> grade in the Trenton school district.**

A **parameter** is a numerical summary of a population.

EX: **The mean IQ score of all 6<sup>th</sup> grade students in the Trenton school district.**

**Sample** – any portion (subset) of a population under consideration

EX: **Fifty 6<sup>th</sup> grade students from the Trenton school district.**

A **statistic** is a numerical summary of a sample

EX: **The mean IQ score for fifty 6<sup>th</sup> grade students from the Trenton school district.**

Note: A parameter goes with a population and a statistics goes with a sample.

**Random Sample**: A sample selected in such a way that every member of the population has an equal chance of being selected. The members of the random sample are picked arbitrarily from the population.

Example:

Consider the following: Population: All students attending Mercer County Community College

Variable: Some measure of mathematical ability

Sample: Students leaving a section of calculus at MCCC.

This is not a random sample from the population of all students at MCCC. From this sample we should not attempt to infer anything about the mathematical ability of all students at MCCC.

Note: A bias in obtaining a sample will destroy the value of the statistical information obtained since statistical inferences made from this information would be invalid.

### **Why use a sample instead of a population?**

- a) *Time*
- b) *Money*
- c) *Population continuously changing.*
- d) *Cannot use the entire population.*

### **Two types of variables**

- 1) **Qualitative or categorical variable** – classification based on some attribute or characteristic of the individual (non-numerical)

EX: **hair color, eye color, gender, race, ethnicity**

- 2) **Quantitative variable** – provides numerical measures of individuals

### **Two types of quantitative variables**

- 1) **Discrete** – has either a finite number of possible values or a countable number of possible values  
(something that can be counted)

EX: **number of cars at a light, number of students in a classroom, number of rooms in a house**

- 2) **Continuous** – has an infinite number of possible values that are not countable  
(something that can be measured)

EX: **height, weight, age, miles per gallon, time**

## Examples

1) Determine whether the number described is a parameter or statistic:

- a) In a recent survey of college graduates, 68% of those who responded said they had more than \$50,000 in student loans. **statistic**
- b) The average age of the employees working at XYZ Company is 37 years. **parameter**
- c) The average GPA of 250 randomly selected students at ABC University is 2.73. **statistic**
- d) Of the students attending Mercer County Community College in 2018, 66% were part time students. **parameter**

2) Determine which of the following is an example of descriptive statistics and which is an example of inferential statistics.

- a) The average weight of all football players on the NY Giants football team is 235 pounds.  
**descriptive**
- b) The average yearly salary of a random sample of 150 minor league baseball players is \$102,000. Therefore, the average yearly salary of all minor league baseball players is \$102,000.  
**inferential**

3) Determine whether or not the sample given represents the given population accurately.

- a) Population: All students attending Mercer County Community College.  
Sample: 100 students selected at random entering the student center at noon on Monday. **No**
- b) Population: All businesses in Mercer County.  
Sample: 75 businesses selected at random from a list of all businesses in Mercer County. **Yes**

4) To determine the average GPA of 500 students who just finished their first year in college, a group of 60 students is randomly selected. It is determined that the average GPA is 2.85.

- a) What is the population for this study? **500 students who just finished their first year in college**
- b) What constitutes the sample? **A group of 60 students**
- c) Based on the sample, what is the statistic for the average GPA of the population? **2.85**

5) The following data set provides information about five college professors.

Name	Specialty	Gender	Age	Height (in)	# of years of teaching	Rank
Allen	Nursing	F	40	65	16	Associate Professor
Backer	Accounting	M	59	71	34	Full Professor
Hughes	Psychology	F	52	69	13	Associate Professor
Ramirez	Mathematics	F	30	68	5	Assistant Professor
Turner	Sociology	M	38	70	9	Assistant Professor

Which variables are qualitative and which are quantitative variables?

Qualitative variables: **Name, Specialty, Gender, Rank**

Quantitative variables: **Age, height, # of years of teaching**