

Business Statistics I

**Introductory Statistics
and Data Analysis**

Learning Objectives

1. Define Statistics.
2. Distinguish Descriptive & Inferential Statistics.
3. Define Population, Sample, Parameter, Statistic, Variable.
4. Identify data types.

What is Statistics?



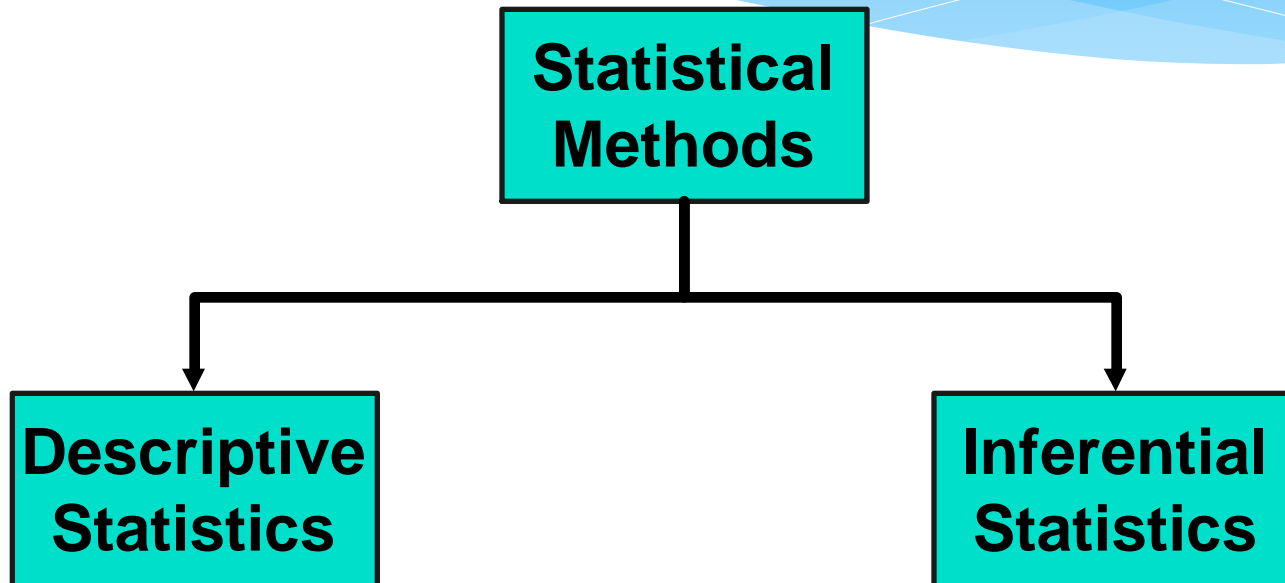
What is Statistics?

- The practice (science?) of data analysis
- Summarizing data and drawing inferences about the larger population from which it was drawn

What is Statistics?

5 7 8 7 6 4 5 5 1 2

Statistical Methods



Branches of Statistics

- ***Descriptive Statistics***

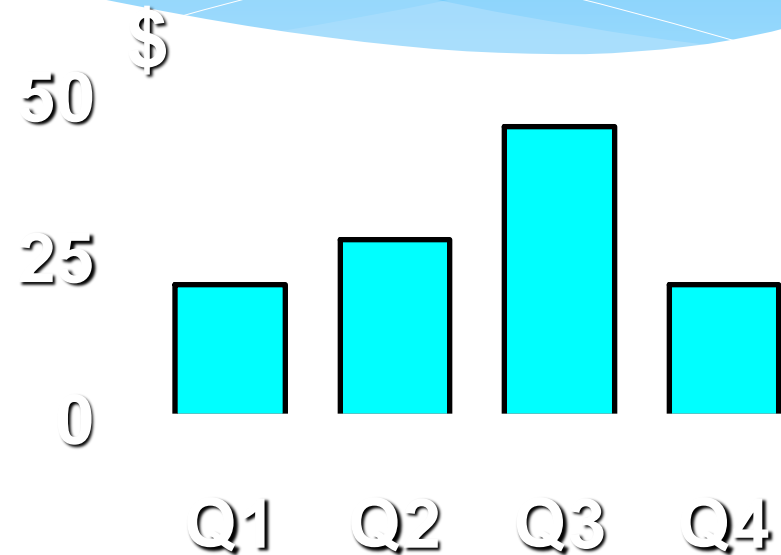
Summary and description of collected data.

- ***Inferential Statistics***

Generalizing from a sample to a population.

Descriptive Statistics

- 1. Involves
 - Collecting Data
 - Presenting Data
 - Characterizing Data
- 2. Purpose
 - Describe Data



$$\bar{X} = 30.5 \quad S^2 = 113$$

Descriptive Statistics

- Sometimes (but rarely) we can enumerate
- the whole population . If so, we need only use:
- **DESCRIPTIVE STATISTICS:**
- Procedures used to summarize and describe the set of measurements.

Inferential Statistics

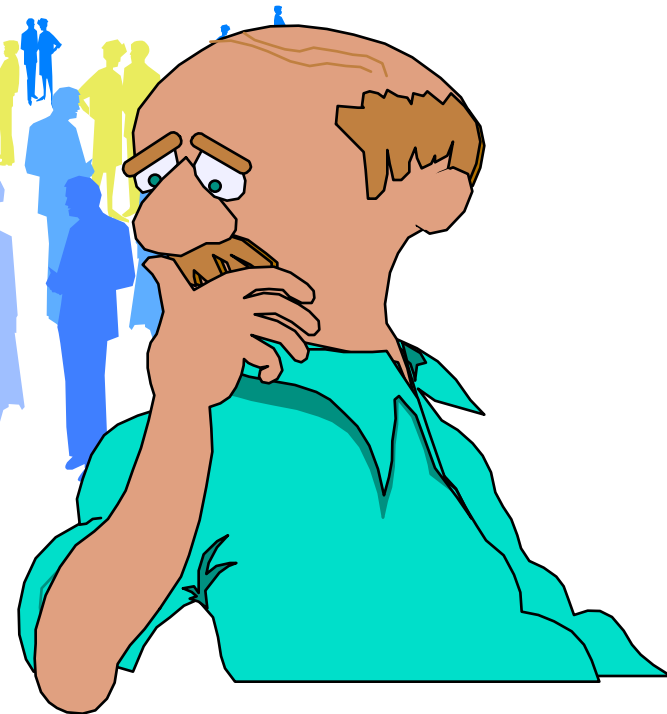
1. Involves

- Estimation
- Hypothesis Testing

2. Purpose

- Make Decisions Based on Population Characteristics

Population?



Inferential Statistics

- When we cannot enumerate the whole population, we use:
- **INFERENTIAL STATISTICS**: Procedures used to draw conclusions or inferences about the population from information contained in the sample.

Key Terms

1. Population

- ✓ All Items of Interest

2. Sample

- ✓ Portion of Population

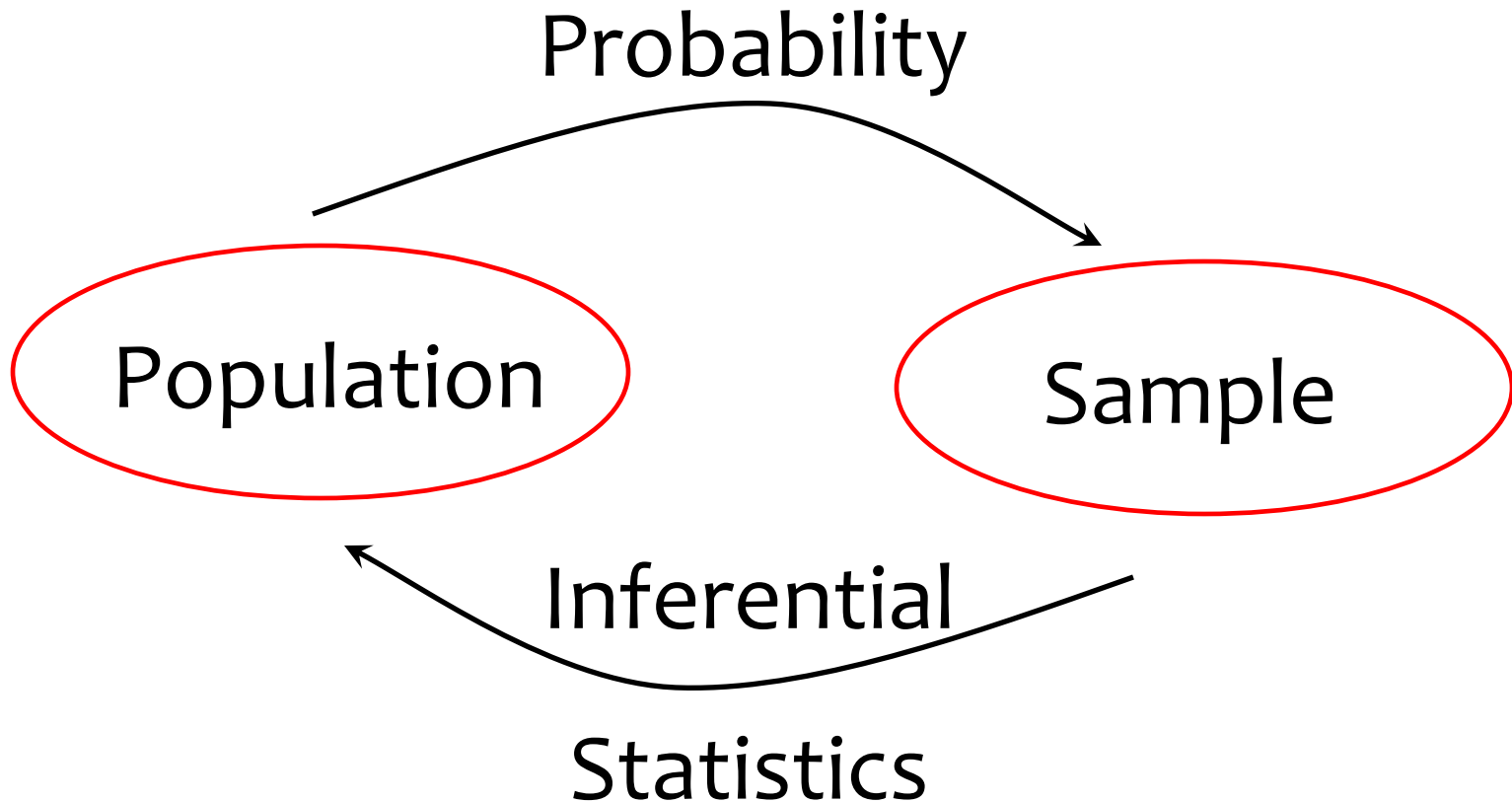
3. Parameter

- ✓ Summary Measure about Population

4. Statistic

- ✓ Summary Measure about Sample

Relationship Between Probability and Inferential Statistics



Parameter

- A numerical measurement on an experimental unit describing some characteristic of a population (e.g., mean)

Population



Parameter

Statistic

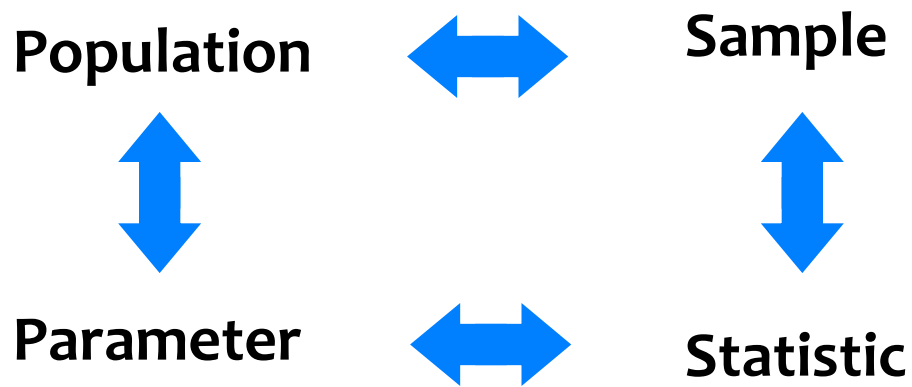
- A numerical measurement describing some characteristic of a sample (e.g., mean)

Sample



Statistic

Relationship Between Population and Sample



Variables and Data

- A *variable* is a characteristic that changes or varies over time and/or for different individuals or objects under consideration.
- **Examples:** Hair color, Age, time to failure of a computer component.

Definitions

- An **experimental unit** is the individual or object on which a variable is measured.
- A **measurement** results when a variable is actually measured on an experimental unit.
- A set of measurements, called **data**, can be either a **sample** or a **population**.

Example

- **Variable**
Hair color
- **Experimental unit**
Person
- **Typical Measurements**
Brown, black, blonde, etc.

Example

- **Variable**
Time until the failure of a computer
- **Experimental unit**
Computer
- **Typical Measurements**
7500 hours, 7535.5 hours, etc.

Level of Measurement

- When we observe and record a variable, it has characteristics that influence the type of statistical analysis that we can perform on it. These characteristics are referred to as the level of measurement of the variable
- The first step in any statistical analysis is to determine the level of measurement; it tells us what statistical tests can and cannot be performed

Data Types(Level of measurement)

- Qualitative
 - ✓ Nominal (categorical): Major in college and race.
 - ✓ Ordinal (rank ordered categories): letter grade on a test
- Quantitative
 - ✓ Interval
 - ✓ Ratio

Qualitative Data

- **Nominal (categorical):** measurement values just "name" the attribute uniquely. No ordering of the cases is implied.
- **Examples:**
 - ✓ Hair color (black, brown, blonde...)
 - ✓ Make of car (Dodge, Honda, Hyundai...)
 - ✓ City of birth (Toronto, Ottawa,....)

Qualitative Data

- **Nominal (categorical):** measurement values just "name" the attribute uniquely. No ordering of the cases is implied.
- **Examples:** uniform numbers in basketball are measures at the nominal level. A player with number 30 is not more of anything than a player with number 15, and is certainly not twice whatever number 15 is.

Qualitative Data

- Ordinal (rank ordered categories): is one where the order matters but not the difference between values.
- Examples:
 - ✓ small (1), medium (2), large (3)
 - ✓ very useful (1), useful (2), moderately useful (3), not very useful (4)

Qualitative Data

- **Example.** On a survey you might code educational attainment as:

0 = less than H.S.;

1 = some H.S.;

2 = H.S. degree;

3 = some college;

4 = college degree;

5 = post college.

- *In this measure, higher numbers mean more education. But is distance from 0 to 1 same as 3 to 4? Of course not. The interval between values is not interpretable in an ordinal measure.*

Types of Variables

- **Quantitative variables**
 - ✓ Interval
 - ✓ Ratio

Interval Variable

- All of the characteristics of ordinal plus the following;
 - ✓ Measurements are on a numerical scale with an arbitrary zero point
 - The “zero” is assigned: it is unphysical and not meaningful
 - Zero does not mean the absence of the quantity that we are trying to measure
 - ✓ Can only meaningfully compare values in terms of the interval between them
 - Cannot compare values by taking their ratios
 - “Interval” is the mathematical difference between the values
 - Equal intervals in the numbers measured represent real equal differences in the variables

Interval Variable

- Example: temperature
 - ✓ 0°C means “cold,” not “no heat”
 - ✓ 20°C is NOT twice as warm as 10°C . But 20°C is 10° warmer than 10°
- Example: GRE, and IQ score

Interval Variable

- Example: temperature
 - ✓ 0° C means “cold,” does not represent the complete absence of temperature
 - ✓ 20° C is NOT twice as warm as 10° C. But 20° C is 10° warmer than 10°
- Example: GRE, and IQ score

Example

Size	Bust (in)	Waist (in)	Hips (in)
8	32	24	35
10	34	26	37
12	36	28	39
14	38	30	41
16	40	32	43
18	42	34	45
20	44	36	47
22	46	38	49
24	48	40	51
26	50	42	53
28	52	44	55

- Standard U.S. women's dress size.
 - ✓ Is there an order to this data?
 - Yes, $8 < 16 < 20 < 24 < 28$.
 - ✓ Are the differences between the data values meaningful?
 - Sure, size 20 is 2 inches more than the size 18.
 - ✓ Does “Zero” mean the absence of the quantity?
 - There is no natural zero point for dress size. A “size 0” dress does not have “zero” material. Instead it would have a 24-inch bust, 16-inch waist and 27-inch hips

Ratio Variable

- All the characteristics of interval plus the following:
 - ✓ Measurements are on a numerical scale with a meaningful zero point
 - Zero means “none” or “nothing”
 - ✓ Values can be compared in terms of their interval and ratio
 - \$30 is \$20 more than \$10
 - \$30 is 3 times as much as \$10
 - \$0 means no money

Example

- Four people are randomly selected and asked how much money they have with them. Here are the results : \$21, \$50, \$65, and \$300.
 - ✓ Is there an order to this data?
 - Yes, $\$21 < \$50 < \$65 < \300 .
 - ✓ Are the differences between the data values meaningful?
 - Sure, the person who has \$50 has \$29 more than the person with \$21.
 - ✓ Does “0” mean “none” or “nothing”?
 - Yes. 0 means no money.
 - ✓ Can we calculate ratios based on this data?
 - Yes because \$0 is the absolute minimum amount of money a person could have with them. The person with \$300 has 6 times as much as the person with \$50.

Example

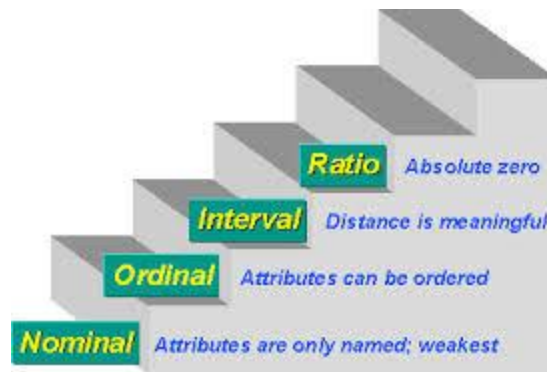
Size	Bust (in)	Waist (in)	Hips (in)
8	32	24	35
10	34	26	37
12	36	28	39
14	38	30	41
16	40	32	43
18	42	34	45
20	44	36	47
22	46	38	49
24	48	40	51
26	50	42	53
28	52	44	55

- Standard U.S. women's dress size.
 - ✓ Is there an order to this data?
 - Yes, $8 < 16 < 20 < 24 < 28$.
 - ✓ Are the differences between the data values meaningful?
 - Sure, size 20 is 2 inches more than the size 18.
 - ✓ Does “0” mean “none” or “nothing”?
 - No, There is no natural zero point for dress size. A “size 0” dress does not have “zero” material. Instead it would have a 24-inch bust,
 - ✓ Can we calculate ratios based on this data?
 - No, If you divide a size 28 by a size 14, you do not get the same answer as dividing a size 20 by 10. Neither ratio is equal to two as the “size” number would suggest.

Ratio Variable

- ✓ In business and finance, most quantitative variables are ratio variables, such as anything to do with money
- ✓ **Examples:** Earnings, profit, loss, age, distance, height, weight, ages of people, scores on exams (graded from 0 to 100), and hours of study for a test.

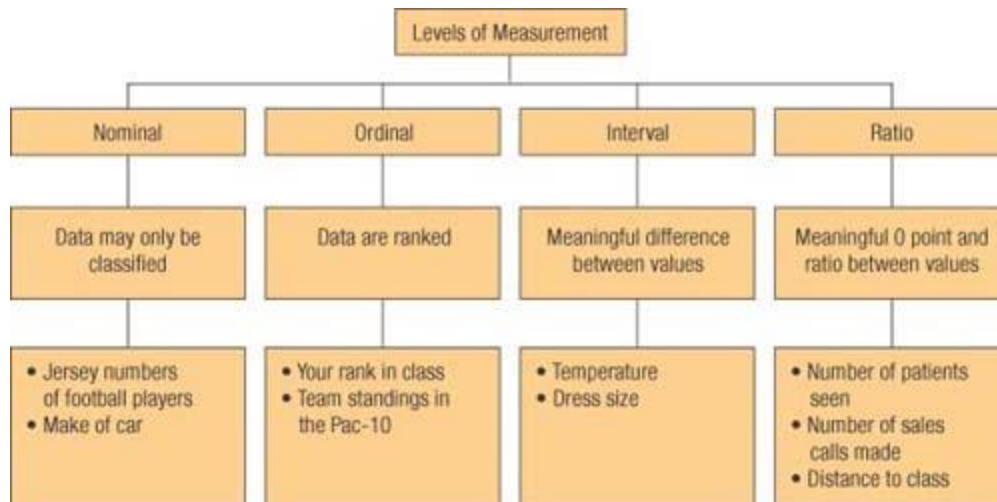
Level of Measurement



Level of Measurement

- At lower levels of measurement, assumptions tend to be less restrictive and data analyses tend to be less sensitive.
- At each level up the hierarchy, the current level includes all of the qualities of the one below it and adds something new.
- In general, it is desirable to have a higher level of measurement (e.g., interval or ratio) rather than a lower one (nominal or ordinal).

Level of Measurement



Why is Level of Measurement Important?

1. Knowing the level of measurement helps you decide how to interpret the data from that variable.
2. Knowing the level of measurement helps you decide what statistical analysis is appropriate on the values that were assigned

Why is Level of Measurement Important?

OK to compute...	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes

Types of Variables

In addition to the various levels of measure, variables can also be either discrete (or discontinuous) or continuous

- **Quantitative variables** measure a numerical quantity on each experimental unit.
 - ✓ **Discrete** if it can assume only a finite or countable number of values.
 - ✓ **Continuous** if it can assume the infinitely many values corresponding to the points on a line interval.

Examples

- For each orange tree in a grove, the number of oranges is measured.
 - ✓ **Quantitative discrete, ratio**
- For a particular day, the number of cars entering Carleton campus is measured.
 - ✓ **Quantitative discrete, ratio**
- Time until a computer failure
 - ✓ **Quantitative continuous, ratio**

Exercise

- Data types
 - ✓ Bacteria count
 - ✓ Occupations of shoppers
 - ✓ Marital status
 - ✓ Time since last auto maintenance

Exercise

- Identify the type of measurement for the following:
military title -- Lieutenant, Captain, Major.
- ✓ nominal
- ✓ Ordinal
- ✓ Interval
- ✓ ratio

Exercise

- Identify the type of measurement for the following categorization of clothing:
hat, shirt, shoes, pants

- ✓ nominal
- ✓ Ordinal
- ✓ Interval
- ✓ ratio

Exercise

- What is the level of measurement for each of the following variables? Student IQ ratings.
 1. Distance students travel to class.
 2. Student scores on the first statistics test.
 3. A classification of students by eye color.
 4. A ranking of students as freshman, sophomore, junior, and senior.
 5. Number of hours students study per week.

Practical Question

Self-Correcting Exercises

- **1.1,1.3,1.5,1.19,1.21,1.22**