Basic Mathematics and Statistics

CHAPTER 9: DATA AND STATISTICS

Dr. Khaing S. Htun

9.1 What is meant by Statistics?

Statistics

1. refers to numerical facts

2. is the art and science of collecting, summarizing, analyzing, presenting and interpreting data, and to make decisions about phenomenon under study

9.2 Why study Statistics?

1. The principles and methodology of statistics are useful in answering questions.

   - What kind and how much data need to be collected?

   - How should we organize and interpret the data?

   - How can we analyze the data and draw conclusions?

2. The numerical information is everywhere, we need to be able to read the charts and graphs and understand the discussion of the numerical information.

3. The statistical techniques are used to make decisions that affect our daily lives.

4. The knowledge of statistical methods will help you understand why decisions are made and give you a better understanding of how they affect you.

9.3 Who uses Statistics

Statistical techniques are used extensively by

- marketing,

- accounting,

- quality control,

- consumers,

- hospital administrators,

- educators,

- politicians,

- physicians,

- etc..

9.3 Who uses Statistics

examples of statistics in everyday life:

Accounting : Public accounting firms use statistical sampling procedures when conducting audits for their clients.

Finance : Financial advisors use a variety of statistical information to guide their investment recommendations

Marketing : Electronic scanners at retail checkout counters are being used to collect data for a variety of marketing research applications.

Production : With today's emphasis on quality, quality control is an important application of statistics in production. A variety of statistical quality control charts are used to monitor the output of a production process.

Economics : Economists are frequently asked to provide forecasts about the future of the economy or some aspect of it. They use a variety of statistical information in making such forecasts.

9.4 Some Statistics definitions

A **population** is the set of all elements—individuals, items, or objects— of interest in a particular study.

The population that is being studied is also called the **target population**.

A *sample* is a subset of the population selected for study.
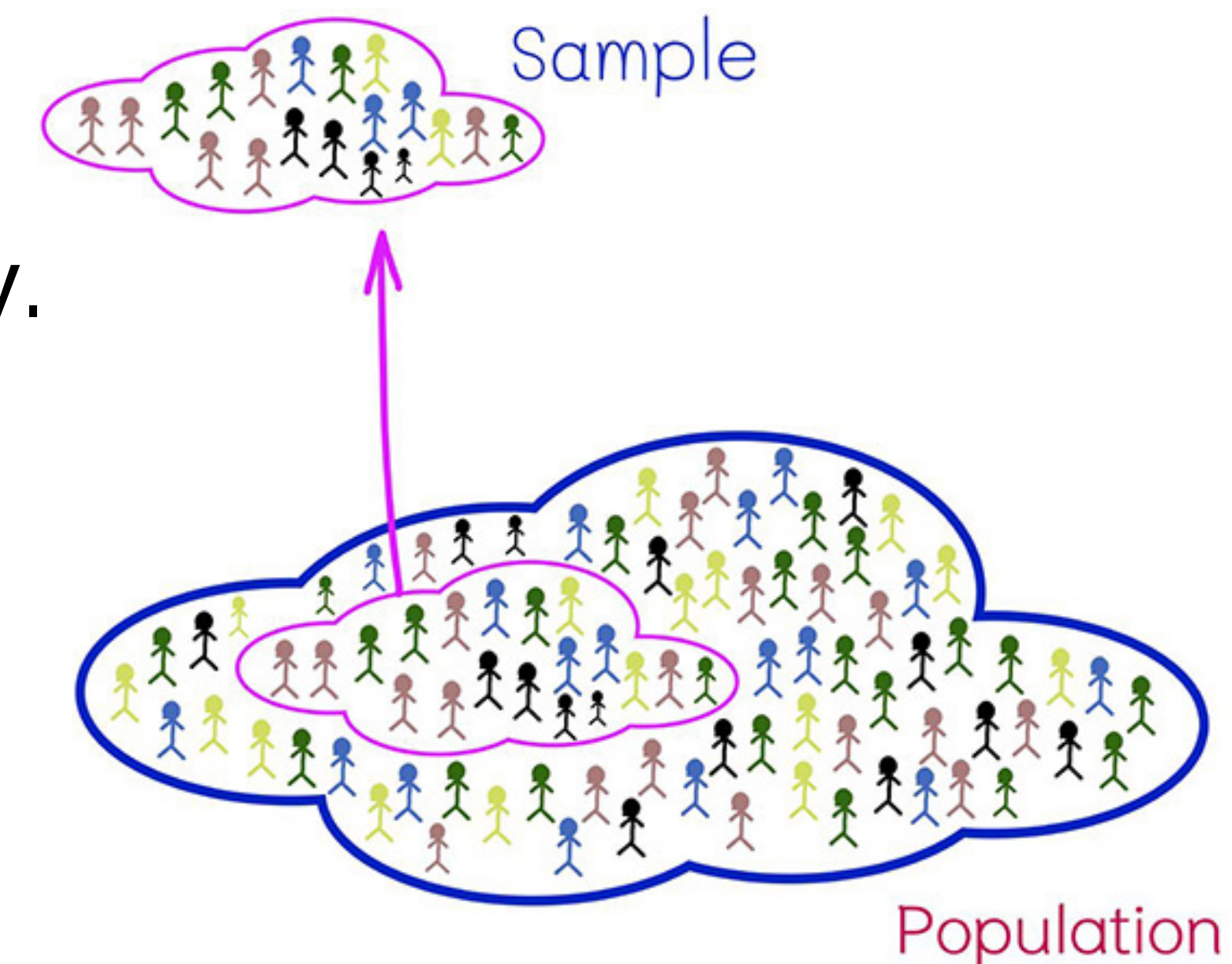
Sample

Population

9.4 Some Statistics definitions

A **survey** is the collection of information from the elements of a population or a sample.

- A **census** is a survey that includes every element of the target population.

- A **sample survey** is a survey that includes every element of the sample.

9.4 Some Statistics definitions

A **parameter** is a summary measure that is computed to describe a characteristic of an entire population.

A **statistics** is a summary measure that is computed to describe a characteristic from only the sample.

Table 9.1 Some Symbols of Parameter and Statistics

Items	Parameter	Statistics
Number of Observations	N	n
Mean	μ (*read mu*)	$\overline{X}$ (*read x-bar*)
Standard Deviation	σ (read sigma)	S
Variance	σ^2	S^2
Proportion	P	p

9.5 What is data?

- **Data** are the facts and figures that are collected, analyzed, and summarized for presentation and interpretation.

- **Data set** are all the data collected in a particular study.

- **Elements** are the entities on which data are collected.

- **Variable** is a characteristic of interest for the elements that assumes different values for different elements.

- **Observation** is the set of measurements obtained for a single element.

9.5 What is data?

variable

elements

observation

1. *Exchange:* Where the stock is traded

2. *Ticker Symbol:* The abbreviation used to identify the stock on the exchange listing.

3. *Annual Sales:* Total sales for the company for the most recent 12 months in millions of dollars.

4. *Earnings per Share:* Total earnings per share for the least 12 months in dollars.

5. *Price/Earning Ratio:* Market price per share divided by the most recent 12 months 'earnings per share.

Table 9.2 A Data Set for 24 Shadow Stocks

Company	Exchange	Ticker Symbol	Annual Sales ($ Million)	Earnings Per Share ($)	Price-Earnings Ratio
1.Advanced Comm. Systems	OTC	ACSC	75.10	0.32	39.10
2.Ag-Chem Equipment Co.	OTC	AGCH	321.10	0.48	23.40
3.Aztec Manufacturing Co.	NYSE	AZZ	79.70	1.18	7.80
4.Cal-Maine Foods, Inc.	OTC	CALM	314.10	0.38	11.70
5.Chesapeake Utilities	NYSE	CPK	174.50	1.13	16.20
6.Dataram Corporation	AMEX	DTM	73.10	0.86	11.00
7.EnergySouth, Inc.	OTC	ENSI	74.00	1.67	13.20
8.Gencor Industries, Inc.	AMEX	GX	263.30	1.96	4.70
9.Industrial Scientific	OTC	ISCX	43.50	2.03	11.50
10.Keystone Consolidated	NYSE	KES	365.70	0.86	9.40
11.LandCare USA, Inc.	NYSE	GRW	111.40	0.33	29.40
12.Market Facts, Inc.	OTC	MFAC	126.70	0.98	26.50
13.Meridian Diagnostics, Inc.	OTC	KITS	36.30	0.46	14.70
14.Merit Medical Systems	OTC	MMSI	67.20	0.27	24.50
15.Met-Pro Corporation	NYSE	MPR	61.90	1.01	12.40
16.Nobility Homes, Inc.	OTC	NOBH	45.80	0.87	14.70
17.Omega Research, Inc.	OTC	OMGA	27.60	0.11	27.30
18.Point of Sale Limited	OTC	POSIF	12.30	0.28	25.40
19.Psychemedics Corp.	AMEX	PMD	17.60	0.13	39.40
20.Roadhouse Grill, Inc.	OTC	GRLL	118.40	0.26	20.80
21.Selas Corp. of America	AMEX	SLS	97.10	0.77	10.70
22.Toymax International, Inc	OTC	TMAX	104.50	1.08	4.70
23.VSI Holdings, Inc.	AMEX	VIS	166.8	0.25	21.00
24.Warrantech Corporation	OTC	WTEC	207.30	0.13	29.80
25.Webco Industries, Inc.	AMEX	WEB	153.50	0.88	7.50

9.5 What is data?

	Company	Stock Exchange	Annual Sales($M)	Earn/Share($)
	Dataram	AMEX	73.10	0.86
	EnergySouth	OTC	74.00	1.67
	Keystone	NYSE	365.70	0.86
	LandCare	NYSE	111.40	0.33
	Psychemedics	AMEX	17.60	0.13

Element Names

Observation

Variables

Data Set

9.6 Types of data

DATA

Categorical;
Qualitative

gender (male, female)

Numerical;
Quantitative

the number of children per family

Discrete
Only Integer Values

Continuous
Any Value in a Particular Range
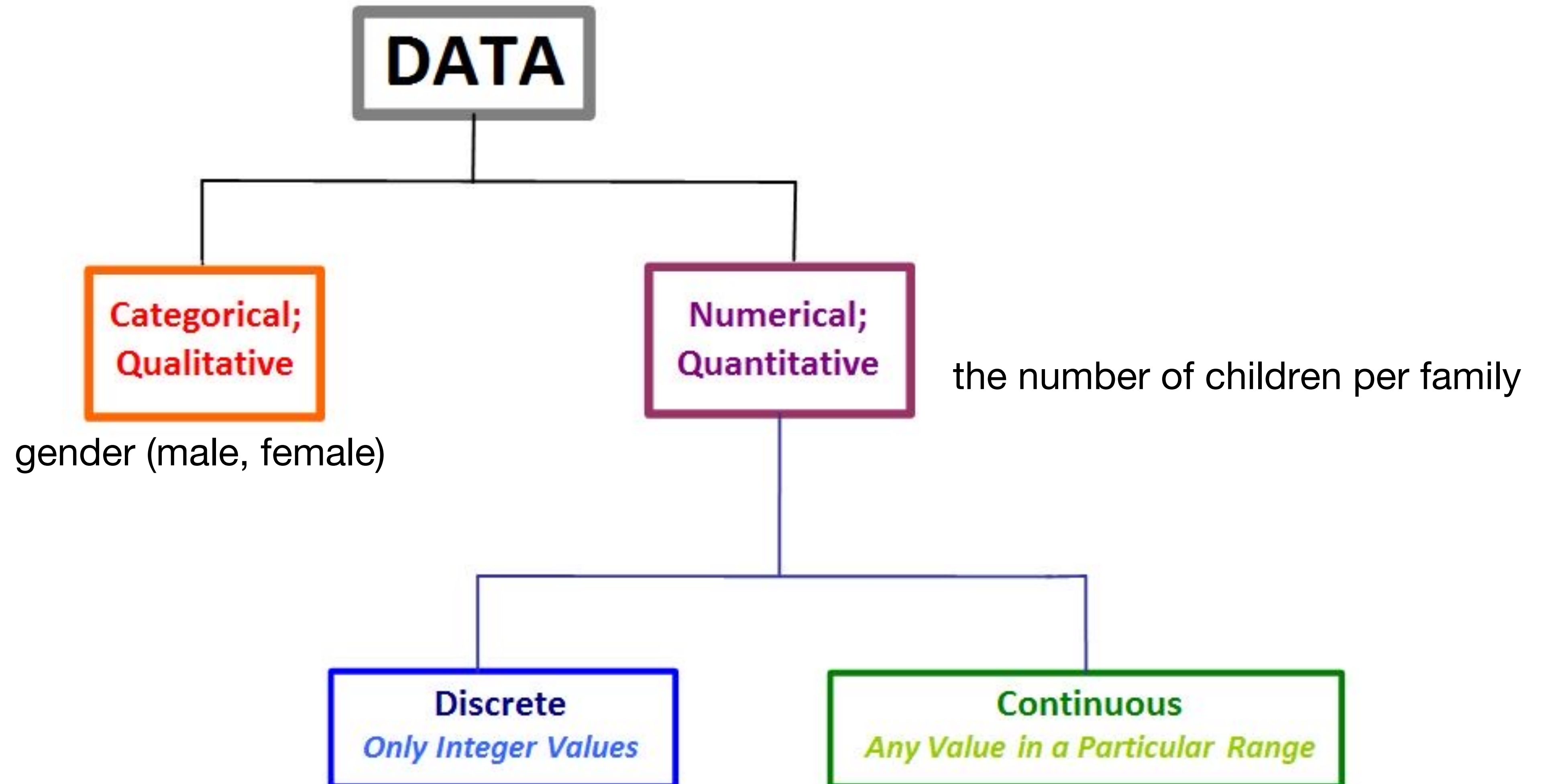
9.6 Types of data

Qualitative (Categorical) Data is the data that only classified in categories and not numerically measured. **Example:-** hair color (blond, brown, black), blood type (O, A, B, AB), the gender of the person (male, female) etc.

Quantitative (Numerical) Data is the data that measured on a numerical scale. **Example:-** the number of children per family, the time taken to complete an examination, number of car owned the income of family etc.

1. **Discrete data** is the data that whose values are countable number. **Example:-** the number of children per family is a discrete data because its value must be 0,1,2,…

2. **Continuous data** is the data that can assume any numerical value over certain interval. **Example**:- the time taken to complete an examination is a continuous data because it can assume any value, let us say, between 30 and 60 minutes. The time taken may be 42.6 minutes, 42.67 minutes, or 42.674 minutes.

9.7 Data sources

data can be obtained in 4 different ways

- **published source** - such as a book, journal, or newspaper.

- **designed experiment** - the researcher exerts strict control over the units (people, objects, or events) in the study. For example, a recent medical study investigated the potential of aspirin in preventing heart attacks.

- **survey** - the researcher samples a group of people, asks one or more questions, and records the responses. Surveys can be conducted through the mail, with telephone interviews, or with in-person interviews. Although in-person interviews are more expensive than mail or telephone surveys, they may be necessary when complex information must be collected.

- **collected observationally** – the researcher observes the experiment units in their natural setting and records the variable(s) of interest.

9.6 Scale of Measurement
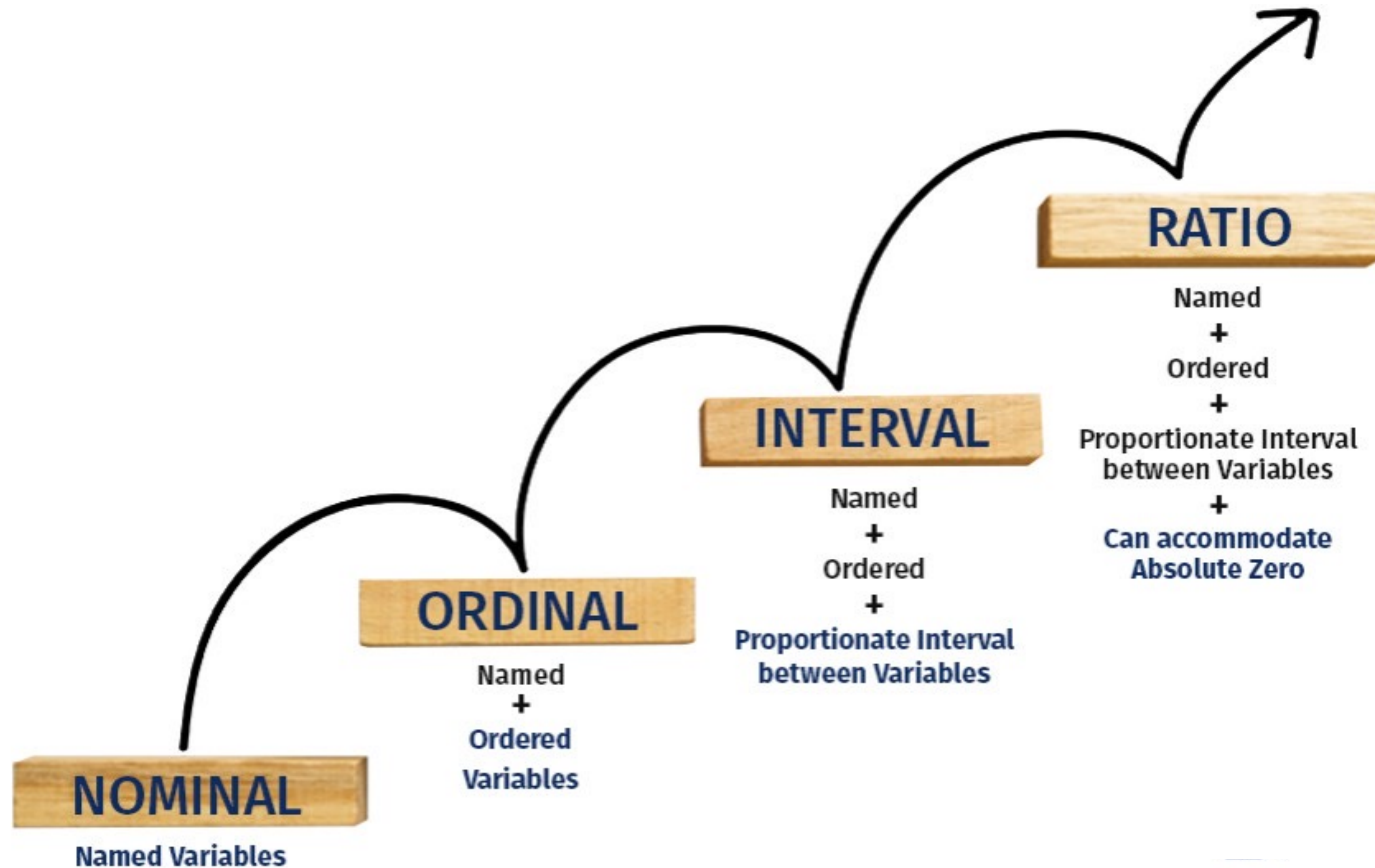
9.6 Scale of Measurement

Nominal Scale : data that can only be classified into categories and cannot be arranged in an ordering scheme. **Example:-** eye color, gender, religious affiliation.

Ordinal Scale : involves data that may be arranged in some order, but differences between data values cannot be determined or are meaningless. **Example:-** grade (A,B,C,D,F), rating scale(poor, good, excellent).

Interval Scale : has the properties of ordinal data with the additional property that meaningful amounts of differences between data values can be determined. There is no natural zero point. Interval data are always numeric. **Example:-** IQ, temperature.

Ratio Scale: has all the properties of interval data with an inherent zero starting point. The ratio of two values is meaningful. Ratio data are always numeric. **Example:-** height, time, salary, age.

9.6 Types of Statistics

Statistics

Descriptive Statistics — Presenting, organizing and summarizing data

Inferential Statistics — Drawing conclusions about a population based on data observed in a sample

Descriptive statistics do not generalize because they deal with the population itself.

Population of 10,0000

Inferential statistics generalize information from a random sample to a population.

Population of 10,0000

Sample of 30 — generalize to → Population of 10,0000

9.6 Types of Statistics

Descriptive Statistics

It can be defined as the methods for collecting, organizing, displaying and describing data by using tables, graphs, or numerical. The most numerical descriptive statistic is the **mean, median, mode, or percentage**.

A data set in its original form is usually very large. It is easier to draw conclusions from summary tables and diagrams than from the original version of a data set. So, we reduce data to a manageable size by constructing tables, drawing graphs, or calculating summary measures.

Example:- according to Consumer Reports, Whirlpool washing machine owners reported 9 problems per 100 machines during 1995. The *statistic 9* describes the number of problems out of every 100 machines.

9.6 Types of Statistics

Inferential Statistics

It can be defined as those methods that make possible the estimation of a characteristic of a population or the making of a decision concerning a population based only on sample results.

Example:- Suppose Norris Electronic attempt to increase the useful life of the lightbulb, the product design group has developed a new lightbulb filament. In this case, the population is defined as all lightbulbs that could be produced with the new filament. To evaluate the advantages of the new filament, the process of statistical inference was used. Figure 9.1 is a graphical summary of the statistical inference process for Norris Electronic.

9.6 Types of Statistics

Inferential Statistics

Figure 9.1 The Process of Statistical Inference for Norris Electronic Example
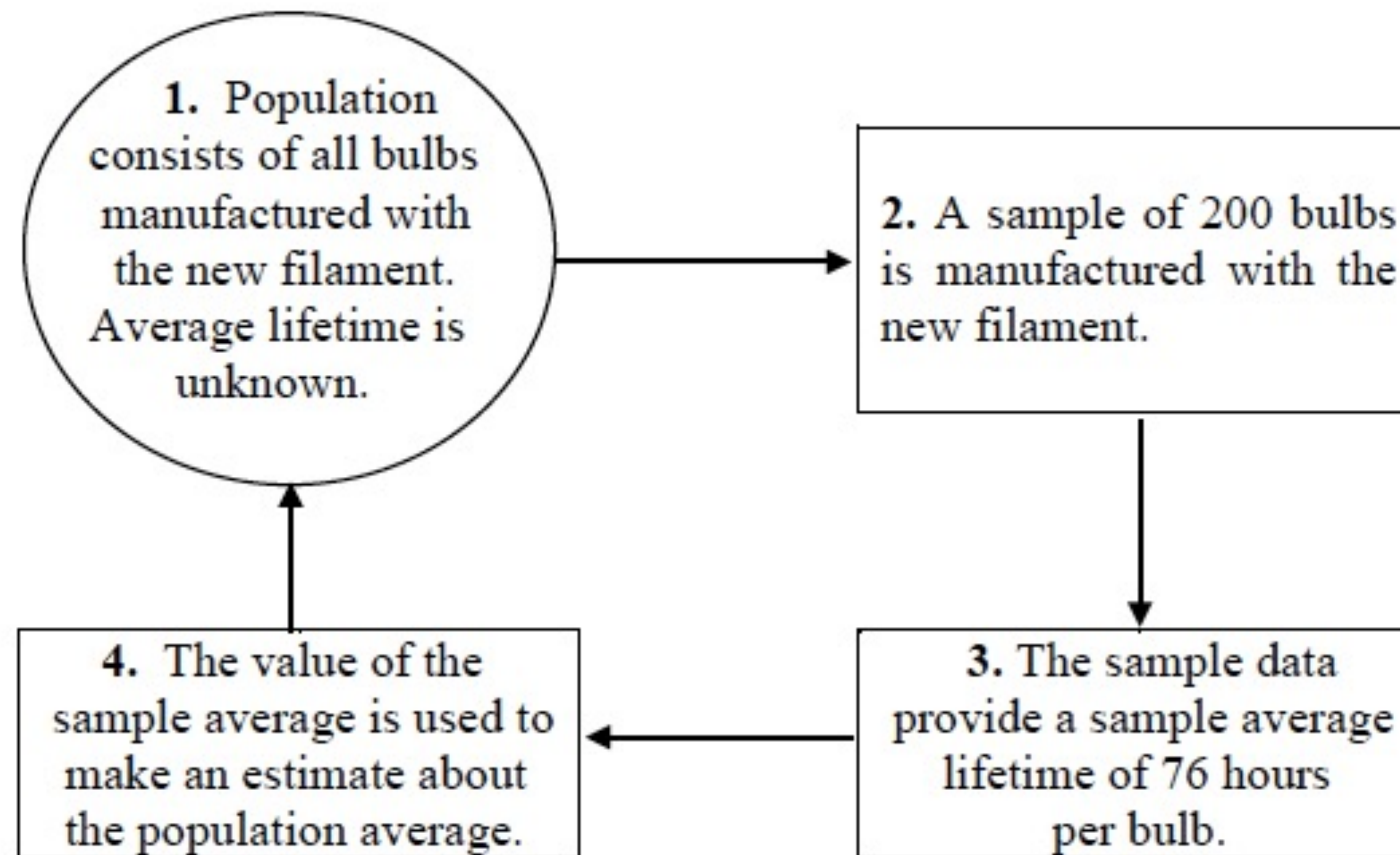
Whenever statisticians use a sample to estimate population characteristics of interest, they usually provide a statement of the quality population, associated with the estimate. For the Norris example, the statistician might state that the estimate of the average lifetime for the population of new lightbulbs is 76 hours with a margin of error of ± 4 hours. Thus, an interval estimate of the average lifetime of all lightbulbs produced with the new filament is 72 hours to 80 hours. The statistician can also state how confident he or she is that the interval 72 hours to 80 hours contains the population average.

1. Population consists of all bulbs manufactured with the new filament. Average lifetime is unknown.

2. A sample of 200 bulbs is manufactured with the new filament.

3. The sample data provide a sample average lifetime of 76 hours per bulb.

4. The value of the sample average is used to make an estimate about the population average.

9.6 Types of Statistics

- **Descriptive**
  - Used to summarize a collection of data in a clear and understandable way

- **Inferential**
  - Used to draw inferences about a population from a sample
  - "generalize to a larger population"
  - Common methods used
    - Estimation
    - Hypothesis testing