[1] DEFINITIONS AND TERMINOLOGY

1.1] Give examples to illustrate that we may all be naturally born statisticians.

1.2] Statistics
Briefly stated, statistics may be described as a branch of mathematical science that studies the methods of collecting, summarizing, organizing, analyzing, presenting, describing, interpreting, explaining data in a manner easy to understand by any audience (even by those with little or no background in statistics).

1.3] Data
Data are sets of observations, measurements, counts, results of experiments, survey responses, etc., obtained from some or all objects (human or nonhuman) being studied.

1.4] Elements, Elementary Units, Experimental Units, Observational Units, Subjects
The objects (human or nonhuman) from which data are obtained for a statistical study. Many times, when the objects are people, we call them subjects. Otherwise, we call them elements, experimental units, elementary units, or observational units.

1.5] Population
The complete set of all objects (elements, elementary units, experimental units, elementary units, subjects) being studied or analyzed. This definition is what some considered the definition of what is called the frame. In such a case, the population is the complete set of the data collected from all members of the frame and the frame is the complete set of all objects (elements, elementary units, experimental units, elementary units, subjects) being studied or analyzed. In here, we will not make any specific distinction, so the terms population and frame may be used interchangeably.

1.6] Statistical Study
The application of statistical methods to study, to analyze, or to learn about particular characteristics, properties, or attributes of a specific population (called target population)

1.7] Census
The process of collecting data from every member (subject or object) of the population (or frame) to be studied.

For many reasons (time, cost, destruction of elements, inaccessible subjects/objects, fatigue, large or infinite populations, ..., etc.) the target population (or frame) may not always be completely accessible. Therefore, it is not always possible or practical to conduct a census. In this kind of situation, a sample from the target population is used (instead of the entire population).

1.8] Sample
A nonempty subset of objects of the target population being studied and it cannot be the entire population. Of course, the selected sample must be a reliable representative of the population being studied because the results of the statistical analysis performed on the sample are used to make reliable inferences, forecasts, predictions, decisions, generalizations, estimates, recommendations, conclusions about characteristics of the relevant population. Sample statistics are customarily used to test claims (hypotheses) about population parameters.
1.9] **Parameters vs Statistics**
A measure that describes characteristics of a population is said to be a *parameter*. A measure that describes characteristics of a sample is said to be a *statistic*. Customarily, we use a sample statistic to estimate a corresponding population parameter. Here, we must not confuse a *statistic* (plural: *statistics*) with the title of this course: *Statistics*

**Example 1**
An academic dean is interested in finding the mean GPA of all students currently attending her university. She accesses the records of all such students, obtains their GPA’s, and computes the mean GPA of 3.07

a) Describe the population. What would be the frame?
b) Was a census conducted? Explain
c) Is the mean GPA of 3.07 a statistic or a parameter? Explain

**Example 2**
To have an estimate of the mean GPA of all 14,000 students currently attending her university, an academic dean selects a random sample of 1000 students currently attending her university, obtains their GPA’s, and computes a mean GPA of 3.04

a) Describe the population. What would be the frame?
b) Was a census conducted? Explain
c) Is the mean GPA of 3.04 a statistic or a parameter? Explain

**Example 3**
In each part, decide whether a statistic or a parameter is being used: 
A) 42% of all supermarket customers pay cash. 
B) In a randomly selected group of 50 customers of a particular supermarket, 48% paid cash. 
C) In a random sample of 120 murder cases, 57% of the murders were alcohol related. 
D) 54% of all murders are alcohol related.

1.10] **Steps of a Statistical Study**
Typical steps to follow when conducting a statistical study:

1.10.1] State the specific purpose of the study. This includes identifying the target population being studied and clearly stating what is to be learned about it.

1.10.2] Decide whether or not it is practical or suitable to conduct a census.

1.10.3] If a census is practical or suitable, 
   a) conduct a census to collect all relevant data from every member of the target population (or frame).
   b) Summarize, organize, analyze, present, interpret, construct models (mathematical equations relating two or more variables)

1.10.4] If a census is not practical and/or not suitable, skip step 1.10.3] and
   a) Choose a *representative sample* from the target population.
   b) Collect raw data from the sample and summarize, organize, present, interpret these data, and compute statistics of interest.
   c) Use the sample statistics to infer the corresponding population parameters.

1.10.5] Draw conclusions, describe what you have learned, and determine or verify whether or not you reached the objective.
1.11] **Representative Sample**: A sample that (at least approximately) exhibits, reflects, or reproduces most characteristics of the target population. In order for inferential statistics results to be reliable, the sample(s) obtained from the target population must be a representative sample(s).

1.12] **Biased Sample**: A sample selected in a way that favors the inclusion (or exclusion) of particular subjects or elements of the target population. This usually forces a particular result, which (of course) is not valid (or not reliable). In this case, we say also that the statistical study suffers from bias. Bias may also arise from not being able to obtain all the required data or from people who answer survey questions untruthfully, etc.

**Example 4**

a) The 1936 presidential election pitted Alfred Landon (Republican governor of Kansas) against the incumbent President, Franklin D. Roosevelt (Democrat). The magazine *Literary Digest* was one of the most respected magazines of the time and had a history of accurately predicting the winners of presidential elections since 1916. Explain the facts that lead *Literary Digest* to the wrong prediction and the facts that lead to the *Gallup Poll* to the correct prediction of the winner of the 1936 presidential election. Incorporate the concepts of representative sample and biased sample.

b) A basketball fan wants to find the most popular basketball team in the U.S. She randomly selects 400 people from Miami, Florida, and asks them “what is your favorite U.S. basketball team?” Is this a representative sample? **Explain.** If you answer “no,” what would be a “reliable” representative sample for this study?

**Example 5**

An investigator is interested in a statistical study of the age and yearly salary of employees working at CG Corporation. She selects a representative sample of four CG employees and records the data shown in the table at right:

<table>
<thead>
<tr>
<th></th>
<th><strong>Employee</strong></th>
<th>Age</th>
<th><strong>Salary</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) What observation was obtained from Xin?</td>
<td>Xin</td>
<td>40</td>
<td>$150,000</td>
</tr>
<tr>
<td>b) What is the data of this statistical study?</td>
<td>Amal</td>
<td>36</td>
<td>$53,000</td>
</tr>
<tr>
<td>c) Describe the frame of this statistical study</td>
<td>Sy</td>
<td>52</td>
<td>$48,000</td>
</tr>
<tr>
<td>d) Describe the population of this statistical study</td>
<td>Xin</td>
<td>40</td>
<td>$150,000</td>
</tr>
<tr>
<td>e) List the elements of this statistical study</td>
<td>Zoyla</td>
<td>33</td>
<td>$66,000</td>
</tr>
<tr>
<td>f) Was a census conducted? <strong>Explain.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example 6**

An investigator is interested in the age, gender, nationality, marital status, career path, number of credits completed, and GPA of all students at *U* university. She randomly selects four students from *U* university and records the following data:

<table>
<thead>
<tr>
<th><strong>Student</strong></th>
<th><strong>Age</strong></th>
<th><strong>Gender</strong></th>
<th><strong>Nationality</strong></th>
<th><strong>Marital St.</strong></th>
<th><strong>Career</strong></th>
<th><strong>No. Credits</strong></th>
<th><strong>GPA</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
<td>18</td>
<td>Female</td>
<td>Canada</td>
<td>Single</td>
<td>Biology</td>
<td>16</td>
<td>3.62</td>
</tr>
<tr>
<td>Ben</td>
<td>26</td>
<td>Male</td>
<td>U.S.</td>
<td>Married</td>
<td>Business</td>
<td>36</td>
<td>3.75</td>
</tr>
<tr>
<td>Rosa</td>
<td>34</td>
<td>Female</td>
<td>Panama</td>
<td>Married</td>
<td>Nursing</td>
<td>42</td>
<td>3.66</td>
</tr>
<tr>
<td>Wo</td>
<td>21</td>
<td>Male</td>
<td>China</td>
<td>Single</td>
<td>Engineering</td>
<td>18</td>
<td>4.00</td>
</tr>
</tbody>
</table>
a) What observation was obtained from Ben? b) What is the data of this study? c) Describe the frame of this statistical study; d) Describe the population of this statistical study; e) List the elements of this statistical study; f) Was a census conducted? Explain

Example 7: Consider the following study:
A U.S. Fish and Wildlife Service investigator wants to update the current list of species of animals and plants that are either Endangered (E) or threatened (T) and that are believed or known to occur in the state of Florida. She selects a random sample of 610 Florida species of animals and 370 Florida species of plants and finds 26 species of animals and 11 species of plants that are believed or known to occur in Florida and are threatened (T). She also finds 44 species of animals and 48 species of plants (also believed or known to occur in Florida) that are endangered (E).

a) Describe the frame and the population of the study.
b) Was a census conducted in the study? Explain.

It turns out that statistics involves more than collecting/organizing/summarizing/describing/presenting/interpreting/analyzing data. It is actually divided into two sub-branches: **descriptive statistics** and **inferential statistics**.

1.13] **Descriptive Statistics**: This is the sub-branch of statistics in charge of the methods of collecting, organizing, summarizing, analyzing, describing, presenting, interpreting data (sample or population data) in a manner easy to understand by any audience (even by those with little or no background in statistics).

1.14] **Inferential Statistics**: This is the sub-branch of statistics in charge of the methods of predicting, forecasting, estimating, making inferences, generalizations, conclusions, recommendations, and testing hypotheses (claims) about characteristics (**parameters**) of a target population, based on results obtained from analysis of **representative samples** collected randomly from that same population.

Example 8: In parts a)-e), state whether descriptive statistics or inferential statistics is illustrated

a) The owner of a local store reported that 81% of her customers pay with credit cards. She concluded this after a sample of 100 of her customers showed that 81 paid with credit cards;
b) In a sample of 100 customers of a local store, 81% paid with credit cards;
c) In 2009, a total of 420,000 children were newly infected with the HIV/AIDS virus;
d) In a random sample of 900 U.S. adult women (over 20), it was found that the mean height was 63.8 inches. This was used to conclude that the average height of adult women in the U.S. is 63.8 inches;
e) In a sample of 600 nannies, it was found that 12 were *mannies*. Therefore, 2% of the selected sample consisted of *mannies*;
f) the average supermarket carries 46,852 items.

1.15] **Two Types of Statistical Studies**

A] **Observational Study**: This is a statistical study in which the researcher simply observes the natural behavior of the experimental units in some environment and collects the relevant data without attempting to modify or influence the characteristics of the experimental units. The researcher has no control over (and does not attempt to control) characteristics of the experimental units.
**Experimental Study** (or *Experiment*)

This is a statistical study in which the researcher has complete control over the characteristics of the experimental units and can modify or influence the characteristics by applying some *treatment* being tested on the experimental units in order to determine the effects of the treatment. Customarily, data are collected from the experimental units before the treatment is applied and after the treatment is applied. In many cases, the set of experimental units is divided into two subsets (groups) called:

**B1) the treatment group:** this group consists of the experimental units that receive the *treatment* being tested (these are the modified subjects or objects).

**B2) the control group:** this group consists of the experimental units that do not receive the *treatment* being tested. The members of the control group receive what is called a *placebo*, which is a substance or object that lacks the active ingredients of the treatment, but it looks or feels identical to the treatment and the experimental units cannot distinguish between whether they are receiving the placebo or the actual treatment.

Data are obtained from each group in order to determine if differences exist between the results from the *treatment group* versus the results from the *control group*. The two groups must be selected randomly and share similar characteristics.

**PLACEBO EFFECT:** This refers to the situation in which members of the control group (the placebo group) improve their symptoms without actually receiving the treatment being tested (this may happen because the members truly believe they are receiving a useful treatment).

**Example 9:** Decide whether an **observational study** or an **experiment** was conducted:

a) In a survey of 360 sport fans, 105 enjoy basketball more than any other sport, 96 enjoy football, 46 enjoy baseball, 48 enjoy soccer, 36 enjoy tennis, and 24 enjoy golf.

b) To test the effectiveness of a new influenza vaccine (flu shot), 400 children over 2 years of age were given the new vaccine, while 366 children (over 2) were given a placebo.

c) A dietary researcher wants to determine whether low-fat restricted-calorie diets are safer and more effective than low-carbohydrate non-restricted-calorie diets. She used a random sample of 984 participants (ages 40 to 65) who needed a diet. Half of the participants were randomly assigned to a low-fat restricted-calorie diet and the other half to a low-carbohydrate non-restricted-calorie diet for a period of one year.

d) In a study conducted by the American Heart Association (AHA), 10,000 people (ages 25 to 50) were asked what contributes more to their happiness: money or love (from family or from a loved partner). 83% responded that it was love.

**REMARK:** Experiments are often preferred over observational studies. However, in some cases, experimental studies are impossible, inappropriate or unethical. For example, medical data suggest that second-hand smoke (from cigarettes, cigars, pipes, etc.) causes respiratory illness in children and people in general. Due to all risks associated with tobacco use, it would be unethical to conduct experiments in which a group of parents (for example) are forced to smoke (cigars, cigarettes, pipes) and their children are forced to breath in that environment (inhale the smoke and wait for the effects of the smoke in their health. In this case, an observational study using (for example) medical records (from doctors, clinics, etc.) would be appropriate and ethical.
1.16 Reasons for Sampling: Recall that sampling is the process of selecting samples. Reasons for sampling include:

A) **Time**: When results are needed quickly or in a very short amount of time, analysis of an entire population may take too long to meet the deadline, especially if the population is relatively large.

B) **Cost**: Time and money follow each other in many cases. Again, when the target population is relatively large, more personnel, or more equipment, or long overtime hours of work may be necessary to complete a statistical study of the entire population (especially if time is restricted). This may lead to additional expenses required to pay for the additional staff or equipment or for the overtime hours of work.

C) **Destruction of Elements**: Many populations consist of elementary units whose use results into natural wearing or destruction (examples: tires of vehicles, light bulbs, batteries, etc.). If a statistical study involves the use of these elementary units and the entire population is analyzed, then the manufacturers of such objects will end up with all their production destroyed. For example, if a researcher wants to find the number of miles that a particular brand of tires last and s/he uses all the produced tires of this particular brand, then at the end of the study, the manufacturer will end up with all her/his production destroyed because the tires must be installed on vehicles and used until they completely wear out in order to record the number of miles that they last.

D) **Inaccessible Elements**: In some statistical studies, many elementary units from the target population may not be accessible. For example, near a presidential election, many television networks, radio networks, magazines, newspapers, etc., attempt to predict exactly who the new president will be among all candidates. The prediction of exactly who the new president will be may require knowing the preferred candidate from all members of the population of voters. This might not be possible (why?)

E) **Fatigue**: Again, if the target population is relatively large, a statistical study may require long hours of work (especially when results are needed within a short period of time). This may cause fatigue, exhaustion, anxiety, desperation to the researcher, who probably wants to finish quickly and go home. Fatigue, exhaustion, anxiety, desperation, and all similar factors combined together may cause brain inefficiency to the researcher. As a result, s/he may make many types of mistakes that may lead to wrong computations, conclusions, decisions, recommendations, etc.

F) **Infinite Populations**: If the population contains an infinite number of elements, it would be virtually and physically impossible to collect an infinite set of data and conduct a study on an infinite number of elementary units.

**NOTE**: Since statistical studies are customarily performed on samples with the objective of making inferences, forecasts, predictions, estimates, generalizations, conclusions, recommendations, and testing hypotheses (claims) about characteristics (or parameters) of the target population, then in order for the results to be reliable, the chosen sample must be a representative sample obtained from that target population.
1.17] **Sampling**: Recall that the process of selecting a sample from a target population is called *sampling*. Two major groups of sampling procedures are: *probabilistic sampling* and *non-probabilistic sampling*.

**a) Probabilistic Sampling**: Samples are obtained in a way that every member of the target population has non-zero *chances (probability)* of being selected for the sample.

**b) Non-Probabilistic Sampling**: Samples are obtained either based on how easy and how convenient the data can be obtained or based on the opinion/experience/judgment of someone who is knowledgeable/expert/familiar with the population being studied.

1.18] **Some Probabilistic Sampling Procedures**

**a) Random Sampling**: Samples are obtained in such a way that (individually) every element of the population has the *same chance* of being selected for the sample. Random sampling provides an efficient method of obtaining a reliable representative sample.

**b) Simple Random Sampling**: Samples of a given size \( n \) are obtained in such a way that every sample of size \( n \) has the *same chance* of being selected for the statistical study. Notice that simple random sampling is a generalization of random sampling. Therefore, we may consider random sampling as a simple random sampling procedure with \( n = 1 \).

**c) Systematic Random Sampling**: From a population of size \( N \), a sample of size \( n < N \) is obtained as follows: **a)** Number the elements from 1 to \( N \); **b)** Compute \( k = N/n \), rounded down to an integer; **c)** Randomly select one element from the first \( k \) elements and then select every \( k^{th} \) element thereafter.

**d) Stratified Random Sampling**: Samples are obtained by first dividing the population into non-overlapping subsets (called *strata*) according to some characteristic like nationality, gender, occupation, etc. Once all the strata are formed, obtain a random sample (or a simple random sample) from each stratum.

**e) Cluster Sampling**: A *cluster* is a subset of elements that are (geographically) located near one another. When in a population, subsets of elements are considered to be too far from one another, then the population is divided into clusters and a number of clusters is selected by random sampling. Once the selection is done, the sample is obtained by conducting a complete census of each selected cluster.

**Example 10**: if we needed a survey of Miami college students, we may randomly select five colleges within the Miami area and obtain data from every student in each of the selected colleges.

**NOTE**: If instead of conducting a census, a sample is obtained from each of the selected clusters, then the procedure is called *two-stage cluster sampling*.

**f) Randomized Response Sampling**: Samples are obtained in a way that encourages individuals to answer sensitive questions truthfully without exposing the identity of the respondents. This method is used when obtaining samples that involve asking sensitive questions that might make respondents embarrassed, uncomfortable, or compromised, so their responses may not be truthful because they may be incriminating or socially undesirable. In most cases, the procedure involves a randomization instrument.
Example 11: If a study on abortion requires a survey in which women are asked how many times they have had an abortion, the survey may be conducted by asking this sensitive question \( S \) together with a non-sensitive question \( N \), as follows:

\[
S: \text{How many times have you had an abortion in your lifetime?}
\]

\[
N: \text{How many magazines do you subscribe to?}
\]

The randomization instrument could be a coin. The respondent flips the coin and if the outcome is Heads, s/he answers (say) question \( S \), but if the outcome is Tails, s/he answers question \( N \). The interviewer receives a response but does not know which question was answered. Even though there is no way to guarantee a truthful answer, this supposedly encourages a truthful response.

Example 12: \( S \) represents “Have you ever used illegal drugs?” and \( N \) represents “Is your birthday between August and December?” The respondent is given a bag containing two black marbles and two white marbles and is instructed to randomly pick one marble from the bag (without showing it to the interviewer). If the selected marble is black, the respondent answers (say) question \( S \) and if the selected marble is white, the respondent answers question \( N \). Again, the interviewer receives a response but does not know which question was answered.

g) Sequential Sampling: In this procedure, the sample size is not known before the sample is collected because the size depends on the results obtained. This is most commonly used when a statistical study involves quality control at a production or manufacturing plant. The researcher must devise a “rule of thumb” that will determine, say, when a manufacturing or production process is not running properly. For example, suppose that it is decided that a process is not running properly if 4 out of 20 outcome items turn out to be defective. Then, if we are sampling 20 items and the first 17 turn out good, the sampling procedure stops at the 17\(^{th}\) item and the conclusion is that the process is running properly. On the other hand, if 4 of the first 5 items turn out defective, then the sampling procedure stops at the 5\(^{th}\) item and the conclusion is that the process is not running properly.

1.18 Some Non-Probabilistic Sampling Procedures

h) Convenience Sampling: Samples obtained in such a way that the elementary units selected for the sample are those that are readily available or easily (conveniently) accessed. For example, when a radio station (or television station) asks listeners (or viewers) to contact the station (by telephone, email, texting…) in order to vote on a particular issue, the station is obtaining a convenience sample.

i) Judgment Sampling: Samples are obtained in such a way that the elementary units (or the observations) selected for the sample are based on the judgment, opinion, recommendation, experience, intuition, credentials, and discretion of an expert in the field relevant to the study or someone who is very familiar or has great experience with the characteristics of the elementary units in the population.
Example 13: Name the sampling methods being used in the following cases: 

A] A TV reporter collects opinions on a recent presidential speech by interviewing customers coming in and out of a supermarket. 

B] A business manager analyzes the stock market by examining changes in the Dow Jones Industrial Index, which is a weighted mean of 30 large publicly owned companies that she chooses. 

C] An inspector checks a batch of 300 recently built computers. She obtains a random sample of 20 computers by first numbering the computers from 1 to 300. Then, she randomly selects a number from 1 to 15 and checks that numbered computer for defects plus every 15th computer thereafter. 

D] To estimate the mean age of students at a particular college, a researcher puts the students’ files in alphabetical order and selects every 50th file. 

E] To estimate the mean GPA of students at a particular college, a researcher forms two groups (females, males) and selects every 50th file from the females and every 50th file from the males. 

F] A political analyst randomly selects three congressional districts and sends a questionnaire to all registered voters in those districts. She later receives responses from 30% of the voters who received the questionnaires. 

G] Students in your class are divided into science majors and non-science majors. A random sample of 4 students is then drawn from each of the two groups. 

H] To estimate the number of hours that residents spend studying, an investigator chooses three dormitories at random and in each dormitory, she chooses one floor (randomly) and interviews all the students on that floor. 

I] A college dean is researching the proportion of college students enrolled as math majors. She accesses the records of all the students at her college and selects every 20th student. 

J] A statistics consultant conducts a political survey among students at a particular university and interviews a group of 60 students who she believes know a lot about politics. 

K] A researcher stands at the entrance of a crowded supermarket and asks the first 100 people who enter to fill out a survey. 

L] An instructor of a large class of 300 students obtains a sample of students by selecting every 8th name from the class roster. 

M] Using a table of random numbers, choose 30 elementary units from a population of 600. 

N] From a group consisting of 10 plumbers, 6 electricians, and 4 architects, randomly select 2 plumbers, 2 electricians, and 2 architects. 

O] From a population consisting of engineering majors, biology majors, statistics majors, business majors, psychology majors, and English majors, randomly choose three majors. If the total number of students in those three categories is less than ten, sample all of them. If the total is more than ten, randomly choose a subsample of ten.

1.19] Variables: a symbol that may assume more than one value (numerical or non-numerical). The values of variables describe, measure, or represent characteristics, properties, or attributes of the elementary units of a statistical study.

Example 14: 

a) In Example 5, what are the variables of the statistical study? 

b) In Example 6, what are the variables of the statistical study? 

NOTE: The counterpart of a variable is a constant (a symbol that assumes exactly one value). Examples: a) under “normal” conditions, the boiling point of water is 100 degrees Celsius (equivalently, 212 degrees Fahrenheit), which is a constant; b) the speed of light in vacuum (300,000 kilometers per second)
1.20] Classification of Variables (and data)

A) by the nature of the values that they assume: qualitative, quantitative
B) by the number of values that they assume: dichotomous, trichotomous, multinomial
C) by levels of measurement: nominal, ordinal, interval, ratio

A) By the nature of the values that they assume

**Qualitative Variables**: These are variables with values that are non-numerical by nature. The values represent characteristics, properties, or attributes all described by names/labels/identification-codes/categories, none of which correspond to actual physical quantities obtained by counting or by measuring. Sometimes, values may consist of numbers that are intended to be identification codes, but do not correspond to actual physical quantities and mathematical computations (addition, multiplication, ratios, average) are meaningless or impossible. **Examples**: social security numbers, personal identification numbers (PIN).

**Example 15**: a) Classification of college students based on number of college-level credit-hours earned (Freshman: 0-29; Sophomore: 30-59; Junior: 60-89; Senior: 90 or more; Special: non-degree seeking students); b) Threat level (Severe: red; High: orange; Elevated: yellow; Guarded: blue; Low: green); c) Sex (Female, Male); d) Political Party Affiliation (Democrat, Republican, Independent, Other); e) Course Letter Grades (A, B, C, D, F); f) Student Number; g) Bank Account Number

**Qualitative variables** may be further classified as **nominal** and **ordinal**.

**Nominal**: These qualitative variables are characterized by values that consist of names/labels/identification-codes/categories, but the values are not comparable. So, they cannot be arranged in any order (except for alphabetical order) and cannot be ranked (high to low, smallest to largest…). Mathematical computations (sum, product, ratio, average) with values of nominal variables are either impossible or meaningless.

**Ordinal**: These qualitative variables are characterized by values that consist of names/labels/categories, but the values are comparable. They can be arranged in some order or rank scheme (high to low, smallest to largest). Mathematical computations (sum, product, ratio, average) with values of ordinal variables are either impossible or meaningless.

**Example 16**: In the final 2015 NCAA College Football Rankings, Clemson was ranked first (rank 1), Alabama second (rank 2), Michigan State third (rank 3), Oklahoma fourth (rank 4), etc. The ranks 1, 2, 3, 4, … show an ordering, but the arithmetic differences between them is meaningless. **Example**: the arithmetic difference of rank 2 minus rank 1 is 1, but the 1 is meaningless because it is not really a quantity that can be compared to other similar differences such as the arithmetic difference of rank 4 minus rank 3, which is also 1. The difference between Clemson (rank 1) and Alabama (rank 2) cannot be quantitatively compared to the difference between Michigan State (rank 3) and Oklahoma (rank 4). We may compare them relatively by other non-numerical methods of describing their differences, but not quantitatively by using numerical magnitudes.

**Example 17**: In **Example 15**, which variables are nominal? Ordinal?
Quantitative Variables: These are variables with values that are numerical by nature and the values represent characteristics, properties, or attributes, all described by numbers that correspond to actual physical quantities obtained by counting or by measuring. Mathematical computations (sum, product, ratio, averages) with values of qualitative variables are possible and have meaning.

Example 18: a) The number of college-level credit-hours earned by college students; b) The number of phone calls a person makes during one day; c) The length of time that children spend on computers each day; d) The height of a building; e) age

Quantitative variables may be further classified as discrete and continuous.

Discrete: These are quantitative variables characterized by values that are obtained by counting. So, the values are generally whole numbers. Discrete variables consist of either a finite number of values or an infinite countable number of values. The number of values may be infinite, but they must be countable, such as the number of rolls of a six-sided die before getting a “5”.

Continuous: These are quantitative variables characterized by values that are obtained by measuring (usually requiring a measuring instrument or device). So, the values are not generally whole numbers. Continuous variables consist of an infinite uncountable number of values, such as the lengths of objects between 50 feet and 200 feet.

Example 19: In Example 18, which variables are discrete? Continuous?

B) by the number of values that they assume

Dichotomous: variables are characterized by having exactly two possible values. Examples: pass/fail; female/male; heads/tails; employed/unemployed; smoker/nonsmoker

Trichotomous: variables are characterized by having exactly three possible values. Examples: yes/no/undecided; optimistic/pessimistic/indifferent; small/medium/large; agree/disagree/no-opinion; heavy/moderate/light

Multinomial: variables are characterized by having more than three possible values. Examples: heights of buildings; nationality; salary of employees; occupation.

C) by levels of measurement

Nominal: (same as the nominal classification of qualitative variables described above)

Ordinal: (same as the ordinal classification of qualitative variables described above)

Interval: These are quantitative variables characterized by values that can be arranged in some order or rank scheme (such as high to low, smallest to largest, etc.) and some mathematical computations (sum, product, average, no ratios) are possible and have meaning. However, there is no absolute zero (no natural zero to mark a starting point or to represent the absence of a property or characteristic). Example: Temperature (in Celsius scale or Fahrenheit scale). If the temperature of a room is 0°F, this does not mean that the room has no temperature, (so “0” does not represent the complete absence of temperature or the complete absence of heat in terms of the kinetic energy of molecules).
Also, temperature values may be negative, so there is no particular starting point for the values of temperature. Since there is no absolute zero, ratios of values are meaningless. The ratio of the values 15°F and 30°F is not the same as the ratio of the values 50°F and 100°F because no physical property of interest is preserved across the two ratios. So, it does not make sense to say that 100°F is twice as hot as 50°F or 15°F is half as cold as 30°F. However, differences between data values are possible and meaningful. So, the 10-degree difference from 5°F to 15°F has the same physical meaning (in terms of the kinetic energy of molecules) as the 10-degree difference from 80°F to 90°F.

**Ratio:** These are quantitative variables characterized by values that can be arranged in some order or rank scheme (high to low, smallest to largest, etc.), all mathematical computations (sum, difference, average, ratio) are possible and have meaning, and there is an **absolute zero** (a **natural zero** to mark a starting point and to represent the absence of a property or characteristic), so ratios of values are meaningful. **Example:** Number of children in a family. A family with zero children is a family with no children and a family with 4 children has twice as many children as a family with 2 children. The ratio of 4 being twice as much as 2 carries the same meaning as the ratio of 100 being twice as much as 50.

**1.21] Statistical Thinking:** The application of the science of statistics and logical reasoning (rational thought) to critically assess processes, data, and inferences (problems and solutions). We must realize and keep in mind that variation exists in all processes and in population data. When using statistical thinking, we must consider context, sources of data, whether population data or sample data were used, and (if sample data were used) what sampling method was used to obtain the sample. Statistical thinking does not require much ability to perform complicated mathematical calculations because nowadays we have the technology that gives us the tools and devices to perform such calculations in an effective manner. Statistical thinking involves critical thinking, the ability to determine whether the results make sense or not, and the ability to determine whether the results have statistical significance and practical significance. **An Example** (from McClave/Sincich, *Statistics*, Custom Edition for Florida International University, 12th Edition; page 15; Example 1.7)

**PROBLEM:** An article in the *New York Times* considered the question of whether motorcyclists should be required by law to wear helmets. In supporting his argument for no helmets, the editor of a magazine for Harley-Davidson bikers presented the results of one study that claimed “nine states without helmet laws had lower fatality rate (3.05 deaths per 10,000 motorcycles) than those that mandated helmets (3.38)” and a survey that found “of 2,500 bikers at a rally, 98% of the respondents opposed such laws.” Based on this information, do you think it is safer to ride a motorcycle without a helmet? What further statistical information would you like?
**SOLUTION:** You can use “statistical thinking” to help you critically evaluate the study. For example, before you can evaluate the validity of the 98% estimate, you would want to know how the data were collected. If a survey was, in fact, conducted, it is possible that the 2,500 bikers in the sample were not selected at random from the target population of all bikers, but rather were “self-selected.” (Remember they were all attending a rally – a rally, likely, for bikers who oppose the law.) If the respondents were likely to have strong opinions regarding the helmet law (e.g., to strongly oppose the law), the resulting estimate is probably highly biased. Also, if the biased sample was intentional, with the sole purpose of misleading the public, then the researchers would be guilty of *unethical statistical practice*.

You would also want more information about the study comparing the motorcycle fatality rate of the nine states without a helmet law to those states that mandate helmets. Were the data obtained from a published source? Were all 50 states included in the study, or were only certain states selected? That is, are seeing sample data or population data? Furthermore, do the helmet laws vary among states? If so, can you really compare the fatality rates?

**LOOK BACK:** Questions such as these led a group of mathematics and statistics teachers attending an American Statistical Association course to discover a scientific and statistically sound study of helmets. The study reported a dramatic *decline* in motorcycle crash deaths after California passed its helmet law.

### 1.22 Statistical Significance versus Practical Significance:

*Statistical Significance* occurs when the application of formal methods of statistical analysis leads to the conclusion that a result (a statistic) obtained from a sample of the relevant population is very unlikely to occur by chance. *An Example:* suppose you want to play a game in which your opponent tosses a coin. If the outcome is Tails, you pay $1 to your opponent. If the outcome is Heads, you receive $1 from your opponent. Before you officially begin the game, you want to make sure that the coin is well-balanced (so there is a 50% chance of getting Tails and a 50% chance of getting Heads). Your opponent and you agree to toss the coin 60 times before the game officially starts. If the result shows 59 Tails out of the 60 tosses of the same coin, this is statistically significant because the application of formal methods of statistical analysis leads to the conclusion that the result of 59 Tails in a sample of 60 tosses of the same coin is very unlikely to occur by random chance. So, the result from the sample is not enough evidence to conclude that the coin is well-balanced and you want to use a different coin! On the other hand, if the result shows 31 Tails out of the 60 tosses of the same coin, this is not statistically significant because the application of formal methods of statistical analysis leads to the conclusion that the 31 Tails in a sample of 60 tosses of the same coin is very likely to occur by random chance. So, the result from the sample is enough evidence to conclude that the coin is well-balanced and you want to use this coin to play the game!
Practical Significance occurs when common sense, logic, and/or reasoning (rather than statistical significance from the application of formal methods of statistical analysis) lead to the justification for using (or not using) a result (a statistic) obtained from a sample of the relevant population for practical purposes. So, statistical significance may be considered as more mathematically objective while practical significance may be considered as subjective and it could be based on other factors such as cost, requirements, goals, etc. Example (from Triola, Elementary Statistics, 12th Edition; page 9; Example 3): Statistical Significance versus Practical Significance:

In a test of the Atkins weight loss program, 40 subjects using that program had a mean weight loss of 2.1 kg (or 4.6 pounds) after one year (based on data from “Comparison of the Atkins, Ornish, Weight Watchers, and Zone Diets for Weight Loss and Heart Disease Risk Reduction,” by Dansinger et al., Journal of the American Medical Association, Vol. 293, No. 1). Using formal methods of statistical analysis, we can conclude that the mean weight loss of 2.1 kg is statistically significant. That is, based on statistical criteria, the diet appears to be effective. However, using common sense, it does not seem very worthwhile to pursue a weight loss program resulting in such relative insignificant results. Someone starting a weight loss program would probably want to lose considerably more than 2.1 kg. Although the mean weight loss of 2.1 kg is statistically significant, it does not have practical significance. The statistical analysis suggests that the weight loss program is effective, but practical considerations suggest that the program is basically ineffective.

Review Items

NOTE: In each item asking to describe or explain or list concepts, answer the question without any examples, graphs, diagrams. Just explain. Read the chapter carefully. Answers are imbedded in the theory described in this chapter.

1] Briefly describe what Statistics is about.
2] Explain the difference between Descriptive Statistics and Inferential Statistics.
3] Explain the difference between Elements, Subjects, and Data.
4] Explain the difference between Population and Census.
5] Explain the difference between Population and Sample.
6] Explain the difference between Representative Sample and Biased Sample.
7] Explain the difference between Parameter and Statistic.
8] Describe the five basic steps of a statistical study.
9] Explain the difference between Observational Study and Experimental Study.
10] Explain the difference between Treatment Group and Control Group.
11] Explain the difference between Treatment and Placebo.
12] Briefly describe the Placebo Effect.
13] Explain the difference between Sampling and Sample.
14] Briefly describe at least 6 reasons for sampling.
15] Explain the difference between Probabilistic Sampling and Non-Probabilistic Sampling.
16] Briefly describe the 7 probabilistic sampling procedures included in this chapter.
17] Briefly describe the 2 non-probabilistic sampling procedures included in this chapter.
18] Briefly describe the 3 ways to classify variables (or data) included in this chapter.
19] Explain the difference between **Qualitative Variables** and **Quantitative Variables**.
20] Explain the difference between **Nominal Data** and **Ordinal Data**.
21] Explain the difference between **Discrete Data** and **Continuous Data**.
22] Explain the difference between **Interval Level Data** and **Ratio Level Data**.
23] Briefly explain what **Statistical Thinking** is.
24] Briefly explain the difference between **Statistical Significance** and **Practical Significance**.

25] A researcher conducts a study on age and salary of employees at C Corporation. She takes a random sample of 4 employees and records the information shown in the table at right (names are fictitious).

<table>
<thead>
<tr>
<th>Employee</th>
<th>Age</th>
<th>$ Salary/yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amal</td>
<td>53</td>
<td>76,000</td>
</tr>
<tr>
<td>Nancy</td>
<td>26</td>
<td>54,000</td>
</tr>
<tr>
<td>Ben</td>
<td>35</td>
<td>63,000</td>
</tr>
<tr>
<td>Rose</td>
<td>43</td>
<td>48,000</td>
</tr>
</tbody>
</table>

a) Describe the target population of the study;  
b) List the subjects;  
c) Describe the data;  
d) What observation was obtained from Rose?  
e) Was a census conducted? Explain;  
f) List the variables;  
g) Which variables are qualitative?  
h) Which variables are quantitative?  
i) Which variables are continuous?  
j) Which variables are discrete?  
k) Which variables are nominal level?  
l) Which variables are ordinal level?  
m) Which variables are interval level?  
n) Which variables are ratio level?  
o) Which variables are dichotomous? trichotomous? multinomial?

26] An investigator conducts a study on price, miles per gallon (MPG), and transmission type of cars sold in the U.S.. She obtains a random sample 6 cars of different makes sold in the U.S. and record the information shown in the table at right.

<table>
<thead>
<tr>
<th>Make</th>
<th>$Price</th>
<th>MPG</th>
<th>Transm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ford</td>
<td>24,500</td>
<td>30</td>
<td>Auto</td>
</tr>
<tr>
<td>Toyota</td>
<td>22,000</td>
<td>32</td>
<td>Manual</td>
</tr>
<tr>
<td>Honda</td>
<td>24,000</td>
<td>34</td>
<td>Manual</td>
</tr>
<tr>
<td>Nissan</td>
<td>26,000</td>
<td>26</td>
<td>Auto</td>
</tr>
<tr>
<td>Chevy</td>
<td>23,500</td>
<td>28</td>
<td>Auto</td>
</tr>
<tr>
<td>BMW</td>
<td>39,000</td>
<td>27</td>
<td>Auto</td>
</tr>
</tbody>
</table>

a) Describe the target population of the study;  
b) List the elementary units;  
c) Describe the data;  
d) What was the observation obtained from the Ford vehicle?  
e) Was a census conducted? Explain;  
f) List the variables;  
g) Which variables are quantitative?  
h) Which variables are quantitative?  
i) Which variables are continuous?  
j) Which variables are discrete?  
k) Which variables are nominal?  
l) Which variables are ordinal level?  
m) Which variables are interval level?  
n) Which variables are ratio level?  
o) Which variables are dichotomous? trichotomous? multinomial?

27] In each part, classify the given variable as **quantitative**, **qualitative**, or **neither**.

   a) marital status;  
   b) income;  
   c) religion;  
   d) occupation;  
   e) number of stars in the sky

28] In each part, classify the given variable as **nominal**, **ordinal**, or **neither**.

   a) political affiliation;  
   b) speed of light in vacuum;  
   c) military rank;  
   d) size of pizza;  
   e) sex

29] In each part, classify the given variable as **discrete**, **continuous**, or **neither**.

   a) height of a building;  
   b) number of daily accidents at a manufacturing plant;  
   c) blood type;  
   d) amount of water needed to put out a fire;  
   e) speed of cars;  
   f) number of phone calls a person receives in one day
30. In each part, classify the given variable as **dichotomous, trichotomous, multinomial, or none of the above**:  
   a) condition of a product (damaged, undamaged);  
   b) number of rooms in a house;  
   c) melting temperature of iron;  
   d) state of alert (high, medium, low);  
   e) hamburger preference (pickles, no pickles);  
   f) year in college;  
   g) drinker type (light, moderate, heavy).

31. In each part, classify the given variable as **nominal, ordinal, interval, ratio, or none of the above**:  
   a) condition of a product (damaged, undamaged);  
   b) number of bedrooms in a house;  
   c) boiling temperature of water under normal conditions;  
   d) state of alert (high, medium, low);  
   e) student’s GPA;  
   f) course letter grade;  
   g) slope of a straight line.

32. Sets of data may be classified as: **univariate** (if observations consist of exactly one component, such as yearly salary);  
   **bivariate** (if observations consist of exactly two components, such as yearly salary and age);  
   **multivariate** (if observations consist of more than two components, such as yearly salary, age, occupation, nationality).

In each part, decide whether the described study requires a **univariate, bivariate, or multivariate** data set:  
   a) a study of the relationship between cholesterol level and high blood pressure in humans;  
   b) a study of the relationship between the price of a house, number of bedrooms, number of bathrooms, square footage, year in which it was constructed, location, whether the house has a garage or no garage, and whether it has a lake view or no lake view;  
   c) a study of the number of packages that UPS delivers daily;  
   d) a study of SAT scores, GPA, and nationality of college students.

33. An Academic Dean is interested in an analysis of GPA, nationality, year in college, and number of credits earned of students currently attending C College, which consists of 10,000 students. She took a sample of 4 students from C College and then collected data are in the table at right.  

<table>
<thead>
<tr>
<th>Student</th>
<th>GPA</th>
<th>Nationality</th>
<th>Year</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amal</td>
<td>3.01</td>
<td>Kuwait</td>
<td>Sophm</td>
<td>60</td>
</tr>
<tr>
<td>Nancy</td>
<td>2.54</td>
<td>Cuba</td>
<td>Junior</td>
<td>90</td>
</tr>
<tr>
<td>Sy</td>
<td>3.47</td>
<td>U.S.</td>
<td>Junior</td>
<td>93</td>
</tr>
<tr>
<td>Ty</td>
<td>2.99</td>
<td>Canada</td>
<td>Freshm</td>
<td>24</td>
</tr>
</tbody>
</table>

   a) Describe the population of the study  
   b) List the variables  
   c) List the subjects;  
   d) describe the observation obtained from Sy;  
   e) Describe the data;  
   f) Did the Academic Dean conduct a census? Explain;  
   g) Which variables are qualitative? Quantitative?  
   h) Which variables are discrete? Continuous?  
   i) Which variables are dichotomous? Trichotomous? Multinomial?  
   j) Which variables are nominal? Ordinal? Interval? Ratio?  
   k) Is the data set univariate? Bivariate? Multivariate?

34. A statistician is hired to study two area colleges: (Western Community College–WCC and Northern State University–NSU). The study is on the age, gender, ethnicity, number of children, marital status (S, M, D), GPA, and college attended (WCC, NSU). She obtains a random sample of 4 students. The collected data are:

<table>
<thead>
<tr>
<th>Student</th>
<th>Age</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>Children</th>
<th>Status</th>
<th>GPA</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamal</td>
<td>19</td>
<td>Male</td>
<td>African American</td>
<td>0</td>
<td>Single</td>
<td>4.00</td>
<td>WCC</td>
</tr>
<tr>
<td>Emilia</td>
<td>25</td>
<td>Female</td>
<td>Hispanic</td>
<td>2</td>
<td>Divorced</td>
<td>3.95</td>
<td>NSU</td>
</tr>
<tr>
<td>Chang</td>
<td>20</td>
<td>Female</td>
<td>Asian</td>
<td>0</td>
<td>Single</td>
<td>3.90</td>
<td>NSU</td>
</tr>
<tr>
<td>Michael</td>
<td>47</td>
<td>Male</td>
<td>European Americ</td>
<td>3</td>
<td>Married</td>
<td>3.75</td>
<td>WCC</td>
</tr>
</tbody>
</table>
a) Describe the study; b) Describe the population of the study; c) List the variables; d) List the subjects; e) what was the observation from Chang? f) describe the data; g) Did the statistician conduct a census? Explain;  h) List the qualitative variables; the quantitative? i) Which variables are discrete? Continuous? j) Which variables are dichotomous? Trichotomous? Multinomial? k) Which variables are nominal? Ordinal? Interval? Ratio? l) Is the data set univariate? Bivariate? Multivariate? Explain why.

A researcher is interested in the analysis of nationality, major, GPA, gender, marital status (married, never married, divorced), year in college, age, and number of credits completed of students at M University. The researcher randomly selects five students from this university and collects the data summarized in the given table.

<table>
<thead>
<tr>
<th>Names</th>
<th>Nationality</th>
<th>Major</th>
<th>GPA</th>
<th>Gender</th>
<th>Year</th>
<th>Age</th>
<th>Crd</th>
<th>Marital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amal</td>
<td>Kuwait</td>
<td>Biology</td>
<td>3.36</td>
<td>Female</td>
<td>Sophomore</td>
<td>22</td>
<td>30</td>
<td>Never M</td>
</tr>
<tr>
<td>Sy</td>
<td>Canada</td>
<td>Business</td>
<td>3.01</td>
<td>Male</td>
<td>Junior</td>
<td>36</td>
<td>93</td>
<td>Married</td>
</tr>
<tr>
<td>Wu</td>
<td>China</td>
<td>Engineering</td>
<td>3.96</td>
<td>Male</td>
<td>Freshman</td>
<td>19</td>
<td>26</td>
<td>Never M</td>
</tr>
<tr>
<td>Nancy</td>
<td>United States</td>
<td>Psychology</td>
<td>2.74</td>
<td>Female</td>
<td>Freshman</td>
<td>18</td>
<td>19</td>
<td>Never M</td>
</tr>
<tr>
<td>Rosa</td>
<td>Panama</td>
<td>Philosophy</td>
<td>2.70</td>
<td>Female</td>
<td>Sophomore</td>
<td>28</td>
<td>96</td>
<td>Divorced</td>
</tr>
</tbody>
</table>

36] A researcher conducts a study of population size (from the 2010 census), population rank (in the U.S.), total area (in square miles), highest temperature recorded, name of a governor (as of 2011), governor’s gender (M or F), and her/his political affiliation (Republican, Democrat, Independent), for the states here in the United States. To do this, she picks a random sample of five states and the recorded data are shown below.

<table>
<thead>
<tr>
<th>STATE</th>
<th>Population (2010 census)</th>
<th>Population U.S. Rank</th>
<th>Area (sq. miles)</th>
<th>Hottest Temperature</th>
<th>Governor</th>
<th>Gender</th>
<th>Political Party</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>37,253,956</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>163,695.57</td>
<td>134°F</td>
<td>Jerry Brown</td>
<td>M</td>
<td>Democrat</td>
</tr>
<tr>
<td>Texas</td>
<td>25,145,561</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>268,580.82</td>
<td>100°F</td>
<td>Rick Perry</td>
<td>M</td>
<td>Republican</td>
</tr>
<tr>
<td>New York</td>
<td>19,378,102</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>54,556.00</td>
<td>84°F</td>
<td>Andrew Cuomo</td>
<td>M</td>
<td>Democrat</td>
</tr>
<tr>
<td>Florida</td>
<td>18,801,310</td>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td>65,754.59</td>
<td>109°F</td>
<td>Rick Scott</td>
<td>M</td>
<td>Republican</td>
</tr>
<tr>
<td>Illinois</td>
<td>12,830,632</td>
<td>5&lt;sup&gt;th&lt;/sup&gt;</td>
<td>57,914.38</td>
<td>117°F</td>
<td>Pat Quinn</td>
<td>M</td>
<td>Democrat</td>
</tr>
</tbody>
</table>

a) Describe the study; b) List the variables; c) List the experimental units; d) what observation was obtained from Texas? e) Describe the data; f) Describe the population; g) Did the researcher conduct a census? Explain;  h) Did the researcher use a sample? i) Which variables are qualitative? Quantitative? j) Which variables are discrete? Continuous? k) Which variables are dichotomous? Trichotomous? Multinomial? l) Which variables are nominal? Ordinal? Interval? Ratio? m) Is the data set univariate? Bivariate? Multivariate? Explain.
Answers To Selected Examples

Example 1: a) The complete set of all students attending the university (NOTE: the complete set of all students attending the university may be considered the frame. In such a case, the population would be the complete set of all GPA values collected from all the students in the frame); b) Yes, because data were collected from every member of the population (or frame); c) Parameter, because 3.07 is the mean GPA of the entire population of students. Example 2: a) The complete set of all 14,000 students attending the university (NOTE: the complete set of all 14,000 students attending the university may be considered the frame. In such a case, the population would be the complete set of all 10,000 GPA values collected from all the students in the frame); b) No, because data were collected from only 1000 students and not from every one of the 10,000 students in the population (or frame); c) Statistic, because 3.04 is the mean GPA of the sample of 1000 students and not the mean GPA of the entire population of 10,000 students.

Example 3: a) Parameter; b) Statistic; c) Statistic; d) Parameter. Example 5: a) Xin is 40 years old and earns $150,000 per year; b) the set \{36; $53,000\}; (52; $48,000); (40; $150,000); (33; $66,000\}; c) the complete set of all employees who work at CG Corporation; d) The complete set of all observations consisting of two components (age, salary) obtained from all CG employees. e) Amal, Sy, Xin, Zoyla; f) No. Data were obtained from only four CG employees, not from all CG employees.

Example 6: a) Ben is 26 years old, male, from the U.S., married, business major, with 36 completed credits, and a GPA of 3.75; b) the set \{(18; female; from Canada; single; biology major; 16 completed credits; 3.62 GPA); (26; male; from U.S.; married; business major; 36 completed credits; 3.75 GPA); (34; female; from Panama; married; nursing major; 42 completed credits; 3.66 GPA); (21; male; from China; single; engineering major; 18 completed credits; 4.00 GPA)\}; c) The complete set of all students attending U university; d) The complete set of all observations consisting of seven components (age, gender, nationality, marital status, career path, number of credits completed, GPA) obtained from all students at U university; e) Ann, Ben, Rosa, Wo; f) No, because data were obtained from only four students, not from all students attending U university.

Example 8: a) Inferential statistics; b) Descriptive statistics; c) Descriptive statistics; d) Inferential statistics; e) Descriptive statistics; f) Descriptive statistics. Example 9: a) Observational study; b) Experimental study; c) Experimental study; d) Observational study. Example 13: a) Convenience sampling; b) Judgment sampling; c) Systematic sampling; d) Systematic sampling; e) Stratified random sampling; f) Cluster random sampling; g) Stratified random sampling; h) Cluster random sampling; i) Systematic sampling; j) Judgment sampling; k) Convenience sampling; l) Systematic sampling; m) Random sampling; n) Stratified random sampling; o) Cluster random sampling. Example 14: a) the age and the salary of employees working at CG Corporation; b) age, gender, nationality, marital status, career path, number of credits completed, and GPA of the students at U university. Example 17: Nominal: c; d; Ordinal: a, b, e. Example 19: a) Discrete; b) Discrete; c) Continuous; d) Continuous; e) Continuous.

ANSWERS TO SELECTED REVIEW ITEMS

25]a) The complete set of all employees working at C Corporation; b) Amal, Nancy, Ben, Rose; c) A set of 4 observations with two components each: age and salary of C Corporation employees; d) Rose is 43 and earns $48,000/yr; e) No. The researcher obtained data from only 4 employees and not from the entire population of all C Corporation employees; f) Age, Salary/yr of C Corporation employees; g) None; h) Age, Salary/yr; i) Age, Salary/yr; j) None; k) None; l) None; m) None; n) Age, Salary/yr; o) None; None; Age, Salary/yr.

26]a) The complete set of all cars sold in the U.S.; b) the selected vehicles from Ford, Toyota, Honda, Nissan, Chevrolet, BMW; c) A set of 6 observations with three components each: dollar price, miles per gallon, and transmission type of cars sold in the U.S.; d) The price of the selected Ford is $24,500, the vehicle has a 30-MPG performance, and automatic transmission; e) No. The researcher obtained data only from a sample of 6 cars and not from the entire population of all cars sold in the U.S.; f) dollar price, miles per gallon, transmission type of cars sold in the U.S.; g) transmission type; h) dollar price, MPG; i) dollar price, MPG; j) None; k) transmission type; l) None; m) None; n) dollar price, MPG; o) transmission type; None; dollar price, MPG. 27]a) qualitative; b) quantitative; c) qualitative; d) qualitative; e) quantitative
A study of the nationality, major, GPA, gender, year in college, age, number of credits completed, and marital status of students attending M University; b) nationality, major, GPA, gender, year in college, age, number of credits completed, and marital status of students attending M University; c) Amal; Sy; Wo; Nancy; Rosa; d) Amal is from Kuwait; biology major; GPA 3.36; female; Sophomore; 22 years old; has 30 credits completed; has never married; e) A set of 5 observations, each with 8 components: nationality, major, GPA, gender, year in college, age, number of credits completed, and marital status of students attending M University; f) the set of all students attending M University; g) No census. The researcher
obtained data from only 5 students and not from *all* students attending *M* University; **h)** Yes, the researcher used a random sample of 5 *M* University students; **i)** Qualitative: Nationality, Major, Gender, Year in College, Marital Status; Quantitative: GPA, Age, Number of Credits Completed; **j)** Discrete: Number of Credits Completed; Continuous: GPA, Age; **k)** Dichotomous: Gender; Trichotomous: Marital Status; Multinomial: Nationality, Major, GPA, Year in College, Age, Number of Credits Completed; **l)** Nominal: Nationality, Major, Gender, Marital Status; Ordinal: Year in College; Interval: GPA; Ratio: Age, Number of Credits Completed; **m)** Since there are 8 variables (more than 2 variables), the data set is Multivariate.