_____is a scientific discipline that provides methods to organize and analyze data.

A _____is the set of all subjects or elements about which we are interested in making inferences.

A list containing all members of the population is referred to as a _____.

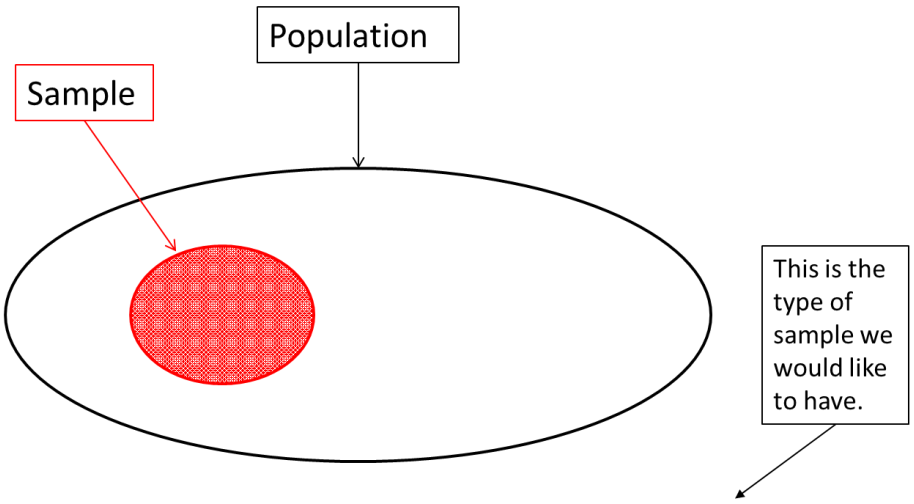
A strict definition of a _____is a survey that includes all the elements or units in the _____.

A _____is a subset of the population which is used to gain insight about the population. Samples are used to represent a larger group, the population.

_____is a group of statistical methods designed to monitor and control processes.

_____of Interest: The entire collection of individuals or objects about which information is desired

_____**:** a subset of the population, selected for study in some prescribed manner



Representative sample: sample which typifies the targeted population

Scientific Process

1.
2.
3. Observe or experiment and collect data (Methods)
4.
5. Make conclusions and recommendations

Do we have good measurements?

1. Is the concept under study adequately reflected by the proposed measurements?
2. Are the data measured accurately?
3. Is there a sufficient quantity of the data to draw a reasonable conclusion?

Recall that, a _____is the set of all subjects or elements about which we are interested in making inferences.

Populations have _____.

_____are facts about the population. Since parameters are descriptions of the population, a population can have many parameters.

Recall that, a _____is a subset of the population which is used to gain insight about the population. Samples are used to represent a larger group, the population.

From samples we get _____.

A _____is a fact or characteristic about a sample.

_____**:** branch of statistics that includes methods for organizing and summarizing data

_____: branch of statistics that involves generalizing from a sample to the population from which it was selected and assessing the reliability of such generalizations

_____**:** an estimate, prediction, or some other generalization concerning a population based on sample information

Sampling

_____**:** each sample of size of n has equal opportunity of being selected

_____**:** population is divided into sets of non-overlapping subgroups, a random sampling procedure is then performed on each subgroup (stratum-singular, strata-plural)

_____**:** population is divided into sets of non-overlapping subgroups, these clusters are then selected at random for inclusion in the sample

_____**:** a specific unit in the population is selected and then a specified kth value is used to select the remaining units for the sample

_____**:** samples which are easily available or convenient but do not necessarily represent the population of interest (voluntary response sampling)

Methods of Data Collection:

1.
2.
3.

_____**:** has at least one independent variable which is manipulated by the researcher while changes in the dependent variable are measured.

_____**:** Group which receives the experimental treatment.

_____**:** Group which does not receive the treatment. Used to make comparisons to the treatment group

_____**:** outcomes are measured but no attempt is made to affect the variable of interest.

_____**:** questionnaire is given to participants, answers to questions represent the data.

Elements of Experimental Design

_____: Certain information usually regarding the variables of interest is withheld from the participants or the researchers but not both

_____: Certain information usually regarding the variables of interest is withheld from the participants and researchers

_____: A treatment which does not affect the dependent variable. “a fake treatment”

_____: A variable which may be affecting the dependent which is not the independent variable. This can bias the measurement of dependent variables.

_____: Removing the effects of variables or making it so that the dependent variable can be measured without the affects of other variables.

_____: Participants are given a questionnaire with closed-form questions on a Likert-scale. The questions are then analyzed by the research to determine patterns in the data.

_____: Participants are observed while at least one dependent variable is measured. Researchers do not manipulate any experimental variables. Researchers simply measure a phenomenon as it occurs.

_____: Has an independent variable(s) and at least one dependent variable. The dependent variable is measured as changes occur in the independent variable.

_____: Type of experimental design that has a lack of randomization

_____: Treatment and control groups are selected without random assignment

_____: Participants are tested prior to beginning the experiment and after the experiment has concluded

_____: Type of randomized block design with two treatment conditions. Subjects are grouped into pairs across a variable and each one of the pair is randomly assigned to a different treatment.

_____: Type of experimental design that has a lack of randomization

_____: Treatment and control groups are selected without random assignment

_____: Measures are recorded for the entire experiment for all groups. Can have single-group or control-group design.

Collecting Data

_____**:** any characteristic which can change across a single individual or object or across individuals or objects

_____data: data or observations on a single attribute

_____data: data or observations on two attributes

_____ data: data or observations on three or more attributes

The _____is the variable of interest in an experiment.

An _____ is a variable that affects the variable of interest (response variable) in an experiment.

Level of measurement: classifies information according to its quality

The terms _____, _____, _____, and _____ are used to describe the level of measurement.

_____: data that only has a specific characteristic, “in name”

_____**:** data which can be ordered

_____**:** data which can be ordered and differences have meaning

_____**:** Ratios of two data points are meaningful

_____ (categorical data also sometimes referred to as nominal data) Examples might include names or labels

_____ (numerical) (Ordinal, Interval, or Ratio)

Quantitative data can be _____ (measurable) or _____ (countable)

Types of Bias and Error

Random Error: Error that can be attributed to chance, variations from one measurement to another which are not attributed to specific characteristics associated with the variable

Systematic Error: attributable to bias within the research design or measurement errors due to instrument failure

Sources of Systematic Error

Selection Bias: part of the population had no opportunity to be selected for the sample

Nonresponse Bias: (sometimes referred to as Response Bias) when researchers cannot obtain sufficient data from a sample to make an inference to the population

Measurement Error: inaccurate values in the data collection, may be due to survey questions, experimenter or completely ambiguous

Displaying Data

Frequency Distribution: A _____ summarizes data into classes and provides in tabular form a list of the classes along with the number of observations in each class.

Relative Frequency: The _____ of any class is the number of observations in the class divided by the total number of observations.

Cumulative Relative Frequency: The _____ is the proportion of observations in a particular class and all preceding classes.

Frequency Distribution Tables and Bar Graphs

A _____ provides information about the number of given measurements for values of a variable in table form.

Bar Graph: The _____ is a simple graphical display in which the length of each bar corresponds to the number of observations in a category. _____ can be used to represent discrete data or categorical data.

Example 1

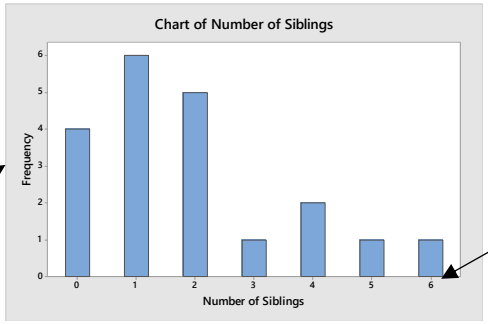
Twenty individuals were asked how many siblings each had. The results are in table 1.

Table 1

Number of Siblings	Frequency
0	4
1	6
2	5
3	1
4	2
5	1
6	1

_____ can be used to represent discrete data or categorical

Frequency of occurrence is listed on the y-axis



Categories are listed on the x-axis

How many individuals had no siblings?

How many individuals had from 2 to 4 siblings?

Example 2

Apple Inc. introduced its third generation iPad in March of 2012. After months of anticipation, owners were thrilled with their new iPads. Two of the main features that separated the 3rd generation iPad from its predecessors were an improved camera and a higher resolution display. In spite of the excitement of the launch, there were still things that owners did not like about the new device. The following table contains the responses from a survey of 30 new iPad owners when asked what they dislike about the iPad.

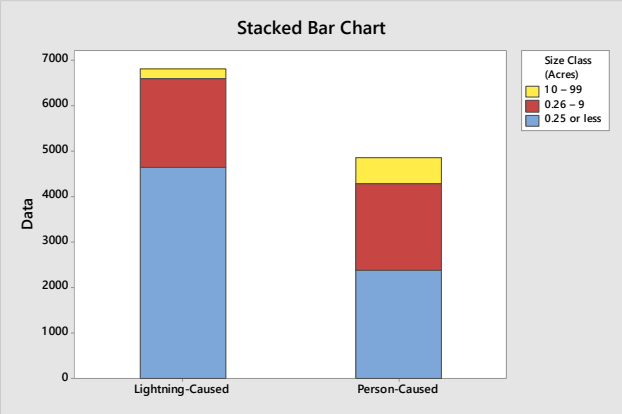
Table 3.4 – Frequency Distribution of Survey Responses

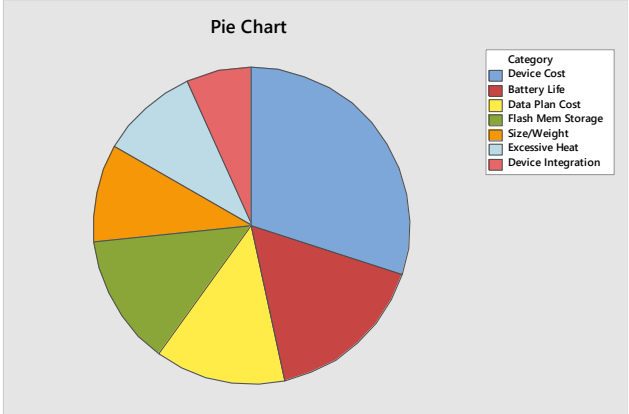
Response	Frequency	Relative Frequency	Cum. Rel. Frequency
Cost of Device	9	.30	.30
Battery Life Too Short	5	.17	.47
Cost of Data Plan	4	.13	.60
Flash Mem Storage	4	.13	.73
Size/Weight	3	.1	.83
Excessive Heat from Device	3	.1	.93
Device Integration	2	.07	1.00

Other Data Displays

_____ are bar charts where different categories are stacked on top of each other. These are useful when there are three components to the data (i.e. subcategories).

A _____ is a graphical representation of qualitative data in which the circle represents the total "pie" available, and the slices are proportional to the amount in each category. Each slice of the pie represents the proportion of total observations belonging to the category.



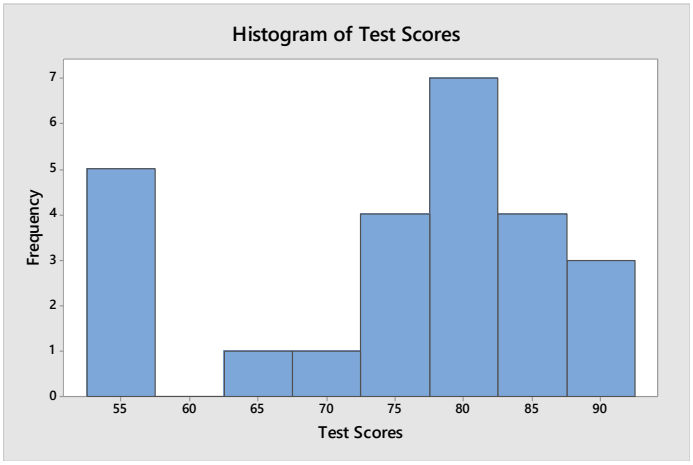


A _____is a bar graph of a frequency or relative frequency distribution in which the height of each bar corresponds to the frequency or relative frequency of each class. _____ represent continuous data.

Example 3

Suppose 25 students took a statistics test and received the following scores.

55	55	55	55	56	67	68	73	74	74	77	79	79	79
80	81	81	82	84	84	85	86	88	90	92			



How many students scored between 77.5 and 82.5?

How many students scored below 67.5?

How many students scored above 67.5?

Boxplots make use of the **five number summary.**

The five number summary is the Minimum, Lower Quartile (Q_1), Median, Upper Quartile (Q_3), and the Maximum.

Boxplots

Boxplots make use of the _____. The five number summary is the Minimum, Lower Quartile (Q_1), Median, Upper Quartile (Q_3), and the Maximum. To find the _____ subtract the minimum value from the maximum value. To find the _____ subtract the first quartile from the third quartile. _____ of the data lies between each value in the five number summary.