

Chapter 1: Introduction to Statistics

Section 1.2

Data – observations that have been collected

Examples: heights, weights, miles per gallon

Statistics – a collection of methods for planning experiments, obtaining data, then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on that data.

****Statistics is the science of making sense of data.**

Population – a complete collection of all elements to be studied

Example 1: the entire student body at Massasoit

Example 2: all residents of Cape Cod

Example 3: all DVDs produced by a given machine

Census – a collection of data from every member of the population

Example: Every ten years, the U.S. government surveys **all** of American citizens

Sample – a subcollection of members selected from a population

Example 1: A group of fifty students at Massasoit are asked how much they spent on textbooks in Fall of 2008.

Example 2: A group of 648 residents of the Cape and islands are asked if they support the windfarm.

Example 3: A sample of 50 DVDss are chosen to see if any of them contain defects.

Parameter – a number that describes a population.

Example 1: The **average** amount spent on textbooks by all Massasoit students in Fall 2007 is \$373.24.

Example 2: The **proportion** of all Cape Cod residents that support the windfarm is 42%.

Note: in these previous examples, I have fabricated the values for the average and proportion.

In statistical practice, the value of a parameter is typically not known because we cannot examine the entire population.

Instead, we choose a sample and compute a statistic.

Statistic – a number that can be computed from sample data. **In practice, we often use a statistic to estimate an unknown parameter.**

Example 1: Take a sample of 50 students, ask them how much they spend on textbooks, and compute the average (mean) for the sample.

Example 2: Take a sample of 648 voters on Cape Cod, ask them whether or not they support the windfarm, and find the proportion for the sample.

**Remember—A Parameter describes a Population and a Statistic describes a Sample

Types of data

Qualitative data—can be separated into different categories

Examples: gender, race, smoker (yes, no),

Quantitative data—consists of numbers representing counts or measurements

Examples: blood pressure (mm of mercury), incomes, heights, weights

There are two types of quantitative data:

Discrete—a countable number of possible values

Example 1: The number of tomatoes produced on a particular hybrid in a season

Example 2: The number of heads in ten tosses of a coin

Continuous—an infinite number of possible values

Example 1: The head circumferences of standard poodles

Example 2: The length of time it takes to complete an obstacle course

1-4 Design of Experiments

Observational study-- observe and measure specific characteristics, but we don't attempt to modify the subjects being studied.

Example: The Coast Guard regularly does observational studies relating to the use of life jackets by recreational boaters.

There are three types of observational studies:

Cross-sectional- data are observed, measured, and collected at one point in time.

Example: Homeless counts—sheltered and unsheltered

Retrospective study - data are collected from the past—used quite a bit in medical studies.

Example: Look at a group of people that have been admitted to the hospital for asthma attacks, and determine what the triggers might be.

Prospective study-- data are collected in the future from groups (called cohorts) sharing common factors.

Example: Framingham Heart Study—50 years of data involving thousands of Framingham residents to determine which biologic and environmental led to

heart disease and stroke. Cholesterol, high blood pressure, diabetes, and cigarette smoking.

Experiment- apply some treatment and proceed to determine its effects on the subjects.

Example: Milgram experiment—measured the willingness of a participant to obey an authority figure even though it conflicts with the participant's conscience. 65% inflicted the final 450 volt shock, and no one stopped below the 30 volt level.

Random sampling – members of a population are selected in such a way that each individual member has an equal chance of being selected.

Example: Put a number into a hat and randomly select one.

Simple random sampling – of size n subjects is selected in such a way that every possible sample of the same size n has the same chance of being chosen. We expect all components of the population to be approximately proportionately represented. (Random samples can be selected by using computers to generate random numbers, or using a random number table)

Systematic sampling – We select some starting point and then select every k th element in the population.

Example: Every 30th iPod is selected from an assembly line to check for defects.

Convenience sampling—use results that are easy to get

Example: Ask your family members for whom they plan to vote in the upcoming election.

Example: Mall survey

Stratified sampling – subdivide the population into at least two different subgroups (or strata) that share the same characteristics, then draw a sample from each subgroup. This ensures that all groups of interest are represented in the study.

Example: To conduct a study on malpractice rates, subdivide doctors into four groups—surgeons, family practitioners, obstetricians, and others. Take a sample of doctors within each group.

Example: To gauge students' attitudes toward alcohol use, divide students into two groups, male and female, and sample 200 students within each group.

Cluster sampling-- Divide the population area into sections (or clusters), then randomly select some of those clusters, then choose all of the members from those selected clusters.

Example: randomly select 10 neighborhoods in a city, and then ask all of the residents their opinion about the new mayor. Instead of having a sample scattered over the entire coverage area, the sample is more localized, which can result in reduced costs, simplified field work and more convenient administration.