

# Chapter 1 Introduction to Statistics

## Statistics Larson and Farber

1-1 pg. 6 1,4,12,14,23,26,31,37,39,46

1-2 pg. 13 3,7,9,16,18,20,21,23,25

1-3 pg. 24 2,5,6,12,14,15,21,23-26,31-34

*Statistics Chapter 1 An overview of Stats**Date: \_\_\_\_\_**Lesson 1-1***Vocabulary:**

Data: Consist of information coming from observations, counts, measurements or responses

Statistics: Is the science of collecting, organizing, analyzing and interpreting data in order to make decisions

Population: Is the collection of all outcomes, responses, measurements, or counts that are of interest

*Example 1: Identifying data Sets*

*In a recent survey, 1708 adults in the US were asked if they think global warming is a problem that requires immediate government action. 939 of the adults said yes. Identify the population and the sample. Describe the data set*

*Population: Consists of the responses of ALL the adults in the US*

*sample: Consists of the 1708 people in the survey the sample is a subset of all the responses in the US*

*Data set: consists of 939 yes's and the 769 No's*

*Another example:*

*the U.S. Department of energy conducts weekly surveys of approximately 800 gas stations to determine the*

*average price per gallon of regular gas. ON February 12, 2007 the average price was \$2.24 per gallon. Find*

*1. Population:*

*2. Sample:*

*3. What does the data set consist of? (how your friend who knows nothing would describe the data:*

Parameter: a numerical description of a population characteristic

Statistic: a numerical description of a sample characteristic

*Example 2: Distinguishing between a parameter and a statistic*

*Decide whether the numerical value describes a population parameter or a sample statistic.*

**EXPLAIN**

*1. A recent survey of a sample of MBAs reported that the average salary for an MBA is more than \$82,000*

*Because the average of \$82,000 is based on a subset of the population, it is a sample statistic*

*2. Starting salaries of the 667 MBA graduates from the University of Chicago Graduate School of Business*

*increased 8.5% from the previous year*

*Because the % increase of 8.5% is based on all 667 graduates' starting salaries, it is a population parameter*

*3. IN a random check of a sample of retail stores, the Food and Drug Administration found that 34% of the*

*stores were not storing fish at the proper temperature*

*Because the % of 34% is based on a subset of the population it is a sample statistic*

Another example:

In 2006, major league baseball teams spent a total of \$2,326,706,685 on players salaries. Does this numerical value describe a population parameter or a sample statistic?

a. Decide whether the numerical value is from a population or a sample first

Parameter Population

b. Specify whether the numerical value is a parameter or a statistic?

### Branches of statistics 1.1 Continued

Descriptive Stats: is the branch of stats that involve the organization, summarization and display of data

Inferential Stats: is the branch of stats that involves using a sample to draw conclusions about a population. A basic tool in the study of inferential stats is probability.

Example 3 Descriptive and inferential stats

decide which part of the study represents the descriptive branch of stats. What conclusion might be drawn from the

study using inferential stats?

1. a large sample of men, aged 48, was studied for 18 years. For unmarried men, approx 70% were alive at age 65. For married men, 90% were alive at age 65.

Descriptive stats: "for unmarried men, approx 70% were alive at 65" And "for Married men, approx 90% were alive at 65"

A possible inference drawn from this would be... that being married is associated with a longer life

2. In a sample of Wall Street analysts, the percentage who incorrectly forecasted high-tech earnings in a recent year was 44%

Descriptive stats: "the percentage of wall street analysts who incorrectly forecasted high-tech earnings in a recent year was 44%"

Inference drawn: that the stock market is difficult to forecast, even for professionals

Another example:

A survey conducted among 1017 men and women by Opinion Research Corporation International found that 76% of women and 60% of men had had a physical exam within the previous year

1. Identify the descriptive aspects of the survey

76% of women and 60% of men

2. What inferences could be drawn from this survey?

## Lesson 1-2 Data Classification

Date: 1/29/14

Vocab: *Categorical*Qualitative Data: Consist of attributes, labels or non numerical entries *eye color*

Quantitative Data: Consist of numerical measurements or counts

Discrete: Quantitative number values are countable *ACT Score*

Continuous: Numerical data infinitely many possible quantitative values not countable

## Example 1: Classifying Data by type

The base price of several vehicles are shown on page 9. Which data are qualitative data and which are quantitative data?

The information can be separated into two data sets. One set contains the names of vehicle models and the other contains the base price of vehicle models. The names are nonnumerical so they are qualitative. The base prices are numerical entires so they are quantitative data

Try it yourself

The population of several US Cities are shown on pg. 9. Which Data are qualitative data and which are quantitative data?

a. Identify the contents of each data Set?

b. Decide whether each data set consists of numerical or non numerical entries.

c. Specify the qualitative and the quantitative data

Levels of Measurement: the level of measurement determines which statistical calculation are meaningful. 4 levels in order from lowest to highest

Nominal, Ordinal, Interval, Ratio

Nominal level of measurement: are qualitative only. Data at this level are categorized using names, labels, or qualities. NO math computation can be made

Ordinal level of measurement: are qualitative or quantitative. Data at this level can be arranged in order, or ranked, but differences between data entries are not meaningful

## Example 2 Classifying Data by level

Two data sets are shown on pg. 12. Which data set consists of data at the nominal level? Which data set consists of data at the ordinal level? Explain

The fist set of data set lists the **RANK** of 5 Tv Shows. They are ranked 1,2,3,4,5.

Because the rankings can be listed in order the data is at Ordinal Level 1-5 has no mathematical meaning

The second set of data consists of the call letters or each network

the call letters are simply the names so they are at the **nominal level**

try it yourself #2

Consider the following data set, decide whether the data are at a nominal level or at an ordinal level

a. Identify what each data set represents

b. Specify the level of measurement and justify your answer

1. The final standings for the Pacific Division of the NBA

2. A collection of phone numbers

Interval level of measurement: Can be ordered, and you can calculate meaningful differences between data entries. At the interval level, a zero entry simply represents a position on a scale, the entry is not an inherent zero

Ratio level of measurement: Are similar to data at the interval level, with the added property that a zero entry is an inherent zero. A ratio of two data values can be formed so that one data value can be meaningfully expressed as a multiple of another

inherent zero: Is a zero that implies "none" FOR example, the amount of money you have in a savings account could be zero dollars. The zero represents NO MONEY; it is called an inherent zero

Example of a zero that is NOT an inherent zero: the temperature 0 degrees, it is a position on the scale, it doesn't mean no heat present

### Example 3 Classifying data by level

Two data sets are shown on pg. 13. Which data consists of data at the interval level? Which data set consists of data at the ratio level? Explain

Both of these data sets contain quantitative data. The Dates of Yankee's World Series victories. You could make sense to find the differences between the years. Like the difference between the first and the last  $2000-1923=77$  years BUT it doesn't make sense to say that one year is a multiple of another SO.... The data is NOT at the interval level

### HOME RUN DATA

You can find the differences and WRITE ratios. From the data you can see Detroit hit 31 more home runs than Seattle hit and that Chicago hit about twice as many home runs as Kansas City SO.... these data are at a RATIO level

### Try it yourself #3

Decide whether the data are at the interval level or at the ratio level

- Identify what each data set represents
- Specify the level of measurement and justify

- The body temperature (in Fahrenheit) of an athlete during an exercise session

0°F Represents a temperature  
Not Inherent Zero Not Ratio (Interval)

- the heart rate (in beats per min) of an athlete during an exercise session

Ratio 0 beats per Min Inherent Zero

The following tables summarize which operations are meaningful at each of the four levels of measurement. When identifying a data set's level of measurement, use the highest level that applies.

Level of measurement	Put data in categories	Arrange data in order	Subtract data values	Determine if one data value is a multiple of another
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

*Stats Lesson 1-3 Data Collection and Experimental Design of a statistical study***Guidelines****Designing a statistical study**

1. Identify the variable(s) of interest (the focus) and the population of the study
2. Develop a detailed plan for collecting data. If you use a sample, make sure the sample is representative of the population
3. Collect the data
4. Describe the data, using descriptive stat techniques
5. Interpret the data and make decisions about the population using inferential stats
6. Identify any possible errors

**Data Collection:**

**\*Do an observational study:** a researcher observes and measures characteristics of interest of part of a population but doesn't change existing conditions. Example: a study in which the researchers observed and recorded the mouthing behavior on nonfood objects of children up to 3 years old.

**\*perform an experiment:** A treatment is applied to part of a population and the responses observed. You may have a

**CONTROL GROUP:** which no treatment is given

**Experimental units:** subjects are called

**Placebo:** Harmless, unmedicated treatment, that is made to look like the actual treatment

The responses of both the control group and treatment group

**\*Use a simulation:** the use of a mathematical or physical model to reproduce the conditions of a situation or process. Allows you to study situations that are impractical or even dangerous to create in real life, and often they are cheaper. You will use computers or calculators to simulate statistical process on a computer.

**\*Use a survey:** An investigation of one or more characteristics of a population. surveys are carried out on people by questioning. Most common interview, mail, telephone and now Internet. You have to be careful to word the questions so you don't have a biased responses.

**Confounding Variable:** Occurs when an experimenter cannot tell the difference between the effects of different factors on the variable.

*Example 1 Deciding on methods of Data Collection*

*Consider the following statistical studies. Which method of data collection would you use to collect data for each study. Explain*

1. A study of the effects of changing flight patterns on the number of plane accidents.

*Very impractical to create the situation, so use a simulation*

2. A study of the effect of eating oatmeal on lowering blood pressure.

*You want to measure the effect a treatment (eating oatmeal) has on patients, so you would perform an experiment*

3. A study on how fourth grade students solve a puzzle

*you want to observe and measure a certain characteristic of part of a population, you could do an observational study*

4. a Study of US residents' approval rating of the US president.

*-Use a survey, Ask "DO you approve of the way the President is handling the job?"*

*Try it yourself 1*

a. First identify the focus of the study

b. Identify the population of the study

c. Choose an appropriate method of data collection

1. A study of the effect of exercise on relieving depression

2. A study of the success of graduates of a large university finding a job within one year of graduation

## 1.3 continued

## Sampling techniques VOCAB

CENSUS: is a count or measure of an ENTIRE population, it provides complete information, but it is very costly and difficult

SAMPLING: is a count or measure of PART of a population, it is more common

UNBIASED DATA: a researcher must ensure that the sample is representative of the population. You must use appropriate techniques to ensure inferences about the population

SAMPLING ERROR: you have to keep in mind that when a study is done with faulty data, the results are questionable and sometimes even with the best methods of sampling error will occur

RANDOM SAMPLE: one in which every member of the population has an equal chance of being selected.

SIMPLE RANDOM SAMPLE: a sample in which every possible sample of the same size has the same chance of being selected.

## Example 3 Using a Random Sample

There are 731 students currently enrolled in Statistics in your school. You wish to perform a sample of 8 students to answer some survey questions. Select the students who will belong to the simple random sample

Math → Prob

USE A RANDOM NUMBER GENERATOR ON CALCULATOR

```
randInt(1,731,8)
{90 39 529 10 3...
```

```
randInt(1,731,8)
...308 225 711 22)
```

Try it yourself #3

A company employs 79 people. Choose a simple random sample of five to survey

Sampling with Replacement and without replacement: You need to decide whether it is acceptable to have the same population member selected more than once. if that is acceptable then the sample is with replacement if not acceptable the it is said to be without replacement

## Common sampling techniques

1. Stratified Sample: use when it is important for the sample to have members from each segment of the population. Depending on the focus of the, study, members of the population the members are divided in to two or more subsets, called STRATA, they share a similar characteristic such as age, gender, ethnicity or even political preference. A sample is then randomly selected from each STRATA. This ensures that each segment of the population is represented

2. Cluster Sampling: When a population falls into naturally occurring subgroups, each having similar characteristics. To select a cluster sample, divide the population into groups called clusters, and select all the members in one or more (BUT NOT ALL) of the clusters. Example pg. 23

3. Systematic Sample: a sample in which each member of the population is assigned a number. The members of the population are ordered in some way, a starting number is randomly selected and then sample members are selected at regular intervals from the starting number.

4. Convenience sample: LEADS to biased studies it consists of only available members of the population... DON'T USE

*Example 3 Identify sampling techniques*

You are doing a study to determine the opinion of students at your school regarding stem cell research.

Identify the sampling technique you are using if you select the sample listed

1. You select a class at random and question each student in the class

Your class is a natural cluster (a subgroup) and you question each student .... CLUSTER SAMPLE

2. You divide the students with respect to majors and randomly select and question some students in each major.

Students are divided into strata (Majors) and a sample is selected and each student has an equal chance of being selected it is .... Stratified Sample

3. You assign each student a number and generate random numbers. You then question each student whose number is randomly selected.

Each sample of the same size has an equal chance of being selected and each student has an equal chance of being selected, so..... Simple random sample

Try it yourself 4

same situation as above

a. Determine HOW the sample is selected

b. Identify the corresponding sampling technique

1. You select students who are in your Stats class

2. You assign each student a number and, after choosing a starting number, question every 25th student



## Case STUDY pg. 17

Nielsen Media Research has been rating television programs for more than 50 years. Nielsen uses several sampling procedures, but its main one is to track the viewing patterns of 10,000 households. These contain more than 30,000 people and are chosen to form a cross section of the overall population. The households represent various locations, ethnic groups, and income brackets. The data gathered from the Nielsen sample of 10,000 households are used to draw inferences about the population of all households in the United States.

TV programs viewed by all households  
in the United States (111.4 million households)

TV programs viewed  
by Nielsen sample  
(10,000 households)

Top-Ranked Programs in Prime Time for the Week of 2/12/07–2/18/07

Rank	Rank Last Week	Program Name	Network	Day, Time	Rating	Share	Audience
1	1	American Idol–Tuesday	FOX	Tues., 8:00 P.M.	17.4	25	19,354,000
2	2	American Idol–Wednesday	FOX	Wed., 9:00 P.M.	16.2	24	18,045,000
3	3	Grey's Anatomy	ABC	Thu., 9:00 P.M.	16.0	23	17,809,000
4	4	House	FOX	Tues., 9:00 P.M.	14.8	22	16,469,000
5	5	CSI	CBS	Thu., 9:00 P.M.	13.8	20	15,323,000
6	7	CSI: Miami	CBS	Mon., 10:00 P.M.	12.7	21	14,093,000
7	8	Desperate Housewives	ABC	Sun., 9:00 P.M.	11.7	18	13,060,000
8	10	Deal or No Deal–Monday	NBC	Mon., 8:00 P.M.	10.0	15	11,167,000
8	8	Two and a Half Men	CBS	Mon., 9:00 P.M.	10.0	14	11,099,000
10	17	Shark	CBS	Thu., 10:00 P.M.	9.8	16	10,909,000

## Exercises

- Rating Points** Each rating point represents 1,114,000 households, or 1% of the households in the United States. Does a program with a rating of 8.4 have twice the number of households as a program with a rating of 4.2? Explain your reasoning.
- Sampling Percent** What percentage of the total number of U.S. households is used in the Nielsen sample?
- Nominal Level of Measurement** Which columns in the table contain data at the nominal level?
- Ordinal Level of Measurement** Which columns in the table contain data at the ordinal level? Describe two ways that the data can be ordered.
- Interval Level of Measurement** Which column in the table contains data at the interval level? How can these data be ordered? What is the unit of measure for the difference of two entries in the data set?
- Ratio Level of Measurement** Which three columns contain data at the ratio level?
- Share** The column listed as “Share” gives the percentage of televisions in use at a given time. Does the Nielsen rating rank shows by rating or by share? Explain your reasoning.
- Inferences** What decisions (inferences) can be made on the basis of the Nielsen ratings?

## Attachments

---

Stats lesson 1.2 pg 14.pdf

Stats Case study 1.2 pg. 17.pdf

1.3 Real stats pg. 34.pdf

Chapter 1 Quiz Questions.pdf

First Day Stats survey.pdf

Chapter 1 Data QUAKE.xls