MATH 110 - INTRODUCTION TO STATISTICS - MODULE

Los Medanos College

LibreTexts™

Los Medanos College

Math 110 - Introduction to Statistics - Module

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

This text was compiled on 03/18/2025

TABLE OF CONTENTS

Licensing

Section 1.1

- 1.1 Definitions of Statistics, Probability, and Key Terms
- Lab Assignment 1.1
- Self-Check 1.1

Section 1.2

- 1.2 Data, Sampling, and Variation in Data and Sampling
- Lab Assignment 1.2
- Self-Check 1.2

Section 1.3

- 1.3 Frequency, Frequency Tables, and Levels of Measurement

Section 1.4

- 1.4 Experimental Design
- Lab Assignment 1.3, 1.4
- Self-Check 1.3, 1.4

Section 2.1

- 2.1 Stem-and-Leaf Plots (Stemplots) and Bar Graphs

Section 2.2

- 2.2 Frequency Table and Histograms
- Lab Assignment 2.1, 2.2
- Self-Check 2.1, 2.2

Section 2.3

- 2.3 Measures of the Location of the Data

Section 2.4

- 2.4 Box Plots
- Lab Assignment 2.3, 2.4
- Self-Check 2.3, 2.4

Section 2.5

- 2.5 Measures of Center of the Data

Section 2.6

- 2.6 Skewness and the Mean, Median, and Mode

Section 2.7

- 2.7 Measures of the Spread of the Data
- Lab Assignment 2.5, 2.6, 2.7
- Self-Check 2.5, 2.6, 2.7

Section 3.1

- 3.1 Terminology
- Lab Assignment 3.1
- Self-Check 3.1

Section 3.2-3.4

- 3.2-3.4 Probability Topics
- Lab Assignment 3.2, 3.3, 3.4
- Self-Check 3.2, 3.3, 3.4

Section 4.1

- 4.1 Probability Distribution Function (PDF) for a Discrete Random Variable

Section 4.2

- 4.2 Mean or Expected Value and Standard Deviation
- Lab Assignment 4.1, 4.2
- Self-Check 4.1, 4.2

Section 6.1

- 6.1 The Standard Normal Distribution
- Lab Assignment 6.1
- Self-Check 6.1

Section 7.1

- 7.1 The Central Limit Theorem for Sample Means
- Lab Assignment 7.1
- Self-Check 7.1

Section 8.1

- 8.1 A Single Population Mean using the Normal Distribution
- Lab Assignment 8.1
- Self-Check 8.1

Section 8.2

- 8.2 A Single Population Mean using the Student t Distribution

Section 8.3

- 8.3 A Population Proportion
- Lab Assignment 8.2, 8.3
- Self-Check 8.2, 8.3

Section 9.1

- 9.1 Null and Alternative Hypothesis

Section 9.4

- 9.4 Rare Events, the Sample, Decision and Conclusion
- Lab Assignment 9.1, 9.4
- Self-Check 9.1, 9.4

Section 9.5

- 9.5 Full Hypothesis Testing with One Sample
- Lab Assignment 9.5
- Self-Check 9.5

Section 10.1

- 10.1 Two Population Means with Unknown Standard Deviations

Section 10.3

- 10.3 Comparing Two Independent Population Proportions
- Lab Assignment 10.1, 10.3
- Self-Check 10.1, 10.3

Section 10.4

- 10.4 Matched or Paired Samples

Section 11.3

- 11.3 Test of Independence
- Lab Assignment 12.3, 12.5, 13.1, 13.2, 13.3, 11.3
- Self-Check 12.3, 12.5, 13.1, 13.2, 13.3, 11.3

Section 12.2, 12.4

- 12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation
- Lab Assignment 10.4, 12.2, 12.4
- Self-Check 10.4, 12.2, 12.4

Section 12.3, 12.5

- 12.3, 12.5 The Regression Equation and Prediction

Section 13.1 - 13.3

- 13.1, 13.2, 13.3 One-Way ANOVA

Index

Glossary

Detailed Licensing

Detailed Licensing

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

CHAPTER OVERVIEW

Section 1.1

1.1 Definitions of Statistics, Probability, and Key Terms
Lab Assignment 1.1
Self-Check 1.1

1.1 Definitions of Statistics, Probability, and Key Terms

Learning Objective:

In this section, you will:

- Develop your understanding of what is statistics.
- Develop your understanding of what is probability.
- Recognize and differentiate between key terms.

What is Statistics?

The science of **statistics** deals with the collection, analysis, interpretation, and presentation of data.
• **Descriptive statistics** is organizing and summarizing data. Two ways to summarize data are by graphing and by using numbers (for example, finding an average).
• **Inferential statistics** is using formal methods for drawing conclusions from "good" data.

What is Probability?

Probability is a number between 0 and 1, inclusive. It gives the likelihood that a specific event will occur.

Key Terms

• A **population** is all individuals, objects, or measurements whose properties are being studied.
• A **sample** is a subset or portion of the population being studied.
• A **parameter** is a number (average, mean, proportion) that is used to represent a characteristic of a population.
• A **statistic** is a number (average, mean, proportion) that represents a characteristic of the sample. The statistic is an estimate of a population parameter.
• A **representative sample** is a subset of the population that has the characteristics as the population.
• A **variable**, notated by capital letters such as X and Y, is a characteristic or measurement of interest for each person or object in a population.
o **Numerical variables** take on values that are indicated by numbers.
o **Categorical variables** take on values that are names or labels.
• **Data** are the actual values of the variable. They may be numbers or they may be words. **Datum** is a single value.

Example 1:

A study was conducted at a local college to analyze the average cumulative GPA's of students who graduated last year. Determine the correct key phrase (population, statistic, parameter, sample, variable and data) for each of the following phrases:

a) all students who attended the college last year

b) the cumulative GPA of one student who graduated from the college last year

c) 3.65, 2.80, 1.50, 3.90

d) a group of students who graduated from the college last year, randomly selected

e) the average cumulative GPA of students who graduated from the college last year

f) all students who graduated from the college last year

g) the average cumulative GPA of students in the study who graduated from the college last year

Example 2:

Determine what the key terms refer to in the following study.
We want to know the average (mean) amount of money first year college students spend at ABC College on school supplies that do not include books. We randomly surveyed 100 first year students at the college. Three of those students spent $150, $200, and $225, respectively.

1. Population –
2. Sample –
3. Parameter –
4. Statistic –
5. Variable –
6. Data –

For more information and examples see online textbook OpenStax Introductory Statistics pages 5-9.

Lab Assignment 1.1

Name:____________________________________Date:____________________Row:________

Lab Assignment 1.1

Determine what the key terms refer to in the following studies.

1. A fitness center is interested in the mean amount of time a client exercises in the center each week.

a. Population –

b. Sample –

c. Parameter –

d. Statistic –

e. Variable –

f. Data –

2. Insurance companies are interested in the mean health costs each year of their clients, so that they can determine the costs of health insurance.

a. Population –

b. Sample –

c. Parameter –

d. Statistic –

e. Variable –

f. Data –

3. A marriage counselor is interested in the proportion of clients she counsels who stay married.

a. Population –

b. Sample –

c. Parameter –

d. Statistic –

e. Variable –

f. Data –

Self-Check 1.1

Name:_________________________________Date:___________________Row:____________

Self-Check 1.1

1. Determine what the key terms refer to in the following study. We want to know the average (mean) amount of money spent on school uniforms each year by families with children at Knoll Academy. We randomly survey 100 families with children in the school. Three of the families spent $65, $75, and $95, respectively.

Population –

Sample –

Parameter –

Statistic –

Variable –

Data –

2. Determine what the key terms refer to in the following study. An insurance company would like to determine the proportion of all medical doctors who have been involved in one or more malpractice lawsuits. The company selects 500 doctors at random from a professional directory and determines the number in the sample who have been involved in a malpractice lawsuit.

Population –

Sample –

Parameter –

Statistic –

Variable –

Data –

CHAPTER OVERVIEW

Section 1.2

1.2 Data, Sampling, and Variation in Data and Sampling
Lab Assignment 1.2
Self-Check 1.2

1.2 Data, Sampling, and Variation in Data and Sampling

Learning Objectives

In this section, you will:

- Recognize and differentiate between types of data.
- Recognize and apply various types of sampling methods.
- Evaluate statistical studies before accepting results.

Types of Data

• **Qualitative data (Categorical data):** data that consists of names or labels that does not represent counts or measurements.
• **Quantitative data:** data that is numerical, representing counts or measurements.

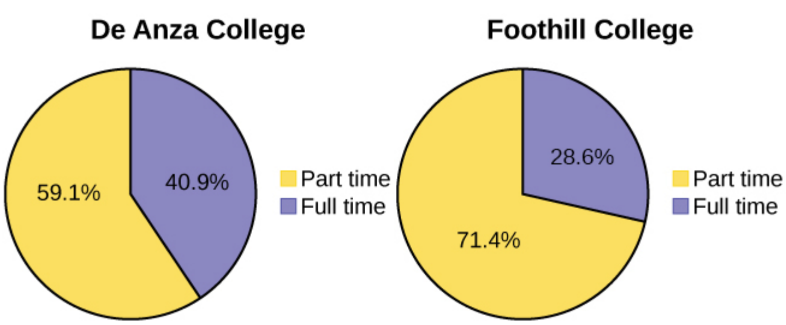
Types of Quantitative data:
• **Discrete:** All data that are the result of counting.
• **Continuous:** All data that are the result of measuring.

Example 1:
Determine the correct data type (quantitative or qualitative). Indicate whether quantitative data are continuous or discrete. Hint: Data that are discrete often start with the words "the number of."
a. the number of pairs of shoes you own
b. the type of car you drive
c. the distance it is from your home to the nearest grocery store
d. the number of classes you take per school year.
e. the type of calculator you use
f. weights of sumo wrestlers
g. number of correct answers on a quiz
h. IQ scores (This may cause some discussion.)

Graphs that are used to display qualitative data

Pie chart

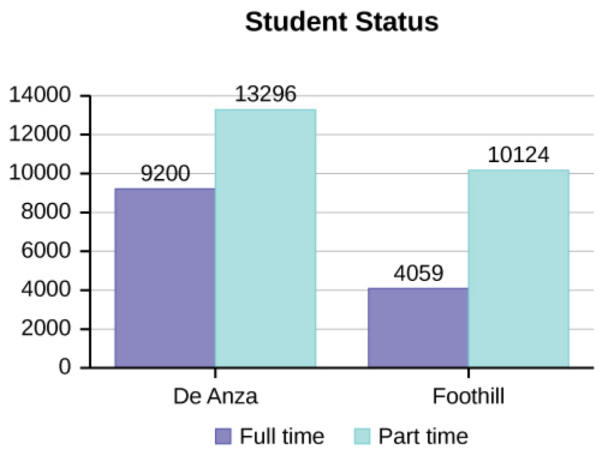
In a **pie chart,** categories of data are represented by wedges in a circle and are proportional in size to the percent of individuals in each category.



Example shows two pie charts. First: De Anza College shows 59.1% of individuals work part time and 40.9% work full time. Second: Foothill College 71.4% work Part time and 28.6% work full time.

Bar graph

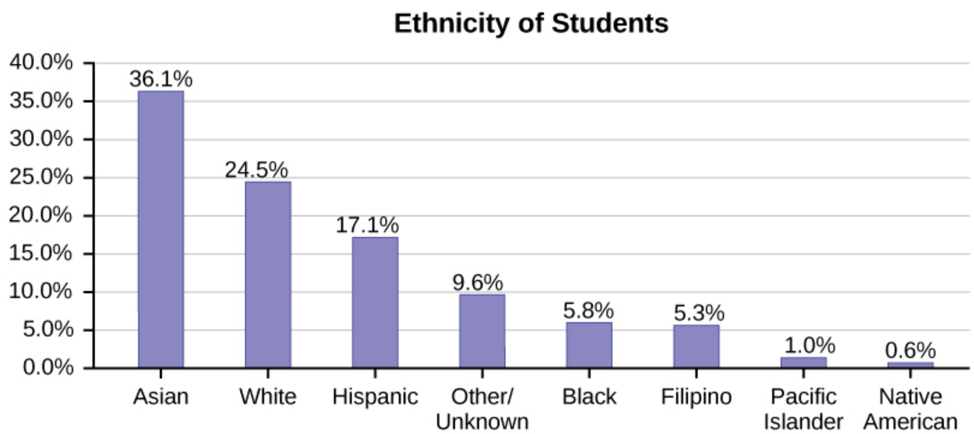
In a **bar graph**, the length of the bar for each category is proportional to the number or percent of individuals in each category. Bars may be vertical or horizontal.



Example is a bar graph depicting student full time and part time status. De Anza shows there are more part time students than full time, and for Foothill there are more Part time than full time.

Pareto Chart

A **Pareto chart** consists of bars that are sorted into order by category size (largest to smallest).



This bar graph is graphed from largest to smallest ethnicity of students with Asian students with the largest number at 36.1%, white 24.5%, Hispanic 17.1%, Other/Unknown 9.6%, Black 5.8%, Filipino 5.3%, Pacific Islander 1.0%, Native American 0.6%.

Sampling Methods

Gathering information about an entire population often costs too much or is virtually impossible. Instead, we use a sample of the population.

• A sample should have the same characteristics as the population it is representing.

• Random sampling is a method of achieving this goal.

There are several different methods of random sampling.

• **Simple random sample.** Each sample of the same size has an equal chance of being selected.

• **Stratified sample.** Divide the population into groups called strata and then take a proportionate number from each stratum.

• **Cluster sample**. Divide the population into clusters (groups) and then randomly select some of the clusters. All the members from these clusters are in the cluster sample.

• To choose a **systematic sample,** randomly select a starting point and take every nth piece of data from a listing of the population.

Convenience sampling is a **non-random method** of selecting a sample. This method selects individuals that are easily available and may result in bias data.

Sampling **with replacement**

• Once a member is picked, that member goes back into the population and thus may be chosen more than once

Sampling **without replacement**

• A member of the population may be chosen only once

Example 2:

Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience).

A. A soccer coach selects six players from a group of boys aged eight to ten, seven players from a group of boys aged 11 to 12, and three players from a group of boys aged 13 to 14 to form a recreational soccer team.

B. A pollster interviews all human resource personnel in five different high tech companies.

C. A high school educational researcher interviews 50 high school female teachers and 50 high school male teachers.

D. A medical researcher interviews every third cancer patient from a list of cancer patients at a local hospital.

E. A high school counselor uses a computer to generate 50 random numbers and then picks students whose names correspond to the numbers.

F. A student interviews classmates in his algebra class to determine how many pairs of jeans a student owns, on the average.

In statistics, **a sampling bias** is created when a sample is collected from a population and some members of the population are not as likely to be chosen as others. When a sampling bias happens, there can be incorrect conclusions drawn about the population that is being studied

Critical Evaluation

We need to evaluate the statistical studies we read about critically and analyze them before accepting the results of the studies. Common problems to be aware of include:

• **Self-selected or Voluntary response samples:** Responses only by people who choose to respond.

• **Small samples:** Samples that are too small may be unreliable

• **Undue influence:** Collecting data or asking questions in a way that influences the response. Loaded question.

• **Non-response** or refusal of subject to participate.

• **Causality:** A relationship between two variables does not mean that one causes the other to occur. Correlation does not imply causation.•

• **Self-funded** or **self-interest studies:** A study performed by a person or organization in order to support their claim.

• **Misleading use of data:** improperly displayed graphs, incomplete data, or lack of context.

Example 3:

Determine whether or not the following samples are representative.

1. To find the average GPA of all students in a university, use all honor students at the university as the sample.

2.To find out the most popular cereal among young people under the age of ten, stand outside a large supermarket for three hours and speak to every twentieth child under age ten who enters the supermarket.

3.To find the average annual income of all adults in the United States, sample U.S. congressmen. Create a cluster sample by considering each state as a stratum (group). By using simple random sampling, select states to be part of the cluster. Then survey every U.S. congressman in the cluster.

4.To determine the proportion of people taking public transportation to work, survey 20 people in New York City. Conduct the survey by sitting in Central Park on a bench and interviewing every person who sits next to you.

5.To determine the average cost of a two-day stay in a hospital in Massachusetts, survey 100 hospitals across the state using simple random sampling.

For more information and examples see online textbook OpenStax Introductory Statistics pages 10-25.

Lab Assignment 1.2

Name:______________________________________Date:_____________________Row:_________

Lab Assignment 1.2

For the following exercises, identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.
Example: number of tickets sold to a concert, Solution: quantitative discrete, 560 tickets

1. percent of body fat

2. favorite baseball team

3. time in line to buy groceries

4. number of students enrolled at Evergreen Valley College

5. most-watched television show

6. brand of toothpaste

7. distance to the closest movie theatre

8. age of executives in Fortune 500 companies

9. number of competing computer spreadsheet software packages

10. Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience).

a. A high school principal polls 50 freshmen, 50 sophomores, 50 juniors, and 50 seniors regarding policy changes for after school activities.

b. The instructor takes her sample by gathering data on five randomly selected students from each Lake Tahoe Community College math class.

c. A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Jose. The first house in the neighborhood around the park was selected randomly and then every eighth house in the neighborhood around the park was interviewed.

11. Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. Suppose an airline conducts a survey. Over Thanksgiving weekend, it surveys six flights from Boston to Salt Lake City to determine the number of babies on the flights. It determines the amount of safety equipment needed by the result of that study.

a. Using complete sentences, list three things wrong with the way the survey was conducted.

b. Using complete sentences, list three ways that you would improve the survey if it were to be repeated.

12. List some practical difficulties involved in getting accurate results from a telephone survey.

13. List some practical difficulties involved in getting accurate results from a mail survey.

14. YouPolls is a website that allows anyone to create and respond to polls. One question posted April 15 asks: “Do you feel happy paying your taxes when members of the Obama administration are allowed to ignore their tax liabilities?” As of April 25, 11 people responded to this question. Each participant answered “NO!” Which of the potential problems with samples discussed in this module could explain this connection?

Self-Check 1.2

Name:____________________________________Date:__________________Row:___________

Self-Check 1.2

For exercises 1-4, identify the type of data (quantitative discrete, quantitative continuous, or qualitative).

1. The data are the number of machines in a gym. You sample five gyms. One gym has 12 machines, one gym has 15 machines, one gym has ten machines, one gym has 22 machines, and the other gym has 20 machines.

2. The data are the areas of lawns in square feet. You sample five houses. The areas of the lawns are 144 sq. feet, 160 sq. feet, 190 sq. feet, 180 sq. feet, and 210 sq. feet.

3. The data are the colors of houses. You sample five houses. The colors of the houses are white, yellow, white, red, and white.

4. The number of cars in a parking lot.

5. A study is done to determine the average tuition that San Jose State undergraduate students pay per semester. Each student in the following samples is asked how much tuition he or she paid for the Fall semester. What is the type of sampling in each case? (simple random, stratified, systematic, cluster, or convenience).

   a. A sample of 100 undergraduate San Jose State students is taken by organizing the students' names by classification (freshman, sophomore, junior, or senior), and then selecting 25 students from each.

   b. A random number generator is used to select a student from the alphabetical listing of all undergraduate students in the Fall semester. Starting with that student, every 50th student is chosen until 75 students are included in the sample.

   c. A completely random method is used to select 75 students. Each undergraduate student in the fall semester has the same probability of being chosen at any stage of the sampling process.

   d. The freshman, sophomore, junior, and senior years are numbered one, two, three, and four, respectively. A random number generator is used to pick two of those years. All students in those two years are in the sample.

   e. An administrative assistant is asked to stand in front of the library one Wednesday and to ask the first 100 undergraduate students he encounters what they paid for tuition the Fall semester. Those 100 students are the sample.

6. A local radio station has a fan base of 20,000 listeners. The station wants to know if its audience would prefer more music or more talk shows. Asking all 20,000 listeners is an almost impossible task. The station uses convenience sampling and surveys the first 200 people they meet at one of the station's music concert events. 24 people said they'd prefer more talk shows, and 176 people said they'd prefer more music. Do you think that this sample is representative of (or is characteristic of) the entire 20,000 listener population?

CHAPTER OVERVIEW

Section 1.3

1.3 Frequency, Frequency Tables, and Levels of Measurement

1.3 Frequency, Frequency Tables, and Levels of Measurement

Learning Objective:

In this section, you will:
• Use rounding rules to round final answers.
• Determine the levels of measurement for data.
• Use frequency tables to organize and analyze data.

Answers and Rounding Off

• Carry your final answer one more decimal place than was present in the original data.
• Round off only the final answer. Do not round off any intermediate results, if possible.

Levels of Measurement

Data can be classified into four levels of measurement.
• **Nominal scale level:** Categories, colors, names, labels, yes or no responses. Nominal scale data are not ordered.
• **Ordinal scale level:** Data can be ordered, but differences are meaningless.
• **Interval scale level:** Data can be ordered, but differences are meaningful. Data does not have a starting point.
• **Ratio scale level:** Ratio scale data is like interval scale data, but it has a 0 point and ratios can be calculated.

Frequency Tables

• A **frequency** is the number of times a value of the data occurs.
• A **relative frequency** is the ratio (fraction or proportion) of the number of times a value of the data occurs in the set of all outcomes to the total number of outcomes.

o To find the relative frequencies, divide each frequency by the total number of students in the sample.

• **Cumulative relative frequency** is the accumulation of the previous relative frequencies.

o To find the cumulative relative frequencies, add all the previous relative frequencies to the relative frequency for the current row

Example 1:

Complete the frequency table with the following information. Twenty students are asked how many hours they worked per day. Their responses, in hours, are as follows: 5, 6, 3, 3, 2, 4, 7, 5, 2, 3, 5, 6, 5, 4, 4, 3, 5, 2, 5, 3.

Hours Worked per Day	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**

- What percent of students work exactly 4 hours?
- What percent of students that work less than 3 hours?

- What is the percent of students that work from 4 to 6 hours?
- Find the number of students that work from 3 to 5 hours?
- What fraction of the students work from 6 to 7 hours?
- What is the frequency of students that work from 3 to 6 hours?
- What is the relative frequency of students that work 3 or less?
- What is the cumulative relative frequency for 4? Explain what this number tells you about the data.

For more information and examples see online textbook OpenStax Introductory Statistics pages 26-35.

CHAPTER OVERVIEW

Section 1.4

1.4 Experimental Design
Lab Assignment 1.3, 1.4
Self-Check 1.3, 1.4

1.4 Experimental Design

Learning Objectives:

In this section, you will:
• Recognize and differentiate between key terms in experimental design.
• Identify proper experimental design.

Experimental Design

The purpose of an experiment is to investigate relationship between two variables.
• **Explanatory variable** is the independent variable in an experiment. This variable causes change in another variable.
• **Response variable** is the dependent variable in an experiment. This variable is measured for change at the end of the experiment.

In a randomized experiment, the researcher manipulates values of the explanatory variable and measures the resulting changes in the response variable.
• **Treatments** are the different values of the explanatory variable.
• **Experimental unit** is a single object or individual to be measured.
• **Lurking variable** is a variable that has an effect on a study even through it is neither an explanatory nor a response variable.
• **Random assignment** is organizing experimental units into treatment groups using random methods.

When participation in a study prompts a physical response from a participant, it is difficult to isolate the effects of the explanatory variable. To counter the power of suggestion, researchers use the following:
• **Control group** is a group that receives an inactive treatment but is otherwise managed exactly as the other group.
• **Placebo** is an inactive treatment that has no effect on the explanatory variable.
• **Blinding** is not telling the participant which treatment a subject is receiving.
• **Double-blinding** is the act of blinding both the subjects and the researchers working with the subjects.

Example 1:

Researchers want to investigate whether taking aspirin regularly reduces the risk of heart attack. Four hundred men between the ages of 50 and 84 are recruited as participants. The men are divided randomly into two groups: one group will take aspirin, and the other group will take a placebo. Each man takes one pill each day for three years, but he does not know whether he is taking aspirin or the placebo. At the end of the study, researchers count the number of men in each group who have had heart attacks.

Identify the following values for this study: population, sample, experimental units, explanatory variable, response variable, treatments.

Example 2:
The Smell & Taste Treatment and Research Foundation conducted a study to investigate whether smell can affect learning. Subjects completed mazes multiple times while wearing masks. They completed the pencil and paper mazes three times wearing floral-scented masks, and three times with unscented masks. Participants were assigned at random to wear the floral mask during the first three trials or during the last three trials. For each trial, researchers recorded the time it took to complete the maze and the subject's impression of the mask's scent: positive, negative, or neutral.

a. Describe the explanatory and response variables in this study.

b. What are the treatments?

c. Identify any lurking variables that could interfere with this study.

d. Is it possible to use blinding in this study?

Ethics

When a statistical study uses human participants, the researcher should be mindful of the safety of their research subjects. Key protections that are mandated by law include the following:

• Risks to participants must be minimized and reasonable with respect to projected benefits.
• Participants must give informed consent. This means that the risks of participation must be clearly explained to the subjects of the study.
• Data collected from individuals must be guarded carefully to protect their privacy.

Example 3:

Describe the unethical behavior in each example and describe how it could impact the reliability of the resulting data. Explain how the problem should be corrected.

A researcher is collecting data in a community.
a. She selects a block where she is comfortable walking because she knows many of the people living on the street.

b.No one seems to be home at four houses on her route. She does not record the addresses and does not return at a later time to try to find residents at home.

c.She skips four houses on her route because she is running late for an appointment. When she gets home, she fills in the forms by selecting random answers from other residents in the neighborhood.

For more information and examples see online textbook OpenStax Introductory Statistics pages 35-39.

"Introduction to Statistics" by OpenStax, used is licensed under a Creative Commons Attribution License 4.0 license.

Lab Assignment 1.3, 1.4

Name:____Date:____________________Row:________

Lab Assignment 1.3, 1.4

1. determine which of the four levels of measurement (nominal, ordinal, interval, ratio) is most appropriate.

a. Types of movies.

b. Microwave voltage measurements.

c. Ratings of restaurant on a scale from 0 starts to 4 starts.

d. Temperature of oven when baking a cake.

2. Sixty adults with gum disease were asked the number of times per week they used to floss before their diagnosis. The incomplete results are shown below:

# Flossing per Week	Frequency	Relative Frequency	Cumulative Relative Frequency
0	27	0.4500	
1	18		
3			0.9333
6	3	0.0500	
7	1	0.0167	

a. Fill in the blanks in the table above.

b. What percent of adults flossed six times per week?

c. What percent flossed at most three times per week?

d. What percent flossed 6 or 7 times per week?

2. Forbes magazine published data on the best small firms in 2012. These were firms which had been publicly traded for at least a year, have a stock price of at least $5 per share, and have reported annual revenue between $5 million and $1 billion. Table shows the ages of the chief executive officers for the first 60 ranked firms.

Age	Frequency	Relative Frequency	Cumulative Relative Frequency
40-44	3		
45-49	11		
50-54	13		
55-59	16		
60-64	10		
65-69	6		
70-74	1		

a. What is the frequency for CEO ages between 54 and 65?

b. What percentage of CEOs are 65 years or older?

c. What is the relative frequency of ages under 50?

d. What is the cumulative relative frequency for CEOs younger than 55?

3. How does sleep deprivation affect your ability to drive? A recent study measured the effects on 19 professional drivers. Each driver participated in two experimental sessions: one after normal sleep and one after 27 hours of total sleep deprivation. The treatments were assigned in random order. In each session, performance was measured on a variety of tasks including a driving simulation. Use key terms from this module to describe the design of this experiment.

a. Explanatory variable:

b. Response variable:

c. Treatments:

d. Experimental unit:

e. Lurking variable:

f. Random assignment:

g. Control group:

f. Blinding:

LibreTexts

Self-Check 1.3, 1.4

Name:____________________________________Date:__________________Row:___________

Self-Check 1.3, 1.4

1. What type of measure scale is being used? Nominal, ordinal, interval or ratio.

   a. Political outlook: extreme left, left-of-center, right-of-center, extreme right
   b. Time of day on an analog watch
   c. The distance in miles to the closest grocery store
   d. The dates 1066, 1492, 1644, 1947, and 1944
   e. The heights of 21–65 year-old women
   f. Common letter grades: A, B, C, D, and F

2. The table shows the amounts, in inches, of annual rainfall in a sample of towns.

Rainfall (Inches)	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
2.95–4.97	6	6/50 = 0.12	0.12
4.97–6.99	7	7/50 = 0.14	0.12 + 0.14 = 0.26
6.99–9.01	15	15/50 = 0/30	0.26 + 0.30 = 0.56
9.01–11.03	8	8/50 = 0.16	0.56 + 0.16 = 0.72
11.03–13.05	9	9/50 = 0.18	0.72 + 0.18 = 0.90
13.05–15.07	5	5/50 = 0.10	0.90 + 0.10 = 1.00
	Total = 50	Total = 1.00	

a. Find the percentage of rainfall that is less than 9.01 inches.

b. Find the percentage of rainfall that is between 6.99 and 13.05 inches.

c. Find the number of towns that have rainfall between 2.95 and 9.01 inches.

d. What fraction of towns surveyed get between 11.03 and 13.05 inches of rainfall each year?

3. Fifty part-time students were asked how many courses they were taking this term. The (incomplete) results are shown below:

# of Courses	Frequency	Relative Frequency	Cumulative Relative Frequency
1	30	0.6	
2	15		
3			

a. Fill in the blanks in the table above.

b. What percent of students take exactly two courses?

c. What percent of students take one or two courses?

4. You are concerned about the effects of texting on driving performance. Design a study to test the response time of drivers while texting and while driving only. How many seconds does it take for a driver to respond when a leading car hits the brakes?

a. Describe the explanatory and response variables in the study.

b. What are the treatments?

c. What should you consider when selecting participants?

d. Your research partner wants to divide participants randomly into two groups: one to drive without distraction and one to text and drive simultaneously. Is this a good idea? Why or why not?

e. Identify any lurking variables that could interfere with this study.

f. How can blinding be used in this study?

CHAPTER OVERVIEW

Section 2.1

2.1 Stem-and-Leaf Plots (Stemplots) and Bar Graphs

2.1 Stem-and-Leaf Plots (Stemplots) and Bar Graphs

Learning Objectives:

In this section, you will:
• Display data by constructing bar graphs and stem-and-leaf plots

Stem-and-Leaf Graphs

A **stem-and-leaf graph** is a way to plot data and look at the distribution. In a stem-and-leaf plot, all data values are visible. The advantage in a stem-and-leaf plot is that all values are listed. To create the plot, divide each observation of data into a stem and a leaf. The leaf consists of a **final significant digit**. Write the stems in a vertical line from smallest to largest. Draw a vertical line to the right of the stems. Then write the leaves in increasing order next to their corresponding stem. An **outlier** is an observation that does not fit the rest of the data.

Example 1:

For Susan Dean's spring pre-calculus class, scores for the first exam were as follows: 33; 42; 49; 49; 53; 55; 55; 61; 63; 67; 68; 68; 69; 69; 72; 73; 74; 78; 80; 83; 88; 88; 88; 90; 92; 94; 94; 94; 94; 96; 100.
Construct a stem and leaf plot for the data.

Example 2:

The data are the distances (in kilometers) from a home to local supermarkets. Create a stem and leaf plot using the data: 1.1; 1.5; 2.3; 2.5; 2.7; 3.2; 3.3; 3.3; 3.5; 3.8; 4.0; 4.2; 4.5; 4.5; 4.7; 4.8; 5.5; 5.6; 6.5; 6.7; 12.3.

Bar Graphs

Bar graphs consist of bars that are separated from each other. The bars are either horizontal or vertical to show comparisons among categories. One axis of the chart shows the specific categories being compared, and the other axis represents the frequency or proportion. Bar graphs are especially useful when categorical data is being used.

Example 3:

By the end of 2011, Facebook had over 146 million users in the United States. The following table outlines three age groups, the number of users in each age group, and the proportion (%) of users in each age group. Construct a bar graph using this data.

Age Group	Number of Facebook users	Proportion (%) of Facebook users
13-25	65,082,280	45%
26-44	53,300,200	36%
45-64	27,885,100	19%

For more information and examples see online textbook OpenStax Introductory Statistics pages 67-77.

CHAPTER OVERVIEW

Section 2.2

2.2 Frequency Table and Histograms
Lab Assignment 2.1, 2.2
Self-Check 2.1, 2.2

2.2 Frequency Table and Histograms

Learning Objectives:

In this section, you will:
• Display data by constructing frequency table and histograms.

Constructing a Frequency Table

The following data represents the amount of Tar (mg) in various nonfiltered cigarettes. Use the data to construct a frequency table and histogram.

11; 18; 18; 18; 18; 18; 18; 19; 19; 19; 20; 20; 20; 20; 21; 21;22;23;23;23;25;25;25;27;28

1. Decide how many classes to represent data. Usually 5 to 15.

Find **range = largest value – smallest value**

2. Calculate class width.

3. Calculate lower class limits.

4. Calculate upper class limits.

5. Find Frequency and Relative Frequency.

Tar (mg) in Nonfiltered Cigarettes	Frequency	Relative Frequency

Constructing a Frequency Table

A **histogram** is a graphic version of a frequency distribution. The graph consists of bars of equal width drawn adjacent to each other. The horizontal scale represents classes of quantitative data values and the vertical scale represents frequencies. The heights of the bars correspond to frequency or proportion. Histograms are typically used for large, continuous, quantitative data sets.

1. Calculate class boundaries.

2. Calculate class midpoints.

3. Graph histogram using the class boundaries and frequency.

4. Graph histogram using the class midpoints and relative frequency in percent.

For more information and examples see online textbook OpenStax Introductory Statistics pages 67-77.

Lab Assignment 2.1, 2.2

Name:____Date:____________________Row:________

Lab Assignment 2.1, 2.2

1. The miles per gallon rating for 30 cars are shown below (lowest to highest). Create a stem plot using the data and identify any outliers: 19, 19, 19, 20, 21, 21, 25, 25, 25, 26,
26, 28, 29, 31, 31, 32, 32, 33, 34, 35, 36, 37, 37, 38, 38, 38, 38, 41, 43, 43.

2.The following data show the distances (in miles) from the homes of off-campus statistics students to the college. Create a stem plot using the data and identify any outliers: 0.5; 0.7; 1.1; 1.2; 1.2; 1.3; 1.3; 1.5; 1.5; 1.7; 1.7; 1.8; 1.9; 2.0; 2.2; 2.5; 2.6; 2.8; 2.8; 2.8; 3.5; 3.8; 4.4; 4.8; 4.9; 5.2; 5.5; 5.7; 5.8; 8.0.

The students in Ms. Ramirez's math class have birthdays in each of the four seasons. The table shows the four seasons, the number of students who have birthdays in each season, and the percentage (%) of students in each group.

Seasons	Number of students	Proportion of Population
Spring	8	24%
Summer	9	26%
Autumn	11	32%
Winter	6	18%

3. Construct a bar graph showing the number of students.

4.Using the data from Mrs. Ramirez's math class, construct a bar graph showing the percentages.

The following data are the shoe sizes of 50 male students. The sizes are continuous data since shoe size is measured. 9; 9; 9.5; 9.5; 10; 10; 10; 10; 10; 10; 10.5; 10.5; 10.5; 10.5; 10.5; 10.5; 10.5; 10.5; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11.5; 11.5; 11.5; 11.5; 11.5; 11.5; 11.5; 12; 12; 12; 12; 12; 12; 12; 12.5; 12.5; 12.5; 12.5; 14.

5. Using this data, create a frequency table. Beginning with a lower class limit of 8.5 and a class width of 1. Also include the relative frequency.

Shoe Size	Frequency	Relative Frequency

6. Construct a histogram and using class midpoint or class boundaries and frequency.

Self-Check 2.1, 2.2

Name:________________________________Date:________________Row:__________

Self-Check 2.1, 2.2

1. For the Park City basketball team, scores for the last 30 games were as follows (smallest to largest): 32; 32; 33; 34; 38; 40; 42; 42; 43; 44; 46; 47; 47; 48; 48; 48; 49; 50; 50; 51; 52; 52; 52; 53; 54; 56; 57; 57; 60; 61. Construct a stem plot for the data.

2. The population in Park City is made up of children, working-age adults, and retirees. The frequency table below shows the three age groups, the number of people in the town from each age group, and the proportion (%) of people in each age group. Construct a bar graph showing the proportions.

Age groups	Number of people	Proportion of population
Children	67,059	19%
Working-age adults	152,198	43%
Retirees	131,662	38%

3.The following data represent the number of employees at various restaurants in New York City. 22; 35; 15; 26; 40; 28; 18; 20; 25; 34; 39; 42; 24; 22; 19; 27; 22; 34; 40; 20; 38; and 28. Using this data, create a frequency table. Beginning with a lower class limit of 10 and a class width of 10. Also include the relative frequency.

Number of Employees	Frequency	Relative Frequency

4. Construct a histogram using class boundaries and frequency.

5. Construct a histogram using class midpoints and relative frequency.

CHAPTER OVERVIEW

Section 2.3

2.3 Measures of the Location of the Data

2.3 Measures of the Location of the Data

Learning Objectives:

In this section, you will:
• Measure the centers of data, including mean, median, and mode

Percentile and Quartiles

Percentiles divide ordered data into hundredths. To score in the 90th percentile of an exam does not mean, necessarily, that you received 90% on a test. It means that 90% of test scoresare the same or less than your score and 10% of the test scores are the same or greater than your test score.

Quartiles divide ordered data into quarters. Quartiles may or may not be part of the data. The first quartile, Q1, is the same as the 25th percentile, and the third quartile, Q3, is the same as the 75th percentile. The median, M, is called both the second quartile, Q2 same as the 50th percentile.

Example 1: Consider the following data. 1; 11.5; 6; 7.2; 4; 8; 9; 10; 6.8; 8.3; 2; 2; 10; 1
Order Data:

a. Find 50th percentile: _____ What is another name for this value?

b.Find 25th percentile: _____Find 75th percentile: _____What is another name for this value?

c.What is another name for this value?

The **interquartile range (IQR)** is a number that indicates the spread of the middle half or the middle 50% of the data. It is the difference between the third quartile (Q3) and the first quartile (Q1).
$IQR = Q_3 – Q_1$

The IQR can help to determine potential outliers. A value is suspected to be a potential outlier if it is less than (1.5)(IQR) below the first quartile or more than (1.5)(IQR) above the third quartile.
Potential outliers: Lower bound = Q_1 – 1.5*IQR and Upper bound = Q_3 + 1.5*IQR

Example 2

For the following 13 real estate prices, calculate the IQR and determine if any prices are potential outliers. Prices are in dollars. 389,950; 230,500; 158,000; 479,000; 639,000; 114,950; 5,500,000; 387,000; 659,000; 529,000; 575,000; 488,800; 1,095,000

A Formula for Finding the kth Percentile
• k = the kth percentile. It may or may not be part of the data.
• i = the index (ranking or position of a data value)
• n = the total number of data
o Order the data from smallest to largest.
o Calculate i = k/100 * n
▪ If i is not a whole number, round up to the next whole number. The value of the k th percentile is the i th value, starting from the lowest value.
▪ If i is a whole number, the percentile is between the i th value and the next data value. Find the mean of those two values.

Example 3

Listed are 28 ages for Academy Award winning best actors in order from smallest to largest. 18; 21; 22; 25; 26; 27; 29; 30; 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69;71; 72; 73; 74; 76

a. Find the 70th percentile.

b. Find the 25th percentile.

Example 4

Fifty statistics students were asked how much sleep they get per school night (rounded to the nearest hour). The results were:

Amount of sleep per school night (hours)	Frequency	Relative Frequency	Cumulative Relative Frequency
4	2	0.04	0.04
5	5	0.10	0.14
6	7	0.14	0.28
7	12	0.24	0.52
8	14	0.28	0.80
9	7	0.14	0.94
10	3	0.06	1.00

a. Find the median. _________

b.c.Find the third quartile. __________

Find the 28th percentile. ________

Formula for Finding the Percentile of a Value in a Data Set

• Order the data from smallest to largest.

• x = the number of data values counting from the bottom of the data list up to but not including the data value for which you want to find the percentile.

• n = the total number of data.

• Calculate x/n * 100. Then round to the nearest whole number.

Example 5

Listed are 29 ages for Academy Award winning best actors in order from smallest to largest. 18; 21; 22; 25; 26; 27; 29; 30; 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77

a. Find the percentile for 58.

b. Find the percentile for 25.

For more information and examples see online textbook OpenStax Introductory Statistics pages 87-96.

CHAPTER OVERVIEW

Section 2.4

2.4 Box Plots
Lab Assignment 2.3, 2.4
Self-Check 2.3, 2.4

2.4 Box Plots

Learning Objectives:

In this section, you will:
• Represent the five-number summary of data using a boxplot

Box plots (also called box-and-whisker plots or box-whisker plots) give a good graphical image of the concentration of the data. They also show how far the extreme values are from most of the data.

A box plot is constructed from five values, **5-Number Summary:** the minimum value, the first quartile, the median, the third quartile, and the maximum value. We use these values to compare how close other data values are to them.

It is important to start a box plot with a scaled number line. Otherwise the box plot may not be useful.

Example 1

Consider, this dataset.

1; 1; 2; 2; 4; 6; 6.8; 7.2; 8; 8.3; 9; 10; 10; 11.5

Using the graphing calculator to find the 5-Number Summary:

• Enter data into the list editor (Press STAT 1:EDIT). If you need to clear the list, arrow up to the name L1, press CLEAR, and then arrow down.
• Put the data values into the list L1.
• Press STAT and arrow to CALC.
• Press 1:1-VarStats.
• Enter L1. Press ENTER.
• Use the down and up arrow keys to scroll.
5-number summary:

Construct a box plot:

Using the graphing calculator to graph box plot:

• Enter data into the list editor (Press STAT 1:EDIT). If you need to clear the list, arrow up to the name L1, Press CLEAR, and then arrow down.
• Clear equations. Press y= and clear.
• Configure plot. Press 2nd and y=.
• Select 1:Polt 1 and Press ENTER.
• Select On and Press ENTER.
• Type: Select the box plot picture and Press ENTER
• Xlist: Enter L1
• Freq: Enter 1
• Color: Select a color
• Press Zoom and Select 9:ZoomStat

Example 2

Graph a box plot for the data values shown.
10; 10; 10; 15; 35; 75; 90; 95; 100; 175; 420; 490; 515; 515; 790
.

For more information and examples see online textbook OpenStax Introductory Statistics pages 96-100.

“Introduction to Statistics” by OpenStax, used is licensed under a Creative Commons Attribution License 4.0 license

Lab Assignment 2.3, 2.4

Name:____Date:____________________Row:________

Lab Assignment 2.3, 2.4

1. Listed are 29 ages for Academy Award winning best actors in order from smallest to largest. 18; 21; 22; 25; 26; 27; 29; 30; 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64;
67; 69; 71; 72; 73; 74; 76; 77.

a. Find the 40th percentile.

b. Find the 78th percentile.

2. Listed are 32 ages for Academy Award winning best actors in order from smallest to largest. 18; 18; 21; 22; 25; 26; 27; 29; 30; 31; 31; 33; 36; 37; 37; 41; 42; 47; 52; 55; 57;
58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77.

a. Find the percentile of 37.

b. Find the percentile of 72.

c. Find the first quartile.

d. Find the second quartile = median = 50th percentile.

e. Find the third quartile.

f. Interquartile range (IQR).

g. Determine if any ages are outliers.

3.Six hundred adult Americans were asked by telephone poll, "What do you think constitutes a middle-class income?" The results are in Table 2.69. Also, include left
endpoint, but not the right endpoint.

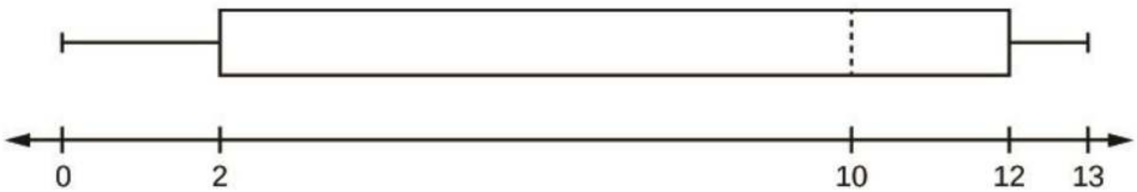
Salary ($)	Relative Frequency
< 20,000	0.02
20,000–25,000	0.09
25,000–30,000	0.19
30,000–40,000	0.26
40,000–50,000	0.18
50,000–75,000	0.17
75,000–99,999	0.02
100,000+	0.01

a. What percentage think that middle-class is from $25,000 to $50,000?

b. Find the 40th percentile.

c. Find the 80th percentile.

4. Given the following box plot:



In this box plot, the first whisker extends from the smallest value, 0, to the first quartile, 2. The box begins at the first quartile and extends to the third quartile, 12. A vertical, dashed line is drawn at the median, 10. The second whisker extends from the third quartile to the largest value, 13.

a. Which quarter has the smallest spread of data? What is that spread?

b. Which quarter has the largest spread of data? What is that spread?

c. Find the interquartile range (IQR).

d.Are there more data in the interval 5–10 or in the interval 10–13? How do you know this?

e.Which interval has the fewest data in it? How do you know this?

i. 0–2
ii. 2–4
iii. 10–12
iv. 12–13
v. need more information

5.Listed are 29 ages for Academy Award winning best actors in order from smallest to largest. 18; 21; 22; 25; 26; 27; 29; 30; 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77.

a. Find the 5-number summary.

b. Construct a box plot.

Self-Check 2.3, 2.4

Name:____Date:____________________Row:________

Self-Check 2.3, 2.4

1. For the following 11 salaries, calculate the IQR and determine if any salaries are outliers. The salaries are in dollars. $33,000; $64,500; $28,000; $54,000; $72,000; $68,500;
$69,000; $42,000; $54,000; $120,000; $40,500.

2.Forty bus drivers were asked how many hours they spend each day running their routes (rounded to the nearest hour).

Amount of time spent on route (hours)	Frequency	Relative Frequency	Cumulative Relative Frequency
2	12	0.30	0.30
3	14	0.35	0.65
4	10	0.25	0.90
5	4	0.10	1.00

a. Find the 65th percentile.

b. Find the third quartile.

3. Listed are 29 ages for Academy Award winning best actors in order from smallest to largest. 18; 21; 22; 25; 26; 27; 29; 30; 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77.

a. Find the 20th percentile.

b. Find the 55th percentile.

4.Listed are 30 ages for Academy Award winning best actors in order from smallest to largest. 18; 21; 22; 25; 26; 27; 29; 30; 31, 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77.

a. Find the percentiles for 47.

b. Find the percentiles for 31.

5. Construct a box plot for the data values shown.

0; 5; 5; 15; 30; 30; 45; 50; 50; 60; 75; 110; 140; 240; 330.

CHAPTER OVERVIEW

Section 2.5

2.5 Measures of Center of the Data

2.5 Measures of Center of the Data

Learning Objectives

In this section, you will:
• Measure the centers of data, including mean, median, and mode.

The "center" of a data set is also a way of describing location. The two most widely used measures of the "center" of the data are the **mean** (average) and the **median.**

Mean and Median

Mean: add all data, divide by the total number of values.

Formulas: $\bar{x} = \frac{\sum x}{n}$ for sample mean or $\mu = \frac{\sum x}{N}$ for population mean.

Median: middle value when the data are placed in order. If there is no middle value, find the mean of the two middle values.
• The median is generally a better measure of the center when there are extreme values or outliers because it is not affected by the precise numerical values of the outliers.
• You can quickly find the location of the median by using the expression (n + 1)/2.

Example 1

Consider the following data: 19; 18; 18; 25; 24; 32; 45; 29; 17; 18; 53; 30; 20; 21.

Find mean.

Find median.

Using the graphing calculator to find the mean and median.

• Clear list L1. Press STAT 4: ClrList. Enter 2nd 1 for list L1. Press ENTER.
• Enter data into the list editor. Press STAT 1: EDIT.
• Put the data values into list L1.
• Press STAT and arrow to CALC. Press 1:1-VarStats. Press 2nd 1 for L1 and then Calculate.
• Press the down and up arrow keys to scroll.

Example 2

AIDS data indicating the number of months a patient with AIDS lives after taking a new antibody drug are as follows (smallest to largest): 3; 4; 8; 8; 10; 11; 12; 13; 14; 15; 15; 16; 16; 17; 17; 18; 21; 22; 22; 24; 24; 25; 26; 26; 27; 27; 29; 29; 31; 32; 33; 33; 34; 34; 35; 37; 40; 44; 44; 47.

Calculate the mean and the median.

Mode

Another measure of the center is the mode. The is the most frequent value. There can be more than one mode in a data set as long as those values have the same frequency and that frequency is the highest. A data set with two modes is called bimodal.

Example 3

Statistics exam scores for 20 students are as follows: 50; 53; 59; 59; 63; 63; 72; 72; 72; 72; 72;
76; 78; 81; 83; 84; 84; 84; 90; 93.

Find the mode.

Calculating the Mean of Frequency Distribution

When only grouped data is available, you do not know the individual data values (we only know intervals and interval frequencies); therefore, you cannot compute an exact mean for the data set. What we must do is estimate the actual mean by calculating the mean of a frequency table.

Mean of frequency table = $\frac{\Sigma(f*m)}{\Sigma f}$, where:

- o f = the frequency of the interval
- o m = the midpoint of the interval

Example 4

A frequency table displaying Professor Blount's last statistic test is shown. Find the best estimate of the class mean.

Grade Interval	Number of Students
50.5-56.4	1
56.5-62.4	0
62.5-68.4	4
68.5-74.4	4
74.5-80.4	2
80.5-86.4	2
86.5-92.4	4
92.5-98.4	1

Using the graphing calculator to find the mean of grouped frequency tables.

- Clear list L1. Pres STAT 4:ClrList. Enter 2nd 1 for list L1. Press ENTER.
- Enter data into the list editor. Press STAT 1:EDIT.
- Put the midpoint values into list L1.
- Put the frequency values into list L2.
- Press STAT and arrow to CALC. Press 1:1-VarStats.
- List: Press 2nd 1 for L1.
- FreqList: Press 2nd 2 for L2. and then Calculate.
- Press the down and up arrow keys to scroll.

For more information and examples see online textbook OpenStax Introductory Statistics pages 100- 106.

CHAPTER OVERVIEW

Section 2.6

2.6 Skewness and the Mean, Median, and Mode

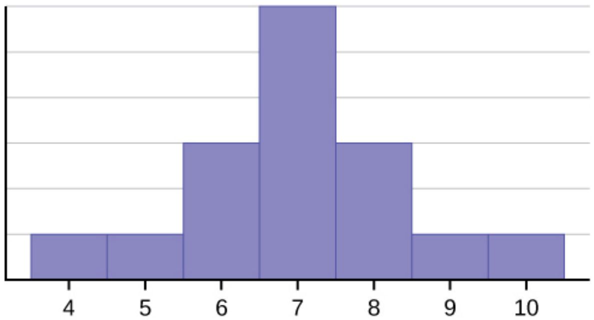
2.6 Skewness and the Mean, Median, and Mode

Learning Objectives:

In this section, you will:
• Measure skewness, including mean, median, and mode

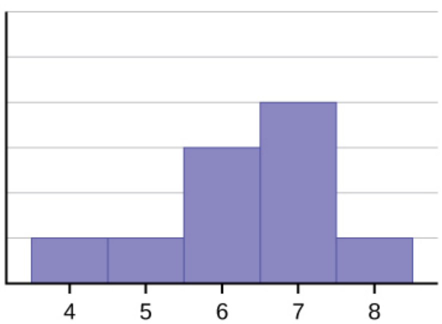
Symmetrical Distribution

Consider the following data set. 4, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 10.



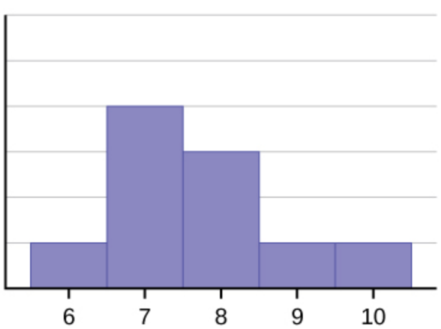
The histogram displays a symmetrical distribution of data. A distribution is symmetrical if a vertical line can be drawn at some point in the histogram such that the shape to the left and the right of the vertical line are mirror images of each other. The mean, the median, and the mode are each seven for these data. **In a perfectly symmetrical distribution, the mean and the median are the same.**

Skewed Distributions



The histogram for the data: 4; 5; 6; 6; 6; 7; 7; 7; 7; 8 is not symmetrical. The right-hand side seems "chopped off" compared to the left side. A distribution of this type is called **skewed to the left** because it is pulled out to the left.

The mean is 6.3, the median is 6.5, and the mode is seven. **Notice that the mean is less than the median, and they are both less than the mode.** The mean and the median both reflect the skewing, but the mean reflects it more so.



The histogram for the data: 6; 7; 7; 7; 7; 8; 8; 8; 9; 10, is also not symmetrical. It is **skewed to the right.**

The mean is 7.7, the median is 7.5, and the mode is seven. Of the three statistics, **the mean is the largest, while the mode is the smallest.** Again, the mean reflects the skewing the most.

For more information and examples see online textbook OpenStax Introductory Statistics pages 106-109.

CHAPTER OVERVIEW

Section 2.7

2.7 Measures of the Spread of the Data
Lab Assignment 2.5, 2.6, 2.7
Self-Check 2.5, 2.6, 2.7

2.7 Measures of the Spread of the Data

Learning Objectives

In this section, you will:
• Measure the spread of data, including standard deviation and variance

An important characteristic of any set of data is the variation in the data. In some data sets, the data values are concentrated closely near the mean; in other data sets, the data values are more widely spread out from the mean. The most common measure of variation, or spread, is the standard deviation.

Standard Deviation: is a number that measures how far data values are from their mean.
• Standard deviation is always positive or zero
• The standard deviation is small when the data are concentrated close to the mean, exhibiting little variation or spread
• The standard deviation is larger when the data values are more spread out from the mean, exhibiting more variation.
• Standard deviation is zero if all data is the same number
• Outliers can have a strong effect on the standard deviation

- $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ sample
- $\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$ population

Range: Highest data value minus lowest data value

Round off rule: Carry one more decimal place than the original data

Example 1

Data set #1: 5, 6, 6, 10, 10, 11 $\bar{x}$ = Range = s=
Data set #2: - 2, 4, 6, 10, 12, 18 $\bar{x}$ = Range = s=

To find values that are a certain number of standard deviations above or below the mean.

Value = mean ± (# of standard deviations)*(standard deviation)

x = $\bar{x}$ ± (# of *STDEV*) ∗ (s)

Variance: is the square of the standard deviation (s^2 or σ^2)

Using the graphing calculator to find variance

• First find standard deviation
• Press VARS
• Press 5:Statistics
• Press 3:Sx
• Press x2
• Press ENTER

Example 2

For data set #1 find s^2

For data set #2 find s^2

Example 3

In a fifth grade class, the teacher was interested in the average age and the sample standard deviation of the ages of her students. The following data are the ages for a SAMPLE of n = 20 fifth grade students. The ages are rounded to the nearest half year:

9; 9.5; 9.5; 10; 10; 10; 10; 10.5; 10.5; 10.5; 10.5; 11; 11; 11; 11; 11; 11; 11.5; 11.5; 11.5;

a. Find the mean and standard deviation on your calculator.

b. Find the value that is one standard deviation above the mean.

c. Find the value that is two standard deviations below the mean.

d. Find the values that are 1.5 standard deviations from the mean.

Example 4

Find the standard deviation and variance of the frequency table

Class	Frequency
0-2	1
3-5	6
6-8	10
9-11	7
12-14	0
15-17	2

Comparing Values from Different Data Sets

Z-score (standardized value) – the number of standard deviations a given value of x is above or below the mean.

Sample, $z = \frac{x-\bar{x}}{s}$ or Population, $z = \frac{x-\mu}{\sigma}$

Example 5

Which is a better score? A score of 76 on a quiz where the mean score was 64 and the standard deviation was 6.7, or a score of 12 on a quiz with a mean of 10 and a standard deviation of 1.1?

Example 6

Two students, John and Ali, from different high schools, wanted to find out who had the highest GPA when compared to his school. Which student had the highest GPA when compared to his school?

Student	GPA	School Mean GPA	School Standard Deviation
John	2.85	3.0	0.7
Ali	77	80	10

For more information and examples see online textbook OpenStax Introductory Statistics pages 110 - 120.

Lab Assignment 2.5, 2.6, 2.7

Name:____Date:____________________Row:________

Lab Assignment 2.5, 2.6, 2.7

1. The most obese countries in the world have obesity rates that range from 11.4% to 74.6%. This data is summarized in the following table.

Percent of Population Obese	Number of Countries
11.45-20.44	29
20.45-29.44	13
29.45-38.33	4
38.45-47.44	0
47.45-56.44	2
56.45-65.44	1
65.45-74.44	0
74.45-83.44	1

a. What is the best estimate of the mean obesity percentage for these countries?

b. The United States has an average obesity rate of 33.9%. Is this rate above average or below?

2. Use the following information to answer the next three exercises: The following data show the lengths of boats moored in a marina. The data are ordered from smallest to largest: 16; 17; 19; 20; 20; 21; 23; 24; 25; 25; 25; 26; 26; 27; 27; 27; 28; 29; 30; 32; 33; 33; 34; 35; 37; 39; 40

a. Calculate the mean.

b. Identify the median.

c. Identify the mode.

d. Calculate the standard deviation.

e. Calculate the variance.

3. The table gives the percent of children under five considered to be underweight.

Percent of Underweight Children	Number of Countries
16.00-21.44	23
21.45-26.89	4
26.90-32.34	9
32.35-37.79	7
37.80-43.24	6
43.25-48.69	1

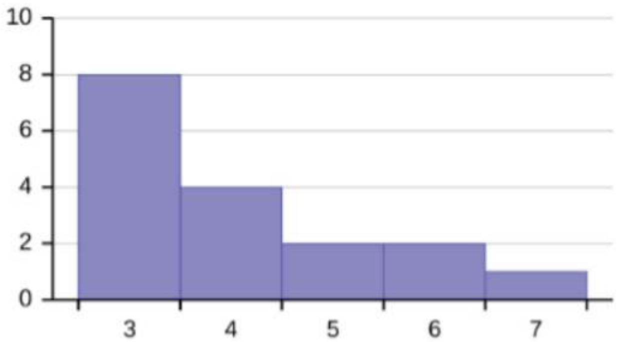
a. What is the best estimate for the mean percentage of underweight children?

b. What is the standard deviation and variance?

4. When the data are skewed left, what is the typical relationship between the mean and median?

5. When the data are symmetrical, what is the typical relationship between the mean and median?

6. Describe the shape of this distribution and describe the relationship between the mean and the median of this distribution.



7. Use the following information to answer the next two exercises: The following data are the distances between 20 retail stores and a large distribution center. The distances are in miles.
29; 37; 38; 40; 58; 67; 68; 69; 76; 86; 87; 95; 96; 96; 99; 106; 112; 127; 145; 150

a. Find the mean and standard deviation and round to the nearest tenth.

b. Find the value that is one standard deviation below the mean.

8. Two baseball players, Fredo and Karl, on different teams wanted to find out who had the higher batting average when compared to his team. Which baseball player had the higher batting average when compared to his team?

Baseball Player	Batting Average	Team Batting Average	Team Standard Deviation
Fredo	0.158	0.166	0.012
Karl	0.177	0.189	0.015

Self-Check 2.5, 2.6, 2.7

Name:____________________________________Date:__________________Row:___________

Self-Check 2.5, 2.6, 2.7

1. The following data show the number of months patients typically wait on a transplant list before getting surgery. The data are ordered from smallest to largest.

3; 4; 5; 7; 7; 7; 7; 8; 8; 9; 9; 10; 10; 10; 10; 10; 11; 12; 12; 13; 14; 14; 15; 15; 17; 17; 18; 19; 19; 19; 21; 21; 22; 22; 23; 24; 24; 24; 24

a. Calculate the mean.

b. Identify the median.

c. Identify the mode.

d. Calculate the standard deviation.

e. Calculate the variance.

2. In a sample of 60 households, one house is worth $2,500,000. Half of the rest are worth $280,000, and all the others are worth $315,000. Which is the better measure of the “center”: the mean or the median?

3. Maris conducted a study on the effect that playing video games has on memory recall. As part of her study, she compiled the following data:

Hours Teenagers Spend on Video Games	Number of Teenagers
0.0-4.4	3
4.5-8.9	7
9.0-13.4	12
13.5-17.9	7
18-22.4	9

a. What is the best estimate for the mean number of hours spent playing video games?

b. What is the standard deviation and variance?

4. On a baseball team, the ages of each of the players are as follows:
21; 21; 22; 23; 24; 24; 25; 25; 28; 29; 29; 31; 32; 33; 33; 34; 35; 36; 36; 36; 36; 38; 38; 38; 40

a. Use your calculator or computer to find the mean and standard deviation.

b. Find the value that is two standard deviations above the mean.

c. Find the value that is one standard deviations below the mean.

5. Two swimmers, Angie and Beth, from different teams, wanted to find out who had the fastest time for the 50 meter freestyle when compared to her team. Which swimmer had the fastest time when compared to her team?

Swimmer	Time (seconds)	Team Mean Time	Team Standard Deviation
Angie	26.2	27.2	0.8
Beth	27.3	30.1	1.4

CHAPTER OVERVIEW

Section 3.1

3.1 Terminology
Lab Assignment 3.1
Self-Check 3.1

3.1 Terminology

Learning Objective:

In this section, you will:
• Understand the fundamentals of probability

Probability is a number between zero and one, inclusive, that gives the likelihood that a specific event will occur.
• The probability of any outcome is the long-term relative frequency of that outcome.
• P(A) = 0 means the event A can never happen.
• P(B) = 1 means the event B is certain to happen.
• P(C) = 0.5 means that event C is equally likely to happen or not happen.

An **experiment** is a planned activity carried out under controlled conditions.

An **outcome** is particular result of an experiment.

The **sample space** is the set of all possible outcomes of an experiment.

An **event** is a collection of results or outcomes of an experiment.

Equally likely means that each outcome of an experiment has the same probability.

To calculate the probability of an event A when all outcomes in the sample space are equally likely, count the number of outcomes for event A and divide by the total number of outcomes in the sample space.

Notation:
• P represents a probability
• A, B, and C represent specific events
• P(A) represents the probability of event A occurring
• P(A) = (number of outcomes for event A)/ (total number of outcomes in the sample space)

The **law of large numbers** states that as the number of repetitions of an experiment is increased, the relative frequency obtained in the experiment tends to become closer and closer to the theoretical probability.

Example 1

Roll a standard six-sided die once.

Sample space: S =

Identify each of the following events with a subset of S and compute its probability.

a. Event A = Roll an even number

b. Event B = Roll a number larger than 4

c. Event C = Roll a 2

d. Event D = Roll a 9

e. Event E = Roll number less than 10

The **complement** of event A consists of all outcomes that are **NOT** in A. Which is denoted as A′ (read "A prime") or $\bar{A}$.

P(A')= P(B')= P(C') =

Notice that P(A) + P(A′) = 1.

"OR" Event

An outcome is in the event **A OR B** if the outcome is in A or is in B or is in both A and B

Example 2: let A = {1, 2, 3, 4, 5} and B = {4, 5, 6, 7, 8}.

A OR B =

"AND" Event

An outcome is in the event **A AND B** if the outcome is in both A and B at the same time.

Example 3: let A and B be {1, 2, 3, 4, 5} and {4, 5, 6, 7, 8}.

A AND B =

"GIVEN" Event

The conditional probability of A given B is written P(A|B). P(A|B) is the probability that event A will occur given that the event B has already occurred. A conditional reduces the sample space. We calculate the probability of A from the reduced sample space B.

Example 4: Suppose we toss one fair, six-sided die. The sample space S = {1, 2, 3, 4, 5, 6}. Let A = face is 2 or 3 and B = face is even (2, 4, 6).

A GIVEN B =

Mutually Exclusive Events

A and B are mutually exclusive events if they cannot occur at the same time. This means that A and B do not share any outcomes.

Example 5: Suppose the sample space S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. Let A = {1, 2, 3, 4, 5}, B = {4, 5, 6, 7, 8}, and C = {7, 9}.

A AND B =

A AND C =

For more information and examples see online textbook OpenStax Introductory Statistics pages 176- 180.

Lab Assignment 3.1

Name:____Date:_____________________Row:________

Lab Assignment 3.1

1. The sample space S is the whole numbers starting at one and less than 20.

a. S =

Let A = the even numbers, Let B= numbers greater than 13.

b. A = B =

c. P(A) = P(B) =

d. A AND B= P(A AND B) =

e. A OR B= P(A OR B) =

f. A′= P(A′) =

g. P(A) +P(A′) =

h. A GIVEN B P(A|B) =

i. B GIVEN A P(B|A) =

2. A box is filled with several party favors. It contains 12 hats, 15 noisemakers, 10 finger traps, and 5 bags of confetti.

- Let H = the event of getting a hat.
- Let N = the event of getting a noisemaker.
- Let F = the event of getting a finger trap.
- Let C = the event of getting a bag of confetti.

a. P(H) =

b. P(N OR F) =

c. P(C’) =

3. You are rolling a fair, six-sided number cube. Let E = the event that it lands on an even number. Let M = the event that it lands on a multiple of three.

a. What does P(E|M) mean in words?

b. What does P(E OR M) mean in words?

4. What is the sum of the probabilities of an event and its complement?

5. What is the probability of drawing a club in a standard deck of 52 cards?

6. What is the probability of rolling a prime number of dots with a fair, six-sided die numbered one through six?

Self-Check 3.1

Name:____________________________________Date:___________________Row:____________

Self-Check 3.1

1. A fair, six-sided die is rolled.

Describe the sample space S =

Identify each of the following events with a subset of S and compute its probability.

a. Event T= the outcome is two.

b. Event A= the outcome is an even number.

c. Event B= the outcome is less than four.

d. The complement of A.

e. A GIVEN B

f. B GIVEN A

g. A AND B

h. A OR B

i. A OR B′

j. Event N = the outcome is a prime number.

k. Event I = the outcome is seven.

LibreTexts

CHAPTER OVERVIEW

Section 3.2-3.4

3.2-3.4 Probability Topics
Lab Assignment 3.2, 3.3, 3.4
Self-Check 3.2, 3.3, 3.4

3.2-3.4 Probability Topics

Section 3.2-3.4 Probability Topics

Learning Objective:

In this section, you will:

- Calculate probabilities of events, including conditional and compound events (and's, or's).
- Read and construct contingency tables to calculate probabilities.

If A and B are defined on a sample space, then: $P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$. If A and B are **mutually exclusive**, then $P(A \text{ OR } B) = P(A) + P(B)$.

Example 1: A box contains 5 white marbles, 2 green, 8 blue, and 4 red marbles. **1 marbles** is selected.

P(green or blue) =

Example 2:

	Male	Female	Total
Scored a C or higher on the exam	28	41	
Scored a D or F on the exam	5	4	
Total			

Let A = Scored a C or higher on the exam

Let B = Scored a D or F on the exam

Let M = Male

Let F = Female

1. P(A) =
2. P(M) =
3. P(A OR M) =
4. P(A AND M) =
5. P (A|M) =

Independent Events: Two events A and B are independent if the knowledge that one occurred does not affect the chance the other occurs. For example, the outcomes of two roles of a fair die are independent events. The outcome of the first roll does not change the probability for the outcome of the second roll.

Sampling with replacement: If each member of a population is replaced after it is picked, then that member has the possibility of being chosen more than once. When sampling is done with replacement, then events are considered to be **independent**, meaning the result of the first pick will not change the probabilities for the second pick.

Sampling without replacement: When sampling is done without replacement, each member of a population may be chosen only once. In this case, the probabilities for the second pick are affected by the result of the first pick. The events are considered to be **dependent** or not independent.

If it is not known whether A and B are independent or dependent, **assume they are dependent until you can show otherwise**.

If A and B are **independent**, then $P(A \text{ AND } B) = P(A)P(B)$

If A and B are **dependent**, then $P(A \text{ AND } B) = P(A)P(B|A) = P(B)P(A|B)$

Example 3: A box contains 5 white marbles, 2 green, 8 blue, and 4 red marbles. **3 marbles** are selected.

1. Find P(All three are red) With replacement:

Without replacement:

2. Find P(The first is green, and the second two are blue) With replacement:

Without replacement:

3. Find P(all three are green) With replacement:

Without replacement:

2

Example 4: Reported use of vaping is 27.8% of 12th graders in the United States. If 4 randomly selected 12th graders are selected, find the probability that they all vape.

Example 5:

	High school	Community College	University
Male	7	13	10
Female	12	8	11

1. P(selecting a student that is in community college)
2. P(selecting a student is male AND in high school)
3. P(selecting a student is male OR in high school)
4. P(selecting male, GIVEN that it is a high school student)
5. P(selecting a high school student, GIVEN that is a male)
6. If four students are selected, find:
7. P(All are university students)

2. P(none are university students)

For more information and examples see online textbook OpenStax Introductory Statistics pages 181-198.

Lab Assignment 3.2, 3.3, 3.4

Name:____Date:_____________________Row:________

Lab Assignment 3.2, 3.3, 3.4

1. The table below describes the distribution of a random sample S of 100 individuals, organized by gender and whether they are right- or left-handed.

	Right-handed	Left-handed
Male	43	9
Female	44	4

Let's denote the events M = the subject is male, F = the subject is female, R = the subject is right-handed, L = the subject is left-handed. Compute the following probabilities:

1. $P(M)$
2. $P(F)$
3. $P(R)$
4. $P(L)$
5. $P(M \text{ AND } R)$
6. $P(F \text{ AND } L)$
7. $P(M \text{ OR } F)$
8. $P(M \text{ OR } R)$
9. $P(F \text{ OR } L)$
10. $P(M')$
11. $P(R|M)$
12. $P(F|L)$
13. $P(L|F)$

2. The table below shows a random sample of musicians and how they learned to play their instruments.

Gender	Self-taught	Studied in School	Private Instruction	Total
Female	12	38	22	72
Male	19	24	15	58
Total	31	62	37	130

1. Find P(musician is a female).
2. Find P(musician is a male AND had private instruction).
3. Find P(musician is a female OR is self taught).
4. If three musicians are randomly selected, with replacement, find the probability they are all self-taught.
5. If three musicians are randomly selected, without replacement, find the probability they are all males.

3. At a college, 72% of courses have a final exam. If we randomly select 5 courses, find the probability that they all have a final exam.
4. The casino game, roulette, allows the gambler to bet on the probability of a ball, which spins in the roulette wheel, landing on a particular color, number, or range of numbers. The table used to place bets contains of 38 numbers, and each number is assigned

to a color and a range.

1. List the sample space of the 38 possible outcomes in roulette.
2. You bet on red. Find P(red).
3. You bet on -1st 12- (1st Dozen). Find P(-1st 12-).
4. You bet on an even number. Find P(even number).
5. Is getting an odd number the complement of getting an even number? Why?

3

5. The table below identifies a group of children by one of four hair colors, and by type of hair.

Hair Type	Brown	Blond	Black	Red	Totals
Wavy	20		15	3	43
Straight	80	15		12	
Totals		20			215

1. Complete the table.
2. What is the probability that a randomly selected child will have wavy hair?
3. What is the probability that a randomly selected child will have either brown or blond hair?
4. What is the probability that a randomly selected child will have wavy brown hair?
5. What is the probability that a randomly selected child will have red hair, given that he or she has straight hair?
6. If B is the event of a child having brown hair, find the probability of the complement of B.
7. If two children are randomly selected **with replacement**, find the probability that they both have red hair?
8. If two children are randomly selected **without replacement**, find the probability that they both have red hair?

Self-Check 3.2, 3.3, 3.4

Name:____________________________________Date:___________________Row:____________

Self-Check 3.2, 3.3, 3.4

1. Table below describes the distribution of a random sample S of 100 individuals, organized by gender and whether they are right- or left-handed.

	Right-handed	Left-handed
Males	43	9
Females	44	4

Let's denote the events M = the subject is male, F = the subject is female, R = the subject is right-handed, L=the subject is left-handed. Compute the following probabilities:

1. P(M)
2. P(F)
3. P(R)
4. P(L)
5. P(M AND R)
6. P(F AND L)
7. P(M OR F)
8. P(M OR R)
9. P(F OR L)
10. P(M')
11. P(R|M)
12. P(F|L)
13. P(L|F)

Self-Check 3.2, 3.3, 3.4

2. The table relates the weights and heights of a group of individuals participating in an observational study.

Weight/Height	Tall	Medium	Short	Totals
Obese	18	28	14	
Normal	20	51	28	
Underweight	12	25	9	
Totals				

1. Find the total for each row and column
2. Find the probability that a randomly chosen individual from this group is Tall.
3. Find the probability that a randomly chosen individual from this group is Obese **and** Tall.
4. Find the probability that a randomly chosen individual from this group is Tall **given** that the individual is Obese.
5. Find the probability that a randomly chosen individual from this group is Obese **given** that the individual is Tall.
6. Find the probability a randomly chosen individual from this group is Tall **and** Underweight.

7. Find the probability a randomly chosen individual from this group is Tall **or** Underweight.
8. If **two individuals** are randomly chosen **with replacement**, find the probability that both individuals from this group are Short.
9. If **two individuals** are randomly chosen **without replacement**, find the probability that both individuals from this group are Short.

CHAPTER OVERVIEW

Section 4.1

4.1 Probability Distribution Function (PDF) for a Discrete Random Variable

4.1 Probability Distribution Function (PDF) for a Discrete Random Variable

Learning Objective:

In this section, you will:

• Understand and apply the fundamentals of random variables and their probability distribution functions

A **random variable** describes the outcomes of a statistical experiment in words.

- Upper case letters such as X or Y denote a random variable.
- Lower case letters like x or y denote the value of a random variable.
-

A discrete **probability distribution function (PDF)** has two characteristics:

1. Each probability is between zero and one, inclusive.
2. The sum of the probabilities is one.

Example 1: A hospital researcher is interested in the number of times the average post-op patient will ring the nurse during a 12-hour shift. For a random sample of 50 patients, the following information was obtained.

X	P(x)
0	
1	
2	
3	
4	
5	

1. Describe the random variable X in words.
2. For this exercise, what are the values of x?

Notes 4.1

3. What is the probability that the number of times a patient rings the nurse is 4?
4. What is the probability that the number of times a patient rings the nurse is at least 4?
5. What is the probability that the number of times a patient rings the nurse does not exceed 4?
6. What is the probability that the number of times a patient rings the nurse is at least 1?

Example 2: Suppose Nancy has classes three days a week. She attends classes three days a week 80% of the time, two days 15% of the time, one day 4% of the time, and no days 1% of the time. Suppose one week is randomly selected.

1. Describe the random variable X in words.
2. For this exercise, what are the values of x?
3. Suppose one week is randomly chosen. Construct a probability distribution table. What does the P(x) column sum to?

For more information and examples see online textbook OpenStax Introductory Statistics pages 244-246.

CHAPTER OVERVIEW

Section 4.2

4.2 Mean or Expected Value and Standard Deviation
Lab Assignment 4.1, 4.2
Self-Check 4.1, 4.2

4.2 Mean or Expected Value and Standard Deviation

Section 4.2 Mean or Expected Value and Standard Deviation

Learning Objective:

In this section, you will:

• Calculate the mean and standard deviation of random variables

The **expected value** is often referred to as the "long-term" average or mean. This means that over the long term of doing an experiment over and over, you would expect this average and is denoted by the Greek letter μ.

Example 1: On average, how many times would you expect a post-op patient will ring the nurse during a 12-hour shift? What is the standard deviation?

X	P(x)
0	
1	
2	
3	
4	
5	

Example 2: A men's soccer team plays soccer zero, one, or two days a week. The probability that they play zero days is 0.2, the probability that they play one day is 0.5, and the probability that they play two days is 0.3. Find the long-term average or expected value, μ, of the number of days per week the men's soccer team plays soccer. Calculate the standard deviation of the variable as well.

Notes 4.2

Example 3: Suppose you play a game with a biased coin. You play each game by tossing the coin once. P(heads) = 2/3 and P(tails) = 1/3. If you toss a head, you pay $6. If you toss a tail, you win $10. If you play this game many times, will you come out ahead?

For more information and examples see online textbook OpenStax Introductory Statistics pages 247-253.

Lab Assignment 4.1, 4.2

Name:____Date:___________________Row:________

Lab Assignment 4.1, 4.2

1. Use the following information to answer the next five exercises: A company wants to evaluate its attrition rate, in other words, how long new hires stay with the company. Over the years, they have established the following probability distribution.

Let X= the number of years a new hire will stay with the company.

Let P(x) = the probability that a new hire will stay with the company x years.

1. Complete table using the data provided.

X	P(x)
0	0.12
1	0.18
2	0.30
3	0.15
4	
5	0.10
6	0.05

2. P(x= 4) =
3. P(x≥ 5) =
4. On average, how long would you expect a new hire to stay with the company?
5. What does the column "P(x)" sum to?

2. A baker is deciding how many batches of muffins to make to sell in his bakery. He wants to make enough to sell everyone and no fewer. Through observation, the baker has established a probability distribution.

X	P(x)
1	0.15
2	0.35
3	0.40
4	0.10

1

1. Define the random variable X.
2. What is the probability the baker will sell more than one batch? P(x > 1) =
3. What is the probability the baker will sell exactly one batch? P(x = 1) =
4. On average, how many batches should the baker make?
5. What is the standard deviation?

3. Use the following information to answer the next three exercises: Ellen has music practice three days a week. She practices for all of the three days 85% of the time, two days 8% of the time, one day 4% of the time, and no days 3% of the time. One week is selected at random.

1. Define the random variable X.
2. Construct a probability distribution table for the data.
3. On average, how many days a week does Ellen practice music?

4. Use the following information to answer the next five exercises: Javier volunteers in community events each month. He does not do more than five events in a month. He attends exactly five events 35% of the time, four events 25% of the time, three events 20% of the time, two events 10% of the time, one event 5% of the time, and no events 5% of the time.

2

1. Define the random variable X.
2. What values does x take on?
3. Construct a PDF table.
4. Find the probability that Javier volunteers for less than three events each month. $P(x< 3) =$
5. Find the probability that Javier volunteers for at least one event each month. $P(x> 0) =$

5. You are playing a game by drawing a card from a standard deck and replacing it. If the card is a face card, you win $30. If it is not a face card, you pay $2. There are 12 face cards in a deck of 52 cards.

1. What is the expected value of playing the game?
2. Should you play the game?

Self-Check 4.1, 4.2

Name:____________________________________Date:___________________Row:____________

Self-Check 4.1, 4.2

1. People visiting video rental stores often rent more than one DVD at a time. The probability distribution for DVD rentals per customer at Video To Go is given in the following table. There is a five-video limit per customer at this store, so nobody ever rents more than five DVDs.

x	P(x)
0	0.03
1	0.50
2	0.24
3	
4	0.07
5	0.04

1. Describe the random variable X in words.
2. For this exercise, what are the values of x?
3. Find the probability that a customer rents three DVDs.
4. Find the probability that a customer rents at least four DVDs.
5. Find the probability that a customer rents at most two DVDs.
6. On average, how many DVDs would you expect a customer rent?
7. What is the standard deviation?

1

Self-Check 4.1, 4.2

2. Suppose you play a game with a spinner. You play each game by spinning the spinner once. P(red) = 25, P(blue) = 25, and P(green) = 15. If you land on red, you pay $10. If you land on blue, you don't pay or win anything. If you land on green, you win $10. Over the long term, what is your expected profit of playing the game?
3. On May 11, 2013 at 9:30 PM, the probability that moderate seismic activity (one moderate earthquake) would occur in the next 48 hours in Iran was about 21.42%. Suppose you make a bet that a moderate earthquake will occur in Iran during this period. If you win the bet, you win $50. If you lose the bet, you pay $20. Let X = the amount of profit from a bet. If you bet many times, will you come out ahead?

CHAPTER OVERVIEW

Section 6.1

6.1 The Standard Normal Distribution
Lab Assignment 6.1
Self-Check 6.1

6.1 The Standard Normal Distribution

Section 6.1 The Standard Normal Distribution

Learning Objective:

In this section, you will:

- Solve and explain real-world applications of probability

The **standard normal distribution** is a normal distribution of **standardized values called *z*-scores**. **A *z*-score is measured in units of the standard deviation** with . = 0 and . = 1. The total area under its density curve is equal to 1.

Z-Score

x−μ

If *X* is a normally distributed random variable and *X* ~ *N(μ, σ)*, then the *z*-score is: z=

σ

- The *z*-score tells you how many standard deviations the value *x* is above (to the right of) or below (to the left of) the mean, *μ*.
- Values of *x* that are larger than the mean have positive *z*-scores, and values of *x* that are smaller than the mean have negative *z*-scores. If *x* equals the mean, then *x* has a *z*-score of zero.
- The *z*-score allows us to compare data that are scaled differently.

Example 1: Suppose X ~ N(5, 6). This says that X is a normally distributed random variable with mean μ = 5 and standard deviation σ= 6. Suppose x= 17. Find the z-score of x.

Example 2: Suppose X ~ N(8, 1). What value of x has a z-score of -2.25?

1

Example 3: Heights of women have a normal distribution with a mean of 161 cm and a standard deviation of 7 cm.

1. What is the z-score for a woman who is 170 cm tall?
2. How many standard deviations away from the mean is a woman who is 170cm tall?

Example 4: Some doctors believe that a person can lose five pounds, on the average, in a month by reducing his or her fat in take and by exercising consistently. Suppose weight loss has a normal distribution. Let X = the amount of weight lost (in pounds) by a person in a month. Use a standard deviation of two pounds. X~N(5, 2).

1. Suppose a person lost ten pounds in a month. Find the z-score.
2. Fill in the blanks: The z-score when x = 10 pounds is z = 2.5. This z-score tells you that x = 10 is

________ standard deviations to the ________ (right or left) of the mean _____ (What is the mean?).

3. Fill in the blanks: Suppose a person gained three pounds (a negative weight loss). Then z =

__________. This z-score tells you that x = –3 is ________ standard deviations to the __________ (right or left) of the mean.

4. Suppose the random variables X and Y have the following normal distributions: X ~ N(5, 6) and Y ~ N(2, 1). If x= 17, then z= 2. (This was previously shown.) If y= 4, what is z?

2

Example 5: Compare data that are scaled differently. Which is a better score? A score of 76 on a quiz where the mean score was 64 and the standard deviation was 6.7, or a score of 12 on a quiz with a mean of 10 and a standard deviation of 1.1?

Empirical Rule

If *X* is a random variable and has a normal distribution with mean *μ* and standard deviation *σ*, then the Empirical Rule says the following:

- About 99.7% of the x values lie between -3σ and $+3\sigma$ of the mean μ Notice that almost all the x values lie within three standard deviations of the mean. The empirical rule is also known as the 68-95-99.7 rule.

Example 6: Heights of women have a normal distribution with a mean of 161 cm and a standard deviation of 7 cm.

1. Approximately what is the percentage of women between 147 cm and 175 cm?

3

2. About 99.7% of women are between what two heights?
3. What percentage of women are between 154 and 182 cm

Example 7: The mean height of 15 to 18-year-old males from Chile from 2009 to 2010 was 170 cm with a standard deviation of 6.28 cm. Male heights are known to follow a normal distribution. Let X = the height of a 15 to 18-year-old male from Chile in 2009 to 2010. Then X~N(170, 6.28).

1. Suppose a 15 to 18-year-old male from Chile was 168cm tall from 2009 to 2010. The z-score when x = 168cm is z =_______. This z-score tells you that x = 168 is ________ standard deviations to the ________ (right or left) of the mean _____ (What is the mean?).
2. Suppose that the height of a 15 to 18-year-old male from Chile from 2009 to 2010 has a z-score of z = 1.27. What is the male's height? The z-score (z=1.27) tells you that the male's height is ________ standard deviations to the __________ (right or left) of the mean.

4

For more information and examples see online textbook OpenStax Introductory Statistics pages 366371.

Lab Assignment 6.1

Name:____Date:_____________________Row:________ **Lab Assignment 6.1**

1. What is the z-score of x= 12, if it is two standard deviations to the right of the mean?
2. What is the z-score of x= 7, if it is 0.133 standard deviations to the left of the mean?
3. Suppose X~N(9, 5). What value of x has a z-score of –0.5?
4. Suppose X~N(4, 2). What value of x is two standard deviations to the right of the mean?
5. Suppose X~N(–1, 2). What is the z-score of x= 2?
6. Suppose X~N(9, 3). What is the z-score of x= 9?
7. In a normal distribution, x = 3 and z = 0.67. This tells you that x = 3 is ____ standard deviations to the ____ (right or left) of the mean.
8. About what percent of the x values from a normal distribution lie within two standard deviations (left and right) of the mean of that distribution?
9. About what percent of x values lie between the mean and three standard deviations?
10. The patient recovery time from a particular surgical procedure is normally distributed with a mean of 5.3 days and a standard deviation of 2.1 days. What is the z-score for a patient who takes ten days to recover?
11. The heights of the 430 National Basketball Association players were listed on team rosters at the start of the 2005–2006 season. The heights of basketball players have an approximate normal distribution with mean, μ=79 inches and a standard deviation, σ=3.89 inches. For each of the following heights, calculate the z-score and interpret it using complete sentences.

1. 77 inches
2. 85 inches
3. If an NBA player reported his height had a z-score of 3.5, what would be his height? Would you believe him? Explain your answer.

Self-Check 6.1

Name:____________________________________Date:__________________Row:___________ **Self-Check 6.1**

1. What is the z-score of x= 9, if it is 1.5 standard deviations to the left of the mean?
2. Suppose X~N(2, 3). What value of x has a z-score of –0.67?
3. Suppose X~N(4, 2). What value of x is 1.5 standard deviations to the left of the mean?
4. Suppose X~N(–1, 2). What is the z-score of x= 2?
5. In a normal distribution, x = –2 and z = 6. This tells you that x = –2 is ____ standard deviations to the ____ (right or left) of the mean.
6. The tallest living man at the time of this writing is Sultan Kosan, who has a height of 251cm. The shortest living man is Chandra Bahadur Dangi, who has a height of 54.6 cm. Heights of men have a mean of 174.12cm and a standard deviation of 7.10 cm. Which of these two men has the height that is more extreme?
7. Blood platelet counts of women have a normal distribution with a mean of 255.1 and a standard deviation of 65.4. Use the empirical rule to approximate the following.

1. Approximately what percentage of women have platelet counts between 124.3 and 385.9?
2. Approximately what percentage of women have platelet counts between 189.7 and 385.9?
3. About 99.7% of women have platelet counts between what two amounts?

CHAPTER OVERVIEW

Section 7.1

7.1 The Central Limit Theorem for Sample Means
Lab Assignment 7.1
Self-Check 7.1

7.1 The Central Limit Theorem for Sample Means

Section 7.1 The Central Limit Theorem for Sample Means

Learning Objective:

In this section, you will:

- Understand and apply the normal probability distribution to real-world situations
- Given an area/percentage/probability, calculate the z-score (or cutoff score) that corresponds to that amount
- Given a range of values on a normal distribution, calculate the probability.
- Solve similar problems using the Central Limit Theorem, when appropriate

Central Limit Theorem

The central limit theorem for sample means says that if you keep drawing larger and larger samples and calculating their means, the sample means form their own normal distribution. The normal distribution has the same mean as the original distribution and a variance that equals the original variance divided by the sample size.

-
-
- μx = μ
-
-

Example 1: An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size n = 25 are drawn randomly from the population.

1. In words, X =
2. In words, X =
3. X ~
4. Find the probability that the sample mean is between 85 and 92.
5. Find the value that is two standard deviations above the expected value, 90, of the sample mean.

Example 2: The length of time, in hours, it takes an "over 40" group of people to play one soccer match is normally distributed with a mean of two hours and a standard deviation of 0.5 hours. A sample of size n = 50 is drawn randomly from the population. 1. In words, X =

2. In words, X =
3. X ~
4. Find the probability that the sample mean is between 1.8 hours and 2.3 hours.
5. Find the 95th percentile for the sample mean hours (to one decimal place).

Example 3: The lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days.

1. If one pregnant woman is randomly selected, find the probability that her length of pregnancy is less than 260 days. Is a length of 260 days significantly low for a pregnancy?
2. 25 women were selected to go on a special diet just before they became pregnant. Find the probability that 25 randomly selected women have a mean pregnancy length that is less than 260 days.

2

For more information and examples see online textbook OpenStax Introductory Statistics pages 400-405.

Lab Assignment 7.1

Name:____Date:_____________________Row:_________

Lab Assignment 7.1

1. An unknown distribution has a mean of 45 and a standard deviation of eight. Samples of size $n = 30$ are drawn randomly from the population. Find the probability that the sample mean is between 42 and 50.
2. According to the Internal Revenue Service, the average length of time for an individual to complete (keep records for, learn, prepare, copy, assemble, and send) IRS Form 1040 is 10.53 hours (without any attached schedules). The distribution is unknown. Let us assume that the standard deviation is two hours. Suppose we randomly sample 36 taxpayers. a. In words, X =

2. In words, $\bar{X}$ =
3. $\bar{X}$~
4. Would you be surprised if the 36 taxpayers finished their Form 1040s in an average of more than 12 hours? Explain why or why not in complete sentences.
5. Would you be surprised if one taxpayer finished his or her Form 1040 in more than 12 hours? In a complete sentence, explain why.

3. Suppose that a category of world-class runners are known to run a marathon (26 miles) in an average of 145 minutes with a standard deviation of 14 minutes. Consider 49 of the races. Let $\bar{X}$ the average of the 49 races.

1. $\bar{X}$~
2. Find the probability that the runner will average between 142 and 146 minutes in these 49 marathons.
3. Find the 80th percentile for the average of these 49 marathons.

4. Cans of a cola beverage claim to contain 16 ounces. The amounts in a sample are measured and the statistics are $n = 34$, $\bar{x}$ = 16.01 ounces. If the cans are filled so that $\mu = 16.00$ ounces (as labeled) and $\sigma = 0.143$ ounces.

1. Find the probability that a sample of 34 cans will have an average amount greater than 16.01 ounces.
1. Do the results suggest that cans are filled with an amount greater than 16 ounces?

Self-Check 7.1

Name:___________________________________Date:__________________Row:___________ **Self-Check 7.1**

1. The length of time taken on the SAT for a group of students is normally distributed with a mean of 2.5 hours and a standard deviation of 0.25 hours. A sample size of n=60 is drawn randomly from the population.

1. In words, X =
2. In words, $\bar{X}$ =
3. $\bar{X}$~
4. Find the probability that the sample mean is between two hours and three hours.
5. Find the 80th percentile for the sample mean hours (two decimal place).

2. A gondola carries skiers to the top of a mountain. It bears a plaque stating that the maximum capacity is 12 people or 2004 lb. That capacity will be exceeded if 12 people have weights with a mean of 2004/12 = 167 lb. Because men tend to weigh more than women, a worst case scenario would be if all 12 passengers were men. Men have weights that are normally distributed with a mean of 172 lb and a standard deviation of 29 lb.

1. Find the probability that if an individual man is randomly selected, his weight is greater than 167 lb.
2. Find the probability that 12 randomly selected men will have a mean that is greater than 167 lb.

3. In a recent study reported Oct. 29, 2012 on the Flurry Blog, the mean age of tablet users is 34 years. Suppose the standard deviation is 15 years. Take a sample of size n= 100.

1. What are the mean and standard deviation for the sample mean ages of tablet users?
2. Find the probability that the sample mean age is more than 30 years.
3. Find the 95th percentile for the sample mean age (to one decimal place).

CHAPTER OVERVIEW

Section 8.1

8.1 A Single Population Mean using the Normal Distribution
Lab Assignment 8.1
Self-Check 8.1

8.1 A Single Population Mean using the Normal Distribution

Section 8.1 A Single Population Mean using the Normal Distribution

Learning Objective:

In this section, you will:

- Apply and interpret point estimates and confidence intervals
- Determine adequate sample sizes needed to estimate population parameters
- Construct and interpret confidence intervals for population means and proportions

We use sample data to make generalizations about an unknown population. This part of statistics is called **inferential statistics**. The sample data help us to make an estimate of a population parameter.

Point Estimate – a single number computed from a sample and used to estimate a population parameter. $\boldsymbol{\mu = mean, \hat{p} = proportion}$

A **confidence interval** is another type of estimate but, instead of being just one number, it is an interval of numbers. The interval of numbers is a range of values calculated from a given set of sample data. The confidence interval is likely to include an unknown population parameter.

$1.8 < \mu < 2.2$ $0.37 < p < 0.43$

Margin of Error (E) – maximum difference between the sample mean and the true value of the population mean.

Confidence Level (CL) - the probability that the confidence interval contains the true population parameter; for example, if the CL=90%, then in 90 out of 100 samples the interval estimate will enclose the true population parameter.

α (alpha) - is related to the confidence level. α is the probability that the interval does not contain the unknown population parameter.

- $\boldsymbol{CL = 1 - \alpha \cdot \alpha = 1 - CL}$
- $\boldsymbol{CL + \alpha = 1}$

Confidence intervals for some parameters have the form:

(point estimate – margin of error, point estimate + margin of error)

1

Confidence interval for population mean ()

Point estimate for population mean: $\bar{x}$ =

Normal requirement: Either n > 30 **or** population is roughly normally distributed

Confidence Interval: $\boldsymbol{\bar{x} \pm E}$ **or** $\boldsymbol{\bar{x} - E} < \mu < \boldsymbol{\bar{x} + E}$ **or** $(\boldsymbol{\bar{x} - E}, \boldsymbol{\bar{x} + E})$

Example 1: Suppose we have collected data from a sample. We know the sample mean but we do not know the mean for the entire population. The sample mean is seven, and the error bound for the mean is 2.5.

If the confidence level (CL) is 95%, then we say that, "We estimate with 95% confidence that the true value of the population mean is between 4.5 and 9.5."

Example 2: Express the interval $230 < \mu < 452$ in the $\bar{x} \pm E$

Example 3: Express the interval 34 ± 108 in the form $\bar{x}$ – E < < $\bar{x}$ + E

Example 4: Given the confidence interval (1819, 3457), find $\bar{x}$ and E

Calculating the confidence interval for population mean () when is known **Using the Graphing calculator TI-84**: STAT, TESTS, 7:ZInterval (Data or Stats)

ZInterval (Stats, σ, $\bar{x}$, n, CL) or Enter data L1, ZInterval (Data, σ, L1, 1, CL)

2

Example 5: Suppose scores on exams in statistics are normally distributed with an unknown population mean and a population standard deviation of three points. A random sample of 36 scores is taken and gives a sample mean of 68. Find a 90% confidence interval for the true (population) mean of statistics exam scores. Write an interpretation.

Example 6: Randomly selected statistics students participated in an experiment to test their ability to determine when 1 minute (60 seconds) has passed. Forty students yielded a sample mean of 58.3 seconds. Assume that σ = 9.5 seconds. Construct a 95% confidence interval estimate of the population mean of all statistics students. Based on the results, is it likely that their estimates have a mean that is reasonably close to 60 seconds?

Example 7: When medical students all measured the blood pressure of the same person, they obtained the following results: 138, 130, 135, 140, 120, 125, 120, 130, 130, 144, 143, 140, 130, 150. Assuming that the population standard deviation is 10 mmHg, construct a 99% confidence interval estimate of the population mean. Write an interpretation.

3

Critical Values – a z-score on the borderline separating sample statistics that are likely to occur by chance from those that are unlikely. $Z_{\alpha/2}$ is the z-score that separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

Example 8: Find the critical value, $Z_{\alpha/2}$ corresponding to a 90% confidence level.

Calculating the Sample Size n for the Mean

If we know our desired margin of error, we must have a large enough sample to guarantee the desired error. $n = \left[\frac{z_{\alpha/2}\cdot\sigma}{E}\right]^2$. Always round up to next whole number when determining sample size.

Example 9: The population standard deviation for the age of Foothill College students is 15 years. If we want to be 95% confident that the sample mean age is within two years of the true population mean age of Foothill College students, how many randomly selected Foothill College students must be surveyed? Write an interpretation.

Example 10: How many daily rainfall amounts in Boston must be randomly selected to estimate the mean daily rainfall amount? We want 90% confidence that the sample mean is within 0.01 in. of the population mean, and the population standard deviation is known to be 0.212 in. Write an interpretation.

For more information and examples see online textbook OpenStax Introductory Statistics pages 443-456.

Lab Assignment 8.1

Name:____Date:_____________________Row:________

Lab Assignment 8.1

1. Among various ethnic groups, the standard deviation of heights is known to be approximately three inches. We wish to construct a 95% confidence interval for the mean height of male Swedes. Forty-eight male Swedes are surveyed. The sample mean is 71 inches. The sample standard deviation is 2.8 inches.

1. $\bar{x}$ =________ σ =________ n =________
2. Construct a 95% confidence interval for the population mean height of male Swedes.
3. State the confidence interval in words.
4. Find the point estimate for mean height of male Swedes
5. Calculate the error bound (E).
6. Express the confidence interval in $\bar{x}$ ± E form.
7. Sketch the graph.

1

2. Suppose that an accounting firm does a study to determine the time needed to complete one person's tax forms. It randomly surveys 100 people. The sample mean is 23.6 hours. There is a known standard deviation of 7.0 hours. The population distribution is assumed to be normal.

1. $\bar{x}$ =________ σ =________ n =________
2. Construct a 90% confidence interval for the population mean time to complete the tax forms.
3. State the confidence interval in words.
4. Sketch the graph.

3. The American Community Survey (ACS), part of the United States Census Bureau, conducts a yearly census similar to the one taken every ten years, but with a smaller percentage of participants. The most recent survey estimates with 90% confidence that the mean household income in the U.S. falls between $69,720 and $69,922.

1. Find the point estimate for mean U.S. household income.
2. Find the error bound for mean U.S. household income.

4. Find the critical value, $Z_{\alpha/2}$ corresponding to a 88% confidence level.

2

5. The cost of homes in the area are listed below. The standard deviation for this data to the nearest hundred is σ= $100,000.

$589,000; $610,000; $765,000; $750,000; $657,000; $475,000; $599,000; $799,950; $499,000;

$629,950

1. Create a 95% confidence interval for the mean cost of homes in the area.
2. Interpret the confidence interval in the context of the problem.

6. The average height of young adult males has a normal distribution with standard deviation of 2.5 inches. You want to estimate the mean height of students at your college or university to within one inch with 90% confidence. How many male students must you measure? Write an interpretation.
7. The population standard deviation for the height of high school basketball players is three inches. If we want to be 95% confident that the sample mean height is within one inch of the true population mean height, how many randomly selected students must be surveyed? Write an interpretation.

Self-Check 8.1

Name:____________________________________Date:__________________Row:___________

Self-Check 8.1

1. Suppose we have data from a sample. The sample mean is 15, and the error bound for the mean is 3.2. What is the confidence interval estimate for the population mean?
2. Suppose average pizza delivery times are normally distributed with an unknown population mean and a population standard deviation of six minutes. A random sample of 28 pizza delivery restaurants is taken and has a sample mean delivery time of 36 minutes.

Find a 90% confidence interval estimate for the population mean delivery time. Write an interpretation.

3. How many adults must be randomly selected to estimate the mean FICO (credit rating) score of working adults in the United States? We want 95% confidence that the sample mean is within 3 points of the population mean, and the population standard deviation is 68.

1

Self-Check 8.1

4. The table below shows a different random sampling of 20 cell phone models. Use this data to calculate a 93% confidence interval for the true mean SAR for cell phones certified for use in the United States. As previously, assume that the population standard deviation is $\sigma = 0.337$. And write an interpretation.

Phone Model	SAR
Blackberry Pearl	1.48
HTC Evo Design 4G	0.8
HTC Freestyle	1.15
LG Ally	1.36
LG Fathom	0.77
LG Optimus Vu	0.462
Motorola Cliq XT	1.36
Motorola Droid Pro	1.39
Motorola Droid Razr M	1.3
Nokia 7705 Twist	0.7
Nokia E71x	1.53
Nokia N75	0.68
Nokia N79	1.4
Sagmen Puma	1.24
Samsung Fascinate	0.57
Samsung Infuse 4G	0.2
Samsung Nextus S	0.51
Samsung Replenish	0.3
Sony W518a Walkman	0.73
ZTE C79	0.869

5. Find the critical value, $Z_{\alpha/2}$ corresponding to a 92% confidence level.

CHAPTER OVERVIEW

Section 8.2

8.2 A Single Population Mean using the Student t Distribution

8.2 A Single Population Mean using the Student t Distribution

Section 8.2 A Single Population Mean using the Student t Distribution

Learning Objective:

In this section, you will:

- Apply and interpret point estimates and confidence intervals
- Construct and interpret confidence intervals for population means

Estimating a population mean (□), when □ is NOT known

Critical Values - If □ is not known, instead of a normal distribution, we use a student t distribution to find the t-score. t□/2 is the t-score that separates an area of □/2 in the right tail of the student t distribution.

In addition to the t-score, we need to consider the **degrees of freedom, n – 1**. The number of degrees of freedom is abbreviated by df. **df = n – 1**.

Example 1: Find the critical value, t□/2, for a sample size of 10, and corresponding to a 99% confidence level.

Calculating the confidence interval for population mean (□) when □ is NOT known

Using the Graphing calculator TI-84: STAT, TESTS, 8:TInterval (Data or Stats)

TInterval (Stats, $\bar{x}$, s, n, CL) or Enter data L1, TInterval (Data, L1, 1, CL)

Example 2: Suppose you do a study of acupuncture to determine how effective it is in relieving pain.

You measure sensory rates for 15 subjects with the results given. Use the sample data to construct a 95% confidence interval for the mean sensory rate for the population (assumed normal) from which you took the data. 8.6; 9.4; 7.9; 6.8; 8.3; 7.3; 9.2; 9.6; 8.7; 11.4; 10.3; 5.4; 8.1; 5.5; 6.9

1

Notes 8.2

Example 3: A sample of 106 body temperatures has a mean of 98.2 degrees Fahrenheit a standard deviation of 0.62 degrees Fahrenheit. Construct a 95% confidence interval estimate of the mean body temperature of all healthy humans. Does the common use of 98.6 degrees Fahrenheit seem to a reasonable estimate of the mean body temperature?

For more information and examples see online textbook OpenStax Introductory Statistics pages 456-460.

CHAPTER OVERVIEW

Section 8.3

8.3 A Population Proportion
Lab Assignment 8.2, 8.3
Self-Check 8.2, 8.3

8.3 A Population Proportion

Section 8.3 A Population Proportion

Learning Objective:

In this section, you will:

- Apply and interpret point estimates and confidence intervals
- Determine adequate sample sizes needed to estimate population parameters
- Construct and interpret confidence intervals for population proportions

Proportion = Probability = Percent – Example: If 28% of scores are higher than yours, then the probability of a score being higher than yours is 0.28, and the proportion of scores higher than yours is 0.28

Point Estimate – the best estimate for a population proportion, *p*, is the sample proportion, $\hat{p}$.

$\hat{p} = x$ $\hat{q} = 1 - \hat{p}$

n

Margin of Error (E) – maximum difference between the sample proportion and the true value of the population proportion.

Confidence Interval – a range of values used to estimate the true value of a population parameter.

$\hat{p} \pm E$ **or** $\hat{p} - E < \mu < \hat{p} + E$ **or** $(\hat{p} - E, \hat{p} + E)$

Confidence level- the probability 1 – (usually expressed as a percentage) that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

Example 1: The 90% confidence interval for the proportion of all students with a GPA over 3.5 is

.0997 < p < .2203.

Interpretation: "We estimate with 90% confident that the true value of the proportion of all students with a GPA over 3.5 is between 0.0997 and .2203." If we construct similar confidence intervals using sample proportions numerous times, we expect that 90% of those intervals would contain the true population proportion.

Calculating Confidence Intervals for population proportions:

Using the Graphing calculator TI-84: STAT, TESTS, A:1-PROPZINT

1-PropZInt(x, n, CL)

Example 2: When Mendel conducted his famous genetics experiments with peas, one sample of offspring consisted of 428 green peas and 152 yellow peas. Find a 95% confidence interval estimate of the percentage of green peas.

Example 2 (continued): Mendel expected that 75% of the peas would be green. Given that the percentage of green peas in our sample is not 75%, do the results contradict Mendel's theory? Why or why not?

Example 3: If 230 out of 600 teenagers plan to see the new Hunger Games movie, find a 90% confidence interval estimate for the percentage of all teenagers planning to see the movie.

Example 3 (continued): The movie theater claims that 25% of teenagers are planning to see the movie. Does their claim appear to be correct?

2

Calculating the Minimum Sample Size

If we know our desired margin of error, we must have a large enough sample to guarantee the desired

$2\ [z\alpha/2]\ \hat{p}\hat{q}$ error. $n = E2$, when $\hat{p}$ and $\hat{q}$ are known

2

$[z\alpha/2]\ 0.25$

If we don't know $\hat{p}$ in advance, we use 0.5 for both $\hat{p}$ and $\hat{q}$. This gives us $n = E2$

Always round up to next whole number when determining sample size.

Example 4: Find the minimum sample size needed if the margin of error must be two percentage points, the confidence level is 99%, and the point estimate for the population proportion is 14%. Write an interpretation.

Example 5: A survey was conducted to determine the percentage of car owners who would pay to put nitrogen in their tires (Nitrogen supposedly leaks out at a slower rate than air, which keeps the tire pressure at the ideal level.) How many randomly selected car owners should be surveyed? Assume that we want to be 95% confident that the sample percentage is within 3% of the true percentage of all car owners who would be willing to pay for nitrogen. Write an interpretation.

3

For more information and examples see online textbook OpenStax Introductory Statistics pages 460-467.

Lab Assignment 8.2, 8.3

Self-Check 8.2, 8.3

Name:____Date:____________________Row:________

Lab Assignment 8.2, 8.3

1. Suppose that a committee is studying whether or not there is waste of time in our judicial system. It is interested in the mean amount of time individuals waste at the courthouse waiting to be called for jury duty. The committee randomly surveyed 81 people who recently served as jurors. The sample mean wait time was eight hours with a sample standard deviation of four hours. Construct a 95% confidence interval for the population mean time wasted. Explain in a complete sentence what the confidence interval means.
2. A pharmaceutical company makes tranquilizers. It is assumed that the distribution for the length of time they last is approximately normal. Researchers in a hospital used the drug on a random sample of nine patients. The effective period of the tranquilizer for each patient (in hours) was as follows: 2.7; 2.8; 3.0; 2.3; 2.3; 2.2; 2.8; 2.1; and 2.4. Construct a 95% confidence interval for the population mean length of time. Explain in a complete sentence what the confidence interval means.

1

3. Insurance companies are interested in knowing the population percent of drivers who always buckle up before riding in a car. They randomly surveyed 400 drivers and found that 320 claimed they always buckle up. Construct a 95% confidence interval for the population proportion who claim they always buckle up. Explain in a complete sentence what the confidence interval means.
4. On May 23, 2013, Gallup reported that 76% of U.S. workers believe that they will continue working past retirement age. The confidence level for this study was reported at 95% with a ±3% margin of error. Determine the sample size. Write an interpretation.

2

5. You plan to conduct a survey on your college campus to learn about the political awareness of students. You want to estimate the true proportion of college students on your campus who voted in the 2012 presidential election with 95% confidence and a margin of error no greater than five percent. How many students must you interview? Write an interpretation of the sample size.
6. Find the critical value, t/2, for a sample size of 32, and corresponding to a 80% confidence level.

Self-Check 8.2, 8.3

Name:___________________________________Date:__________________Row:___________

Self-Check 8.2, 8.3

1. You do a study of hypnotherapy to determine how effective it is in increasing the number of hours of sleep subjects get each night. You measure hours of sleep for 12 subjects with the following results. Construct a 95% confidence interval for the mean number of hours slept for the population (assumed normal) from which you took the data. Write an interpretation.

8.2; 9.1; 7.7; 8.6; 6.9; 11.2; 10.1; 9.9; 8.9; 9.2; 7.5; 10.5

2. A random sample of 15 statistics students were asked to estimate the total number of hours they spend watching television in an average week. The sample mean is 6.1 hours and sample standard deviation is 5.5 hours. Construct a 98% confidence interval for the mean number of hours statistics students will spend watching television in one week. Write an interpretation.
3. Suppose 250 randomly selected people are surveyed to determine if they own a tablet. Of the 250 surveyed, 98 reported owning a tablet. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of people who own tablets.

1

Self-Check 8.2, 8.3

4. Find the critical value, t□/2, for a sample size of 21, and corresponding to a 95% confidence level.
5. Suppose an internet marketing company wants to determine the current percentage of customers who click on ads on their smartphones. How many customers should the company survey in order to be 90% confident that the estimated proportion is within five percentage points of the true population proportion of customers who click on ads on their smartphones?
6. Suppose a mobile phone company wants to determine the current percentage of customers aged 50+ who use text messaging on their cell phones. How many customers aged 50+ should the company survey in order to be 90% confident that the estimated (sample) proportion is within three percentage points of the true population proportion of customers aged 50+ who use text messaging on their cell phones, and the point estimate for the population proportion is 75%.

CHAPTER OVERVIEW

Section 9.1

9.1 Null and Alternative Hypothesis

9.1 Null and Alternative Hypothesis

Section 9.1 Null and Alternative Hypothesis

Learning Objective:

In this section, you will:

• Understand the general concept and use the terminology of hypothesis testing

I claim that my coin is a fair coin. This means that the probability of heads and the probability of tails are both 50% or 0.50.

1. Out of 200 flips of the coin, tails is tossed 102 times. What can we conclude about my claim?
2. Out of 200 flips of the coin, tails is tossed 21 times. What can we conclude about my claim?

Hypothesis is a claim about the value of a population parameter.

Hypothesis Testing is a procedure for determining whether the hypothesis stated is a reasonable statement and should not be rejected, or is unreasonable and should be rejected.

Hypothesis testing begins by considering two hypotheses. They are called the **null hypothesis** and the **alternative hypothesis**. These hypotheses contain opposing viewpoints.

1. The **null hypothesis**, typically denoted with **H0**. The null is not rejected unless the hypothesis test shows otherwise. The null statement must always contain some form of equality **(=, ≤ or ≥)**
2. The **alternative hypothesis**, typically denoted with **Ha or H1**, using less than, greater than, or not equals symbols, **(≠, >, or <).**
3. If we reject the null hypothesis, then we can assume there is enough evidence to support the alternative hypothesis.
4. Never state that a claim is proven true or false. Keep in mind the underlying fact that hypothesis testing is based on probability laws; therefore, we can talk only in terms of non-absolute certainties.

Example 1: We want to test whether the mean GPA of students in American colleges is different from 2.0 (out of 4.0). The null and alternative hypotheses are:

Example 2: We want to test if college students take less than five years to graduate from college, on the average. The null and alternative hypotheses are:

Example 3: In an issue of U.S. News and World Report, an article on school standards stated that about half of all students in France, Germany, and Israel take advanced placement exams and a third pass. The same article stated that 6.6% of U.S. students take advanced placement exams and 4.4% pass. Test if the percentage of U.S. students who take advanced placement exams is more than 6.6%. State the null and alternative hypotheses.

For more information and examples see online textbook OpenStax Introductory Statistics pages 505-508.

CHAPTER OVERVIEW

Section 9.4

9.4 Rare Events, the Sample, Decision and Conclusion
Lab Assignment 9.1, 9.4
Self-Check 9.1, 9.4

9.4 Rare Events, the Sample, Decision and Conclusion

Section 9.4 Rare Events, the Sample, Decision and Conclusion

Learning Objective:

In this section, you will:

- Understand the general concept and use the terminology of hypothesis testing
- Interpret real-world problems and apply the five-step process of hypothesis testing to make statistical conclusions

Rare Event Rule: If under a given assumption, the probability of an event is extremely small, we conclude that the assumption was not correct.

P-value is the probability that an event will happen purely by chance assuming the null hypothesis is true. The smaller the p-value, the stronger the evidence is against the null hypothesis.

A systematic way to make a decision of whether to reject or not reject the null hypothesis is to compare the p-value and a preset or preconceived **α, significance level**.

Decision

- If the p-value is low, the null must go.
- If the p-value is high, the null must fly. Or
- If P-value is less than α, Reject Ho
- If P-value is greater than α, Fail to reject Ho

Example 1: Find the **p-value** for each test, and state the conclusion about the null hypothesis Note 1. H1: $p > ¼$, test statistic is $z = 2.79$, $\alpha = 0.01$

1

2. Two-tailed test, test statistic is $z = 1.75$, $\alpha = 0.05$

Conclusion: After you make your decision, write a thoughtful conclusion about the hypotheses in terms of the given problem.

Example 2: If the original claim is that the percentage of Americans who know their credit score is less than 20%, and our hypothesis test tells us to reject the null hypothesis, the final conclusion is:

Example 3: Original claim: The percentage of nonsmokers exposed to secondhand smoke is 41%. The decision is to fail to reject the null. State the final conclusion.

2

Example 4: Original claim: The mean GPA of LMC students is higher than 3.1. The decision is to fail to reject the null hypothesis. State the final conclusion.

For more information and examples see online textbook OpenStax Introductory Statistics pages 511-514.

Lab Assignment 9.1, 9.4

Name:____Date:____________________Row:________

Lab Assignment 9.1, 9.4

1. You are testing that the mean speed of your cable Internet connection is more than three Megabits per second. State the null and alternative hypothesis.
2. A sociologist claims the probability that a person picked at random in Times Square in New York City is visiting the area is 0.83. You want to test to see if the proportion is actually less. State the null and alternative hypotheses.
3. In a population of fish, approximately 42% are female. A test is conducted to see if, in fact, the proportion is less. State the null and alternative hypotheses.
4. Find the P-value for each test, assume the significant level α = 0.05. State the conclusion about the null hypothesis.

1. The test statistic of $z = -2.00$ is abtained when testsing the claim that $p < 0.5$.
2. The test statistic of $z = 1.50$ is abtained when testsing the claim that $p \neq 0.25$.

5. It is believed that the mean height of high school students who play basketball on the school team is 73 inches.

1. State the null and alternative hypothesis.
2. Use a 0.05 significance level and the *p*-value of 0.005 to state the conclusion about the null hypothesis.
3. State the conclusion that addresses the original claim.

6. Conduct a hypothesis test to determine if the population mean time on death row is 15 years.

1. State the null and alternative hypothesis.
2. Use a 0.01 significance level and the *p*-value of 0.02 to state the conclusion about the null hypothesis.
3. State the conclusion that addresses the original claim.

Self-Check 9.1, 9.4

Name:____________________________________Date:___________________Row:____________

Self-Check 9.1, 9.4

1. We want to test whether the mean height of eighth graders is 66 inches. State the null and alternative hypotheses.
2. On a state driver's test, about 40% pass the test on the first try. We want to test if more than 40% pass on the first try.
3. We want to test the claim that the mean is greater than 12.
4. Find the **p-value** for each test, state the conclusion about the null hypothesis a. $\alpha = 0.05$, test statistic = 1.15

H0: p = 0.4

H1: $p \neq 0.4$

b. $\alpha = 0.01$, test statistic = 2.5

H0: $\mu = 12$ H1: $\mu > 12$

5. It's a Boy Genetics Labs claim their procedures improve the chances of a boy being born. The results for a test of a single population proportion are as follows: H0: p= 0.50, H1: p> 0.50, α= 0.01, p-value = 0.025

1. State the conclusion about the null hypothesis
2. State the conclusion that addresses the original claim.

6. Suppose a baker claims that his break height is more than 15cm.

1. State the null and alternative hypothesis.
2. Use a 0.01 significance level and the *p*-value of 0.0013 to state the conclusion about the null hypothesis.
3. State the conclusion that addresses the original claim.

LibreTexts

CHAPTER OVERVIEW

Section 9.5

9.5 Full Hypothesis Testing with One Sample
Lab Assignment 9.5
Self-Check 9.5

9.5 Full Hypothesis Testing with One Sample

Section 9.5 Full Hypothesis Testing with One Sample

Learning Objective:

In this section, you will:

- Interpret real-world problems and apply the five-step process of hypothesis testing to make statistical conclusions
- Apply the process to problems concerning a single proportion, and a mean (with σ known and unknown)

Testing a claim about a proportion: Use TI-84 calculator 5:1-PropZTest

Example 1: In a poll of 745 randomly selected adults, 591 said that it is morally wrong to not report all income on tax returns. Use a 0.01 significance level to test the claim that 75% of adults say that it is morally wrong to not report all income on tax returns.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

1

Example 2: A recently televised broadcast of *60 minutes* had a 15 share, meaning that among

5000 monitored households with TV sets in use, 15% of them were tuned to *60 minutes*. Use a 0.05 significance level to test the claim of an advertiser that among the household with TV sets in use, less than 20% were tuned to *60 minutes.*

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Testing a claim about a mean if σ is unknown: Use TI-84 calculator 2:TTest

Example 3: A simple random sample of 25 filtered 100mm cigarettes is obtained and the tar content of each cigarette is measured. The sample has a mean tar content of 13.2 mg and a standard deviation of 3.7 mg. Use a 0.05 significance level to test the claim that the mean tar content of filtered 100 mm cigarettes is less than 21.1 mg. Unfiltered king size cigarettes have a mean tar content of 21.1 mg.

1. Null and Alternative Hypothesis
2. Calculator Work

2

3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Example 4: The National Highways Traffic Safety Administration conducted crash tests of child booster seats for cars. Listed below are results from those tests, with the measurements given in hic (standard head injury condition units). The safety requirement is that the hic measurement should be less than 1000 hic. Use a 0.01 significance level to test the claim that the sample is from a population with a mean less than 1000 hic.

774 649 1210 546 431 612

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

3

Testing a claim about a mean if is known: Use TI-84 calculator 1:ZTest

Example 5: A simple random sample of 40 salaries of NCAA football coaches has a mean of $415,953. The standard deviation of all NCAA football coaches is $463,364. Use a 0.05 significance level to test the claim that the mean salary of a football coach in the NCAA is less than $500,000.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Example 6: A simple random sample of FICO scores is listed below. The mean FICO score is reported to be 678. Assuming that the standard deviation of all FICO scores is known to be 58.3, use a 0.01 significance level to test the claim that these sample FICO scores come from a population with a mean equal to 678.

714 751 664 789 818 779 698 836 753 834 693 802 1) Null and Alternative Hypothesis

4

2) Calculator Work

3)Test Statistic and P-Value

4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

For more information and examples see online textbook OpenStax Introductory Statistics pages 514-530.

Lab Assignment 9.5

Name:____Date:____________________Row:________

Lab Assignment 9.5

1. From generation to generation, the mean age when smokers first start to smoke varies.

However, the standard deviation of that age remains constant of around 2.1 years. A survey of 40 smokers of this generation was done and the sample mean was 18.1. Do the data support the claim at the 5% significance level, that the mean starting age is at least 19.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

2. The mean number of sick days an employee takes per year is believed to be about ten. Members of a personnel department do not believe this figure. They randomly survey eight employees. The number of sick days they took for the past year are as follows: 12; 4; 15; 3; 11; 8; 6; 8. Use a 0.05 significance level to test the claim that the mean number of sick days an employee takes per year is ten.

1. Null and Alternative Hypothesis

1

2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

3. Your statistics instructor claims that more than 60 percent of the students who take her Elementary Statistics class go through life feeling more enriched. For some reason that she can't quite figure out, most people don't believe her. You decide to check this out on your own. You randomly survey 64 of her past Elementary Statistics students and find that 34 feel more enriched as a result of her class. Use a 0.05 significance level to test the claim that more than 60 percent of the students who take her Elementary Statistics class go through life feeling more enriched.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis

2

5. Final conclusion that addresses the original claim

4. Toastmasters International cites a report by Gallop Poll that 40% of Americans fear public speaking. A student believes that less than 40% of students at her school fear public speaking. She randomly surveys 361 schoolmates and finds that 135 report they fear public speaking. Use a 0.05 significance level to conduct a hypothesis test to determine if the percent at her school is less than 40%.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis

5. Final conclusion that addresses the original claim

3

5. Registered nurses earned an average annual salary of $69,110. For that same year, a survey was conducted of 41 California registered nurses to determine if the annual salary is higher than $69,110 for California nurses. The sample average was $71,121 with a sample standard deviation of $7,489. Conduct a hypothesis test using a 0.05 significance level.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Self-Check 9.5

Name:___________________________________Date:__________________Row:___________

Self-Check 9.5

1. In a survey of 703 randomly selected workers, 61% got their jobs through networking. Use a

0.01 significance level to test the claim that most workers get their jobs through networking. 1) Null and Alternative Hypothesis

2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

2. Researchers collected a simple random sample of the times that 81 college students required to earn their bachelor's degrees. The sample has a mean of 4.8 years and a standard deviation of 2.2 years. Use a 0.05 significance level to test the claim that the mean time for all college students is greater than 4.5 years.

1. Null and Alternative Hypothesis
2. Calculator Work

1

Self-Check 9.5

3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

3. In the manual "How to Have a Number One the Easy Way," it is stated that a song "must be no longer than three minutes and thirty seconds (210 seconds)". A simple random sample of 40 current hit songs results in a mean length of 252.5 seconds. Assume that the standard deviation of song lengths is 54.5 sec. Use a 0.05 significance level to test the claim that the sample is from a population of songs with a mean greater than 210 seconds. 1) Null and Alternative Hypothesis

2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

CHAPTER OVERVIEW

Section 10.1

10.1 Two Population Means with Unknown Standard Deviations

10.1 Two Population Means with Unknown Standard Deviations

Section 10.1 Two Population Means with Unknown Standard Deviations

Learning Objective:

In this section, you will:

• Apply hypothesis testing and calculate confidence intervals to real-world problems about two population means

1. The two independent samples are simple random samples from two distinct populations.
2. For the two distinct populations:
   - if the sample sizes are small, the distributions are important (should be normal)
   - if the sample sizes are large, the distributions are not important (need not be normal)

Two population means from independent samples where the population standard deviations are not known

- Random Variable: $X1 - X2$ = the difference of the sampling means
- Distribution: Student's *t*-distribution with degrees of freedom (variances not pooled)
-
-

Example 1a: The average amount of time boys and girls aged seven to 11 spend playing sports each day is believed to be the same. A study is done and data are collected, resulting in the data in table below. Each population has a normal distribution.

	Sample Size	Average Number of Hours Playing Sports Per Day	Sample Standard Deviation
Boy	9	2	0.866
Girl	16	3.2	1.00

Is there a difference in the mean amount of time boys and girls aged seven to 11 play sports each day? Test at the 5% level of significance.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value

1

4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Example 1b: Using the date from 1a, construct the corresponding confidence interval estimate for the difference between the mean amount of time boys and girls aged seven to 11 play sports each day. What does the result suggest about the two means?

Example 2: A professor at a large community college wanted to determine whether there is a difference in the means of final exam scores between students who took his statistics course online and the students who took his face-to-face statistics class. He believed that the mean of the final exam scores for the online class would be lower than that of the face-to-face class. Was the professor correct? The randomly selected 30 final exam scores from each group are listed below.

Online Class exam scores: 67.6; 41.2; 85.3; 55.9; 82.4; 91.2; 73.5; 94.1; 64.7; 64.7; 70.6; 38.2;

61.8; 88.2; 70.6; 58.8; 91.2; 73.5; 82.4; 35.5; 94.1; 88.2; 64.7; 55.9; 88.2; 97.1; 85.3; 61.8; 79.4; 79.4

Face-to-Face Class exam scores: 77.9; 95.3; 81.2; 74.1; 98.8; 88.2; 85.9; 92.9; 87.1; 88.2; 69.4;

57.6; 69.4; 67.1; 97.6; 85.9; 88.2; 91.8; 78.8; 71.8; 98.8; 61.2; 92.9; 90.6; 97.6; 100; 95.3; 83.5;

92.9; 89.4

1. Null and Alternative Hypothesis

2

2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

For more information and examples see online textbook OpenStax Introductory Statistics pages 567-576.

CHAPTER OVERVIEW

Section 10.3

10.3 Comparing Two Independent Population Proportions
Lab Assignment 10.1, 10.3
Self-Check 10.1, 10.3

10.3 Comparing Two Independent Population Proportions

Learning Objective:

In this section, you will:

• Apply hypothesis testing and calculate confidence intervals to real-world problems about two population proportions

When conducting a hypothesis test that compares two independent population proportions, the following characteristics should be present:

1. The two independent samples are simple random samples that are independent.
2. The number of successes is at least five, and the number of failures is at least five, for each of the samples.
3. Growing literature states that the population must be at least ten or 20 times the size of the sample. This keeps each population from being over-sampled and causing incorrect results.

Hypothesis testing of two population proportions from independent samples.

- Random Variable: $\hat{p}1 - \hat{p}2$ = the difference between the two estimated proportions
- Distribution: Normal distribution
-
-

Example 1a: Two types of medication for hives are being tested to determine if there is a difference in the proportions of adult patient reactions. Twenty out of a random sample of 200 adults given medication A still had hives 30 minutes after taking the medication. Twelve out of another random sample of 200 adults given medication B still had hives 30 minutes after taking the medication. Test at a 1% level of significance.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis

1

5. Final conclusion that addresses the original claim

Example 1b: Using the date from 1a, construct the corresponding confidence interval estimate for the difference between the proportions of adult patient reactions to medication A and medication B. What does the result suggest about the two proportion?

Example 2: Researchers conducted a study of smartphone use among adults. A cell phone company claimed that iPhone smartphones are more popular with whites (non-Hispanic) than with African Americans. The results of the survey indicate that of the 232 African American cell phone owners randomly sampled, 5% have an iPhone. Of the 1,343 white cell phone owners randomly sampled, 10% own an iPhone. Test at the 5% level of significance. Is the proportion of white iPhone owners greater than the proportion of African American iPhone owners?

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis

2

178

5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

For more information and examples see online textbook OpenStax Introductory Statistics pages 579-584.

10.3 Comparing Two Independent Population Proportions is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Lab Assignment 10.1, 10.3

Name:____Date:____________________Row:________

Lab Assignment 10.1, 10.3

1. The U.S. Center for Disease Control reports that the mean life expectancy was 47.6 years for whites born in 1900 and 33.0 years for nonwhites. Suppose that you randomly survey death records for people born in 1900 in a certain county. Of the 124 whites, the mean life span was 45.3 years with a standard deviation of 12.7 years. Of the 82 nonwhites, the mean life span was 34.1 years with a standard deviation of 15.6 years. Conduct a hypothesis test to see if the mean life spans in the county were the same for whites and nonwhites. Test their hypothesis at a 1% significance level.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

2. At Rachel's 11th birthday party, eight girls were timed to see how long (in seconds) they could hold their breath in a relaxed position. After a two-minute rest, they timed themselves while jumping. The girls thought that the mean difference between their jumping and relaxed times would be zero. Test their hypothesis at a 1% significance level.

Relaxed time (seconds): 26; 47; 30; 22; 23; 45; 37; 29

Jumping time (seconds):21; 40; 28; 21; 25; 43; 35; 32

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

3. We are interested in whether the proportions of female suicide victims for ages 15 to 24 are the same for the whites and the blacks races in the United States. We randomly pick one year, 1992, to compare the races. The number of suicides estimated in the United States in 1992 for white females is 4,930. Five hundred eighty were aged 15 to 24. The estimate for black females is 330. Forty were aged 15 to 24. Test their hypothesis at a 5% significance level.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

4. While her husband spent 2.5 hours picking out new speakers, a statistician decided to determine whether the percent of men who enjoy shopping for electronic equipment is higher than the percent of women who enjoy shopping for electronic equipment. The population was Saturday afternoon shoppers. Out of 67 men, 24 said they enjoyed the activity. Eight of the 24 women surveyed claimed to enjoy the activity. Test their hypothesis at a 5% significance level.

1. Null and Alternative Hypothesis
2. Calculator Work

3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

Self-Check 10.1, 10.3

Name:____________________________________Date:__________________Row:___________

Self-Check 10.1, 10.3

1. Weighted alpha is a measure of risk-adjusted performance of stocks over a period of a year. A high positive weighted alpha signifies a stock whose price has risen while a small positive weighted alpha indicates an unchanged stock price during the time period. Weighted alpha is used to identify companies with strong upward or downward trends. The weighted alpha for the top 30 stocks of banks in the northeast and in the west as identified by Nasdaq on May 24, 2013 are listed below.

Northeast: 94.2; 75.2; 69.6; 52.0; 48.0; 41.9; 36.4; 33.4; 31.5; 27.6; 77.3; 71.9; 67.5; 50.6; 46.2;

38.4; 35.2; 33.0; 28.7; 26.5; 76.3; 71.7; 56.3; 48.7; 43.2; 37.6; 33.7; 31.8; 28.5; 26.0

West: 126.0; 70.6; 65.2; 51.4; 45.5; 37.0; 33.0; 29.6; 23.7; 22.6; 116.1; 70.6; 58.2; 51.2; 43.2;

36.0; 31.4; 28.7; 23.5; 21.6; 78.2; 68.2; 55.6; 50.3; 39.0; 34.1; 31.0; 25.3; 23.4; 21.5

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

1

Self-Check 10.1, 10.3

2. Two types of valves are being tested to determine if there is a difference in pressure tolerances. Fifteen out of a random sample of 100 of Valve *A* cracked under 4,500 psi. Six out of a random sample of 100 of Valve *B* cracked under 4,500 psi. Test at a 5% level of significance.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

CHAPTER OVERVIEW

Section 10.4

10.4 Matched or Paired Samples

10.4 Matched or Paired Samples

Section 10.4 Matched or Paired Samples

Learning Objective:

In this section, you will:

• Apply hypothesis testing and calculate confidence intervals to real-world problems about the mean of two dependent samples (matched pairs)

When using a hypothesis test for matched or paired samples, the following characteristics should be present:

1. Simple random sampling is used.
2. Sample sizes are often small.
3. Two measurements (samples) are drawn from the same pair of individuals or objects.
4. Differences are calculated from the matched or paired samples.
5. The differences form the sample that is used for the hypothesis test.
6. Either the matched pairs have differences that come from a population that is normal or the number of differences is sufficiently large so that distribution of the sample mean of differences is approximately normal.

A hypothesis test for matched or paired samples

- Random Variable: $\bar{x}d$ = mean of the differences
- Distribution: Student's-t distribution with *n* – 1 degrees of freedom
-
-

Example 1: A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Results for randomly selected subjects are shown in the table below. A lower score indicates less pain. The "before" value is matched to an "after" value and the differences are calculated. The differences have a normal distribution. Are the sensory measurements, on average, lower after hypnotism? Test at a 5% significance level.

Subject	A	B	C	D	E	F	G	H
Before	6.6	6.5	9.0	10.3	11.3	8.1	6.3	11.6
After	6.8	2.4	7.4	8.5	8.1	6.1	3.4	2.0

First find the differences:

1. Null and Alternative Hypothesis

1

2. Calculator Work

3. Test Statistic and P-Value

4. Conclusion about the null hypothesis

5. Final conclusion that addresses the original claim

6. Test the above claim by constructing an appropriate confidence interval.

Example 2: Seven eighth graders at Kennedy Middle School measured how far they could push the shot-put with their dominant (writing) hand and their weaker (non-writing) hand. They thought that they could push equal distances with either hand. The data were collected and recorded in the table below.

Distance (in feet) using	Student 1	Student 2	Student 3	Student 4	Student 5	Student 6	Student 7
Dominant Hand	30	26	34	17	19	26	20

Weaker Hand	28	14	27	18	17	26	16

Conduct a hypothesis test to determine whether the mean difference in distances between the children's dominant versus weaker hands is significant.

1. Null and Alternative Hypothesis

2

2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

For more information and examples see online textbook OpenStax Introductory Statistics pages 584-590.

CHAPTER OVERVIEW

Section 11.3

11.3 Test of Independence
Lab Assignment 12.3, 12.5, 13.1, 13.2, 13.3, 11.3
Self-Check 12.3, 12.5, 13.1, 13.2, 13.3, 11.3

11.3 Test of Independence

Section 11.3 Test of Independence

Learning Objective:

In this section, you will:

• Conduct and interpret chi-square test of independence hypothesis tests

A **test of independence** is used to determine whether two factors are independent or not. We use the graphing calculator, $\boldsymbol{C}:\boldsymbol{\chi 2-Test}$, with test statistic $\chi^2 = \Sigma\frac{(O-E)^2}{E}$, with df = (# of rows – 1)(# of columns – 1), O = Observed values, E = expected values

Enter data in a matrix [A] using a graphing calculator, **Matrix (2nd x-1)**, Edit, input number of rows and columns. Note that the $\boldsymbol{\chi 2-Test}$ will create a matrix of the expected values and place it in matrix [B].

Example 1: In a volunteer group, adults 21 and older volunteer from one to nine hours each week to spend time with a disabled senior citizen. The program recruits among community college students, four-year college students, and nonstudents. In the table below is a sample of the adult volunteers and the number of hours they volunteer per week.

Type of Volunteer	**1-3 Hours**	**4-6 Hours**	**7-9 Hours**
Community College Student	111	96	48
Four-Year College Student	96	133	61
Nonstudents	91	150	53

Is the number of hours volunteered independent of the type of volunteer? Test at a 5% significance level.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Notes 11.3

Example 2: De Anza College is interested in the relationship between anxiety level and the need to succeed in school. A random sample of 400 students took a test that measured anxiety level and need to succeed in school. The table below shows the results. De Anza College wants to know if anxiety level and need to succeed in school are independent events. Test at a 5% significance level.

Need to Succeed in School	High Anxiety	Med-high Anxiety	Medium Anxiety	Med-low Anxiety	Low Anxiety
High Need	35	42	53	15	10
Medium Need	18	48	63	33	31
Low Need	4	5	11	15	17

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

For more information and examples see online textbook OpenStax Introductory Statistics pages 633-638.

Lab Assignment 12.3, 12.5, 13.1, 13.2, 13.3, 11.3

Name:____Date:_____________________Row:_________

Lab Assignment 12.3, 12.5, 13.1, 13.2, 13.3, 11.3

For each of the following problems, if no significance level is given, use 0.05.

1. SCUBA divers have maximum dive times they cannot exceed when going to different depths. The data in the table below show different depths with the maximum dive times in minutes.

Use your calculator to find the regression line and predict the maximum dive time for 110 feet.

x (depth in feet)	*y* (maximum dive time)
50	80
60	55
70	45
80	35
90	25
100	22

1. Calculator work
2. Find a, b
3. Find the regression line
4. Find Critical Value and compare with r and state conclusion about linear correlation
5. Find the best predicted maximum dive time for 110 feet

2. The following table shows economic development measured in per capita income PCINC. Find the regression line and best estimate PCINC for 1905.

Year	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960
PCNIC	340	499	592	757	927	1050	1170	1364	1836	2132

1. Calculator work
2. Find a, b
3. Find the regression line
4. Find Critical Value and compare with r and state conclusion about linear correlation
5. Find the best estimate PCINC for 1905.

3. MRSA, or *Staphylococcus aureus*, can cause a serious bacterial infections in hospital patients. The table below shows various colony counts from different patients who may or may not have MRSA.

Conc = 0.6	Conc = 0.8	Conc = 1.0	Conc = 1.2	Conc = 1.4
9	16	22	30	27
66	93	147	199	168
98	82	120	148	132

Test whether the mean number of colonies are the same or are different.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

4. Four sororities took a random sample of sisters regarding their grade means for the past term.

Sorority 1	Sorority 2	Sorority 3	Sorority 4
2.17	2.63	2.63	3.79
1.85	1.77	3.78	33.45
2.83	3.25	4.00	3.08
1.69	1.86	2.55	2.26
3.33	2.21	2.45	3.18

Using a significance level of 1%, is there a difference in mean grades among the sororities?

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

5. Transit Railroads is interested in the relationship between travel distance and the ticket class purchased. A random sample of 200 passengers is taken. The table below shows the results. The railroad wants to know if a passenger's choice in ticket class is independent of the distance they must travel.

Traveling Distance	**Third class**	**Second class**	**First class**
1-100 miles	21	14	6
101-200 miles	18	16	8
201-300 miles	16	17	15
301-400 miles	12	14	21
401-500 miles	6	6	10

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Self-Check 12.3, 12.5, 13.1, 13.2, 13.3, 11.3

Name:___________________________________Date:__________________Row:___________

Self-Check 12.3, 12.5, 13.1, 13.2, 13.3, 11.3

1. Find the regression line and the best predicted cost of a ticket purchased one day in advance, given that the cost of the ticket is $300 if purchased 30 days in advance.

30 Days	244	260	264	264	278	318	280
One Day	456	614	567	943	628	1088	536

1. Calculator work
2. Find a, b
3. Find the regression line
4. Find Critical Value and compare with r and state conclusion about linear correlation
5. Find the best predicted cost of a ticket purchased one day in advance, given that the cost of the ticket is $300 if purchased 30 days in advance.

2. As part of an experiment to see how different types of soil cover would affect slicing tomato production, Marist College students grew tomato plants under different soil cover conditions. Groups of three plants each had one of the following treatments:

- bare soil
- a commercial ground cover
- black plastic
- straw
- compost

All plants grew under the same conditions and were the same variety. Students recorded the weight (in grams) of tomatoes produced by each of the n = 15 plants:

Bare	Ground Cover	Plastic	Straw	Compost
2625	5348	6583	7285	6277
2997	5682	8560	6897	7818
4915	5482	3830	9230	8677

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

3. The Bureau of Labor Statistics gathers data about employment in the United States. A sample is taken to calculate the number of U.S. citizens working in one of several industry sectors over time. The table below shows the results:

Industry Sector	**2000**	**2010**	**2020**
Non-agriculture wage and salary	13,243	13,044	15,018
Goods-producing, excluding agriculture	2,457	1,771	1,950
Services-providing	10,786	11,273	13,068

Agriculture, forestry, fishing, and hunting	240	214	201
Non-agriculture self-employed and unpaid family worker	931	894	972
Secondary wage and salary jobs in agriculture and private household industries	14	11	11
Secondary jobs as a self-employed or unpaid family worker	196	144	152

We want to know if the change in the number of jobs is independent of the change in years. Test at a 5% significance level.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

CHAPTER OVERVIEW

Section 12.2, 12.4

12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation
Lab Assignment 10.4, 12.2, 12.4
Self-Check 10.4, 12.2, 12.4

12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation

Section 12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation

Learning Objective:

In this section, you will:

- Understand and interpret linear correlation
- Apply hypothesis testing about linear correlation

Scatter plots are particularly helpful graphs when we want to see if there is a linear relationship among data points. They indicate both the direction of the relationship between the *x* variables and the *y* variables, and the strength of the relationship. We calculate the strength of the relationship between an independent variable and a dependent variable using linear regression.

A **correlation** exists between two variables when the values of one variable are associated with the values of the other variable.

The **linear correlation coefficient, r,** measures the strength of the linear correlation between the independent variable *x* and the dependent variable *y*.

What the VALUE of *r* tells us:

- The value of *r* is always between –1 and +1: $-1 \le r \le 1$.
- The size of the correlation *r* indicates the strength of the linear relationship between *x* and

y. Values of *r* close to –1 or to +1 indicate a stronger linear relationship between *x* and *y*.

- If $r = 0$ there is absolutely no linear relationship between *x* and *y* **(no linear correlation)**.
- If $r = 1$, there is perfect positive correlation. If $r = -1$, there is perfect negative correlation. In both these cases, all of the original data points lie on a straight line. Of course, in the real world, this will not generally happen.

What the SIGN of *r* tells us

- A positive value of *r* means that when *x* increases, *y* tends to increase and when *x* decreases, *y* tends to decrease **(positive correlation)**.
- A negative value of *r* means that when *x* increases, *y* tends to decrease and when *x* decreases, *y* tends to increase **(negative correlation)**.

Example 1a: In Europe and Asia, m-commerce is popular. M-commerce users have special mobile phones that work like electronic wallets as well as provide phone and Internet services. Users can do everything from paying for parking to buying a TV set or soda from a machine to banking to checking sports scores on the Internet. For the years 2000 through 2004, was there a relationship between the year and the number of m-commerce users? Construct a scatter plot. Let *x* = the year and let *y* = the number of m-commerce users, in millions.

x (year)	*y* (# of users in millions)
2000	0.5
2002	20.0
2003	33.0
2004	47.0

Finding the linear correlation coefficient:

We can calculate the linear correlation using the graphing calculator, **F:LinRegTTest**. First enter the data into two lists (L1 and L2), then under STAT-TESTS, find F:LinRegTTest. Enter your lists, and choose the "not equal" option. Scroll down to find r.

We can draw a conclusion about the relationship between the entire populations by using a hypothesis testing.

- Method 1: Using the p-value
- Method 2: Using a table of critical values.

Example 1b: Use the data from example 1a, and use 0.01 significance level to test the claim that there is a linear correlation between the year and the number of m-commerce users 1) Null and Alternative Hypothesis

2. Calculator Work
3. Test Statistic, P-Value and Linear correlation coefficient r
4. Conclusion about the null hypothesis

Method 1:

Method 2:

5. Final conclusion that addresses the original claim

Example 2: A random sample of 11 statistics students produced the following data, where *x* is the third exam score out of 80, and *y* is the final exam score out of 200. Use a 0.05 significance level to test the claim that there is a linear correlation between the third exam score and the final exam score.

x (third exam score)	*y* (final exam score)
65	175
67	133
71	185
71	163
66	126
75	198
67	153
70	163
71	159
69	151
69	159

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic, P-Value and Linear correlation coefficient r
4. Conclusion about the null hypothesis

Method 1:

Method 2:

5. Final conclusion that addresses the original claim

For more information and examples see online textbook OpenStax Introductory Statistics pages 682-685, 691-696.

Lab Assignment 10.4, 12.2, 12.4

Name:____Date:____________________Row:________

Lab Assignment 10.4, 12.2, 12.4

1. A new prep class was designed to improve SAT test scores. Five students were selected at random. Their scores on two practice exams were recorded, one before the class and one after. The data recorded in the table below. Are the scores, on average, higher after the class? Test at a 5% level.

SAT Score	Student 1	Student 2	Student 3	Student 4
Score before class	1840	1960	1920	2150
Score after class	1920	2160	2200	2100

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

1

2. Five ball players think they can throw the same distance with their dominant hand (throwing) and off-hand (catching hand). The data were collected and recorded in the table below. Conduct a hypothesis test to determine whether the mean difference in distances between the dominant and off-hand is significant. Test at the 5% level.

	Player 1	Player 2	Player 3	Player 4	Player 5
Dominant Hand	120	111	135	140	125
Off-Hand	105	109	98	111	99

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

2

3. The following table shows the poverty rates and cell phone usage in the United States. Use a 0.05 significance level to test the claim that there is a linear correlation between the poverty rate and the cellular usage per capita.

Year	Poverty Rate	Cellular Usage per Capita
2003	12.7	54.67
2005	1206	74.19
2007	12	84.86
2009	12	90.82

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic, P-Value and Linear correlation coefficient r
4. Conclusion about the null hypothesis

Method 1:

Method 2:

5. Final conclusion that addresses the original claim

3

4. The table below shows the life expectancy for an individual born in the United States in certain years. Use a 0.05 significance level to test the claim that there is a linear correlation between the life expectancy for an individual born in the United States in certain years.

Year of Birth	Life Expectancy
1930	59.7
1940	62.9
1950	70.2
1965	69.7
1973	71.4
1982	74.5
1987	75
1992	75.7
2010	78.7

) Null and Alternative Hypothesis

2. Calculator Work
3. Test Statistic, P-Value and Linear correlation coefficient r
4. Conclusion about the null hypothesis

Method 1:

Method 2:

5. Final conclusion that addresses the original claim

Self-Check 10.4, 12.2, 12.4

Name:____________________________________Date:__________________Row:___________

Self-Check 10.4, 12.2, 12.4

1. A study was conducted to investigate how effective a new diet was in lowering cholesterol. Results for the randomly selected subjects are shown in the table. The differences have a normal distribution. Are the subjects' cholesterol levels lower on average after the diet? Test at the 5% level.

Subject	A	B	C	D	E	F	G	H	I
Before	209	210	205	198	216	217	238	240	222
After	199	207	189	209	217	202	211	223	201

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim
6. Test the above claim by constructing an appropriate confidence interval.

2. Amelia plays basketball for her high school. She wants to improve to play at the college level. She notices that the number of points she scores in a game goes up in response to the number of hours she practices her jump shot each week. She records the following data:

x (hours practicing jump shot)	*y* (points scored in a game)
5	15
7	22
9	28
10	31
11	33
12	36

Use a 0.05 significance level to test the claim that there is a linear correlation between the number of hours she practices her jump shot each week and the number of points she scores in a game.

1. Null and Alternative Hypothesis
2. Calculator Work
3. Test Statistic, P-Value and Linear correlation coefficient r
4. Conclusion about the null hypothesis

Method 1:

Method 2:

5. Final conclusion that addresses the original claim

CHAPTER OVERVIEW

Section 12.3, 12.5

12.3, 12.5 The Regression Equation and Prediction

12.3, 12.5 The Regression Equation and Prediction

Section 12.3, 12.5 The Regression Equation and Prediction

Learning Objective:

In this section, you will:

• Calculate, interpret, and appropriately apply the regression line between two sets of data

If a correlation exists between two variables, then a **regression line** (or regression equation) can be used to make predictions about those variables.

The equation: $\hat{y} = \boldsymbol{a} + \boldsymbol{b}\boldsymbol{x}$ describes the relationship between *x* and *y*. *b* is the slope of the line, and *a* is the y-intercept.

You can find the regression line using the same **F:LinRegTTest** that we use to find *r*.

Example 1a: A random sample of ten professional athletes produced the following data where *x* is the number of endorsements the player has and *y* is the amount of money made (in millions of dollars).

x	0	3	2	1	5	5	4	3	0	4
y	2	8	7	3	13	12	9	9	3	10

Draw a scatter plot of the data and find the regression line and graph it on the same axes.

Predictions: If we find that there **is a linear correlation** between the two variables, we can use the regression line to predict y-values for given x-values**.** If there is **no linear correlation** between the two variables, then the best prediction for y is **the mean value, $\bar{y}$**

Example 1b: What is the best estimate for a y-value corresponding to an x-value of 6?

1

Example 2: Concerns about global warming have led to studies of the relationship between global temperatures and the concentration of carbon dioxide. Listed below are the concentrations of carbon dioxide and temperatures for different years. Find the regression line. What is the best prediction for the global temperature when the carbon dioxide concentration is 320?

Carbon Dioxide	314	317	320	326	331	339	346	354	361	369
Temperature	13.9	14.0	13.9	14.1	14.0	14.3	14.1	14.5	14.5	14.4

1. Calculator work
2. Find a, b
3. Find the regression line
4. Find Critical Value and compare with r and state conclusion about linear correlation
5. What is the best prediction for the global temperature when the carbon dioxide concentration is 320?

Example 3: Listed below are brain sizes (in cm3) and IQ scores of subjects. Find the regression line. If a person has a brain size of 1020 cm3, find the best predicted IQ score for that person.

Brain Size	965	1029	1030	1285	1049	1077	1037	1068	1176	1105	1029	1030
IQ	90	85	86	102	103	97	124	125	102	114	86	87

1. Calculator work
2. Find a, b
3. Find the regression line
4. Find Critical Value and compare with r and state conclusion about linear correlation

2

5. If a person has a brain size of 1020 cm3, find the best predicted IQ score for that person.

For more information and examples see online textbook OpenStax Introductory Statistics pages 685-691, 696-697.

CHAPTER OVERVIEW

Section 13.1 - 13.3

13.1, 13.2, 13.3 One-Way ANOVA

13.1, 13.2, 13.3 One-Way ANOVA

Section 13.1, 13.2, 13.3 One-Way ANOVA

Learning Objective:

In this section, you will:

• Determine, using an ANOVA hypothesis test, whether means of three or more populations are equal or not

ANOVA is short for analysis of variance. The purpose of a one-way ANOVA test is to determine the existence of a statistically significant difference among several group means. The test actually uses **variances** to help determine if three or more means are equal or not. In order to perform a one-way ANOVA test, there are five basic assumptions to be fulfilled:

1. Each population from which a sample is taken is assumed to be normal.
2. All samples are randomly selected and independent.
3. The populations are assumed to have equal standard deviations (or variances).
4. The factor is a categorical variable.
5. The response is a numerical variable.

The calculations are very complicated. We will use our calculators to perform the computations using the F-test.

To perform and ANOVA test, first enter all the raw data into lists, one list for each sample. Then select STAT–TESTS and choose **H:ANOVA**. You will then enter the names of all your lists. If you have four populations, you would enter your data into L1 through L4, and choose ANOVA(L1,L2,L3,L4).

Use the p-value to make your conclusion.

Example 1: Listed below is head injury data from crash test dummies used in the small, medium, and large cars. These measurements are in hic (a standard head injury criterion). Use a 0.01 significance level to test the claim that the different car categories have the same mean.

Small Cars	290	406	371	544	374	501	376	499	479	475
Medium Cars	245	502	474	505	393	264	368	510	296	349
Large Cars	342	216	335	698	216	169	608	432	510	332

1. Null and Alternative Hypothesis

2. Calculator Work

1

3. Test Statistic and P-Value

4. Conclusion about the null hypothesis

5. Final conclusion that addresses the original claim

6. Do these data suggest that larger cars are safer?

Example 2: The table below lists weights (in kilograms) of poplar trees that were given different treatments at different sites. Use a 0.05 significance level to test the claim that the trees with different treatments have the same mean weight.

None	Fertilizer	Irrigation	Fertilizer and Irrigation
0.15	1.34	0.23	2.03
0.02	0.14	0.04	0.27
0.16	0.02	0.34	0.92
0.37	0.08	0.16	1.07

0.22	0.08	0.05	2.38

1. Null and Alternative Hypothesis
2. Calculator Work

2

206

3. Test Statistic and P-Value
4. Conclusion about the null hypothesis
5. Final conclusion that addresses the original claim

Does is appear that the treatment affects the weights of the trees? If so, which treatment appears to be best?

For more information and examples see online textbook OpenStax Introductory Statistics pages 743-756.

Index

D

dire

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Math 110 - Introduction to Statistics - Module

Webpages: 107

All licenses found:

- Undeclared: 100% (107 pages)

By Page

- Math 110 - Introduction to Statistics - Module - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - Section 1.1 - *Undeclared*
    - 1.1 Definitions of Statistics, Probability, and Key Terms - *Undeclared*
    - Lab Assignment 1.1 - *Undeclared*
    - Self-Check 1.1 - *Undeclared*
  - Section 1.2 - *Undeclared*
    - 1.2 Data, Sampling, and Variation in Data and Sampling - *Undeclared*
    - Lab Assignment 1.2 - *Undeclared*
    - Self-Check 1.2 - *Undeclared*
  - Section 1.3 - *Undeclared*
    - 1.3 Frequency, Frequency Tables, and Levels of Measurement - *Undeclared*
  - Section 1.4 - *Undeclared*
    - 1.4 Experimental Design - *Undeclared*
    - Lab Assignment 1.3, 1.4 - *Undeclared*
    - Self-Check 1.3, 1.4 - *Undeclared*
  - Section 2.1 - *Undeclared*
    - 2.1 Stem-and-Leaf Plots (Stemplots) and Bar Graphs - *Undeclared*
  - Section 2.2 - *Undeclared*
    - 2.2 Frequency Table and Histograms - *Undeclared*
    - Lab Assignment 2.1, 2.2 - *Undeclared*
    - Self-Check 2.1, 2.2 - *Undeclared*
  - Section 2.3 - *Undeclared*
    - 2.3 Measures of the Location of the Data - *Undeclared*
  - Section 2.4 - *Undeclared*
    - 2.4 Box Plots - *Undeclared*
    - Lab Assignment 2.3, 2.4 - *Undeclared*
    - Self-Check 2.3, 2.4 - *Undeclared*
  - Section 2.5 - *Undeclared*
    - 2.5 Measures of Center of the Data - *Undeclared*
  - Section 2.6 - *Undeclared*
    - 2.6 Skewness and the Mean, Median, and Mode - *Undeclared*
  - Section 2.7 - *Undeclared*
    - 2.7 Measures of the Spread of the Data - *Undeclared*
    - Lab Assignment 2.5, 2.6, 2.7 - *Undeclared*
    - Self-Check 2.5, 2.6, 2.7 - *Undeclared*
  - Section 3.1 - *Undeclared*
    - 3.1 Terminology - *Undeclared*
    - Lab Assignment 3.1 - *Undeclared*
    - Self-Check 3.1 - *Undeclared*
  - Section 3.2-3.4 - *Undeclared*
    - 3.2-3.4 Probability Topics - *Undeclared*
    - Lab Assignment 3.2, 3.3, 3.4 - *Undeclared*
    - Self-Check 3.2, 3.3, 3.4 - *Undeclared*
  - Section 4.1 - *Undeclared*
    - 4.1 Probability Distribution Function (PDF) for a Discrete Random Variable - *Undeclared*
  - Section 4.2 - *Undeclared*
    - 4.2 Mean or Expected Value and Standard Deviation - *Undeclared*
    - Lab Assignment 4.1, 4.2 - *Undeclared*
    - Self-Check 4.1, 4.2 - *Undeclared*
  - Section 6.1 - *Undeclared*
    - 6.1 The Standard Normal Distribution - *Undeclared*
    - Lab Assignment 6.1 - *Undeclared*
    - Self-Check 6.1 - *Undeclared*
  - Section 7.1 - *Undeclared*
    - 7.1 The Central Limit Theorem for Sample Means - *Undeclared*
    - Lab Assignment 7.1 - *Undeclared*
    - Self-Check 7.1 - *Undeclared*
  - Section 8.1 - *Undeclared*
    - 8.1 A Single Population Mean using the Normal Distribution - *Undeclared*

- Lab Assignment 8.1 - *Undeclared*
- Self-Check 8.1 - *Undeclared*
- Section 8.2 - *Undeclared*
  - 8.2 A Single Population Mean using the Student t Distribution - *Undeclared*
- Section 8.3 - *Undeclared*
  - 8.3 A Population Proportion - *Undeclared*
  - Lab Assignment 8.2, 8.3 - *Undeclared*
  - Self-Check 8.2, 8.3 - *Undeclared*
- Section 9.1 - *Undeclared*
  - 9.1 Null and Alternative Hypothesis - *Undeclared*
- Section 9.4 - *Undeclared*
  - 9.4 Rare Events, the Sample, Decision and Conclusion - *Undeclared*
  - Lab Assignment 9.1, 9.4 - *Undeclared*
  - Self-Check 9.1, 9.4 - *Undeclared*
- Section 9.5 - *Undeclared*
  - 9.5 Full Hypothesis Testing with One Sample - *Undeclared*
  - Lab Assignment 9.5 - *Undeclared*
  - Self-Check 9.5 - *Undeclared*
- Section 10.1 - *Undeclared*
  - 10.1 Two Population Means with Unknown Standard Deviations - *Undeclared*
- Section 10.3 - *Undeclared*
  - 10.3 Comparing Two Independent Population Proportions - *Undeclared*
  - Lab Assignment 10.1, 10.3 - *Undeclared*
  - Self-Check 10.1, 10.3 - *Undeclared*
- Section 10.4 - *Undeclared*
  - 10.4 Matched or Paired Samples - *Undeclared*
- Section 11.3 - *Undeclared*
  - 11.3 Test of Independence - *Undeclared*
  - Lab Assignment 12.3, 12.5, 13.1, 13.2, 13.3, 11.3 - *Undeclared*
  - Self-Check 12.3, 12.5, 13.1, 13.2, 13.3, 11.3 - *Undeclared*
- Section 12.2, 12.4 - *Undeclared*
  - 12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation - *Undeclared*
  - Lab Assignment 10.4, 12.2, 12.4 - *Undeclared*
  - Self-Check 10.4, 12.2, 12.4 - *Undeclared*
- Section 12.3, 12.5 - *Undeclared*
  - 12.3, 12.5 The Regression Equation and Prediction - *Undeclared*
- Section 13.1 - 13.3 - *Undeclared*
  - 13.1, 13.2, 13.3 One-Way ANOVA - *Undeclared*
- Back Matter - *Undeclared*
  - Index - *Undeclared*
  - Glossary - *Undeclared*
  - Detailed Licensing - *Undeclared*
  - Detailed Licensing - *Undeclared*

LibreTexts

Detailed Licensing

Overview

Title: Math 110 - Introduction to Statistics - Module

Webpages: 107

All licenses found:

- Undeclared: 100% (107 pages)

By Page

- Math 110 - Introduction to Statistics - Module - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - Section 1.1 - *Undeclared*
    - 1.1 Definitions of Statistics, Probability, and Key Terms - *Undeclared*
    - Lab Assignment 1.1 - *Undeclared*
    - Self-Check 1.1 - *Undeclared*
  - Section 1.2 - *Undeclared*
    - 1.2 Data, Sampling, and Variation in Data and Sampling - *Undeclared*
    - Lab Assignment 1.2 - *Undeclared*
    - Self-Check 1.2 - *Undeclared*
  - Section 1.3 - *Undeclared*
    - 1.3 Frequency, Frequency Tables, and Levels of Measurement - *Undeclared*
  - Section 1.4 - *Undeclared*
    - 1.4 Experimental Design - *Undeclared*
    - Lab Assignment 1.3, 1.4 - *Undeclared*
    - Self-Check 1.3, 1.4 - *Undeclared*
  - Section 2.1 - *Undeclared*
    - 2.1 Stem-and-Leaf Plots (Stemplots) and Bar Graphs - *Undeclared*
  - Section 2.2 - *Undeclared*
    - 2.2 Frequency Table and Histograms - *Undeclared*
    - Lab Assignment 2.1, 2.2 - *Undeclared*
    - Self-Check 2.1, 2.2 - *Undeclared*
  - Section 2.3 - *Undeclared*
    - 2.3 Measures of the Location of the Data - *Undeclared*
  - Section 2.4 - *Undeclared*
    - 2.4 Box Plots - *Undeclared*
    - Lab Assignment 2.3, 2.4 - *Undeclared*
    - Self-Check 2.3, 2.4 - *Undeclared*
  - Section 2.5 - *Undeclared*
    - 2.5 Measures of Center of the Data - *Undeclared*
  - Section 2.6 - *Undeclared*
    - 2.6 Skewness and the Mean, Median, and Mode - *Undeclared*
  - Section 2.7 - *Undeclared*
    - 2.7 Measures of the Spread of the Data - *Undeclared*
    - Lab Assignment 2.5, 2.6, 2.7 - *Undeclared*
    - Self-Check 2.5, 2.6, 2.7 - *Undeclared*
  - Section 3.1 - *Undeclared*
    - 3.1 Terminology - *Undeclared*
    - Lab Assignment 3.1 - *Undeclared*
    - Self-Check 3.1 - *Undeclared*
  - Section 3.2-3.4 - *Undeclared*
    - 3.2-3.4 Probability Topics - *Undeclared*
    - Lab Assignment 3.2, 3.3, 3.4 - *Undeclared*
    - Self-Check 3.2, 3.3, 3.4 - *Undeclared*
  - Section 4.1 - *Undeclared*
    - 4.1 Probability Distribution Function (PDF) for a Discrete Random Variable - *Undeclared*
  - Section 4.2 - *Undeclared*
    - 4.2 Mean or Expected Value and Standard Deviation - *Undeclared*
    - Lab Assignment 4.1, 4.2 - *Undeclared*
    - Self-Check 4.1, 4.2 - *Undeclared*
  - Section 6.1 - *Undeclared*
    - 6.1 The Standard Normal Distribution - *Undeclared*
    - Lab Assignment 6.1 - *Undeclared*
    - Self-Check 6.1 - *Undeclared*
  - Section 7.1 - *Undeclared*
    - 7.1 The Central Limit Theorem for Sample Means - *Undeclared*
    - Lab Assignment 7.1 - *Undeclared*
    - Self-Check 7.1 - *Undeclared*
  - Section 8.1 - *Undeclared*
    - 8.1 A Single Population Mean using the Normal Distribution - *Undeclared*

- Lab Assignment 8.1 - *Undeclared*
- Self-Check 8.1 - *Undeclared*
- Section 8.2 - *Undeclared*
  - 8.2 A Single Population Mean using the Student t Distribution - *Undeclared*
- Section 8.3 - *Undeclared*
  - 8.3 A Population Proportion - *Undeclared*
  - Lab Assignment 8.2, 8.3 - *Undeclared*
  - Self-Check 8.2, 8.3 - *Undeclared*
- Section 9.1 - *Undeclared*
  - 9.1 Null and Alternative Hypothesis - *Undeclared*
- Section 9.4 - *Undeclared*
  - 9.4 Rare Events, the Sample, Decision and Conclusion - *Undeclared*
  - Lab Assignment 9.1, 9.4 - *Undeclared*
  - Self-Check 9.1, 9.4 - *Undeclared*
- Section 9.5 - *Undeclared*
  - 9.5 Full Hypothesis Testing with One Sample - *Undeclared*
  - Lab Assignment 9.5 - *Undeclared*
  - Self-Check 9.5 - *Undeclared*
- Section 10.1 - *Undeclared*
  - 10.1 Two Population Means with Unknown Standard Deviations - *Undeclared*
- Section 10.3 - *Undeclared*
  - 10.3 Comparing Two Independent Population Proportions - *Undeclared*
  - Lab Assignment 10.1, 10.3 - *Undeclared*
  - Self-Check 10.1, 10.3 - *Undeclared*
- Section 10.4 - *Undeclared*
  - 10.4 Matched or Paired Samples - *Undeclared*
- Section 11.3 - *Undeclared*
  - 11.3 Test of Independence - *Undeclared*
  - Lab Assignment 12.3, 12.5, 13.1, 13.2, 13.3, 11.3 - *Undeclared*
  - Self-Check 12.3, 12.5, 13.1, 13.2, 13.3, 11.3 - *Undeclared*
- Section 12.2, 12.4 - *Undeclared*
  - 12.2, 12.4 Scatter Plots and Testing the Significance of the Correlation - *Undeclared*
  - Lab Assignment 10.4, 12.2, 12.4 - *Undeclared*
  - Self-Check 10.4, 12.2, 12.4 - *Undeclared*
- Section 12.3, 12.5 - *Undeclared*
  - 12.3, 12.5 The Regression Equation and Prediction - *Undeclared*
- Section 13.1 - 13.3 - *Undeclared*
  - 13.1, 13.2, 13.3 One-Way ANOVA - *Undeclared*
- Back Matter - *Undeclared*
  - Index - *Undeclared*
  - Glossary - *Undeclared*
  - Detailed Licensing - *Undeclared*
  - Detailed Licensing - *Undeclared*