8-2006

Statistics: Raising the Bar for the Seventh Grade Classroom.

Sherry Lynn Mullins
East Tennessee State University

Statistics: Raising the Bar for the Seventh Grade Classroom

A thesis
presented to
the faculty of the Department of Mathematics
East Tennessee State University

In partial fulfillment
of the requirements for the degree
Master of Science in Mathematical Sciences

---

by
Sherry Lynn Mullins
August 2006

Robert Price, Ph.D., Chair
Edith Seier, Ph.D.
Yali Liu, Ph.D.

Keywords: Middle School Statistics, Graphs, Statistical Estimation,
Virginia Standards of Learning

ABSTRACT

Statistics: Raising the Bar for the Seventh Grade Classroom

by

Sherry Lynn Mullins

After recognizing the need for a more thorough concentration of statistics at the seventh grade level, the author concluded that it would be a good idea to include statistics that cover both seventh and eighth grade Virginia Standards of Learning. Many years of administering the SOL mathematics test at the eighth grade level led the author to the understanding that some of the more advanced seventh graders would be missing some key concepts taught in eighth grade because those advanced students would be taking algebra in the eighth grade. In this thesis, the author has developed four units that she feels are appropriate for this level and will fill the gap.

DEDICATION

This thesis is dedicated to our Lord because without him none of this would have been possible and I owe Him everything. To my husband and best friend, Joe Mullins, who holds my whole world in his hand, I would not have been able to do this without your love and support. I'll always love you! To our son, Eric Mullins, who believes that his Mom can do anything. You and your Dad are my reasons for living. Thank you for always being there for me. To my parents, Garrett and Lena Barton, your love and prayers have helped me to be who I am and I am grateful. To Velma Bostic Crabtree for showing me that I could do math, making me feel special, and planting a seed in me to become a math teacher, I am forever grateful.

ACKNOWLEDGEMENTS

I would like to thank Dr. Bob Price for the materials, all of his encouragement and for creating three of my graphs.

CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

1 INTRODUCTION

1.1 The Idea

Individuals need an understanding of statistics to function in our society. Such items as consumer reports, cost of living, and surveys are a part of everyday lives. Competency with the utilitarian aspects of statistics will help students process the mounds of data they receive everyday. It is with this idea in mind that we shall approach the teaching of statistics at the seventh grade level.

During the fall of 1991, the author was hired as a middle school math teacher for Sugar Grove Combined School in Smyth County, Virginia. She soon realized that this school was very unique. The school contained K-8 grades with slightly more than two hundred students. The more advanced eighth grade students would attend another middle school in the county for one-half of the day to take Algebra, thereby skipping math eight with all of its Standards of Learning (SOLs). These students were still responsible for passing the Math 8 SOL test at the end of the school year. This led the author to the idea of teaching more than just the basic seventh grade SOLS to the seventh grade. She decided to include more eighth grade material, but what should it be? The decision was simple. She would compare the Standards of Learning of seventh to eighth grade to see what those students would be missing the most. This led to the idea of creating units in statistics for seventh grade so they would have a better chance of passing the eighth grade requirements.

1.2 Available Options

The author first researched and compared the two lists of Standards of Learning for seventh and eighth graders [5] that are statistical in nature:

7.16 The student will create and solve problems involving the mean, median, mode and range of a set of data.

7.17 The student will display data, using frequency distributions, line plots, stem-and-leaf plots, box-and-whisker plots and scatter grams.

7.18 The student will make inferences and predictions based on a set of data that the student(s) collect.

7.19 The student will represent, analyze and generalize a variety of patterns, including arithmetic and geometric sequences with tables, graphs, rules and word in order to investigate and describe functional relationships.

8.11 The student will analyze problem situations, such as games of chance, board games or grading scales, and make predictions using knowledge of probability.

8.12 The student will use information displayed in line, bar, circle, or picture graphs and histograms to make inferences, comparisons, and predictions.

8.13 The student will use a matrix to organize and describe data.

After closely comparing the number of questions from each strand covered by the eighth grade Math Standards of Learning test, the author realized that approximately 20% of the questions came from the probability and statistics strand.

Upon learning this information, the author wanted to see if other middle schools in her county had the same problem. By talking with other middle school math teachers, she came to the conclusion that all seventh grade teachers were suffering from the same dilemma. The teachers needed some way of covering a few more of the eighth grade standards for those advanced placement students.

In the middle school classroom, finding the time to add more to the curriculum can be a problem. Most of the teachers did explain that, at the end of the year after their grade level SOLS were covered, they had two to three weeks that they could use for statistics beyond their textbook.

1.3 Selecting the Units

The teachers' responses led the author to create units of statistics that can be used at the seventh grade level when extra time permits. The units, except Unit Two, are designed to be covered in approximately one to two weeks. Unit One is designed to be taught before any of the others. This unit contains basic definitions and terminology in statistics, and the laying of background for the other units.

The author decided that this unit should contain the measures of central tendency since these are usually taught near the beginning of the statistics section in any middle school math textbook. The Graphing or Display Unit is usually a favorite of middle school students because it allows them to collect their own data and display the data in numerous ways. The unit on Analyzing Graphs will guide students by teaching them to read graphs and make predictions about future

outcomes. This unit will also teach students to recognize when statistics and graphs are misleading or not telling the whole truth. The last unit on Sampling is a very basic beginners approach to surveys and to estimating populations.

In all of these units, an Introduction Section defines new terminology for the particular unit along with examples and objectives. An In Class Section follows, which gives teachers examples for the students to work. The Application Section will contain material for a real life data or hands-on approach using manipulatives. The Homework Section allows the student to work or collect data individually. It also allows the teacher to check for understanding of the concepts taught the previous day(s).

1.4 Statistics

What is statistics? A common definition is that statistics is (1) a collection of numerical data and (2) the science of handling that data. Statistics is usually divided into two branches: descriptive statistics and inferential statistics. Descriptive statistics includes techniques designed to describe the main features of numerical data. For example, suppose all ETSU students are measured and their average height was 5'9". This would be an example of descriptive statistics because it describes a population parameter. Suppose 50,000 people are to vote for Mayor of the City, a random sample of the 50,000 people is conducted and 40% will vote for James Doe and 60% will vote for May Adams. This is an example of inferential statistics. The inference can be made that May Adams will become the Mayor. The political preference of the total population can be inferred from the examination of a sample of the total population since all 50,000 people were not polled. This type of statistical inference moves beyond the data at hand to draw conclusions about a wider universe.

Applications of statistics in everyday life are not hard to find. The average person needs to be aware of the why, how, and where of both the collection of data and the processing of that data [6].

Too often Statistics is confused with Probability. Consider the following examples:

(1) Suppose in a sample of 100 people it is known that 42 are boys and 58 are girls. We can find the probability that a person chosen at random is a boy:

P(boy)=42/100 =.42

This is a problem in probability; from a known population we have the chances of selecting a particular sample.

(2) Suppose that the break down of boys to girls is not known. We could select 20 people at random and determine the number of boys and the number of girls. If we find that the number of boys is 8 and the number of girls is 12, we may estimate that the same ratio holds for the entire group. Therefore, we can say that there are 40 boys and 60 girls. This is a problem in statistics, and we drew our conclusion of the gender distribution of the population from a random sample.

2 MEASURES OF CENTRAL TENDENCY

2.1 Introduction

The Measures of Central Tendency in mathematics are three common ways to summarize data using a single number. These are the mean, median, and mode. All are types of averages. The goal of this lesson is to introduce the students to the concepts of mean, median, and mode.

Virginia Standards of Learning Objective [5]:

7.16 The student will create and solve problems involving the mean, median, mode and range of a set of data.

Key Terms:

This unit will introduce students to the following terms:

1. mean
2. median
3. mode
4. total

The mean or arithmetic average is found by adding the data points and dividing by the number of data points [1].

The median is the middle number in a set of data when they are ordered from the smallest to the largest. In other words, it is the midpoint of distribution, the number such that half the data lie below it and the other half above it [1].

The mode is the number(s) or item(s) that occurs the most [1].

Materials Needed:

Four-function calculator

Paper

Pencil

2.2 In Class Examples

Example 1:

A group of ten seventh grade math students took a chapter test. Their test scores were:

75, 60, 90, 83, 75, 70, 77, 85, 80, 75

1. Find the mean or arithmetic average for the data.
2. Find the median for the data.
3. Find the mode for the data.

Solution:

1. Using your calculator, add the list of numbers. The total is 770. Divide the total by how many numbers are in your list; 770 divided by 10 is 77. Then 77 is the mean score of the data.
2. To compute the median score, first arrange the scores in ascending order: 60, 70, 75, 75, 75, 77, 80, 83, 85, 90. There are 10 scores therefore, the median lies between the 5^{th} and 6^{th} scores which are 75 and 77. Since 76 lies exactly between 75 and 77, then 76 is the median.
3. To compute the mode**,** scan the list to determine the score that appears the most often. Seventy five appears with the most frequency, therefore 75 is the mode.

Example 2: The wage and distribution from a small company is given in the following frequency table.

EMPLOYEE SALARIES FOR 2004-05

Annual Income		Number Receiving This Income
$100,000		1
50,000		1
15,000		7
12,000		10
9,000		8
6,000		13

1. Find the mean salary of the workers.
2. Find the median salary of the workers.
3. Find the mode salary of the workers.
4. If you were a union leader, which average would you use in a salary negotiation?
5. Which statistic would a company owner be more likely to use in salary negotiations?
6. Which of the statistics would you, as an employee, prefer to be used to determine your salary?

Solution:

1. To find the mean, you must be able to read the frequency table. Notice there is only 1 person whose salary is $100,000, etc. To find the mean, multiply the number of employees receiving the income and the amount of income received. Next, total the salaries. Now divide, by the number of employees.

   (1)100,000 + (1)50,000 + (7)15,000 + (10)12,000 + (8)9,000 + (13)6,000 =$525,000

   $525,000/40 = $13,125 mean salary

2. Find the median. Notice, there are 40 employees listed on the table. Therefore, the median salary will be located between the 20^{th} and 21^{st} salary. Looking back at the table, this will be within the section of $9,000 salary. $9,000 is the median salary.
3. Find the mode. Using the frequency table, we can determine that more employees receive a $6,000 salary. $6,000 is the mode.
4. If you were a union leader, you would want to use the mode for salary negotiations because it is the lowest salary and more people received this salary.
5. A company owner for salary negotiations would use the mean because it is the highest. He/She would say that since the average salary is $13,125, your figure is too low.
6. An employee would rather have the mean in this situation when given a choice. The mean amount is much larger than the mode or median.

2.3 Application

"Mean Chocolate"

Objective: Using M & M candies, the students will do an activity to find the mean, median and mode

Materials:

1. One regular size bag of M & M's plain for each student
2. Calculators
3. Pencil and Paper for recording data

Procedure:

1. Give each student one bag of candy.
2. Have each student estimate the number of candies contained in the bag without feeling the bag. Each should record his or her answer.
3. Each student should open their bag and count the candies. Record the answer and compare it to the estimate.
4. The teacher should allow each student to write their number of candies on the board.
5. After each has completed the task, the student should list the number of candies of all the students on his/her paper in numerical order.

Assignment:

1. Find the mean.
2. Find the median
3. Find the mode.
4. Did each bag contain the same number of candies? If not, why not?
5. How many candies should a consumer expect in a bag? On which statistic did you base your answer?

Solutions:

1, 2 and 3 answers will vary.

4. Not all bags will contain the same number. Candy is sold by weight not count.
5. The consumer should probably expect the number of candies to be close to the mean.

2.4 Homework Exercises

Problem 1:

Grade Point Averages of 20 ETSU Freshmen

1.5	2.9	2.2	2.5
2.6	2.8	2.3	3.5
1.4	3.0	1.7	2.7
2.0	1.9	2.2	2.5
1.8	2.4	1.6	2.3

Using the data collected from 20 ETSU Freshmen find:

1. Mean
2. Median
3. Mode
4. Describe the overall GPA, grade point average, of these 20 freshmen.

Problem 2:

Use the following table to answer the questions at the bottom of this page.

Virginia's Rainfall for the Last Ten Years

Year	Rainfall in Inches
1	47
2	45
3	38
4	42
5	52
6	37
7	17
8	40
9	25
10	40

1. Find the Mean rainfall for VA over the last ten years.

2. Find the Median rainfall for VA over the last ten years.

3. Find the Mode rainfall for VA over the last ten years.

Bonus:

Materials: Access to Computer with an Internet Connection

Using the website: http://countrystudies.us/united-states/weather/virginia/marion/htm

Find the mean for the average high temperature for Marion, Virginia for Jan. – Dec.

2.5 Answer Key

Worksheet 1: GPA of 20 ETSU Freshmen

1. Mean: 2.29
2. Median: 2.3
3. Mode: 2.2, 2.3, 2.5
4. Most ETSU Freshmen's grades are between 2.2 and 2.3.

Worksheet 2: VA's Rainfall for the Last Ten Years

1. Mean: 38.3
2. Median: 40
3. Mode: 40

BONUS:

The mean of the average high temperature for Marion, Virginia is 72 degrees.

3 FREQUENCY TABLES

3.1 Introduction

Data can be displayed in numerous ways. One way to display large amounts of information is in a frequency table. A frequency table is a listing of the possible values for a variable and the count of how often each occurs. For a categorical variable, each observation falls in one of the categories. For a quantitative variable, a frequency table usually divides the possible values into a set of intervals and displays the number of values in each interval.

Virginia Standards of Learning [5]:

7.16 The student will display data, using frequency distributions, line plots, stem-and-leaf plots, box-and-whisker plots and scatter grams.

8.13 The student will use a matrix to organize and describe data.

Key Terms:

1. Frequency Table
2. Tally
3. Scale
4. Interval
5. Range

A Frequency Table is a table for organizing a set of data that shows the number of times each item or number appears.

A Tally is a mark made in the "tally' column for each time a specific piece of information appears.

Scale is the set of all possible values of a given measurement, including the least and greatest numbers in the set, separated by the intervals used.

Interval is the difference between successive values on a scale.

Range is the difference between the greatest value and the least value in a set of data. It is a simple measure of spread.

Materials Needed**:**

Paper

Pencil

3.2 In Class Examples

Example 1:

A group of eighth grade students was asked how many hours they used the computer each week.

The results were:

1	0	2	2	3
1	1	0	3	0
4	3	2	0	0
2	2	1	3	4
2	3	2	2	0

1. Make a frequency table to display the results.
2. What is the most common amount of time used?
3. Give one advantage of using a frequency table.

Solution:

1.

Hrs. on Computer	Tally	Frequency
0	1 1 1 1 1 1	6
1	1 1 1 1	4
2	1 1 1 1 1 1 1 1	8
3	1 1 1 1 1	5
4	1 1	2

Figure 1: Frequency Table Example

2. The most common amount of time spent on the computer is 2 hours.

3. One advantage of using a frequency table is that it provides a quick summary of information.

Example 2**:**

A list of the lengths of TV Commercials in seconds:

25	30	10	20	60	10	10
20	60	15	20	20	30	45
20	10	60	20	35	30	30

1. Choose a scale and interval for the table.
2. Why do you believe you need to use an Interval other than one?
3. Find the range of data.
4. Construct a frequency table.

Solutions:

1. The scale will be 10-69 with an interval of 10.

2. An Interval of 10 is needed because the smallest number is 10 with the largest being 60. Using ones would make the table too large.

3. The range is 60 – 10 = 50.

4.

Lengths of Commercials (secs.)	Tally	Frequency
10-19	1 1 1 1 1	5
20-29	1 1 1 1 1 1	6
30-39	1 1 1 1 1 1	6
40-49	1	1
50-59		0
60-69	1 1 1	3

3.3 Application

"Money Matters"

Objective: Conducting a Survey. Each student will collect his/her own information and construct a frequency table using a proper scale and interval.

Materials:

1. Paper
2. Pencil

Procedure:

1. Each student will **survey** 20 other students concerning the amount of money each spends during a one week period.
2. The results will be written along with the person's name.
3. Using the results, each student will construct a frequency table using an interval of $1.00.

Assignment:

1. Survey 20 students concerning the amount of money spent within one week and record the results.
2. Construct a frequency table using an interval of $1.00.
3. Find the mean, median, mode and range.
4. What is the most common interval of money spent?

Solution:

All answers will vary, depending upon the people surveyed.

3.4 Homework Exercises

Using the following information, answer the questions at the bottom of the page.

Eighth Grade Exam Grades for 25 Students

60	55	80	90	80
75	90	95	80	88
73	65	72	77	79
77	81	99	74	96
49	37	84	22	15

1. Determine a scale and interval for the information.
2. Construct a frequency table.
3. If 70 and up is a passing score, how many students passed the exam?
4. Most students scored within which interval?

3.5 Answer Key

1. We see that the scores range from 15 to 99.

   A good scale would be from 10 to 99, with an interval of 10.

2.

Scores	Tally	Frequency
10-19	1	1
20-29	1	1
30-39	1	1
40-49	1	1
50-59	1	1
60-69	11	2
70 -79	1111111	7
80-89	111111	6
90-99	11111	5

3. If 70 and above is a passing score, 18 students passed the exam.

4. The most common Interval is 70-79.

.

3.6 Optional Homework

Materials:

Magazines

Newspapers

Objective: The student will research the distribution of highway miles per gallon for cars using ads found in magazines and newspapers.

Procedure:

1. Using a scale of 10-59 mph with an Interval of 5, construct a frequency table using the information found in the car ads from newspapers and magazines. Find the mileage of at least 20 different vehicles.
2. Which car gets the best gas mileage?
3. Within which Interval do most vehicles fall?

3.7 Optional Homework Solution

1. Construct a Frequency Table. Results will vary.

Miles per Gallon Distribution

MPG	Tally	Frequency
10 – 14		
15 – 19		
20 – 24		
25 – 29		
30 – 34		
35 – 39		
40 – 44		
45 – 49		
50 – 54		
55 – 59		

2. Answers will vary, depending upon ads used.

3. Answers will vary, depending upon ads used.

4 BAR GRAPHS AND HISTOGRAMS

4.1 Introduction

The most basic graph in statistics would be the bar graph. The bar graph uses bars to make general comparisons. The histogram is a special kind of bar graph that uses intervals.

Virginia Standards of Learning Objective [5].

7.17 The student will make inferences and predictions based on the analysis of a set of data that the student(s) collect.

8.12 The student will use information displayed in line, bar, circle, picture graphs and histograms to make comparisons, predictions and inferences.

Key Terms:

This unit will introduce students to the following terms.

1. bar graph
2. histogram
3. “x” axis
4. “y” axis

A bar graph is a simple graph for categorical variables. The height of the bar is usually presented as a percent.

A histogram is a special type of bar graph that displays the frequency of information for a quantitative variable. Equal intervals are used to display all possible values of data leaving no gaps between the bars.

The “x” axis is the horizontal axis or line on which the intervals are placed.

The “y” axis is the vertical axis or line that shows the frequency of the intervals on the “x” axis**.**

Materials:

Paper

Colored Pencils

Rules

Pencil

4.2 In Class Examples

Example 1:

City Middle School sold candy in the month of August. The 6^{th} grade sold \$400, the 7^{th} grade sold \$350, and the 8^{th} grade sold \$500.

1. Which information would you place on the "x" axis?

2. Which information would you place on the "y" axis?

3. Construct a Bar Graph using the information. Be sure to give the graph a title and color the bars.

4. Which grade sold the most candy?

Solution**:**

1. The information placed on the "x" axis is the grade of the students.
2. The information placed on the "y" axis is the amount sold by each grade.
3. The graph is on the next page
4. The 8^{th} grade sold the most by \$100.

3. Candy Sales for City Middle School
(August)

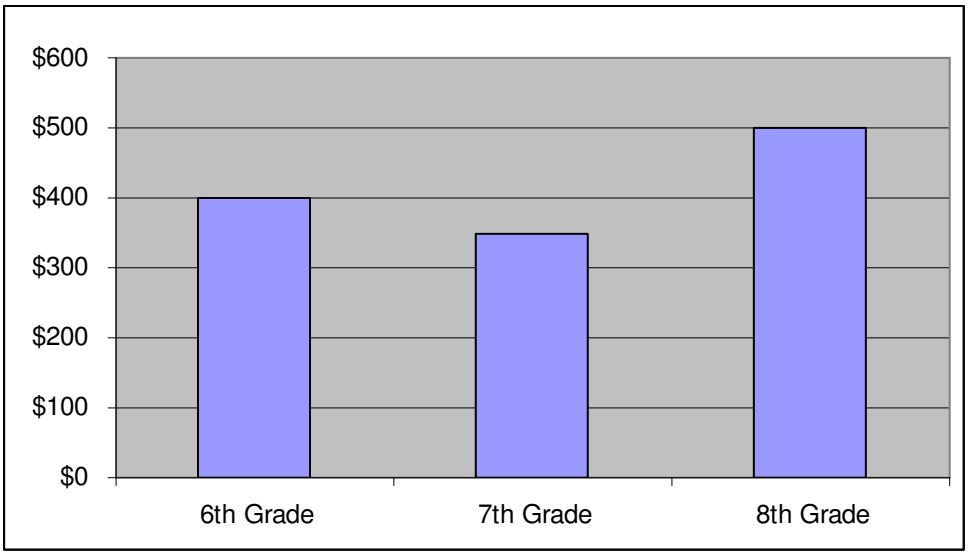


Figure 2: Bar Graph Example

Example 2**:**

A survey was taken by Mary outside of Wal-Mart on a Friday afternoon. She asked the first 50 people their ages and recorded their responses in a frequency table.

Age Group	Frequency
0-9	2
10-19	6
20-29	7
30-39	15
40-49	12
50-59	2
60-69	5
70-79	1

1. Using Mary's information, construct a histogram.

2. Most of Wal-Mart's shoppers are between what ages?

3. Could this kind of information be useful to a company?

Solutions:

1. Histogram

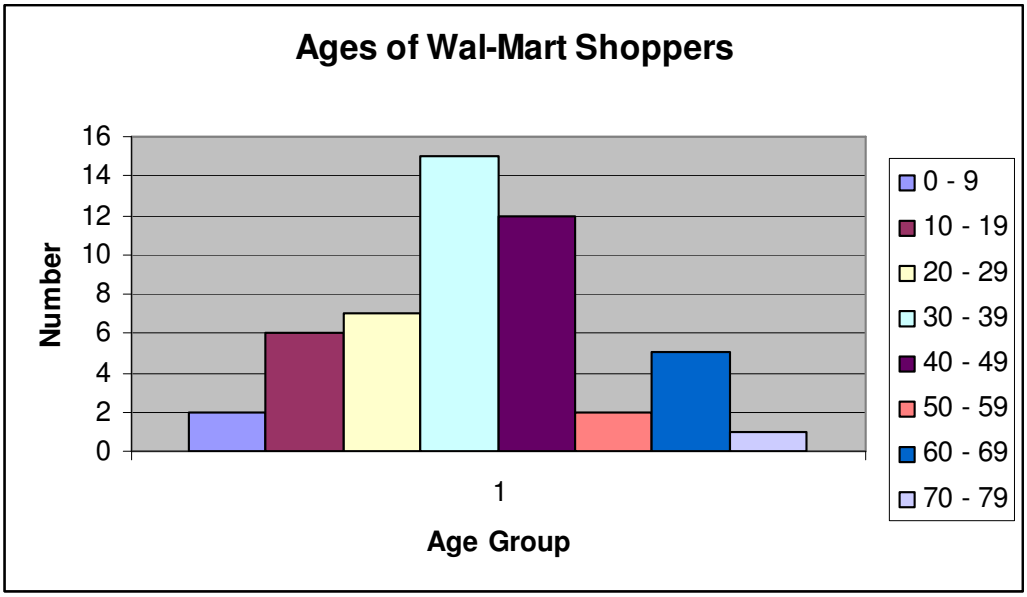


Figure 3: Histogram Example

2. Most of the shoppers are between 30 and 39. This group is followed closely by the 40-49.

3. This information is very useful to companies. It is used in the retail marketing to determine what items the store should stock.

4.3 Application

"Animal Crackers on Parade"

Objective: Using Animal Crackers, the student will create a bar graph with the number of each type of animal in the box.

Materials**:**

1. One regular size box of animal crackers for two people.
2. Paper
3. Colored Pencils

Procedure**:**

1. Divide the students into groups of two.
2. Give each pair a box of crackers.
3. Next, have them count the number of each animal in the box, ignoring the broken ones.

Assignment:

1. Students should make a table with the results.
2. Construct a bar graph using the table. Be sure to label and color the graph.
3. Does each animal appear the same number of times?
4. Could this information have been displayed in a histogram?

Solution**:**

1. Answers will vary depending upon the box of crackers.
2. The graph should match the table that is created in answer one.
3. It is highly unlikely that the animals appear in the same number.
4. This information does not involve intervals**,** so it could not be displayed in a histogram.

4.4 Homework Exercises

Johnson Middle School
Seventh Grade's Favorite Sports.

Sport	Frequency
Volleyball	10
Soccer	15
Baseball	4
Football	12
Basketball	8
Track	1

Using the above table from Johnson Middle School complete the following:

1. Construct a bar graph**.**

The Heights of Sugar Grove Eighth Grade Male Students in Inches

55 60 52 67 50

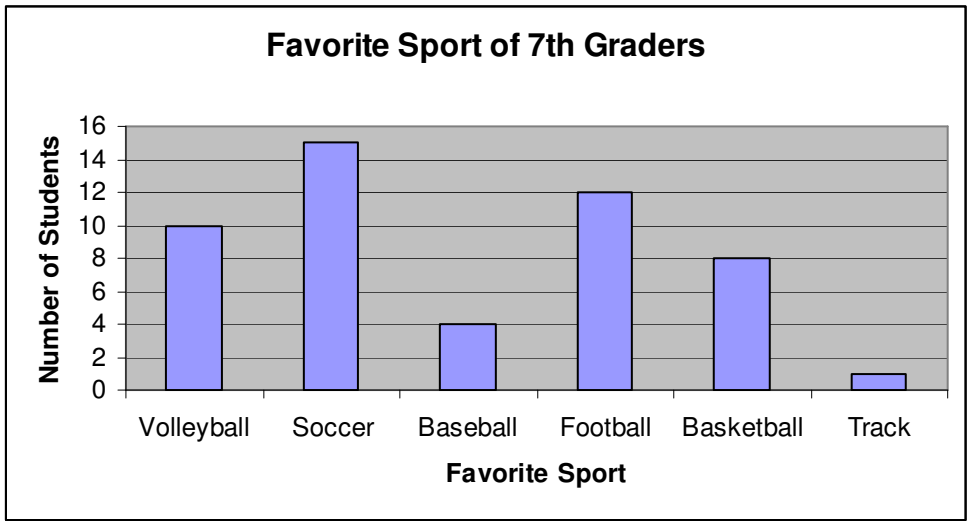
72 69 48 58 65

52 59 60 68 56

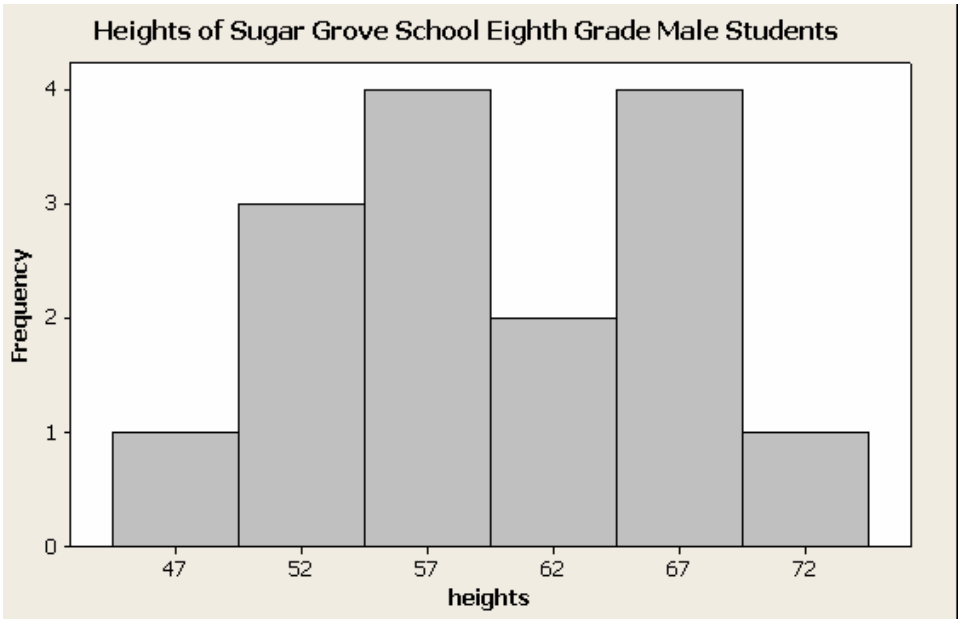
2. Construct a histogram using the heights of the students. Use an interval of 5.
3. Most of the students are in which interval.
4. In your opinion, are most of the students tall?
5. Find the mean, median and mode.

4.5 Answer Key

1. Johnson Middle School (Seventh Grade's Favorite Sport)



2. Heights of Sugar Grove Eighth Grade Male Student in Inches



3. Most of the students fall into the 55-59 and 65-59 interval.

4. Most of the students are of average height, but opinions will vary.

5. The mean is 59.4 inches. The median is 59 inches. The modes are 60 and 52.

5 CIRCLE GRAPHS

5.1 Introduction

A Circle Graph is a type of graph used to compare parts to the whole. It is often known as a Pie Graph. Each section of the circle represents a percent of the whole.

Virginia Standards of Learning Objective**:**

8.12 The student will use information displayed in line, bar, circle, picture graphs and histograms to make comparisons, predictions and inferences.

8.13 The student will use a matrix to organize and describe data.

Key Terms:

This unit will introduce students to the following terms.

1. Circle Graph
2. Degree

A Circle Graph is a type of statistical graph that displays parts in comparison to the whole.

A Degree is the most common unit of measurement for an angle. A circle has 360 degrees.

Materials Needed:

1. Calculator
2. Protractor
3. Compass
4. Paper
5. Colored Pencils

5.2 In Class Examples

Example 1:

Middle School Students Favorite Sports

(25 people surveyed)

Sport	Number of Votes
Volleyball	5
Basketball	8
Football	6
Softball	3
Soccer	1
Track	2

1. Construct a Circle Graph using the information in the table.

Solution Step by Step**:**

1. On your calculator, total the number of votes. The total is 25.

2. Next, divide the number of votes each sport received by the total number of votes. For example, for volleyball divide 5 by 25=.20

3. Multiply each of those answers by 360. This will give you the size of the area on the circle that each sport will have. Using a compass, draw a circle with a diameter of approximately 3 inches. Using the protractor, measure each angle using the center of the circle as the vertex for the angle. Color each section a different color for maximum visual comparison.

Sport	Number of Votes	Decimal	Degree
Volleyball	5	.20	72
Basketball	8	.32	115.2
Football	6	.24	86.4
Softball	3	.12	43.2
Soccer	1	.04	14.4
Track	2	.08	28.8
Total	**25**	**1.00**	**360**

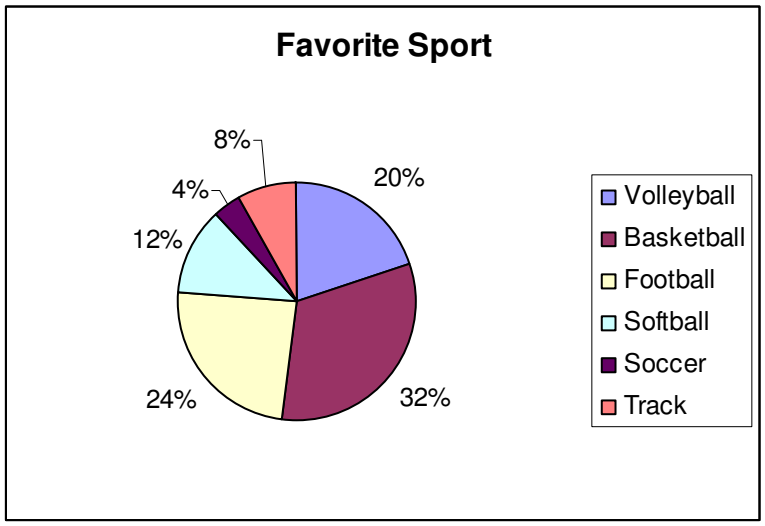


Figure 4: Circle Graph Example

Example 2:

On Saturday, John went to the beach for the day. He spent $8.00 for goggles, $9.25 for a hat, $10.50 for lunch, $11.75 for a souvenir and $13.00 for a T-shirt.

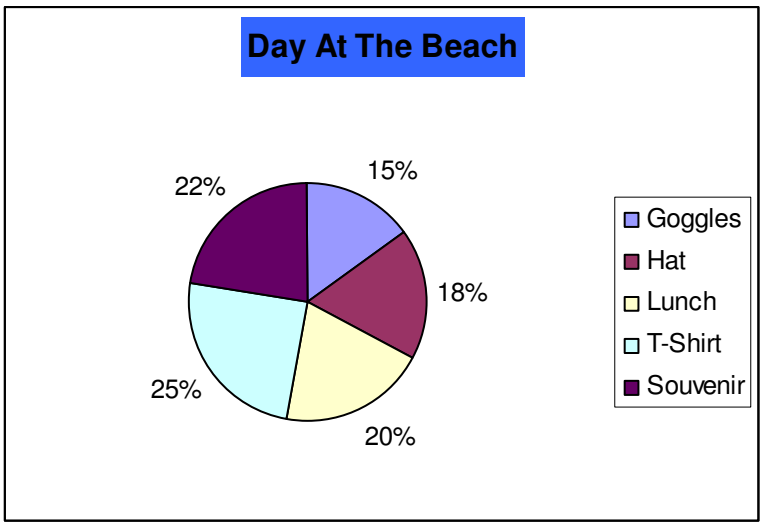
1. Construct a table showing his expenses for the day.
2. Construct a circle graph to compare those expenses.

Solution:

1. Table

Item Purchased	Amount	Decimal	Degree
Goggles	$8.00	0.15	54
Hat	$9.25	0.18	64.8
Lunch	$10.50	0.20	72
T-shirt	$13.00	0.247	88.92
Souvenir	$11.75	0.223	80.28
Total	$52.50	1.00	360

2. Circle Graph



5.3 Application

Objective**:** Using M & M's, the students will do an activity and create a circle graph.

Materials**:**

1. One regular size bag of M & M candies (plain)
2. Paper
3. Colored Pencils

Procedure:

1. Give each student a bag of candy.
2. Each student should make a chart similar to the one below and record the number of each color present in the bag.

Color	Number
Brown	
Orange	
Red	
Blue	
Green	
Yellow	

3. Using the information obtained from the table, the students should create a circle graph.

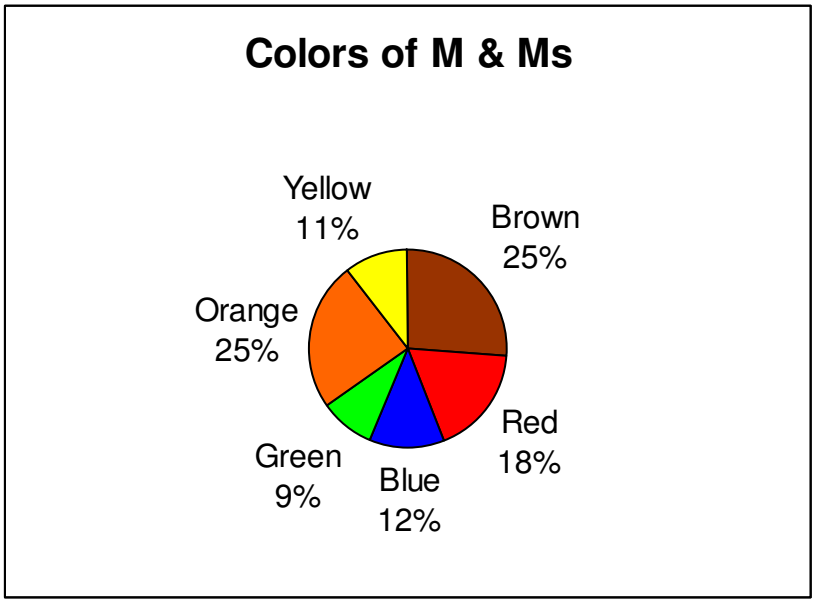
Solution:

Tables will vary due to color and numbers.

Example:

Color	Number	Decimal	Degree
Brown	15	0.25	90
Orange	14	0.25	90
Red	10	0.18	64.8
Blue	7	0.12	43.2
Green	5	0.09	32.4
Yellow	6	0.11	39.6
Total	57	1.00	360

The circle graph would look like:



5.4 Homework Exercises

Signer of the Declaration of Independence and Their Birthplaces [8]

Location	Number of Signers
Connecticut	5
Delaware	2
Maine	1
Maryland	5
Massachusetts	9
New York	3
New Jersey	3
Pennsylvania	5
Rhode Island	2
South Carolina	4
United Kingdom	8
Virginia	9

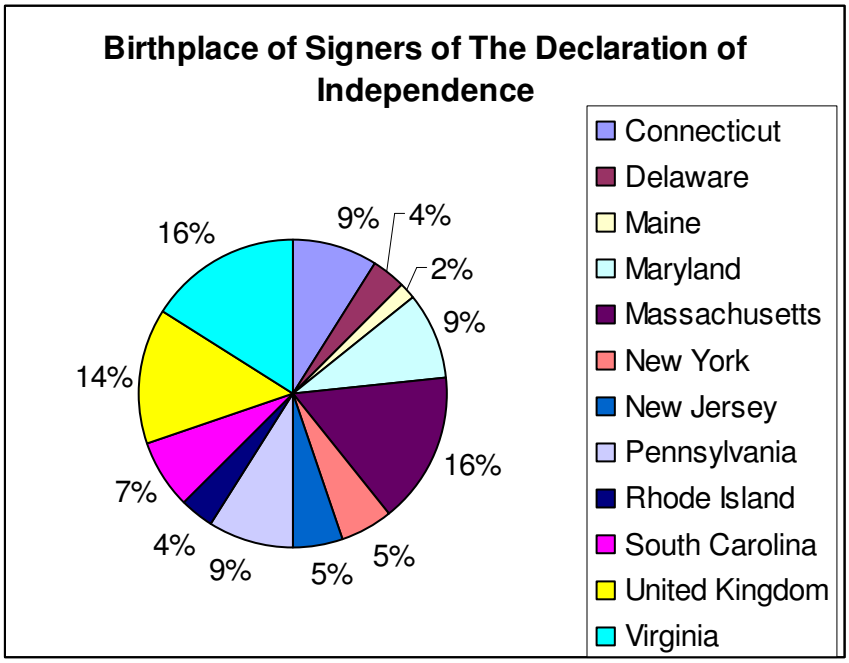
Assignment:

1. Construct a table showing your decimal ratio and degrees.
2. Construct a circle graph.

Location	Number of Signers	Decimal	Degree
Connecticut	5	0.09	32.4
Delaware	2	0.04	14.4
Maine	1	0.02	7.2
Maryland	5	0.09	32.4
Massachusetts	9	0.16	57.6
New York	3	0.05	18
New Jersey	3	0.05	18
Pennsylvania	5	0.09	32.4
Rhode Island	2	0.04	14.4
South Carolina	4	0.07	25.2
United Kingdom	8	0.14	50.4
Virginia	9	0.16	57.6
Total	56	1.00	360

1. Construct a table showing decimal ratios and degrees.

2. Construct a circle graph.



6 BOX-AND-WHISKER PLOTS

6.1 Introduction

The Box-and-Whisker Plot, often called the Box Plot, is a diagram that summarizes data using the median, the upper and lower quartiles, and the upper and lower extremes (or minimum and maximum values). It is called a Box-and-Whisker Plot because a box is drawn around the quartiles and whiskers or lines are extended to the extremes. The median is marked within the box.

Virginia Standards of Learning [5]:

7.17 The student will display data, using frequency distributions, line plots, stem-and-leaf plots, box-and-whisker plots and scatter grams.

8.12 The student will use information displayed in line, bar, circle, picture graphs and histograms to make comparisons, predictions, and inferences.

Key Terms:

1. Variation
2. Quartile
3. Interquartile Range
4. Upper Quartile
5. Lower Quartile
6. Outlier
7. Lower Extreme (Minimum Value)
8. Upper Extreme (Maximum Value)

Key Terms Defined

1. Variation is the spread in values of the data.
2. Quartiles are values that divide the data into four equal parts.
3. Upper Quartile has three-fourths of the observations below it.
4. Lower Quartile has one-fourth of the observations below it.
5. Interquartile Range is the range of the middle half of the data or the Upper Quartile minus the Lower Quartile.
6. A point is suspected to be an Outlier if it lies more than 1.5 times the interquartile range from the upper or lower quartiles
7. Lower Extreme is minimum value in the set of data.
8. Upper Extreme is the maximum value in the set of data.

Materials:

1. Calculator
2. Ruler
3. Pencil
4. Paper

6.2 In Class Examples

Example 1:

6^{th} Grade's Spelling Test Grades

78 76 81 83 78

72 83 85 89 84

90

1. Find the median.
2. Find the upper and lower quartiles
3. Find the upper and lower extremes.
4. Find the interquartile range.
5. Construct a Box-and-Whisker Plot.

Solution:

First, list the data in numerical order.

72 76 78 78 81 83 83 84 85 89 90

1. The median is the center number when the data is in numerical order. It is 83.
2. The upper and lower quartiles are found by finding the median of the upper half and lower half of data. The center of the upper half is 85. 85 is the upper quartile. The center of the lower half is 78. 78 is the lower quartile.

3. The upper and lower extremes are merely the maximum value and minimum value. The maximum is 90, upper extreme. The minimum is 72, lower extreme.
4. The interquartile range is found by subtracting the lower quartile from the upper quartile. This would be 85-78= 7.
5. Construct a Box-and-Whisker Plot**.**

Sixth Grade Spelling Scores

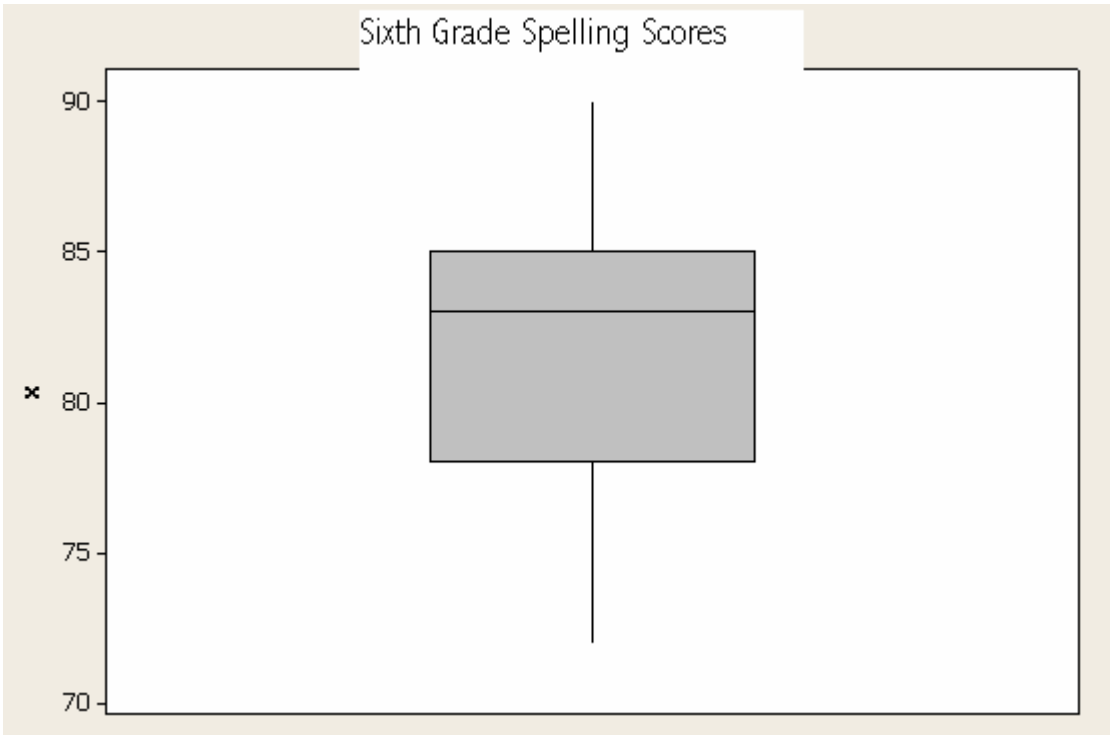


Figure 5: Box Plot Example

6.3 Application

This list shows the ages of the Academy Award winners for the past 25 years [8].

29 31 31 35 36 36 37

37 38 39 41 41 42 44

47 48 50 51 52 53 53

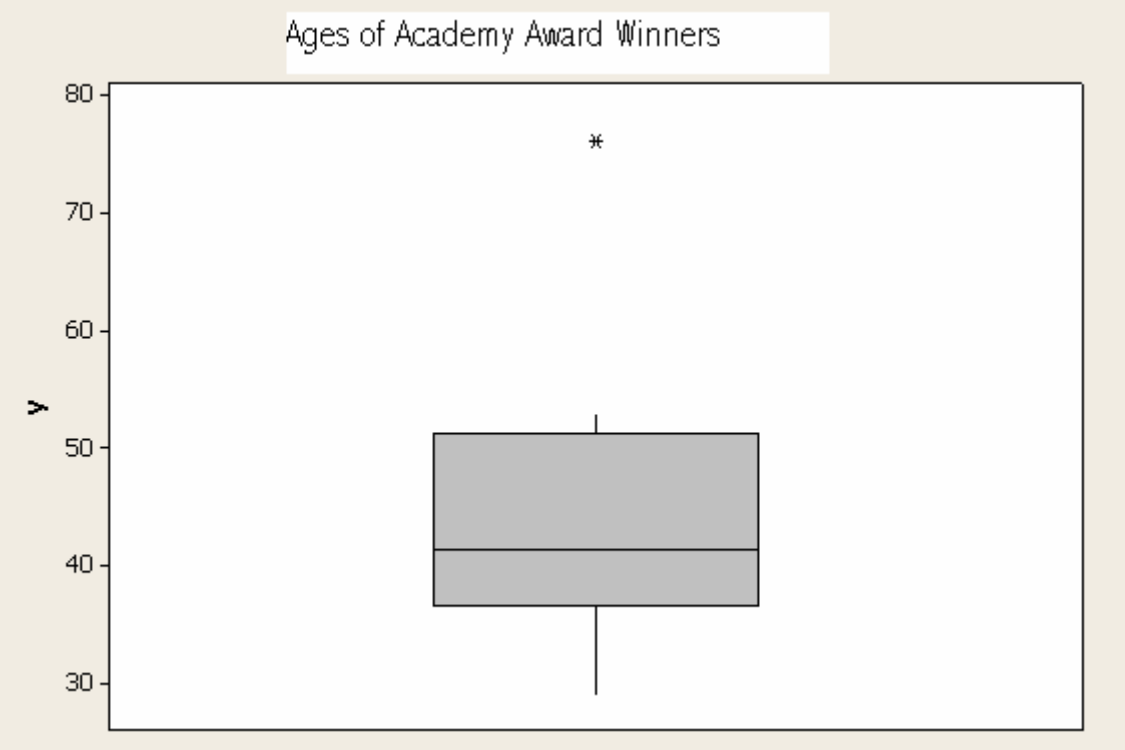
55 60 61 76

Assignment:

1. Find the median.
2. Find the upper and lower quartiles.
3. Find the upper and lower extremes.
4. Find the interquartile range.
5. Construct a Box-and-Whisker Plot.

First, check for numerical order. The ages are in numerical order from least to greatest.

1. The median is 42.
2. The upper quartile is 52.5; the lower quartile is 36.5.
3. The upper extreme is 76; the lower extreme is 29.
4. The interquartile range is 16.
5. Box-and-Whisker Plot:



6.4 Homework Exercises

Jose surveyed 14 people as they left the grocery store to find the number of hours they spent each week watching television.

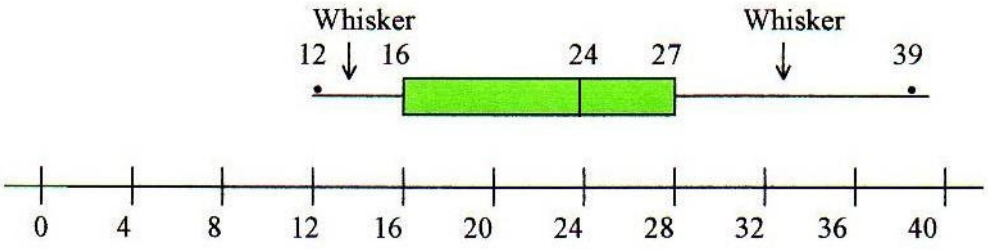
Hours per week: 20, 23, 14, 22, 39, 26, 26, 26, 16, 12, 27, 27

Assignment:

1. Find the median.
2. Find the upper and lower extremes.
3. Find the upper and lower quartiles.
4. Find the Interquartile range.
5. Construct a Box and Whisker Plot.

6.5 Answer Key

1. The median is 24.
2. The upper and lower extremes are 12 and 39.
3. The upper and lower quartiles are 16 and 27.
4. The interquartile range is 9.
5. Construct a Box and Whisker Plot.



7 SCATTER PLOTS (SCATTER GRAMS)

7.1 Introduction

A Scatter Plot or Scatter Gram is a graph that shows the general relationship between two quantitative variables measured on the same individuals.

Virginia Standards of Learning [5]:

7.17 The student will display data, using frequency distributions, line plots, stem-and-leaf plots, box-and-whisker plots and scatter grams.

Key Terms:

1. Scatter Plot or Scatter Gram
2. Positive Relationship
3. Negative Relationship
4. Linear Relationship

Scatter Plots show the relationship between two sets of data. This relationship can be a positive relationship, negative relationship, or linear relationship. When the ordered pairs are plotted on a graph, if the line slopes upward the relationship is positive. If the line has a downward slope, the relationship is negative. A linear relationship is when the points on a scatter plot show a straight line pattern. If no line can be drawn, then the two sets of data have no relationship.

7.2 In Class Examples

Have you ever wondered if the amount of time one spends studying has an affect on the test score?

Let's decide if studying longer is related to your grade.

Mr. Smith's Algebra students studied for their exam. The length of time each spent studying is recorded in the chart along with the grade each received. What is the relationship between study time and test scores for Mr. Smith's class?

Study Hrs.	**Test Scores**
3	80
5	90
2	75
6	80
7	90
1	50
2	65
7	85
1	40
7	100

First, plot the ordered pairs on graph paper. The ordered pairs are (3,80), (5,90), (2,75), (6,80), (7,90), (1,50), (2,65), (7,85), (1,40), (7,100).

Reminder: Just because a set of data shows a particular relationship, this does not mean that the relationship will be the same for all sets of data. It depends on the data being examined.

If the slope of the line rises from left to right, there is a positive correlation between the number of hours studied and the grade. If the slope of the line falls from left to right, there is a negative correlation between the number of hours studied and the grade.

In Mr. Smith's class, there is a positive association between study time and scores on the exam.

Hours Studied and Grades

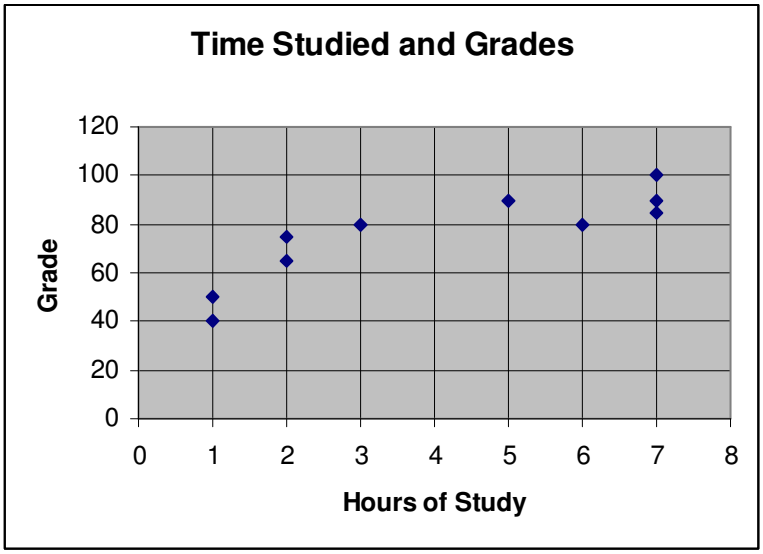


Figure 6: Scatter Plot Example

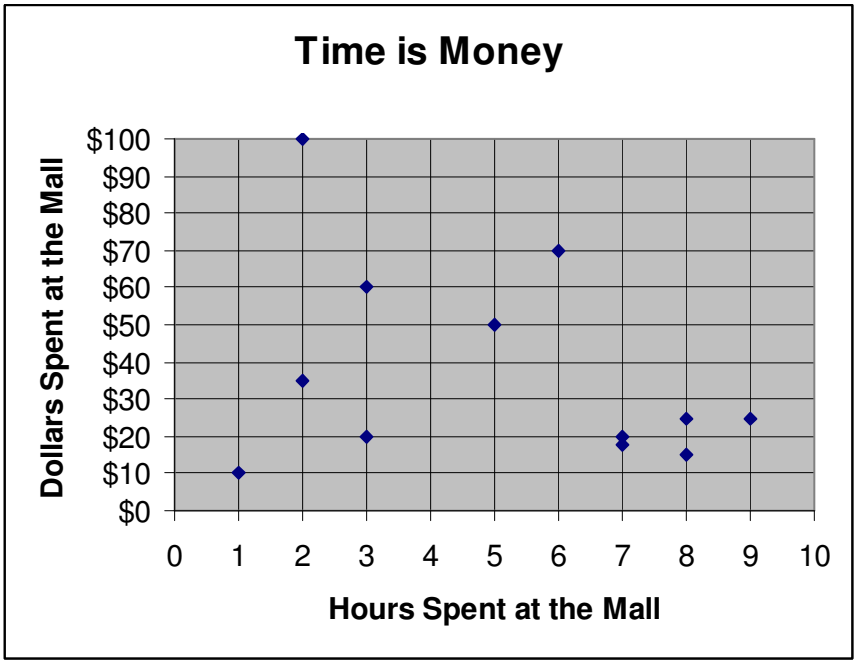
7.3 Application

What relationship exists between the number of hours spent in a Mall and the amount of money spent?

Judy surveyed 12 people leaving the Bristol Mall to determine if the amount of money spent and the time in the Mall had a positive correlation. Here are her results:

Dollars Spent	Hrs. Spent
$10	1
$20	3
$100	2
$60	3
$20	7
$35	2
$70	6
$25	8
$15	8
$18	7
$25	9
$50	5

Solution to Application



After plotting the ordered pairs, there is no relationship between the dollar amounts spent and the hours spent at the Mall. (This conclusion is for Judy's set of data and may be different for another survey.) We note that if the outlying point (2, 100) is removed from the graph we would see a curved relationship. An outlier is an observation that lies outside the overall pattern and it may influence your conclusion.

7.4 Homework Exercises

What relationship exists between a person's shoe size and their age?

Assignment:

1. Survey 20 people to obtain their age and their shoe size. Plot these on the scatter plot. The x-axis will be their age and the y-axis will be their shoe size. The shoe sizes may include half sizes.
2. Determine the relationship.
3. Would this relationship change from group to group? If so, why?
4. What could cause outliers?

7.5 Answer Key

1. Scatter Plot will vary due to differences in people surveyed.
2. There will probably not be a relationship. Many people with the same age wear different sizes of shoes.
3. The data will change from survey to survey, but the relationship probably will not.
4. Outliers could be caused by a young person wearing a large shoe or by an older person wearing a very small shoe.

8 STEM AND LEAF PLOTS

8.1 Introduction

A stem-and-leaf plot is a display that organizes data to show its shape and distribution. A stem plot looks like a histogram turned on end. The goal of this lesson is to have students be able to construct and interpret stem-and-leaf plots.

Virginia Standards of Learning Objective:

7.17 The student will display data, using frequency distributions, line plots, stem-and-leaf plots, box-and-whisker plots and scatter grams.

Key Terms:

This unit will introduce students to the following terms:

1. Stem: All digits to the left of the last digit
2. Leaf: The last digit in increasing order out from the stem.
3. Stem-and-Leaf Plot is a display that organizes data to show its shape and distribution.

8.2 In Class Examples

Example 1:

Test Scores on a 7th Grade English Test:

64, 82, 85, 99, 96, 81, 97, 80, 81, 80, 84, 87, 98, 75, 86, 88, 82, 78, 81, 86, 80, 50, 84, 88, 83, 82

The first step is to place the grades in order from least to greatest.

50, 64, 75, 78, 80, 80, 80, 81, 81, 81, 82, 82, 82, 83, 84, 84, 85, 86, 86, 87, 88, 88, 96, 97, 98, 99

1. Create your graph by using the last number as the leaves and the first number for the stem.
2. Find the median, mode and range.

Solution:

1. Create the graph**:**

7th Grade English Test Scores

Stem	Leaf
5	0
6	4
7	58
8	000111222344566788
9	6789

Figure 7: Stem-and-Leaf Plot Example

2. The median is 82. The modes are 80, 81 and 82. The range is 49.

Example 2:

Math test scores:

70 72, 76, 80, 82, 82, 88, 90, 90, 94, 98, 99, 100, 100, 100

These are already in numerical order.

1. Create a stem-and-leaf plot of the test scores.
2. Find the median, mode and range.

Solution:

1. Stem-and-Leaf Plot

Stem	Leaf
7	026
8	0228
9	00489
10	000

7 0 represents 70

2. The median is 90. The mode is 100. The range is 30.

8.3 Application

Below is a table of data from a middle school class showing the amount of water used by each student in one week.

Students' Water Use

Student initials	Gallons per student
RE	380
TW	420
HW	299
WE	334
GK	266
DJ	218
MJ	246
WD	246
MA	241
LR	206
FP	247
HA	197
TB	313
CH	188
ME	231
JW	228
PR	211
NP	273
BH	202
EB	189
PJ	182
HJ	160
HM	185
JZ	247

Assignment:

1. Make a stem-and-leaf plot of the gallons used per student.
2. Find the mean number of gallons of water used in a week by each student. Use this information to estimate the number of gallons of water used by all 260,000,000 Americans in a week.
3. How do the mean and median number of gallons compare?

Application Solution

1. Stem-and-Leaf Plot

Stem	Leaf
16	0
17	
18	2589
19	7
20	26
21	18
22	8
23	1
24	16677
25	
26	6
27	3
28	
29	9
30	
31	3
32	
33	4
34	
35	
36	
37	
38	0
39	
40	
41	
42	0

2. The mean gallons of water used by each student is 246.2 gallons per week. All 260,000,000 Americans would use approximately 64,012,000,000 gallons of water per week based on the amount used by the students.

3. The median gallons of water used by the students is 236 gallons. The mean is higher by ten gallons.

8.4 Homework Exercises

Suppose you and your best friend are in different math classes but have the same teacher. On the last quiz, both classes had the same median, mean and mode**.** Does this mean that the grades were identical?

The test scores for the two classes are:

Class 1: 35, 72, 98, 28, 38, 50, 60, 89, 96, 20, 45, 88, 60, 75, 91

Class 2: 43, 52, 55, 60, 60, 72, 78, 85, 45, 81, 75, 68, 60, 56, 54

Assignment:

1. Construct a stem-and-leaf plot for each class.
2. Compare the two plots. Were the grades identical?

8.5 Answer Key

1. Construct a **stem-and-leaf plot** for each class.

Class 1

Stem	Leaf
2	08
3	58
4	5
5	0
6	00
7	25
8	89
9	168

Class 2

Stem	Leaf
2	
3	
4	35
5	2456
6	0008
7	258
8	15
9	

2. The grades were not identical. For both classes, the modes and the medians were 60. The mean was 63 even with the grades being different. The range for class one is 78 and class two is 42. The mean or median do not tell the whole story. Shape, center and spread describe the overall pattern. Be sure to look for outliers.

9 LINE PLOTS

9.1 Introduction

A line plot is a type of graph that is used to summarize quantitative variables usually fewer than 50 data values. This graph consists of a horizontal number line with each value represented by a letter, usually an x, above the corresponding value on the number line. The number of x's above each value denotes the number of times that value occurred [7].

Virginia Standards of Learning Objective [5]:

7. 17 The student will display data, using frequency distributions, line plots, stem-and-leaf plots, box-and-whisker plots and scatter grams.

7.18 The student will make inferences and predictions based on the analysis of a set of data that the student(s) have collected.

Key Terms:

The student will be introduced to the following terms:

1. Clusters are isolated groups of points.
2. Gaps are large spaces between points
3. Outliers are data points whose values are significantly larger or smaller than other values [8].

9.2 In Class Examples

Example 1:

Elm Street Apartments have 30 tenants. These are their ages:

58, 30, 37, 36, 34, 49, 35, 40, 47, 47, 39, 54, 47, 48, 54, 50, 35, 40, 38, 47, 48, 34, 40, 46, 49, 47, 35, 48, 47, 46

The first step is to place the data in numerical order.

30, 34, 34, 35, 35, 35, 36, 37, 38, 39, 40, 40, 40, 46, 46, 47, 47, 47, 47, 47, 47, 48, 48, 48, 49, 49, 50, 54, 54, 58

1. Create a line plot.

2. Find any gaps, clusters and outliers.

Solution to Example 1:

1.

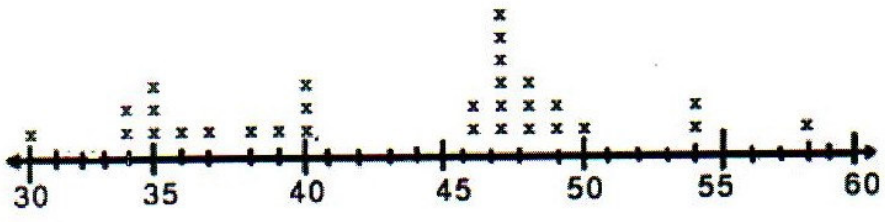


Figure 8: Line Plot Example

2. The large spaces between 41 and 45, 31 and 33, 51 and 54, 55 and 57 are gaps. The largest group of data that lies from 46 to 50 is a cluster. The outliers are 30 and 58.

Example 2:

The ages of students at Lincoln Middle School are as follows:

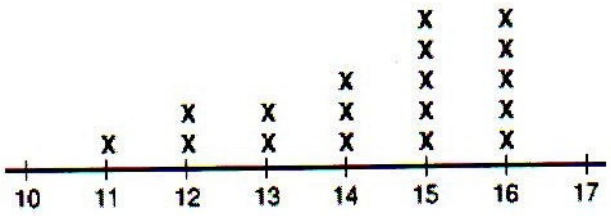
11, 12, 12, 13, 13, 14, 14, 14, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16

Assignment:

1. Construct a line plot.
2. Find any gaps, outliers or clusters.

Solution:

1. Line Plot for Lincoln Middle School



2. There are no gaps or outliers**.** The spread of the data is from 11 to 16. The shape of the distribution is skewed to the left, with no outliers nor clusters.

9.3 Application

The National Football League began choosing its champion in the Super Bowl in 1967.

The list shows the margin of victory and the winning league for all 40 Super Bowl Games [8].

A=AFC, N=NFC

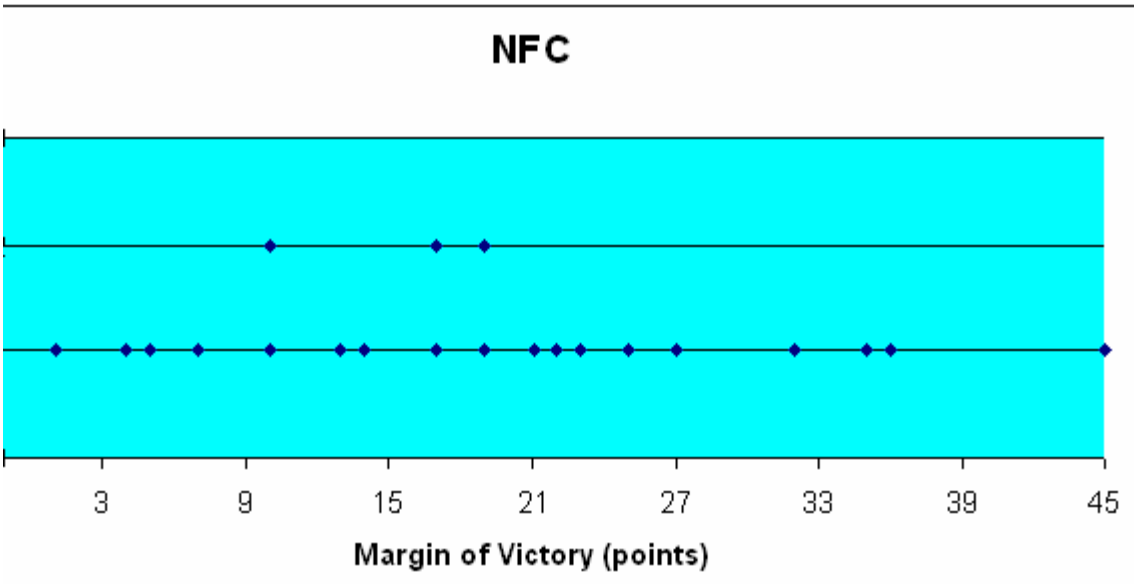
25-N	19-N	9-A	16-A	3-A	21-N	7-A	17-A
10-A	4-A	18-A	17-N	4-A	12-A	17-A	5-N
10-N	29-A	22-N	36-N	19-N	32-N	4-N	45-N
1-N	13-N	35-N	17-N	10-N	23-N	14-N	7-A
15-A	7-N	27-A	3-A	27-N	3-A	23-A	11-A

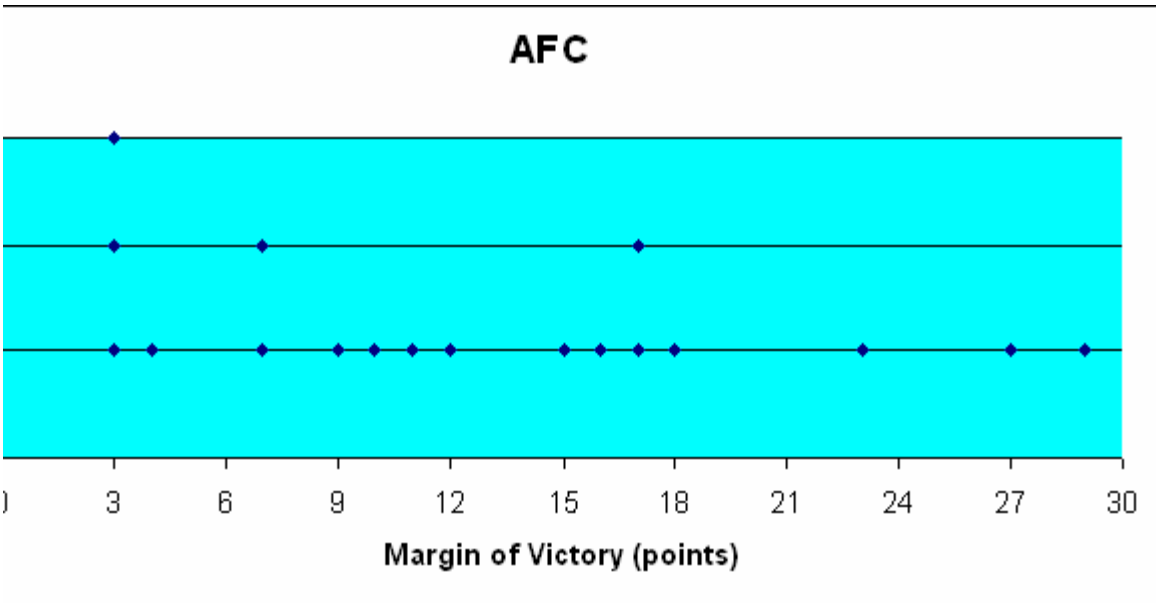
Assignment:

1. Make line plots using the data.
2. What do you observe about the winning margins?

Answer Key:

1. Line Plots for NFC and AFC





2. The National Football Conference has won more Super Bowls. The margins tend to be high, with the NFC winning by larger margins than the AFC teams.

9.4 Homework Exercises

Make a line plot for each set of data.

1. 4, 6, 4, 5, 6, 6, 3, 11, 6, 7
2. 20, 25, 15, 30, 35, 20, 25, 15, 25, 30
3. 85, 75, 80, 55, 70, 70, 70

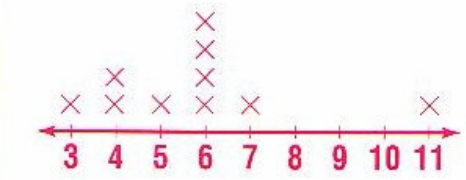
Pamela's allowance is $6 per week, and she wants to ask her parents to increase it. She surveyed other students in her class. The results are shown below:

$15, $0, $5, $10, $3, $4, $10, $6, $5, $10, $0, $6, $8, $8, $15

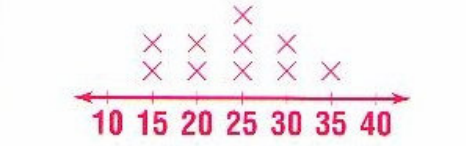
4. Make a line plot of the survey results.
5. Find the mean, median and mode.
6. Which should Pamela use to argue that she needs more allowance?

Make a line plot for each set of data.

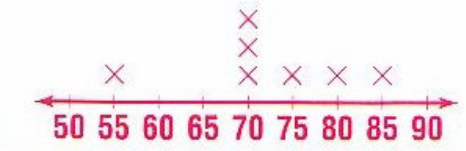
1. 4, 6, 4, 5, 6, 6, 3, 11, 6, 7



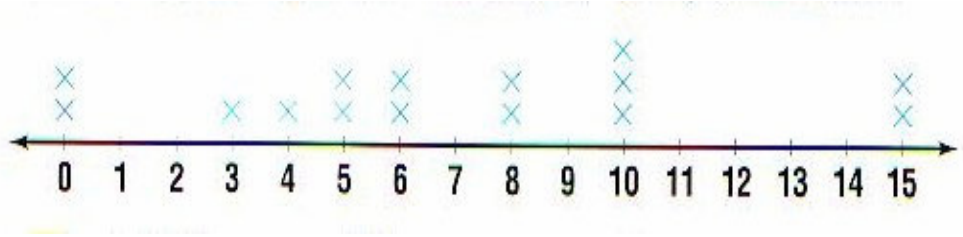
2. 20, 25, 15, 30, 35, 20, 25, 15, 25, 30



3. 85, 75, 80, 55, 70, 70, 70



4. Pamela's survey results on a line plot.



5. The mean is $7. The median is $6. The mode is $10.

6. Pamela would probably want to use the mode to argue for a larger allowance.

10 LINE GRAPHS

10.1 Introduction

A line graph is similar to a scatter plot. On a line graph, points are graphed and then connected by using a line. The purpose of a line graph is to show changes over a time period or to compare sets of data.

Virginia Standard of Learning Objective [5]:

8.12 The student will use information displayed in line, bar, circle, picture graphs and histograms to make comparisons, predictions and inferences.

Key Terms:

This unit will introduce the student to the following terms:

1. line graph
2. independent variable
3. dependent variable
4. horizontal axis (x-axis)
5. vertical axis (y-axis)

10.2 In Class Examples

Example 1:

John wishes to purchase a pizza. He wants to compare the prices of pizza according to the number of toppings. The prices are in the chart below:

Number of Toppings	Pizza Price
1	$8.00
2	$9.50
3	$10.50
4	$11.75
5	$13.00

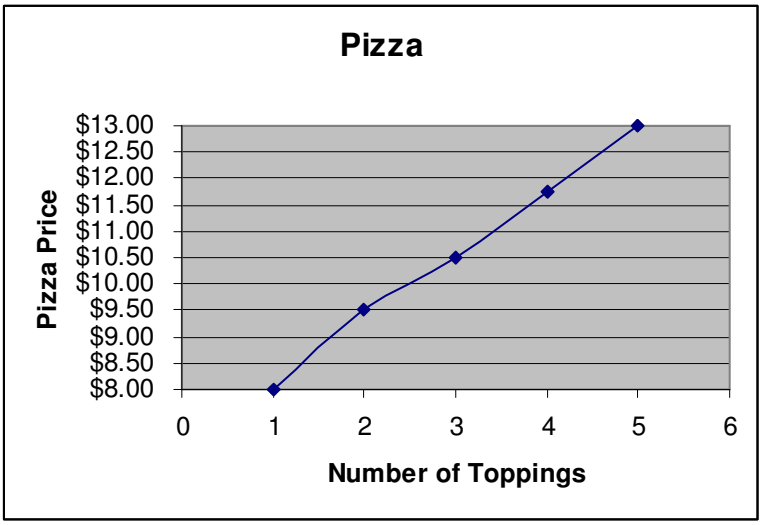
John knows that the price of the pizza is dependent upon the number of toppings. The price is the dependent variable and the number of toppings is the independent variable.

The dependent variable is graphed on the vertical axis (y-axis). The independent variable is graphed on the horizontal axis (x-axis).

Assignment:

1. Construct a line graph of the pizza information.

Solution:



The price of pizza depends upon the number of toppings.

Example 2:

Use a line graph to predict the time it takes Mary to deliver newspapers in the morning. Construct a line graph and predict the total time Mary needs to deliver 50 papers. Determine the dependent variable and the independent variables.

Number of Papers	Time (Min.)
5	12
10	23
15	37
20	48
25	60
30	71
35	85

Solution:

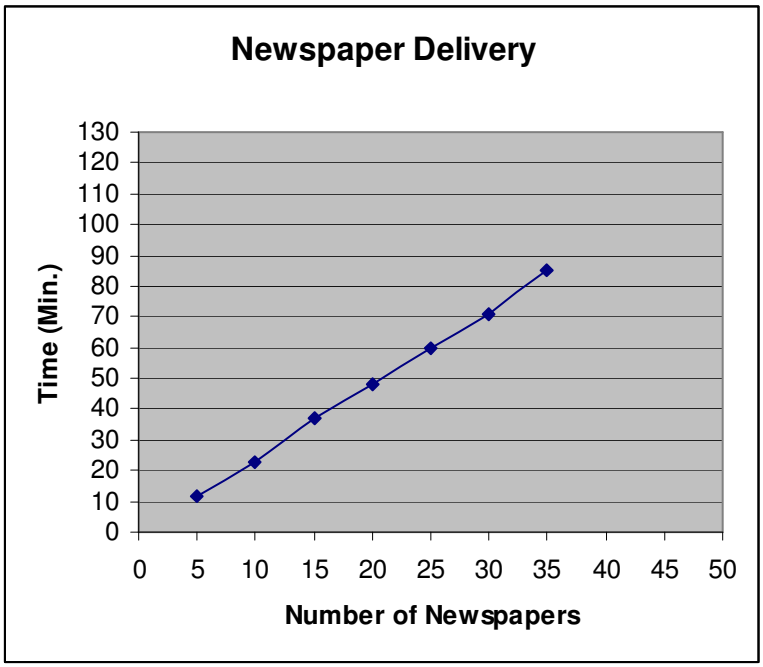


Figure 9: Line Graph Example

The dependent variable is the time and the independent variable is the number of newspapers. To predict the amount of time it would take to deliver 50 papers, extend the line on the graph. It would take Mary about 120 minutes to deliver 50 newspapers.

10.3 Application

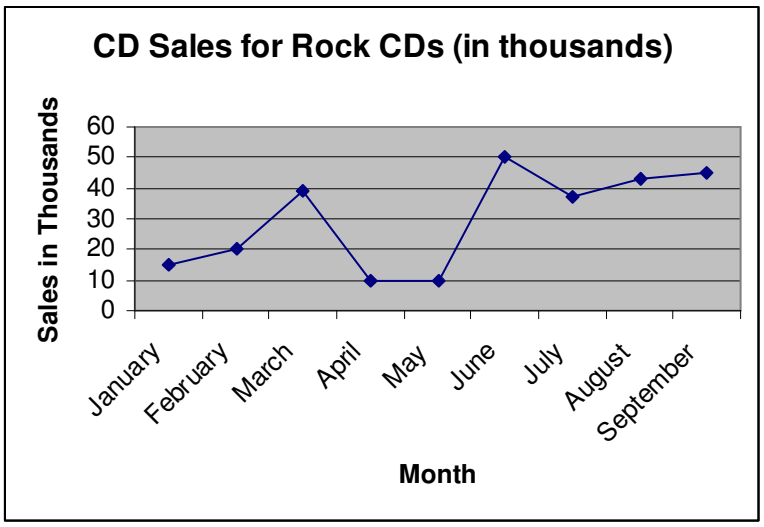
RCA wants to compare its sales of Rock CDs over the past 9 months. The information is displayed in thousands of copies.

January	February	March	April	May	June	July	August	September
15	20	39	10	10	50	37	43	45

1. Using the above information, construct a line graph to display the changes over the nine month period.
2. What month had the greatest sales?
3. What was the amount of sales for that month?
4. Sales dropped to its lowest during what month?
5. From May to June, how much did sales increase?

Answer Key for Application

1. Line Graph



2. The greatest sales were during the month of June.

3. The company sold 50,000 CDs during the month of June.

4. The sales dropped to its lowest during the month of April to May.

5. Sales increased 40,000 from May to June.

10.4 Homework Exercises

Age	**Height (cm)**
5	76
16	170
21	180
30	181
49	180
65	177
75	174

Assignment:

Section 1

1. Determine the dependent and independent variables.
2. Construct a line graph.
3. How does a person's height change according to age?
4. At what age, does a person's height begin to decrease?
5. At age 80, what do you predict the height will be?

Section 2

For questions 6-8, <u>underline</u> the dependent variable.

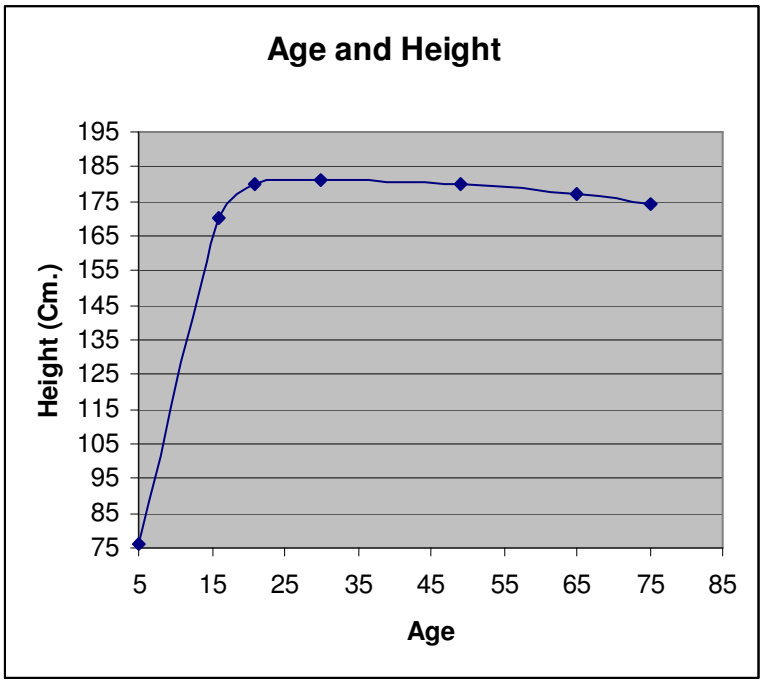
6. City Temperature
7. Number of Moons Planet
8. Height Mountain

10.5 Answer Key

Section 1

1. Since the height depends upon the age, the height is the dependent variable and the age is the independent variable**.**

2. Construct a line graph.



3. How does a person's height change according to age? (Answers will vary.)

A person's height increases until age 30 and then it decreases slightly.

4. At what age does a person's height begin to decrease? According to the graph, height begins to decrease around the age of 30.

5. At age 80, what do you predict the height will be? Answers will vary, but approximately 170 cm.

Section 2

Underline the dependent variable.

6.	City	<u>Temperature</u>
7.	<u>Number of Moons</u>	Planet
8.	<u>Height</u>	Mountain

11 ANALYZING GRAPHS

11.1 Introduction

A case for teaching statistics in schools, as early as possible, has been set forth by numerous textbook authors [3]. These authors have pointed out that people need to learn about statistics just to be able to make reasonable decisions as consumers, voters, or even when choosing a career. It is difficult to be well informed without a good background knowledge of statistics. In order to develop adequate statistical literacy in our students, statistics should be taught as early in school as possible. Misuse of statistics can affect the human decision making process, which in turn affects the course of human lives.

Virginia Standards of Learning [5]:

7.17 The student will make inferences and predictions based on the analysis of a set of data that the student(s) collected.

7.19 The student will represent, analyze, and generalize a variety of patterns, including arithmetic and geometric sequences with tables, graphs, rules and words in order to investigate and describe functional relationships.

8.12 The student will use information displayed in line, bar, circle, picture graphs and histograms to make comparisons, predictions and inferences.

Key Terms:

This unit will introduce the student to the following terms.

1. analyze
2. predict
3. inference

1. Analyze is to take apart in order to see how certain information fits together.
2. To predict is to foretell beforehand, or to make a prediction about the future from analyzing data.
3. An inference is a conclusion based on facts and reasoning gained from data.

11.2 In Class Examples

Example 1:

The manager of a computer store analyzed software sales by category for the past month. The results are shown in the table below. If the manager plans to order 150 software packages for the upcoming month, how many of them should be in the home productivity category?

Software Sales	
Category	Percent of Sales
Games	24
References	13
Education	17
Home Productivity	25
Business	21

Solution:

By analyzing the information displayed in the table, the manager should purchase 38 home productivity software packages. This answer was found by finding 25% of 150.

Example 2:

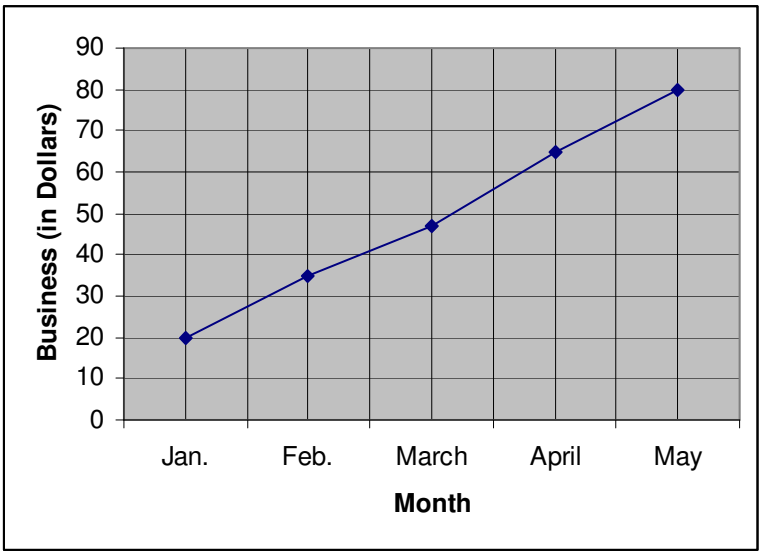
Ben randomly surveyed one hundred people in his city about their main type of exercise. Of those asked, 21% walk or jog to stay fit. If 65,000 people live in Ben's city, how many people could be expected to walk or jog for exercise?

Solution:

If 21% walk or jog to stay fit, then 21% of 65,000 people could be expected to walk or jog for exercise. The answer would by 21% of 65,000 or 13,650 people.

Example 3:

Using the following graph, how much money do you predict Classic Comics will make on the Internet in July?



Solution:

If the trend continues, you can infer from the graph that the company will probably make $100-$105. This is if, and only if the trend continues.

11.3 Application [2]

The table below gives data for the 15 North Carolina counties with the greatest 1993 hog populations. For each county, the table gives the 1993 hog population, the percent growth from 1983-1993 and the number of hogs per square mile. Hog farming is an important business in North Carolina. In fact, one area of the state is known as the hog belt.

In areas with large populations of people and animals, finding ways to dispose of waste can be a problem. Any area must deal with the storage, treatment, and disposal of garbage and hazardous chemical wastes.

North Carolina Hog Population

County	1993 hog population	Growth from 1983	Hogs per square mile
Sampson	1,152,000	363%	1218
Duplin	1,041,000	349%	1273
Wayne	333,000	265%	603
Bladen	271,000	1178%	310
Greene	231,000	82%	870
Pitt	193,000	128%	297
Lenoir	159,000	312%	399
Johnston	129,000	61%	163
Robeson	124,000	81%	131
Onslow	115,000	261%	150
Jones	105,000	999%	223
Beaufort	103,000	51%	125
Pender	97,000	588%	111
Halifax	82,000	18%	113
Northampton	81,000	53%	151

Source: U.S. Department of Agriculture, as reported in the *Raleigh News and Observer,* 19–26 February 1995.

Assignment:

1. Write at least three statements comparing the hog data for Johnston County with the hog data for Bladen County.
2. Choose two different pairs of counties from the table. For each pair, write at least three statements comparing the hog data for the two counties.
3. Find the mean and the median for the 1993 hog populations for the 15 counties. How do the mean and median compare?
4. Why is it important for the state to be aware of the growth of the hog population?

Solution:

1. Answers will vary.
2. Answers will vary.
3. The mean of the 1993 hog population is 281,066.7. The median is 129,000. The mean is more than twice the median.
4. Answers will vary. One reason it is important for the state to be aware of the growth of the hog population is to provide the necessary funding for wastewater disposal plants.

11.4 Homework Exercises

The table below shows the change in the "hog inventory" from 1993 to 1994 for the five North Carolina counties with the highest pig populations.

County	1994 hog inventory (millions)	Growth from 1993
Duplin	1.47	41%
Sampson	1.45	26%
Wayne	0.41	23%
Bladen	0.41	50%
Greene	0.28	22%

Note: The hog inventory is the number of hogs and pigs on farms as of December 1, 1994
Source: North Carolina Department of Agriculture.

Assignment:

1. If the percent change for each county was the same from 1994 to 1995 as it was from 1993-1994, how many hogs were in each county in 1995? Show your results in a table.
2. Use the information in the table to figure out what the hog inventory for Bladen County was in 1993. How does this number compare with the total 1993 hog population of Bladen County given in the table?

11.5 Answer Key

1. 1995 Hog Inventory if same growth as from 1993-94.

County	1995 Hog Inventory (millions)
Duplin	2.07
Sampson	1.83
Wayne	.50
Bladen	.62
Greene	.34

2. The hog population in Bladen County in 1993 was approximately .274 million or 274,000. The hog population for Bladen County in 1993 according to Application was 271,000. Ours is a close estimate.

12 MISLEADING GRAPHS AND STATISTICS

12.1 Introduction

Have you ever heard the expression, “Two out of three dentists recommend this product”? Or “… 42% more relief from heartburn”?

Perhaps you suspected these claims were not completely true. It is wise to be suspicious, because statistics and numbers can be manipulated to make anything sound or look believable and convincing. When used improperly, statistics can be manipulated to show just about anything.

There are many ways to misuse statistics. One way is to take nonrandom surveys until you get the desired results. An example would be to survey five families about their favorite grocery store. Suppose you want to claim that 4 out of 5 families prefer Store X. You ask those five families, if you do not get the desired results just forget about that sample and survey five more families. Eventually, you may get the results you want. The results are often misleading, but are they lies? We as consumers must be aware of many tactics to protect ourselves. Statistics is a great tool when used correctly.

There are five types of sampling: Random, Systematic, Convenience, Cluster and Stratified. Random Sampling is analogous to putting everyone’s name in a hat and drawing out several names. Each element in the population has an equal chance of being drawn. If the population is large, this can be difficult to do. Systematic Sampling is easier than random sampling. With this type of sampling, every nth element is chosen from a population. It is similar to lining up children and numbering them “1, 2, 3; 1, 2, 3; etc.” and choosing those numbered 3. Probably the worst technique to use would be to do Convenience sampling.

The surveyor surveys everyone he or she runs into. Cluster sampling is accomplished by dividing the population into groups. This is usually done geographically with every element in the group used. Stratified sampling is also completed by dividing the population into groups called strata, but this time some characteristic is used such as male or female.

We should always ask who is funding the study and why. How was the data collected? Where did the data come from?

The most common way to mislead with data is to cite correct numbers that do not mean quite what they appear to say because we are not told the whole story. An example of this is a news broadcast that reports a snowstorm over the weekend has caused 28 minor traffic accidents. In this small town in good weather, there are usually 48 minor accidents over the weekend. The storm may have prevented 20 accidents. What else are they not telling us?

Virginia Standards of Learning Objective [5]:

7.18 The student will make inferences and predictions based on the analysis of a set of data.

8.12 The student will use information displayed in line, bar, circle, and picture graphs and histograms to make comparisons, predictions, and inferences.

Key Term:

This unit will introduce the student to the following term:

1. Misleading Statistics

Misleading Statistics is a method of manipulation found usually in advertising.

12.2 In Class Examples

Example 1:

A substitute teacher wants to know how the students did on their last math test. She asks only the 10 students sitting in the front of the room to report how they did on the test. Identify the population, the sample and any problems with choosing the sample in this manner.

Solution to example 1:

The population consists of all members of the class. The sample consists of only the 10 students seated in front. Problem: Students seated in front of the room usually show more interest in the class and their scores tend to be higher. These students are not a good representation of the entire class.

Example 2:

The following table shows seven different companies and their prices for detailing a car. Company A claims to have prices that are less than the prices of their competitors. Is this an accurate statement? Why?

Company	Price for Car Detailing
A	$92
B	$91
C	$90
D	$93
E	$93
F	$87
G	$85

Comparing the prices of all Companies, Company A's price is not the lowest although it is lower than the mode. The mode is $93. In this case, the mode could be misleading.

Example 3:

Look at the following two graphs. Both graphs show the high temperature Monday through Friday. Which graph could be misleading? Why?

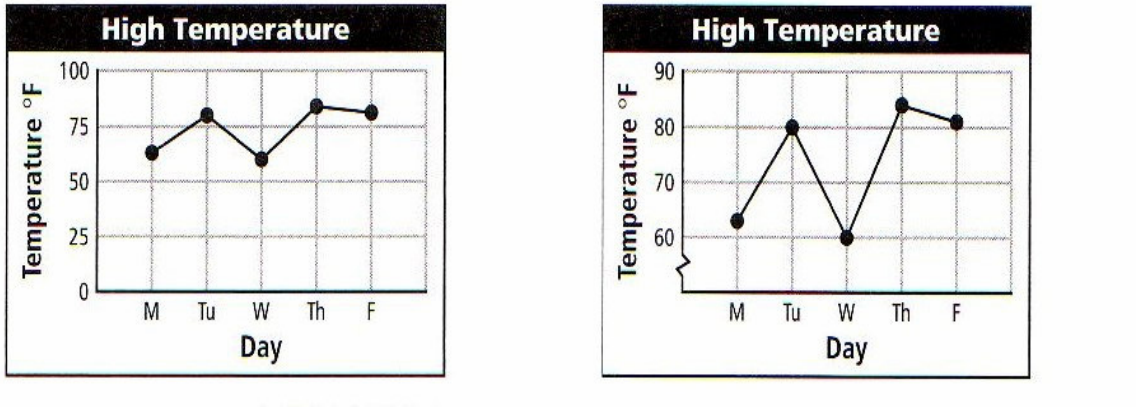
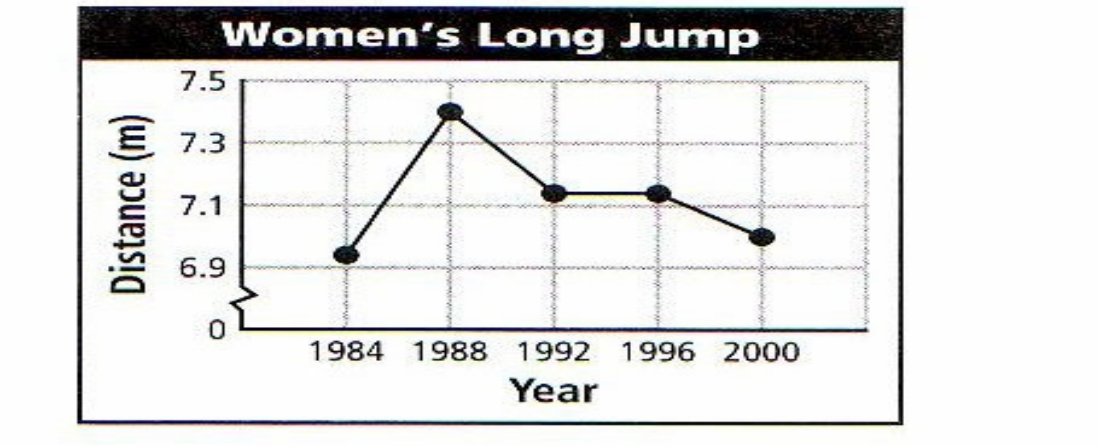


Figure 10: Misleading Graphs Example

When looking for misleading graphs, look at the scale of the y-axis. The graph on the right is missing part of its scale. Notice the broken line where 0 to 60 should be displayed. This y-axis is broken, so differences in temperature appear much larger. Information on the y-axis should begin with 0 and maintain equal intervals to be accurate and not misleading.

12.3 Application

Using the following graph, explain how you could redraw the graph so it would not be misleading. Please redraw the graph.



Answer:

The y-axis would need to include the numbers from 0-6.9.

Application [2]

Use the information in these reports of well-known disasters to answer questions 1-5.

The Blizzard of 1888 From March 11 through March 14 of 1888, a huge blizzard covered the eastern U.S. with as much as five feet of snow. The storm caused 400 deaths and $20 million in damage.

The Tropical Storm of 1972 In June 1972 Tropical Storm Agnes moved up the East Coast of the U.S., causing flash floods that killed 129 people, left 115,000 people homeless, and caused $3.5 billion in damage.

The Flood of 1993 From June through August 1993, heavy rains in the Midwestern U.S. led to flooding that caused 50 deaths, left 70,000 homeless, and damaged $12 billion worth of crops and building.

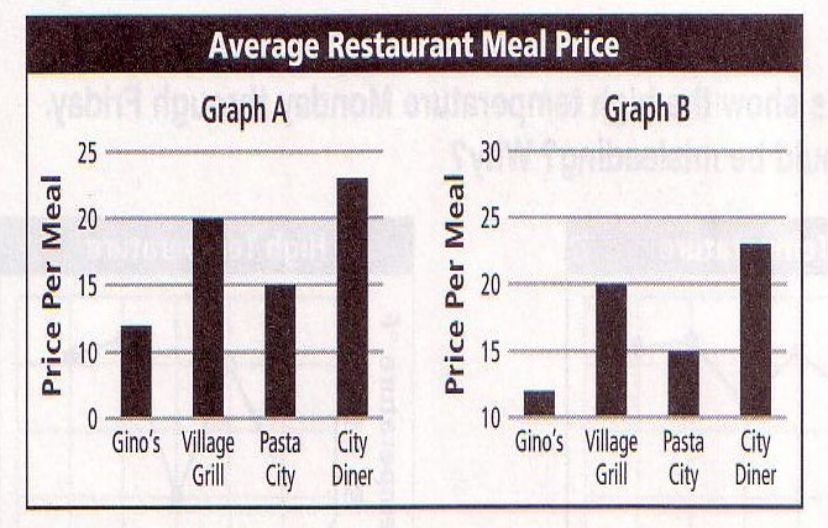
1. Which numbers in these reports are probably accurate**?** Which are probably estimates?
2. Which disaster caused the greatest financial loss?
3. Compare the loss of life from the Blizzard of 1888 with the loss of life from Tropical Storm Agnes.
4. Compare the loss of life from Tropical Storm Agnes with the loss of life from the Midwest floods of 1993.
5. Compare the loss of life from the Blizzard of 1888 with the loss of life from the Midwest floods of 1993.

Answer Key

1. The dates, numbers of deaths (except for the Blizzard), and the snow depths are probably accurate; the dollars of damage and the numbers of homeless are probably estimated.
2. The greatest financial loss was the floods of 1993 because $12 billion is greater than $3.5 billion or $20 million. This is a comparison of different years and inflation probably should be considered.
3. The blizzard caused 271 more deaths than did the tropical storm. The ratio is 400 to 129 or about 3 to 1.
4. The tropical storm caused 79 more deaths than the floods. The ratio of deaths is 129 to 50, or about 2.5 times as many deaths.
5. The blizzard caused 350 more deaths than did the floods. The ratio of deaths from the blizzard to the floods is 400 to 50 or 8 to 1.

When statistics are estimated, they often seem worse or better than they are. Inflation would also play a role in estimating damages from year to year.

12.4 Homework Exercises



Assignment:

1. Using the graphs of the Average Restaurant Meal Price, determine if either graph is misleading. If so, how could the graph be corrected?
2. Which restaurant do you believe published this comparison?
3. Find a graph or other statistic that is misleading. Describe what makes it misleading and what you could do to correct it. (Use newspapers, magazines or internet)

Problem 4:

A physical education teacher is interested in how many cartwheels the average eighth grade student can do. Seven volunteers come forward. After observing their performances, the teacher concludes that the average eighth grade student can do five cartwheels without stopping. Identify the population, the sample, and any possible problems.

12.5 Answer Key

1. The graph on the right is misleading. The scale on the y-axis begins at 10 not 0. This would distort the information making one of the restaurants appear much cheaper. The graph would be corrected if it looked like the one on the left.
2. The graph was probably published by Gino's Restaurant since their price appears cheaper.
3. Answers will vary according to information found during research.
4. The population is the entire eighth grade class. The sample was just those five students who volunteered. The problem with the sample is that those students who volunteered are more likely to be able to do cartwheels than those who did not. We do not know if the volunteers were girls or boys, and gender could affect the outcome.

13 STATISTICAL ESTIMATION

13.1 Introduction

One of the component branches of statistics is inferential statistics. Inferential statistics involve those methods or techniques whereby estimates of a general nature are made on the basis of knowledge about a part or sample of the general population. For example, in the manufacture of shoes one must be concerned with the distribution of shoe sizes among all possible customers. The proportion of shoes produced of any given size will depend on the proportion of the population of potential customers who wear that size. Shoe manufactures cannot measure the shoe size of each person in the country. However, they could determine the shoe size of a few people and try to predict from such a sample how many of each size to make.

One factor to consider is how the sample is to be selected. If the sample consists entirely of women, there will be no information on men's shoe sizes. The company must ensure that the sample represents the population. The sample must be selected in a way that each member of the population has as good a chance of being included as any other member. A sample selected in this manner is called a random sample.

Virginia Standards of Learning [5]:

7.18 The student will make inferences and predictions based on the analysis of a set of data that the student(s) collect.

Key Terms:

The student will be introduced to the following terms:

1. Sample
2. Population

A sample is a small group of items chosen to represent a larger group [4].

A population is the larger group.

Materials Needed:

1. Box with small opening
2. 200 Beans, 40 green and 160 red
3. 4 function calculator for percentages
4. Chart for recording samples and percentages

13.2 In Class Example

Students can get a grasp of inferential statistics by carrying out lab experiments with concrete objects.

Procedure:

1. Prior to the beginning of class, place 40 green and 160 red beans (population) into a small box. Do not allow the students to see into the box.
2. At the beginning of class, shake the box. Allow the students to guess what it may contain. This may also lead to guesses concerning the number of beans.
3. After the guesses, tell the students that it contains 200 beans, but do not tell the colors.
4. Reach into the box and draw out a sample of 5 beans. Let's say it contains 3 red. From this the students may conclude that since 3 out of the 5 were red, then 60% of the beans are red. They should record the sample and percentage. Replace the five beans drawn at random.
5. Next, reach into the box and draw out another sample of 5 beans. Let's say this time there are no red beans. Let the students arrive at their own conclusions, first. One conclusion may be that an average of the first and second sample would be a good estimate.
6. Now, draw a sample of size ten. Explain that this will give considerably more information about the population of the box. Record the number of red and the percentage.
7. Draw a sample of 50 beans from the box. Record the results.
8. This time, look in the box and draw out a sample. The students should comment that this method would not produce a random sample.

9. Allow the students to count the number of red and green beans in the box and compare these to the samples. The class can now note the accuracy of their predictions from the samples. .

Example 2:

A video store is thinking about opening a new section that offers only cartoons. They decide to take a survey to determine whether there is enough interest in this type of video. For two Saturdays, they ask people at the mall if they would rent an animated movie at least every four months. The following table gives the results.

	Yes	No
First Saturday	98	75
Second Saturday	187	116

1. How many people were in the survey? Total 476
2. What percent said they would rent an animated movie? 285/476 = 60%
3. Out of 2,000 customers, how many animated movies should the video store expect to rent? .60 x 3,000 = 1,200
4. Would the survey have a random sample if they asked people coming out of a theater that was showing an animated movie? No, the opinion could be biased.

13.3 Application

Objective: To estimate the grasshopper population in a given area by using the capture-recapture technique [3].

Materials Needed:

1. Q-tips
2. Small amounts of paint (bright colors)
3. Pint or Quart Jars (preferably plastic)

Time: Two 55 minute class periods

We must assume that all the grasshoppers in a given area have an equal chance of being caught. Marked grasshoppers distribute themselves equally throughout the population so that both marked and unmarked grasshoppers have the same opportunity to be captured in the second go-around.

Procedure:

(Day One)

1. Divide the students into groups of 3 or 4.
2. Allot each group an area of the school yard about 100 feet x 100 feet in which to capture grasshoppers. Provide each group with a different color of paint and several Q-tips for using as brushes.
3. Upon capture, place your grasshoppers into the jar.
4. Other members of the team can begin to mark the grasshoppers by lightly placing a drop of paint directly behind the head of the grasshopper.
5. Record the number of grasshoppers that you have marked as you go.
6. Complete the sweep of the area assigned.

7. After all grasshoppers have been marked and recorded, release them into the same area where they were captured.

(Day Two)

1. Repeat steps 1-3 and 5-6 skipping step 4 from yesterday.
2. This time keep track of the total number of both marked and unmarked grasshoppers as you go. Once you have taken the count, release the grasshoppers into the same area.
3. Each group should now have the data to plug into the following formula to estimate the total population of grasshoppers.
    - If N/M = n/R, then N = (M)(n)/R
    - Where N is the total population size to be estimated.
    - M is the sample of grasshoppers that was captured, marked, and released.
    - n is the number of grasshoppers captured the second day.
    - R is the number of marked grasshoppers that were recaptured on the second day.

(Example: M =32 grasshoppers caught, marked & released on the first day, n = 40 grasshoppers caught on the second day, R = 20 recaptured, marked grasshoppers. Therefore N/32 = 40/20 and N = (32)(40)/20 or 64.

Discussion Questions:

1. What were your results and conclusions for this experiment?
2. What could you infer based on your conclusions?
3. How would you design and/or conduct this experiment differently next time?

13.4 Homework Exercises

Assignment 1:

Repeat the application assignment at home using your lawn as the area.

Record your results and compare those to the ones your group recorded at school.

Questions:

1. Did changing the geographical location change the results? Why or why not?
2. What inferences can you draw from your additional conclusions?
3. What external factors may have contributed to different capture/recapture results?
4. What similarities existed between both experiments?

Assignment 2:

Objective: Using sampling to predict in the business world

The Southwest Middle School 4-H club is opening a book store. All incoming students are required to use a 3-ring binder. The binders come in four colors; red, green, blue, or yellow. The students who run the store decide to survey 50 students to find out their favorite color. The information will be used to order 450 binders to sell in the fall.

1. From an alphabetical list, the students survey every 9^{th} student. Is this a good sample? Explain.
2. Of the students surveyed, 25 chose red, 10 chose green, and 2 choose yellow. How many chose blue?
3. For the 450-binder order, how many of each color should be ordered?

13.5 Answer Key

Assignment One:

Answers will vary depending upon location of experiment.

Assignment Two:

1. This is a good sample. The students are chosen at random, every 9^{th} student. This method is called systematic sampling.
2. If 25 choose red, 10 chose green, and 2 chose yellow, then 13 chose blue.
3. If 450 binders are ordered, 225 should be red, 90 green, 18 yellow, and 117 blue.

14 SUMMARY

Because the Statistics Units were designed to prepare seventh grade students for the Math 8 Virginia Standards of Learning test, the author used these with her own seventh grade math students this year. These students have not been tested. In past years, the author has used many of the ideas in the units for the same purpose and has had very good results. In the year 2004-05, 92% of the author's eighth grade students passed the math 8 test. This number includes those students who were advanced and took Algebra I instead of Math 8. These are the students for whom the units were created.

The author has received positive responses regarding the units from her students. Their favorite unit is the Statistical Estimation. Prior to the teaching of this unit, the students had not been exposed to sampling. Most of the students live in close proximity to Jefferson National Forest and the discussion of animal population was of great interest. They enjoyed capture and recapture of caterpillars as an example of how population is counted.

The author considers these units as a work in progress and plans to add ideas to the units as the Virginia Standards of Learning change. Hopefully, the author's students will continue to reap the benefits of these units and any additional units added.

BIBLIOGRAPHY

[1] G. Lappen, J. Fey, W. Fitzgerald, S. Friel, E. Phillips, *Data Around Us*, Pearson/Prentice Hall, New Jersey, 2004.

[2] K. Smith, *Mathematics: It's Power and Utility*, 4th ed. Brooks/Cole Publishing, California, 1994.

[3] R. Millard and J. Turner, *Activities and Projects for the W. H. Freeman Statistics Series*, Freeman and Company, New York, 2004.

[4] M. Fligner and W. Notz, *Study Guide for Moore's The Basic Practice of Statistics*, 3rd ed, Freeman and Company, New York, 2003.

[5] *Mathematics Curriculum Standards*, Virginia Department of Education, viewed 26 August 2005, <http://www.pen.12.va.us/VDOE/superintendent/sols/mathsol2001.pdf >.

[6] National Council of Teachers of Mathematics, Inc., *Teaching Statistics and Probability*, Yearbook/National Council of Teachers of Mathematics, 1981.

[7] W. Collins, L. Dritsas, P. Frey, A. Howard, K. McClain, D. Molina, B. Moore-Harris, J. Ott, R. Pelfrey, J. Price, B. Smith, and P. Wilson, *Course 2 Mathematics Applications and Connections*, Glencoe and McGraw-Hill, New York, 1999.

[8] W. Collins, L. Dritsas, P. Frey, A. Howard, K. McClain, D. Molina, B. Moore-Harris, J. Ott, R. Pelfrey, J. Price, B. Smith, and P. Wilson, *Course 3 Mathematics Applications and Connections*, Glencoe and McGraw-Hill, New York, 1999.

VITA

Sherry Lynn Mullins

398 Apple Valley Road

Marion, Virginia 24354

sherrymullins@scsb.org

Education

- Masters of Science Degree in Pre-collegiate Mathematics, East Tennessee State University, August 2006.
- Bachelors of Science Degree in Elementary Education, Clinch Valley College, 1984.

Experience

- V-Quest Math Lead Teacher, Sugar Grove Combined School, Sugar Grove, VA 08/92-05/95
- Smyth County Schools, Marion, VA 08/91 - Present