MCAS

A Resource for Instructors – Data Analysis

David Stanley and Marvin Stick
with contributions from Amanda Rothrock

July 29, 2008
University of Massachusetts Lowell

In cooperation with the Lawrence, MA Public Schools

Focus on Mathematics

Boston University – Education Development Center, Inc.
Lesley University – University of Massachusetts Lowell
And Participating School Districts:
Arlington – Chelsea – Lawrence – Waltham – Watertown

Funded by the National Science Foundation under grant no. 0314692

Acknowledgements

The UMass Lowell Department of Mathematical Sciences has been a partner in Focus on Mathematics (FoM), a partnership funded by the National Science Foundation. The goal of FoM has been to improve student achievement by providing mathematics teachers with the content knowledge and skills valuable in their profession. To this end, the project is developing a mathematics community running across several school districts, universities and educational organizations.

UMass Lowell mathematicians have joined mathematicians from Boston University, Education Development Center, and Worcester Polytechnic Institute (2003-2007). A team from Lesley University has conducted program evaluations. As part of the program, each mathematician works with teachers in a school in one of the FoM districts: Arlington, Chelsea, Lawrence, Waltham and Watertown.

Preparation of this MCAS instructors' resource document dealing with data analysis is a follow-up to the documents previously prepared. One dealt with geometry and measurement MCAS related topics and was used in study groups for middle school teachers in Lawrence. The second document dealt with fractions related material and the continued fractions material has been used in Lawrence High School study groups. Seven years of MCAS examination problems and resources from referenced texts have been used to supplement developed materials for both the geometry and fractions documents. Copies have been generated for all mathematics teachers in the Lawrence district. This would not have been possible without the foresight of Donna Chevaire, mathematics principal for the Lawrence district. During the development of those documents, the authors were fortunate to receive comments from Anne Cook, Lawrence Curriculum Facilitator Mathematics 6-12, and from UMass Lowell mathematicians (Al Doerr, James Graham-Eagle and Ken Levasseur) involved with the FoM partnership.

The authors represent the Department of Mathematical Sciences at UMass Lowell. David Stanley was a UMass Lowell graduate student when the document was prepared and is now an instructional technology integration specialist with JFY Networks. As a mathematics graduate student, Amanda Rothrock did summer research to assemble material for this document. Marvin Stick is a mathematics department faculty member.

Preface

This instructors' resource was developed to provide teachers with a concise reference for data analysis related topics, as well as professional development material. TI graphing calculator material has been included to help facilitate graph construction. MCAS problems were included to help improve student achievement.

We hope this document provides the users with a valuable resource. On-line versions are available at http://www.focusonmath.org/ and at http://faculty.uml.edu/mstick.

David Stanley, Marvin Stick

Updated Sept 26, 2011

Table of Contents

Measure of Central Tendency 1
- Choosing the Best Measure 1
- Mean 2
- Median 5
- Mode 7
- MCAS Problems 10

Measure of Variability 23
- Range 23
  - MCAS Problems 25

Data Representation 27
- Bar Graphs 27
- Double-Bar Graphs 31
  - MCAS Problems 32
- Line Plots 33
  - MCAS Problems 35
- Stem and Leaf Plots 37
  - MCAS Problems 42
- Scatterplots 47
  - Correlation 50
  - MCAS Problems 52
- Box and Whisker Diagrams 55
  - Deciles and Percentiles 58
  - MCAS Problems 60
- Frequency Tables 65
  - MCAS Problems 69
- Histograms 73
  - MCAS Problems 76
- Circle Graphs 79
  - MCAS Problems 84
- Venn Diagrams 89
  - MCAS Problems 94

Regression 97
- MCAS Problems 100

References 105

Measure of Central Tendency

If you were to pick a number that best describes all the data in a set, what number would you pick? People most often choose a number somewhere in the middle of the data ordered from least to greatest, or a number with a lot of data clustered around it. This number is known as an average, or a measure of central tendency.

An **average** is a **measure of central tendency**, a number that summarizes a set of data by giving some sense of the typical (often the middle) value. Measures of central tendency include the **mean**, **median**, and **mode**.

The average that you choose to use depends on the data and your purpose.

- For sets of data with no unusually high or low numbers, *mean* is a good measure to use.
- For sets of data with some points that are much higher or lower than most of the others, *median* may be a good choice.
- For sets of data with many data points that are the same, *mode* may be the best choice.

Look at the question you are trying to answer to decide which type of average to use.

Choosing the Best Measure

One measure of central tendency may be better than another to describe data. For example, consider the eight hourly wage rates shown in Table 1. Here are the measures of central tendency.

Table 1

Mode: $5.50
Mean: $7.50
Median: $6.10

Employees' Hourly Wages	
$5.50	$6.20
$5.50	$6.30
$5.50	$8.00
$6.00	$17.00

The mode is the lowest wage listed. So the mode does not describe the data well.

The mean is above the hourly wage of all but two workers. The mean is influenced by the outlier, $17. See page 33 for a description of outliers.

The median is the best measure of central tendency here since it is not influenced by the size of the outlier.

Example 1

Which measure of central tendency best describes each situation? Explain.

a.) The favorite movies of students in the eighth grade
Mode; since the data are not numerical, the mode is the appropriate measure. When determining the most frequently chosen item, or when the data are not numerical, use the mode.

b.) The daily high temperatures during a week in July
Mean; since daily high temperatures in July are not likely to have an outlier, mean is the appropriate measure. When the data have no outliers, use the mean.

c.) The distances students in your class travel to school
Median; since one student may live much farther from school than the majority of students, the median is the appropriate measure. When an outlier may significantly influence the mean, use the median.

Example 2

Which measure of central tendency best describes each situation? Explain.

a.) The favorite radio stations of teenagers in your neighborhood
Mode; data is not numerical, so the mode is the best to use.

b.) The numbers of videos owned by students in your class
Median; one student may have a lot more videos than the rest of the class. The median is best to use, since there may be an outlier.

c.) The prices of 8-oz containers of yogurt at six local grocery stores.
Mean; it is not likely that there would be an outlier, so the mean is best to use.

Mean

What does it mean when we find the mean? In statistics, mean is a type of average. To find the mean, you add all the numbers in a set and divide by the number of values.

Mean = (sum of the values) ÷ (number of values)

Case 1 – When all the data are close to each other, the mean is close to all the data.

Example 1

What is the mean number of books sold per day at Beautiful Books?

Beautiful Books				
Mon.	Tues.	Wed.	Thurs.	Fri.
105	96	90	108	116

1. Find the sum of the numbers.
   105 + 96 + 90 + 108 + 116 = 515

2. Divide by the number of numbers.
   515 ÷ 5 = 103

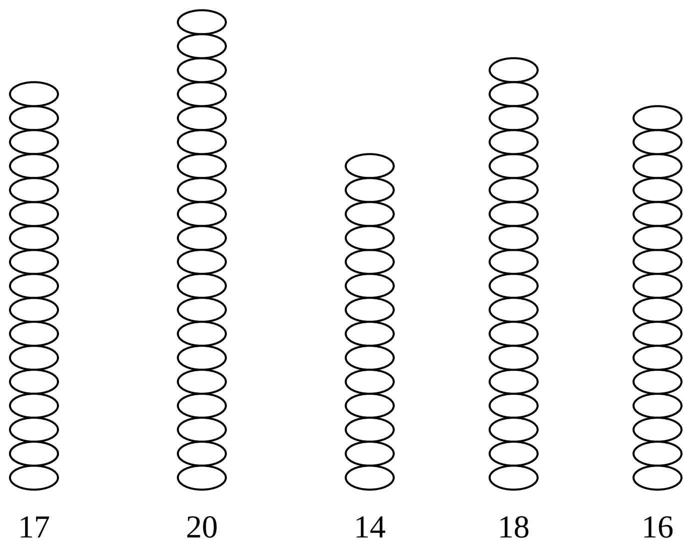
The book store's mean sales were 103 books per day. This means that if the sales were not the same every day, but the weekly total stayed at 515 books, each day's sales would be about 103 books. See page 55 for an example of how to find the mean using a TI-84 calculator.

Example 2

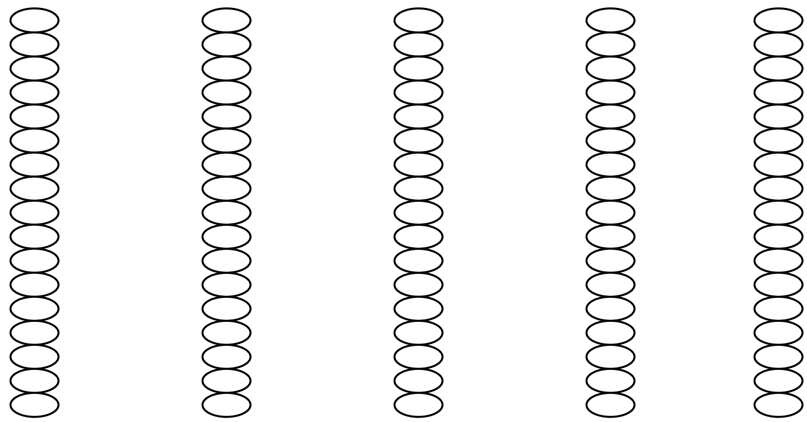
In five basketball games, you score 17, 20, 14, 18, and 16 points. What are your average points per game (the mean of your scores)?

<u>One way to solve the problem:</u> Find the mean by evening out the scores.

1. Make a stack of counters for each score.



2. Move counters from stack to stack until each stack has the same number of counters.



17 17 17 17 17

<u>*Another way to solve the problem:*</u> Find the mean by computing it.

1. Add to find the sum of the numbers.
   17 + 20 + 14 + 18 + 16 = 85

2. Divide the sum by the number of addends.
   85 ÷ 5 = 17

Either way, you scored an average of 17 points per game.

Case 2 – When a piece of data is much bigger or smaller than the rest, it is called an **outlier**. It can move the mean away from the main group of data toward the outlier.

Example 3

Referring to Example 1 data, suppose on the next day, the book store sells only 7 books. What happens to the mean?

1. Find the sum of the numbers.
   515 + 7 = 522

2. Divide by the number of numbers.
   522 ÷ 6 = 87

One day with only 7 books sold brought the mean sales per day down by 16 books. It made the mean smaller than any of the other pieces of data, so the mean no longer describes a typical day of sales.

Example 4

What would happen to the mean in Example 2 if you scored 50 points at the next game?

1. Add to find the sum of the numbers.
   17 + 20 + 14 + 18 + 16 + 50 = 135

2. Divide the sum by the number of addends.
   135 ÷ 6 = 22.5

Just one high score game brought your average points-per-game up by more than 5 points! It made the mean bigger than any of the other data points. So the mean no longer describes a typical number in the data.

Median

The median is a type of average. It is the number that falls exactly in the middle when a set of data is arranged in order from least to greatest. If there aren't any big gaps in the middle of the data but there are outliers at either end, the median may be a better number than the mean to describe the data.

Case 1 – When there is an odd number of pieces of data, the median is the middle number.

Example 1

On August 15, 1999, the top 15 nonfiction bestsellers had been on the bestseller list the following numbers of weeks: 95, 35, 7, 6, 30, 10, 3, 122, 4, 10, 17, 37, 19, 8, and 17. What is the median number of weeks on the bestseller list? Why might the median be a better measure of central tendency than the mean for these data?

1. Arrange the numbers in order from least to greatest.
   3, 4, 6, 7, 8, 10, 10, 17, 17, 19, 30, 35, 37, 95, 122

2. Find the middle number.
   3, 4, 6, 7, 8, 10, 10, <u>17</u>, 17, 19, 30, 35, 37, 95, 122

   17 is the middle number (it's underlined)

The median for this set of data is 17. The mean for this set of data, 28, is greater than 10 of the 15 values. The two outliers, 95, and 122 are not typical of the most of the list, and, since they do not affect the median as much as they affect the mean, the median better represents the typical values in the list. See page 55 for an example of how to find the median using a TI-84 calculator.

Example 2

During one 7-hour shift, The Sandwich Barn kept track of the number of customers served each hour.

Hours	7-8	8-9	9-10	10-11	11-12	12-1	1-2
Number of Customers Served	7	16	13	11	14	50	17

What is the median number of customers per hour? Why is the median a good average for these data?

1. Arrange the numbers in order from least to greatest.
   7, 11, 13, 14, 16, 17, 50

2. Find the middle number.
   7, 11, 13, 14, 16, 17, 50

The median for the set of data is 14. So the typical number of customers in an hour is 14. The mean for this set of data is about 18.29. The slow first hour and the very hectic lunch hour are not typical of the rest of the shift and they do not affect the median as much as they affect the mean. Hence, the median is a good average for the data.

Case 2 – When there is an even number of data points in a set of data, there are two middle numbers. In this case you need to find the mean of the two middle numbers. That will then give you the median number.

Example 3

Ted gets the test scores: 82, 86, 39, 91, 84, and 80. What is Ted's median test score?

1. Arrange the numbers in order from least to greatest. Find the two middle numbers.
   39, 80, 82, 84, 86, 91
   The two middle numbers are 82 and 84.

2. Find the mean of the two middle numbers.
   $(82 + 84) \div 2 = 83$

The median of Ted's test scores is 83.

Example 4

Rolando is The Sandwich Barn's best customer. Last week he went there six times and spent the amounts shown below. What is his typical lunch bill (the median amount)?

Amount spent: $3.95, $5.02, $4.26, $25.90, $5.36, $4.58

1. Arrange the numbers in order from least to greatest. Find the two middle numbers.
   3.95, 4.26, 4.58, 5.02, 5.36, 25.90

2. Find the mean of the two middle numbers.
   (4.58 + 5.02) ÷ 2 = 4.80

Rolando's typical bill is $4.80

Mode

The mode is the value that occurs most often in a set of data.

Sometimes, the best way to describe what is typical about a set of data is to use the value that occurs most often. This value is called the **mode**. For example, in the set of data 2, 3, 5, 5, 6, the mode is 5.

Case 1 – Sometimes there is one value that occurs more often than any other.

Example 1

For the week of May 3 to 9, 1999, Variety listed the maximum weekend ticket prices for 35 Broadway shows. What is the mode? Why is the mode a good measure of central tendency for these data?

Maximum Weekend Ticket Price	Number of Shows
$55	2
$60	6
$65	4
$67.50	1
$75	15
$80	4
$85	2
$100	1

The maximum weekend price that occurs most often is $75, so $75 is the mode for this set of data. It is a good measure of central tendency in this situation because it is a very typical maximum price; nearly half of the shows had a maximum price of $75.

Example 2

According to the American Trucking Association's Web site, the 50 states and Washington, D.C., have different maximum speed limits for trucks:

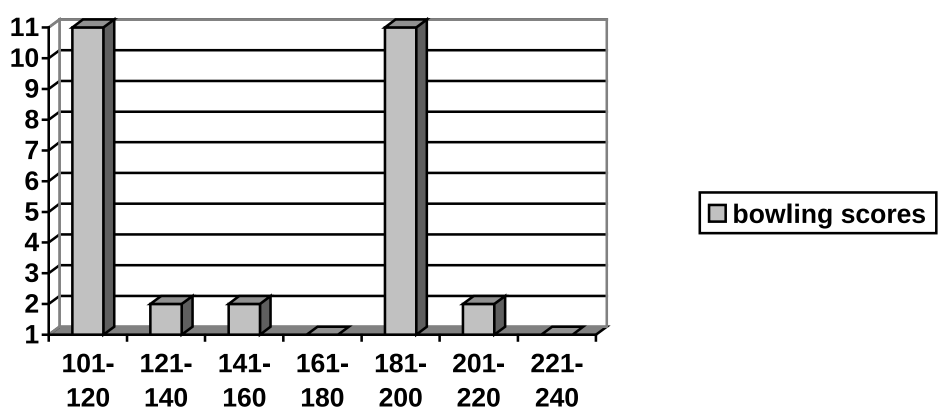
Maximum Speed Limits	Number of States
55 mph	9
60 mph	3
65 mph	21
70 mph	8
75 mph	10

If you look at the data, far more states have a 65 mph speed limit for trucks than have any other speed limit. So you could say the typical maximum speed limit for trucks is 65 mph.

Case 2 – Sometimes there is more than one mode. In this case, all these highest values are modes for the set of data.

Example 3

The chart shows Janine's bowling scores for one season. What is the mode? Why is the mode a good measure of central tendency for this set of data?



Janine's scores are bimodal (two modes) because she has 11 scores from 101 to 120 and 11 scores from 181 to 200. The data suggests that Janine has on games and off games, and infrequently has an average game that falls in the middle of the scores. The two modes give a good idea of how Janine bowls.

Example 4

According to Weather Post at The Washington Post's Web site, Sacramento, California has an interesting pattern for average daily temperatures:

Average Daily Temperature	Number of Months
In the 40s	2
In the 50s	4
In the 60s	2
In the 70s	4

You could say that Sacramento, California has a **bimodal** (two modes) weather pattern because there are 4 months with averages in the 50s and 4 months with averages in the 70s.

Case 3 – Sometimes there is no value that occurs more often than the others. In this case, there is no mode and it might be a good idea to use the mean or median to describe the set of data.

Example 5

How many modes, if any, does each have?

a.) $1.50, $2.00, $2.25, $2.40, $3.50, $4.00

No values are the same, so there is no mode.

b.) 2, 3, 6, 8, 8, 10, 11, 12, 14, 14, 18, 20

Both 8 and 14 appear the same number of times. So there are two modes.

c.) grape, grape, banana, nectarine, strawberry, strawberry, strawberry, orange, watermelon

Strawberry appears most often, so there is one mode.

MCAS Problems

2006

http://www.doe.mass.edu/mcas/2006/release/g6math.pdf

2006 Grade 6 #31

31 Katie will take a total of 5 mathematics tests. She has taken 4 mathematics tests so far. The scores on her first 4 tests are shown in the table below.

Katie's Mathematics Test Scores

Test	Score
1	94
2	98
3	86
4	92
5	?

a. What is the median of Katie's first 4 mathematics test scores? Show or explain how you got your answer.

b. What is the mean of Katie's first 4 mathematics test scores? Show or explain how you got your answer.

c. What score must Katie get on her 5th test in order to have a mean score of 90 on all 5 of her mathematics tests? Show or explain how you got your answer.

Answer for a: Put the grades in numerical order.
86, 92, 94, 98

The median is the mean of the two middle numbers.
92 + 94 = 186
186 / 2 = 93
The median is 93.

Answer for b: 94 + 98 + 86 + 92 = 370
370 / 4 = 92.5
The mean is 92.5.

Answer for c: The mean in part (b) is 92.5 with four grades. To get a mean of 90 on all five grades, you must pick a number x to arrive at a mean of 90.
The equation to solve for x is:

$$94 + 98 + 86 + 92 + x = 90 \times 5 = 450$$
$$x = 450 - 370 = 80$$

The mean is 90 when Katie gets an 80 on her fifth test.

http://www.doe.mass.edu/mcas/2006/release/g8math.pdf

2006 Grade 8 #19

19 Lorna swam 30 laps per day for the first 6 days of swim practice. She swam 40 laps per day for the next 4 days of practice. What was the mean number of laps that Lorna swam per day for these 10 days?

$$\frac{\text{Total number of laps}}{\text{Total number of days}} = \frac{(30 \cdot 6)+(40 \cdot 4)}{6+4} = \frac{180+160}{10} = \frac{340}{10} = 34$$

Answer: 34

2006 Grade 8 #30

30 The prices of some of the comic books sold in a collectors' catalogue are listed below.

$5.00	$20.00	$4.50	$3.00	$3.50
$3.00	$5.50	$3.00	$6.00	$4.00

What is the mean price of these books?

A. $3.00

B. $4.25

C. $5.00

D. $5.75

$$\frac{\$5.00+\$20.00+\$4.50+\$3.00+\$3.50+\$3.00+\$5.50+\$3.00+\$6.00+\$4.00}{10} = \frac{\$57.50}{10} = \$5.75$$

Answer: D

2005

http://www.doe.mass.edu/mcas/2005/release/g6math.pdf

2005 Grade 6 #15

15. Mr. Young wrote five numbers on the board in his classroom. After class, one of the numbers was erased. Four of the five numbers are shown below.

18 25 30 17 _?_

If the median of the five numbers that Mr. Young wrote on the board was 18, which of the following could be true?

A. The number that was erased was greater than 30.

B. The mode of the five numbers Mr. Young wrote on the board was 24.

C. The mean of the five numbers Mr. Young wrote on the board was 22.6.

D. The number that was erased was less than or equal to 18.

To find the median, the numbers must be arranged in ascending order. The four numbers that were not erased, listed in ascending order, are:

17 18 25 30

With the erased number inserted, there are 5 numbers. The median is the middle number which is the 3rd number. Therefore the 3rd number must be 18 since the median is 18.

The erased number must be in the 1st or 2nd position in order to move 18 to the 3rd position. Any number less than or equal to 18 will do this.

Answer: D

2005 Grade 6 #29

29 Ticket sales for the sixth-grade banquet are shown in the table below.

Banquet Tickets Sold

Homeroom	**Number of Tickets Sold**
Ms. Sanchez	22
Ms. Blake	28
Mr. Chang	21
Mr. Williams	25

What is the **mean** number of tickets sold per homeroom?

$$\frac{\text{Total number of tickets sold}}{\text{Total number of homerooms}} = \frac{22+28+21+25}{4} = \frac{96}{4} = 24$$

Answer: 24

http://www.doe.mass.edu/mcas/2005/release/g8math.pdf

2005 Grade 8 #29

29 Glenn bowls in a bowling league every Saturday morning. Last Saturday, the scores from Glenn's first 3 bowling games were 141, 128, and 157.

a. What is the mean of the scores from Glenn's first 3 games? Show or explain how you got your answer.

b. Glenn will bowl a fourth game. What will he have to bowl in his fourth game to have a mean of 150 for the 4 games? Show or explain how you got your answer.

c. Each player in Glenn's bowling league is given a handicap, which allows players of different abilities to compete equally. A player's handicap is determined with the following formula.

A player's handicap is equal to 80 percent of the difference between the player's average (mean) and 220.

Miguel is Glenn's teammate. If Miguel's average (mean) is 130, what is his handicap? Show or explain how you got your answer.

Answer for a: 141 + 128 + 157 = 426
426 / 3 = 142
The mean is 142.

Answer for b: $426 + x = 150 \times 4 = 600$
$x = 600 - 426 = 174$

Answer for c: handicap = 80% × (player's average – 220)
= 80% × (130 – 220)
= 80% × (– 90)
= –72

http://www.doe.mass.edu/mcas/2005/release/g10math.pdf

2005 Grade 10 #22

22 Four major underwater tunnels were constructed in New York City between 1925 and 1950. The tunnels and their lengths are listed in the chart below.

New York City Tunnel Lengths

Tunnel Name	Length (kilometers)	Year Completed
Holland	2.6	1927
Lincoln	2.5	1937
Queens-Midtown	1.9	1940
Brooklyn-Battery	2.8	1950

Which of the following is closest to the mean of these four lengths?

A. 2.20 kilometers

B. 2.35 kilometers

C. 2.45 kilometers

D. 2.55 kilometers

$$\frac{2.6+2.5+1.9+2.8}{4}=2.45$$

Answer: C

2004

http://www.doe.mass.edu/mcas/2004/release/g6math.pdf

2004 Grade 6 #39

39 The chart below shows the number of minutes five students spent exercising yesterday.

Student	Minutes Spent Exercising
Bill	34
Hassan	32
Jenna	0
Monty	20
Beth	34

Based on the data given in the chart, what is the **median** number of minutes the five students spent exercising?

A. 24 minutes

B. 30 minutes

C. 32 minutes

D. 34 minutes

Find the median by arranging the numbers in ascending order:

0   20   <u>32</u>   34   34

The middle number is the median.

Answer: C

http://www.doe.mass.edu/mcas/2004/release/g10math.pdf

2004 Grade 10 #18

18 Latrice plans to ride her bicycle a mean of 80 miles per week. During the last four weeks, she has recorded distances of 76, 80, 82, and 74 miles. How many miles must Latrice ride this week to obtain a 5-week mean of 80 miles?

$$80 = \frac{76 + 80 + 82 + 74 + x}{5}$$

$$80 = \frac{312 + x}{5}$$

$$400 = 312 + x$$

$$x = 82$$

Answer: 82 miles.

2004 Grade 10 #39

39 Keesha and her family visit her grandparents once a month. Keesha recorded the driving time to her grandparents' house for 6 trips in the chart below.

Driving Times

Month	Time
January	3 hours, 23 minutes
February	3 hours, 5 minutes
March	3 hours, 50 minutes
April	3 hours, 52 minutes
May	3 hours, 15 minutes
June	3 hours, 35 minutes

Based on the data in the chart, what is the mean driving time to Keesha's grandparents' house?

A. 3 hours, 23 minutes

B. 3 hours, 26 minutes

C. 3 hours, 30 minutes

D. 3 hours, 35 minutes

$$\frac{3\frac{23}{60}+3\frac{5}{60}+3\frac{50}{60}+3\frac{52}{60}+3\frac{15}{60}+3\frac{35}{60}}{6}=\frac{18\frac{180}{60}}{6}=\frac{21\text{ hours}}{6}=3.5\text{ hours}$$

or 3 hours, 30 minutes

Answer: C

Note that since the hours are all the same, we only need to average the minutes.

2004 Grade 10 #31

31 In an experiment, Sue and Helise asked each of 30 students in a random sample of the juniors at their school to record the number of minutes they watched television on a Saturday and Sunday in April. The results, rounded to the nearest 30 minutes, are shown in the table.

Minutes of Television Watching

Total Number of Minutes of Television Watched on Saturday and Sunday	Number of Junior Students
0	1
60	3
90	6
120	5
180	5
240	2
300	5
420	1
540	2

a. What number of minutes spent watching television should the girls report as the mode for this group of students? Justify your answer.

b. Helise said that the median number of minutes for this group of students is 180, but Sue disagreed. Do you agree with Sue or Helise? Justify your answer.

c. Suppose that Sue and Helise had used the entire class of 185 juniors as their sample. Based on the results from their smaller sample, what total number of the 185 juniors would probably have reported watching 300 minutes of television on that weekend? Show or explain how you obtained your answer.

Answers:

a. The largest frequency is for 6 students watching TV 90 minutes. Therefore, the mode is 90.

b. Determine the median by find the center number. There are 30 students so the average must be taken of the 15th and 16th students. The 15th student is in the 120 minute interval, the 16th student is in the 180 minute interval. So the median is:

$\frac{120+180}{2} = \frac{300}{2} = 150$ minutes and Helise is incorrect.

c. There were 5 students of the 30 who watch TV 300 minutes, which when expressed as a fraction is $\frac{5}{30} = \frac{1}{6}$. For 185 students this is $\frac{1}{6} \cdot 185 \simeq 31$.

2003

http://www.doe.mass.edu/mcas/2003/release/g6math.pdf

2003 Grade 6 #3

3 Emil, Logan, Stacey, and Stephanie took a test. Their test scores were 95, 70, 81, and 78, respectively. What was the **mean** of their test scores?

A. 79.5

B. 81

C. 88.5

D. 324

$$\frac{95+70+81+78}{4}=\frac{324}{4}=81$$

Answer: B

http://www.doe.mass.edu/mcas/2003/release/g8math.pdf

2003 Grade 8 #25

25 To sell their house, the Fords placed an advertisement in the local newspaper. For one week, they recorded the number of calls they received each day in response to the advertisement in the following chart.

Responses to Advertisement

S	M	T	W	Th	F	Sa
2	5	2	8	10	4	11

What is the **mean** number of calls for the days shown in the chart?

A. 2

B. 5

C. 6

D. 8

$$\frac{2+5+2+8+10+4+11}{7}=\frac{42}{7}=6$$

Answer: C

http://www.doe.mass.edu/mcas/2003/release/g10math.pdf

2003 Grade 10 #37

37 The mean exam score for 31 students in a geometry class was 79. The median exam score for the same set of students was 75. Two additional students took the exam at a later time and scored 65 and 93. How did the mean and median change when these two additional scores were included?

A. The median increased and the mean stayed the same.

B. The median stayed the same and the mean increased.

C. The median and the mean both stayed the same.

D. The median and the mean increased.

Since the mean is 79 for 31 students, the total of all students' scores is:

$$\text{mean} = \frac{\text{total of scores}}{\text{number of students}} \Rightarrow 79 = \frac{\text{total of scores}}{31} \Rightarrow \text{total of scores} = 79 \cdot 31 = 2449$$

Find the mean with the additional scores of 65 and 93:

$$\text{mean} = \frac{\text{total of scores}}{\text{number of students}} = \frac{2449 + 65 + 93}{31 + 2} = \frac{2607}{33} = 79 \text{ which is the same.}$$

The median (75) of the 31 students is the center score. Adding the scores of 65 and 93 keeps the median at 75. This is because 65 is before and 93 is after 75 in the list of scores and 75 remains the center score of the 33 scores.

Answer: C

Measure of Variability

Range

Sometimes it's useful to know how much numbers in a set of data vary. One way to summarize variation is to find the **range**, the difference between the **maximum** (highest) value and the **minimum** (lowest) value.

A singer who can sing very high and very low notes is said to have a wide range. A shortstop who can cover a lot of ground has a large range. In statistics, the range is the difference between the greatest and least numbers in a set of data. Knowing the range can help you decide whether the differences among your data are important.

Example 1

The greatest recorded difference in temperature during a single day occurred in Browning, Montana in January of 1916. During one 24-hour period, the low temperature was -56°F and the high temperature was 44°F. What was the range of the temperatures?

Compare the temperatures by subtracting them.

$$\begin{aligned} \text{Range} &= \text{maximum} - \text{minimum} \\ &= 44^o - (-56^o) \\ &= 44^o + 56^o \\ &= 100^o \end{aligned}$$

The range of the temperature was 100°F. See page 55 for an example of how to find the minimum and maximum using a TI-84 calculator.

Example 2

The table shows some of the top money-making movies of all time as of October 1998. What can you say about the range of the top-10 earners compared to the range of the next 140 earners?

Movie	Amount Earned in the U.S. (rounded to the nearest million dollars)
1. Titanic	601,000,000
2. Star Wars	461,000,000
3. E.T.	400,000,000
4. Jurassic Park	357,000,000
5. Forrest Gump	330,000,000
6. The Lion King	313,000,000
7. Return of the Jedi	309,000,000
8. Independence Day	306,000,000
9. The Empire Strikes Back	290,000,000
10. Home Alone	286,000,000
***********************************	***********************************
149. George of the Jungle	105,000,000

The range of earnings for the top-10 movies is $315,000,000 ($601,000,000 - $286,000,000 = $315,000,000). What makes these data really interesting is that the earnings drop $315 million from number 1 to number 10, then drop *only* another $181 million from number 10 to number 149.

Example 3

Find the range for San Francisco's mean monthly temperatures. Why might some people like San Francisco's climate?

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
51°	54°	55°	56°	56°	58°	59°	60°	62°	62°	57°	52°

Range = maximum value – minimum value
= 62° - 51°
= 11°

The range is 11°. That's a pretty narrow range, especially when you compare it to a place like Bismarck, North Dakota where the mean monthly temperature has a range of over 60°. Maybe people like San Francisco's weather because it is mild and doesn't vary very much.

MCAS Problems

2006

http://www.doe.mass.edu/mcas/2006/release/g8math.pdf

2006 Grade 8 #39

39 The individual weights, in pounds, of the members of a school's wrestling team are shown in the box below.

180	163	165	165
171	177	191	168
180	203	196	175
162	155	178	195

a. What is the range of the weights? Show or explain how you got your answer.

The range is the difference of the minimum and maximum weights. The minimum weight is 155. The maximum weight is 203. Therefore the range is 203 lbs – 155 lbs = 48 lbs.

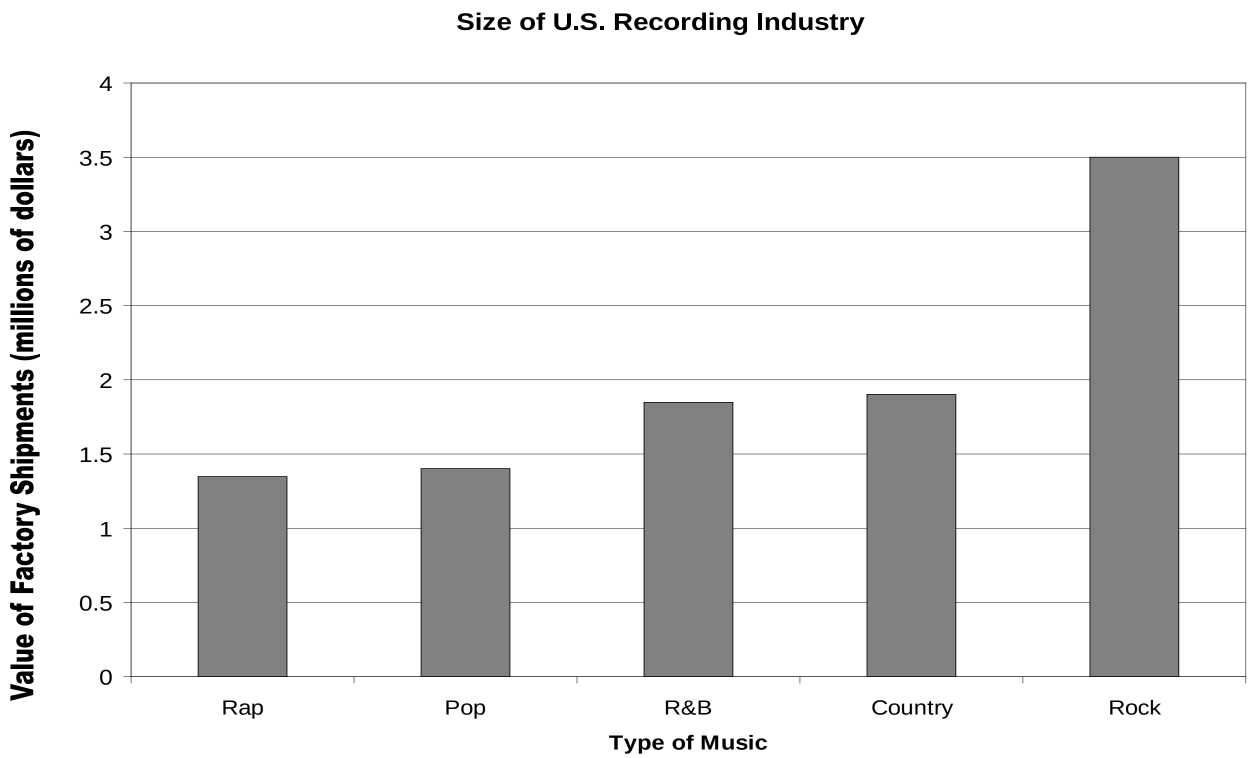
Data Representation

Bar Graphs

A bar graph uses the length of solid bars to represent numbers and compare data. The longer the bar, the greater the quantity. With just one glance at a bar graph, you can see how quantities compare.

Example 1

Does this bar graph help you explain why more radio stations have a country format than have a rock format?

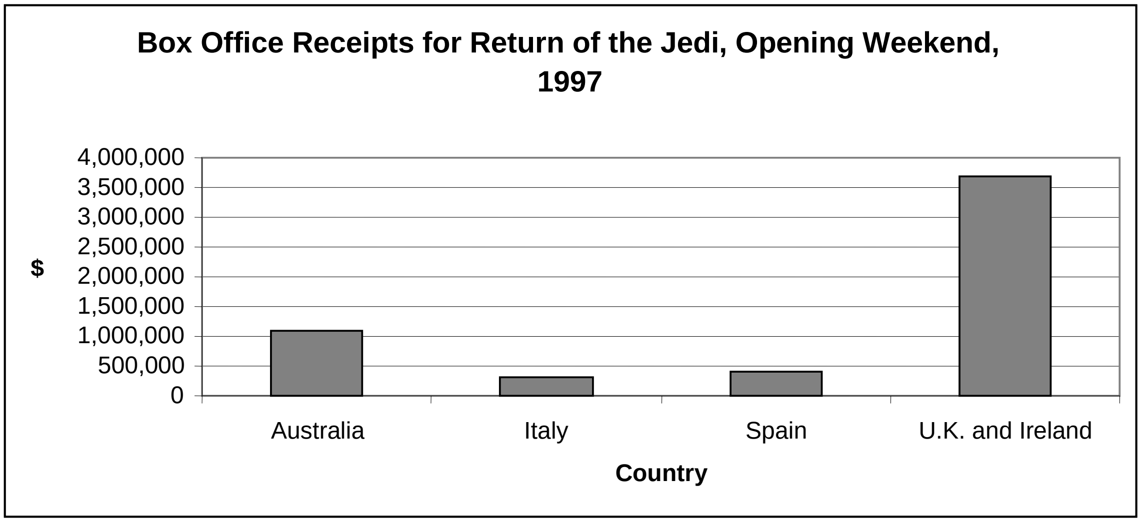


The tallest bar, by far, is for rock music, so you would expect that more radio listeners would want to listen to rock music than to country music.

The bar graph does not provide an explanation for why more radio stations have a country format than have a rock format.

Example 2

Many movies made in the United States are also popular in other countries. One such movie is *Return of the Jedi*. It was first shown in 1983 and then again in 1997 in theaters around the world. The graph shows the receipts for *Return of the Jedi* in some countries during the movie's opening weekend in 1997. What comparisons can you make from the graph?

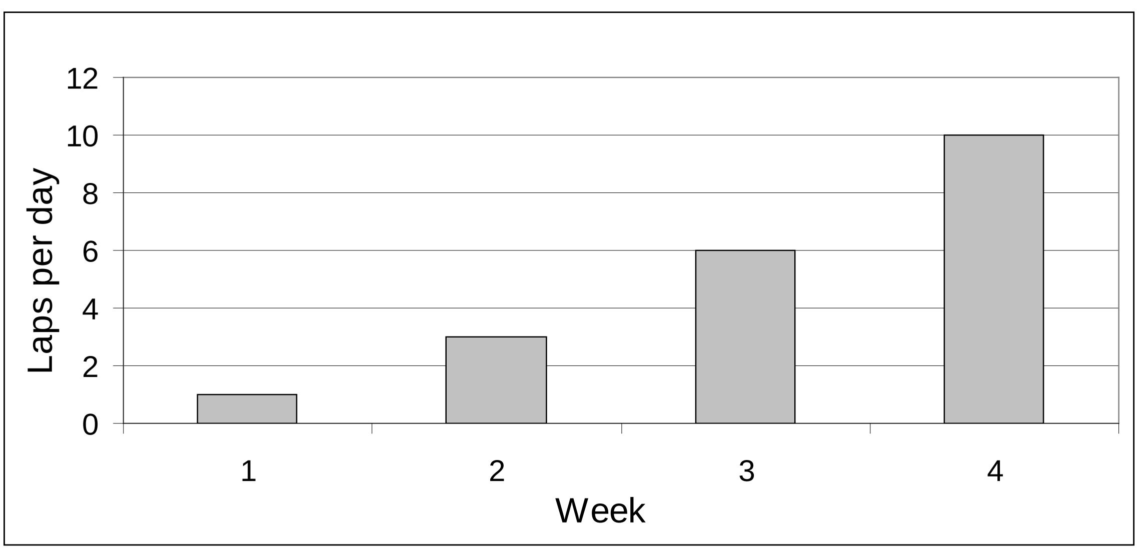


The tallest bar is for the United Kingdom and Ireland. So, *Return of the Jedi* made the most money in the United Kingdom and Ireland.

Where was *Return of the Jedi* most popular? To answer this, you need more information – how many people live in each country and what the ticket price was in each country.

Example 3

Caroline is training for a swim meet. The graph shows the number of laps per day she swims each week. If she stays with this training pattern, how many laps per day will Caroline swim in week 8?



Between week 1 and week 2, she increased her laps by 2. Then between week 2 and week 3, she increased her laps by 3. Between week 3 and week 4 she increased her laps by 4. So by week 5 she would increase her laps by 5. By week 6 she will increase her laps by 6, by week 7 she would increase her laps by 7 and then by week 8 she would increase her laps by 8.

To generate the laps per day, we can use the sequence defined by $a_n = \begin{cases} 1, n = 1 \\ a_{n-1} + n, n \geq 2 \end{cases}$.

A summary of the results is:

Week	Laps per day
1	1
2	3
3	6
4	10
5	15
6	21
7	28
8	36

Double-Bar Graphs

Sometimes, placing bars side-by-side in pairs makes it easier to display the kinds of comparisons you want to show. This kind of graph is called a **double-bar graph**.

Example 1

Table 2 and the graph in Figure 1 show where teenagers get their spending money. What comparisons can you make between the two age ranges? In Figure 1, Ages 12-14 are represented by the lighter shaded bars.

Table 2

Sources of Spending Money for Teenagers		
Source	Ages 12-14	Ages 15-17
From parents	88%	79%
Odd jobs	74%	70%
Regular allowance	54%	29%
Full or part-time job	13%	33%

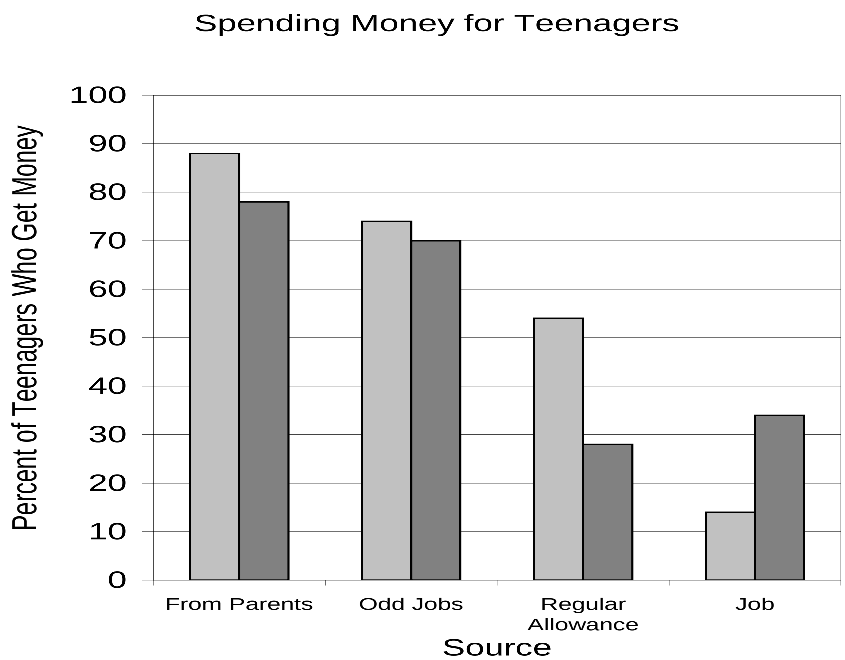


Figure 1

The most common source for both age groups is their parents. The percent of teenagers who have jobs is much greater for older than for younger teenagers. The percent of teenagers who get regular allowances is much greater for younger than for older teenagers.

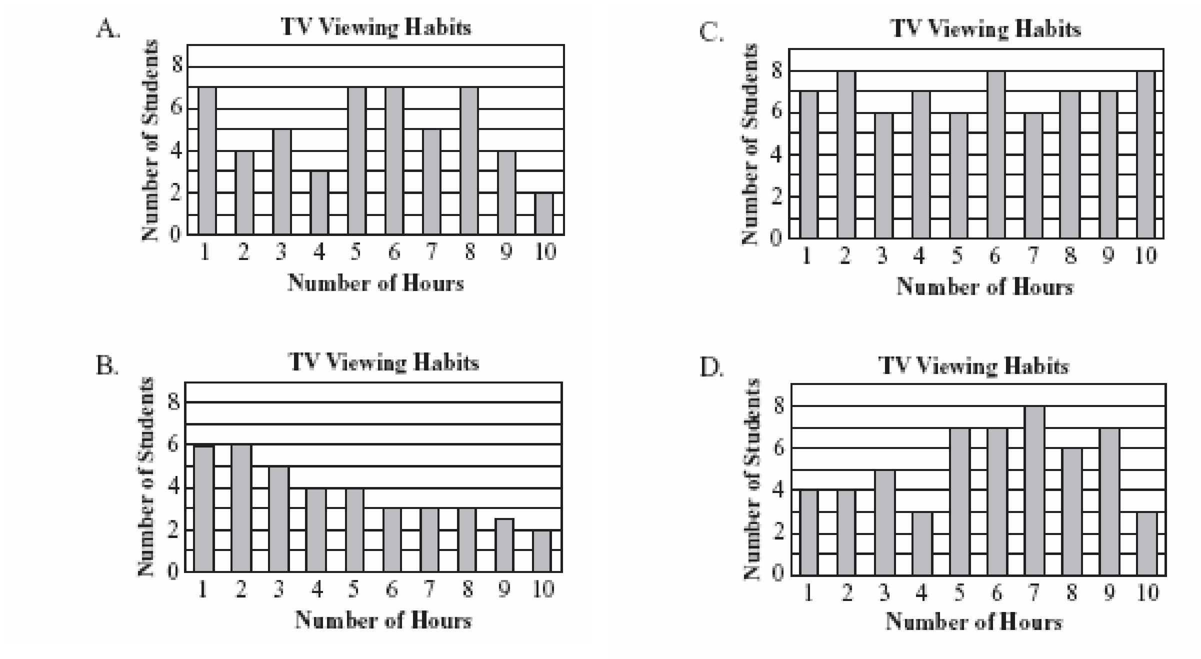
MCAS Problems

2006

http://www.doe.mass.edu/mcas/2006/release/g8math.pdf

2006 Grade 8 #38

38 Which bar graph below shows a mode of 7 hours of television viewed per week?



The mode is the number that shows up the most. Since we want a mode of 7 hours, we look for the bar graph where the bar for 7 hours is the tallest.

Answer: D

Line Plots

A **line plot** displays data with X marks above a number line.

Suppose you want to show how data points are spread out. One way to do this is to make a **line plot**. When data are shown on a line plot, it's easy to see the mode, the range and any unusually high or low data points. These very high or low points are called **outliers**.

Sometimes, instead of comparing data or showing trends, you want to show the spread of the data. You can do this with a line plot. When you look at a line plot, you can quickly identify the range, the mode, and any outliers of the data. Often, line plots are used when you want to see the mode – how often one number occurs in a set of data.

Example 1

In 1999, the roster for the WNBA's Phoenix Mercury listed 12 players. Here are their heights, in inches: 77, 74, 68, 73, 75, 73, 73, 64, 69, 80, 67, and 70. Show the data on a line plot. Then identify the mode, the range, and any outliers.

1. Title your plot.
2. Draw a horizontal line.
3. Make a scale of numbers. The numbers should include the greatest value and the least value in the set of data.
4. For each piece of data, draw an X above the corresponding number.

Heights on Phoenix Mercury (in inches)

```
                                x
                                x
          x    x  x  x  x       x  x x   x       x
|---------------|---------------|---------------|---------------|
60             65              70              75              80
```

The mode for the data is 73 inches because 73 appears most frequently in the line plot. Since the maximum value is 80 and the minimum value is 64, the range is 80 – 64, or 16 inches. Both 64 and 80 are separated from the rest of the data, but not significantly. So there are no outliers.

Example 2

Twenty-five students in a school hallway were asked how many books they were carrying. The frequency table (Table 3) shows their responses. Display the data in a line plot. Then find the range.

Table 3

"How many books are you carrying?"	
Number	**Frequency**
0	3
1	7
2	6
3	2
4	4
5	3

Students Carrying Books

```
          x
          x         x
          x         x
          x         x                   x
x         x         x                   x         x
x         x         x         x         x         x
x         x         x         x         x         x
|---------|---------|---------|---------|---------|
0         1         2         3         4         5
```

The greatest value in the data set is 5 and the least value is 0. So the range is 5 - 0, or 5.

Example 3

Sixteen students estimated how much television they watched each week, to the nearest hour. Here are their results: 14, 16, 12, 14, 14, 11, 20, 12, 8, 10, 16, 15, 17, 5, 15, and 10. Show these results on a line plot. Then identify the mode of the data and any outliers that you see.

Hours of Television Watched

```
                                                      x
                              x           x           x     x     x
x                 x           x     x     x           x     x     x     x                 x
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
5                             10                            15                            20
```

The most X's are above 14. So, the mode is 14. Both 5 and 20 are separated from the rest of the data. So, 5 and 20 are outliers.

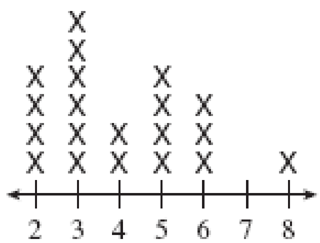
MCAS Problems

2005

http://www.doe.mass.edu/mcas/2005/release/g8math.pdf

2005 Grade 8 #31

31 The line plot below shows the number of people in each student's household for a class of students.



Number of People in Student Households

What is the mean number of people in households for this class of students?

A. 3

B. 3.5

C. 4

D. 6

$$\text{mean} = \frac{(4)(2)+(6)(3)+(2)(4)+(4)(5)+(3)(6)+(1)(8)}{4+6+2+4+3+1} = \frac{8+18+8+20+18+8}{20} = \frac{80}{20} = 4$$

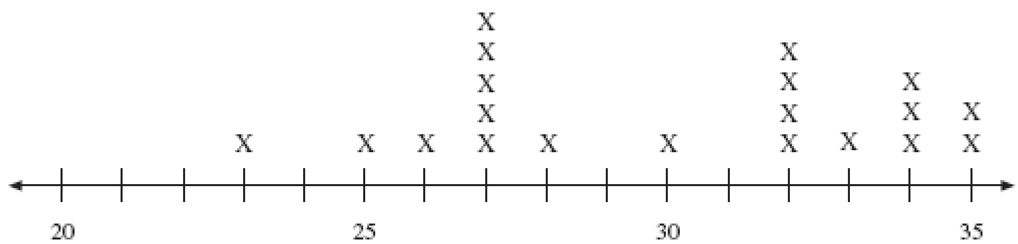
Answer: C

2003

http://www.doe.mass.edu/mcas/2003/release/g8math.pdf

2003 Grade 8 #15

15 The line plot below represents the number of raisins that Janika's class counted in each of 20 boxes of cereal.



What was the median number of raisins in a box?

A. 27

B. 29

C. 30

D. 31

Answer: D

Since there are 20 cereal boxes, the median is between the 10^{th} and 11^{th} boxes. Counting from the left, the 10^{th} box had 30 raisins and the 11^{th} box had 32 raisins. The median is the average of 30 and 32 which is 31.

Stem and Leaf Plots

A **stem-and-leaf plot** organizes data by showing the items in order. The leaf is placed as the last digit to the right. The stem is the remaining digit or digits, and is placed to the left of the vertical line.

Stem-and-leaf plots allow you to see easily the greatest, least, and median values in a set of data. They are very similar to histograms (see page 73) and grouped frequency tables (see page 67). Stem-and-leaf plots give you a quick way of checking how many pieces of data fall in various ranges. They also let you see something the other two displays don't: the value of every piece of data.

In this type of display (stem-and-leaf plot), the raw data values are incorporated into a frequency distribution. This method is illustrated for the following set of thirty scores.

83	71	92	79	74	80	63	86	84	74
100	81	94	98	79	62	50	82	56	67
75	86	83	96	57	100	87	67	98	44

First the stems, derived by using the ten's digit from each score, are written in order to the left of a vertical line. For each score the leaf, or one's digit, is then recorded to the right of the corresponding stem. For the score of 84, for example, the leaf 4 is recorded to the right of the stem 8. The leaves are separated by a space, using equal space for each leaf.

```
 4 | 4
 5 | 0 6 7
 6 | 3 2 7 7
 7 | 1 9 4 4 9 5
 8 | 3 0 6 4 1 2 6 3 7
 9 | 2 4 8 6 8
10 | 0 0
```

Example 1

Use the table below to construct a stem-and-leaf plot for the number of new shows per season. Then find the median, mode, and range.

Broadway Productions	
Season	**New Shows**
1993-1994	37
1994-1995	29
1995-1996	38
1996-1997	37
1997-1998	33
1998-1999	39
1999-2000	37
2000-2001	28

Choose the stems. For this data set, use the values in the tens place. Draw a line to the right of the stems.

2 |
3 |

Leaves are single digits, so for this data set the leaves will be the values in the ones place.

2 | 9 8
3 | 7 8 7 3 9 7

Arrange the leaves on each stem from least to greatest. Include a title and a key that shows how to read your stem-and-leaf plot.

New Shows per Season
2 | 8 9
3 | 3 7 7 7 8 9

Key: 2 | 8 means 28

Since the data items are in order, the median is the midpoint. The median is the mean of the fourth and fifth values, or 37. The mode corresponds to the most repeated leaf. The mode is 37. The range is the difference of the greatest and least values, or 11.

Example 2

As of 1997, the following are the ages, in chronological order, at which U.S. Presidents were inaugurated: 57, 61, 57, 57, 58, 57, 61, 54, 68, 51, 49, 64, 50, 48, 65, 52, 56, 46, 54, 49, 50, 47, 55, 55, 54, 42, 51, 56, 55, 51, 54, 51, 60, 62, 43, 55, 56, 61, 52, 69, 64, and 46.

Use a stem-and-leaf plot to help you summarize the data.

To make the stem-and-leaf plot:

1. Title your plot.

2. Write the data in order from least to greatest:

   42, 43, 46, 46, 47, 48, 49, 49, 50, 50, 51, 51, 51, 51, 52, 52, 54, 54, 54, 54, 55, 55, 55, 55, 56, 56, 56, 57, 57, 57, 57, 58, 60, 61, 61, 61, 62, 64, 64, 65, 68, 69

3. Find the least and greatest values: 42 and 69.

4. Choose stem values that will include the extreme values. For this graph, it makes sense to use tens: 4 tens, 5 tens, and 6 tens.

5. Write the tens vertically from least to greatest. Draw a vertical line to the right of the stem values.

6. Separate each number into **stems** (tens) and **leaves** (ones). Write each leaf to the right of its stem in order from least to greatest.

7. Write a key that explains how to read the stems and leaves.

Ages of U.S. Presidents when Inaugurated

4 | 2 3 6 6 7 8 9 9
5 | 0 0 1 1 1 1 2 2 4 4 4 4 5 5 5 5 6 6 6 7 7 7 7 8
6 | 0 1 1 1 2 4 4 5 8 9

Key: 4 | 2 represents 42 years.

You can see from the plot that:

~ all the U.S. Presidents entered office in their 40s, 50s, and 60s;
~ most of them were inaugurated during their 50s;
~ more Presidents were sworn in while in their 60s than in their 40s.

Example 3

Here are the top 40 earnings, in millions of dollars, from the Forbes 1998 Top 40 Entertainers list: 225, 200, 175, 125, 115, 77, 65, 58, 57, 56.5, 56, 55.5, 55, 54, 53.5, 53, 53, 52, 49.5, 49, 48, 47.5, 47, 45.5, 45, 44, 42, 41, 40.5, 40, 40, 38.5, 38, 37, 35, 34, 32.5, 32, 31, 28. Repeating the steps to create the stem-and-leaf plot:

1. Title your plot.

2. If the data are not already written in order, do so. For this example, we will order the stems from high to low.

3. Choose stem values that will include the extreme values. For this graph, it makes sense to use ten millions: 2, 3, 4,…,22.

4. Write the stems vertically from greatest to least. Draw a vertical line to the right of the values.

5. Separate each number into **stems** (ten millions) and **leaves** (millions). Write each leaf to the right of its stem in order from least to greatest.

6. Write a key that explains how to read the stems and leaves.

Earnings of the Top 40 Entertainers (in millions of dollars)

Stems	Leaves
22	5
21	
20	0
19	
18	
17	5
16	
15	
14	
13	
12	5
11	5
10	
9	
8	
7	7
6	5
5	2 3 3 3.5 4 5 5.5 6 6.5 7 8
4	0 0 0.5 1 2 4 5 5.5 7 7.5 8 9 9.5
3	1 2 2.5 4 5 7 8 8.5
2	8

You can see from the plot:

~ Most of the top earners earned between $30,000,000 and $50,000,000.
~ Only five entertainers earned more than $200,000,000.

A **back-to-back stem-and-leaf plot** uses two sets of data. The side-by-side display makes the data easier to compare.

Example 4

Draw a back-to-back stem-and-leaf plot for the winning times in the Olympic 100-m dash. Find each median and mode.

Winning Times, 100-m Dash (seconds)		
Year	**Men**	**Women**
1960	10.2	11.0
1964	10.0	11.4
1968	9.9	11.0
1972	10.1	11.1
1976	10.1	11.1
1980	10.3	11.1
1984	10.0	11.0
1988	9.9	10.5
1992	10.0	10.8
1996	9.8	10.9
2000	9.9	10.8

Use seconds for the stem and tenths of seconds for the leaves. Put the leaves in ascending order starting at the stem.

Winning Times, 100-m Dash

Men's Times (tenths of second)	Stem (seconds)	Women's Times (tenths of second)
9 9 9 8	9	
3 2 1 1 0 0 0	10	5 8 8 9
	11	0 0 0 1 1 1 4

The median of the times for men is 10.0 s. The median of the times for women is 11.0 s. The modes of the times for men are 9.9 s and 10.0 s. The modes of the times for women are 11.0 s and 11.1 s. The stem-and-leaf plot does not tell us the mean. A separate computation is required to show that the men's mean time is approximately 10.02 seconds and the women's mean time is approximately 10.97 seconds.

MCAS Problems

2006

http://www.doe.mass.edu/mcas/2006/release/g8math.pdf

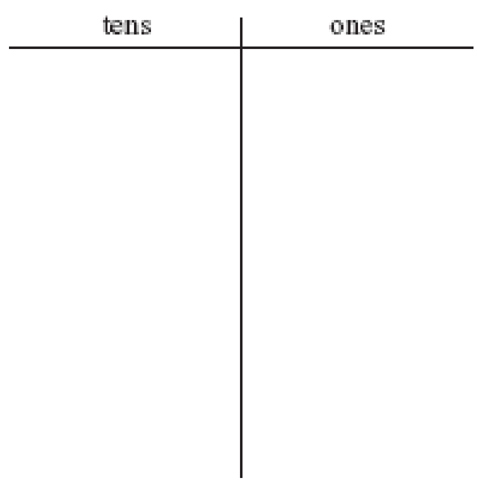
2006 Grade 8 #39

39 The individual weights, in pounds, of the members of a school's wrestling team are shown in the box below.

180	163	165	165
171	177	191	168
180	203	196	175
162	155	178	195

a. What is the range of the weights? Show or explain how you got your answer.

b. Copy the diagram below into your Student Answer Booklet. Use the diagram to make a stem-and-leaf plot of the data above. Be sure to title your plot and provide a key.



c. What is the median weight for the data in your stem-and-leaf plot from part (b)? Show or explain how you got your answer.

Answer for a: Put the weights in numerical order:

155, 162, 163, 165, 165, 168, 171, 175, 177, 178, 180, 180, 191, 195, 196, 203

The range is 230 lbs – 155 lbs or 48 lbs.

Answer for b: **Wrestling team in lbs.**

15 | 5
16 | 2 3 5 5 8
17 | 1 5 7 8
18 | 0 0
19 | 1 5 6
20 | 3

Key: 15 | 5 represents 155 lbs.

Answer for c: Counting the number of leaves, there are 16 and the two middle numbers are 175 and 177.

Find the mean of the two middle numbers to get the median. It is 176.

2006 Grade 8 #21

21 The stem-and-leaf plot below shows the number of laps walked by 15 students in a walk-a-thon.

Number of Laps Walked

1	2 4 4 6 7
2	0 0 0 3 5 8
3	0 2 6
4	2

Key
1 | 8 represents 18

What is the total number of students who walked more than 29 laps?

We need to count the number of leaves in the 3rd and 4th stem row since they represent numbers greater than 29.

Answer: 4

2005

http://www.doe.mass.edu/mcas/2005/release/g8math.pdf

2005 Grade 8 #3

3 The stem-and-leaf plot below shows the number of people using a skateboard park on 13 different days.

Number of Skateboard Park Users

3	0 2
4	2 3 5 6
5	1 4 4 6
6	1 2 4

Key
4 | 3 represents 43

What is the range of the data in the stem-and-leaf plot?

A. 29

B. 31

C. 32

D. 34

The range is difference between the largest number and the smallest number. The largest number is 64 and the smallest number is 30. So, 64 – 30 = 34.

Answer: D

2004

http://www.doe.mass.edu/mcas/2004/release/g8math.pdf

2004 Grade 8 #20

20 The stem-and-leaf plot below shows the ages of the people who bought skateboards at a store during a sale.

Ages of People

Stem	Leaf
1	1 3 4 5 5 6 6 6 8
2	0 1 7 8
3	9
4	3 6
5	
6	5 5
7	1

Key
6 | 2 represents 62

What is the range of ages of people who bought skateboards during the sale?

The range is difference between the largest number and the smallest number. The largest number is 71 and the smallest number is 11. So, 71 – 11 = 60.

Answer: 60

2003

http://www.doe.mass.edu/mcas/2003/release/g8math.pdf

2003 Grade 8 #8

8 The stem-and-leaf plot below shows the ages of the people who bought skateboards at a store during a sale.

Ages of People

Stem	Leaf
1	1 3 4 5 5 6 6 6 8
2	0 1 7 8
3	9
4	3 6
6	5 5
7	1

Key
6|2 = 62

What is the median age of the people who bought skateboards during the sale?

We find that there are 19 ages listed by counting the numbers in the leaf column. The median is the 10^{th} number since this is the middle of the 19 numbers. The 10^{th} number is the 1^{st} one in stem row 2 which is 20.

Answer: 20 years old

Scatterplots

A scatterplot is a graph that shows the relationship between two sets of data. To make a scatterplot, graph the data as ordered pairs.

Suppose we want to analyze two sets of data to see how closely they are related. One way to do it is to use a **scatterplot** (also called a **scatter diagram** or **scattergram**). On a scatterplot, we plot corresponding numbers from two sets of data as ordered pairs. When we examine a scatterplot, we should look at the pattern of the plotted points. If there is a pattern, note its pattern. It may be a straight line or some other type of curve such as a quadratic, an exponential, etc. A pattern may indicate that as one variable increases, the other variable increases or perhaps it decreases.

Researchers often wish to compare pairs of data, such as the following semester mathematics (M) and history (H) averages collected from twenty students.

Example 1

M	83	90	64	49	73	81	71	60	79	62	85	72	78	66	93	81	74	85	66	76
H	76	85	75	67	88	90	88	78	90	57	93	71	96	71	91	54	81	78	81	82

Scatterplots are often used to show researchers whether a relationship exists between two sets of measurements. Researchers can then base predictions on the patterns they observe in these relationships. Each pair of values in the table above can be plotted on a scatterplot to investigate the relationship between the mathematics and history averages.

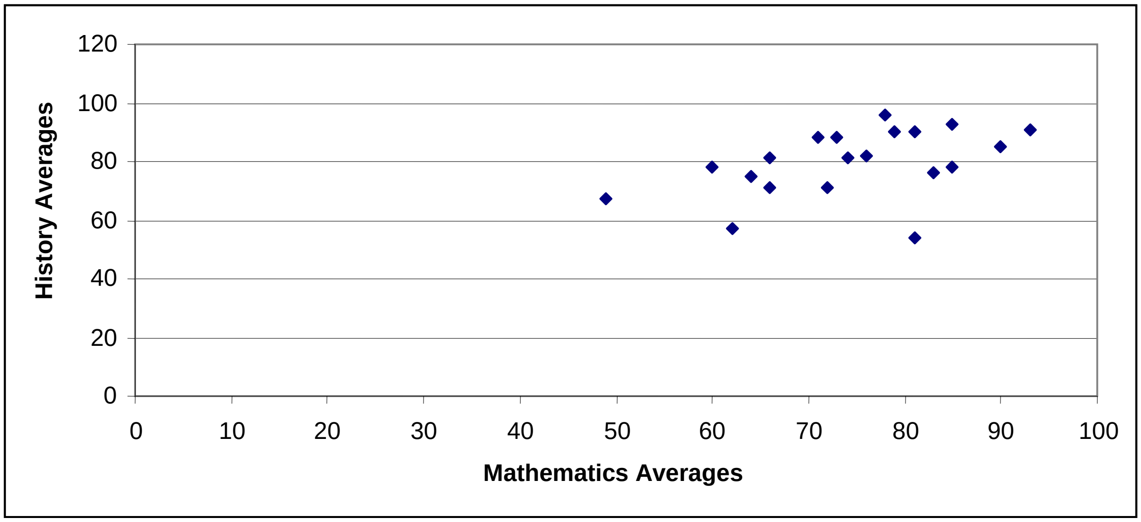


Figure 2

In Figure 2, the x axis represents Mathematics Averages and the y axis represents History Averages. There appears to be a positive linear trend to the data, i.e. in general as x increases, y also increases. There are also a few data points that appear not to follow the overall trend. These are called outliers and exist at (62, 57) and (81, 54). We will soon discuss correlation coefficients in general, but in the **Regression** section, we will discuss creating models for the data in this example and will examine the computed correlation coefficients to evaluate how well the models fit the data. In general, a correlation coefficient R can take on values $-1 \le R \le 1$. When R=0, there is no relationship between the *x* and *y* data. R=1 is a perfect positive correlation and R=-1 is a perfect negative correlation. We will see in the **Regression** section that correlations can be good whether the relationship between *x* and *y* is linear or nonlinear.

In Figure 3, we have created a scatterplot for data showing a relationship between time watching TV (the *x* axis) and time spent on homework (the *y* axis).

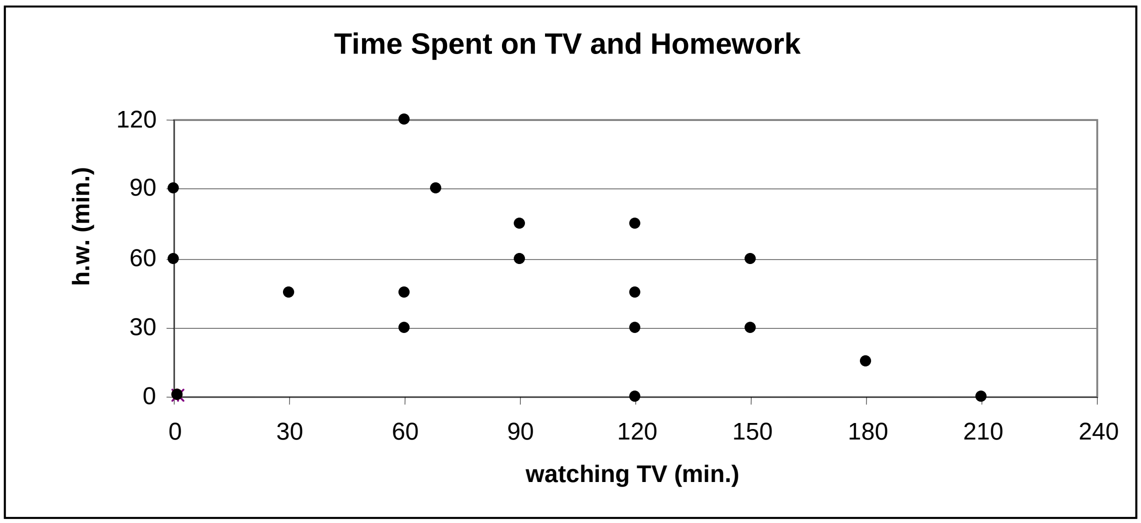


Figure 3

Example 2

Referring to the scatterplot in Figure 3:
a) How many student responses are shown on the scatterplot?
Answer: There are 18 student responses because there are 18 data points on the scatterplot.
b) How many students reported doing homework for exactly 30 minutes?
Answer: Homework time is shown by the distance above the horizontal axis. Three students did exactly 30 minutes of homework. They are shown by the three data points on the horizontal line for 30.
c) How many students watched at least 120 minutes of television?
Answer: “At least 120” means 120 or more. Television time is shown by distance to the right of the origin. Points showing 120 minutes of TV time are in a vertical line in the middle of the graph. Look for points on this line or to the right of it. There are 8 such points. So 8 students watched at least 120 minutes of TV.

Example 3

The following chart shows total public secondary-school enrollment in the United States. Draw a scatterplot of the data and plot years on the horizontal axis.

Year	Number of Students
1955	8,521,000
1965	15,504,000
1975	19,151,000
1985	15,219,000
1995	16,431,000

1. Title your plot.
2. Decide which set of numbers you will plot on each axis and label the axes.
3. Choose a scale for each axis.
4. Plot corresponding numbers as ordered pairs.

Number of Students in each Year of School

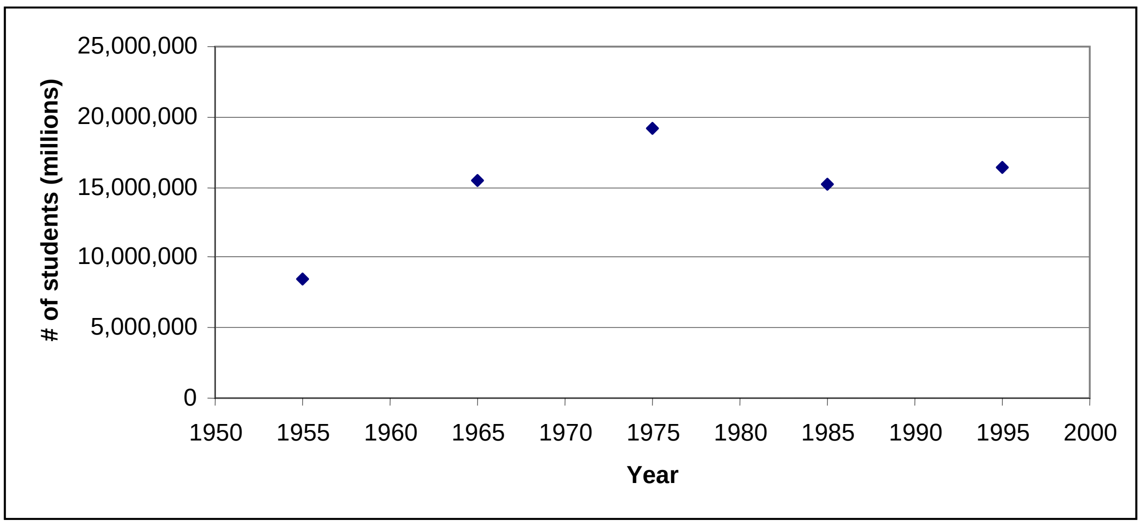
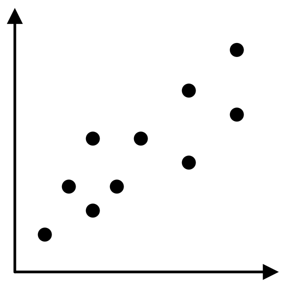


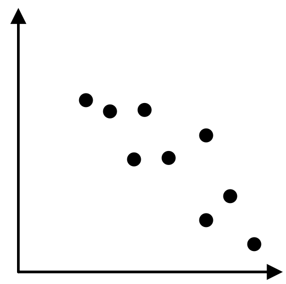
Figure 4

Correlation

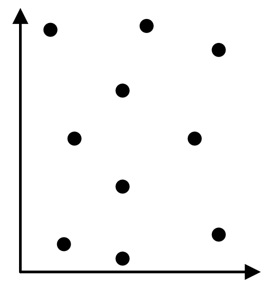
We can use scatterplots to look for trends in the data. The next three scatterplots show the types of relationships two sets of data may have.



Positive correlation
As one set of values increases, the other set tends to increase.



Negative correlation
As one set of values increases, the other set tends to decrease.



No correlation
The values show no relationship.

Example 4

Using the scatterplot in Figure 5, is there a positive correlation, a negative correlation, or no correlation between the years and the winning times? Explain.

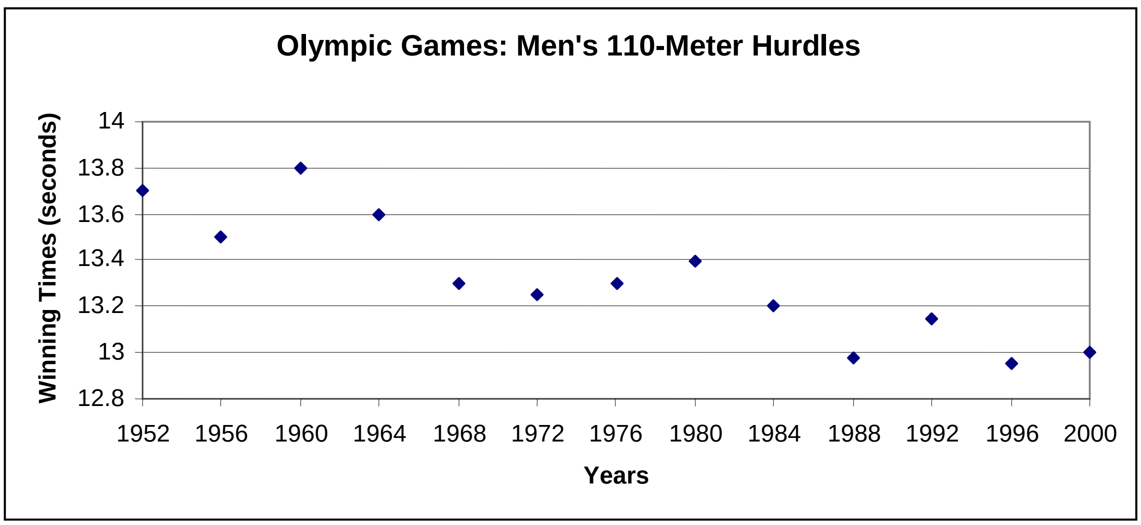


Figure 5

Answer: Since 1952, the winning times have generally decreased. There is a negative correlation.

Example 5

A student kept track of the number of hours she studied for tests and the grade on each test. The results are in the following table. Make a scatterplot for the data and then describe the relationship between the amount of time she studied and the grade she received.

Study Time (in hours)	1.5	1	3	2.5	1.5	4	3.5
Grade	75	71	88	86	80	97	92

As done in Example 3, make the scatterplot by following the procedure:

1. Title your plot.
2. Decide which set of numbers you will plot on each axis and label the axes.
3. Choose a scale for each axis. We are selecting study time to be *x* and grade to be *y*. In essence, the assumption here is that *x* is the independent variable and it is a predictor for the dependent variable *y*.
4. Plot corresponding numbers as ordered pairs. For example, in the table above, plot (1.5, 75), (1, 71), (3,88), and so on.

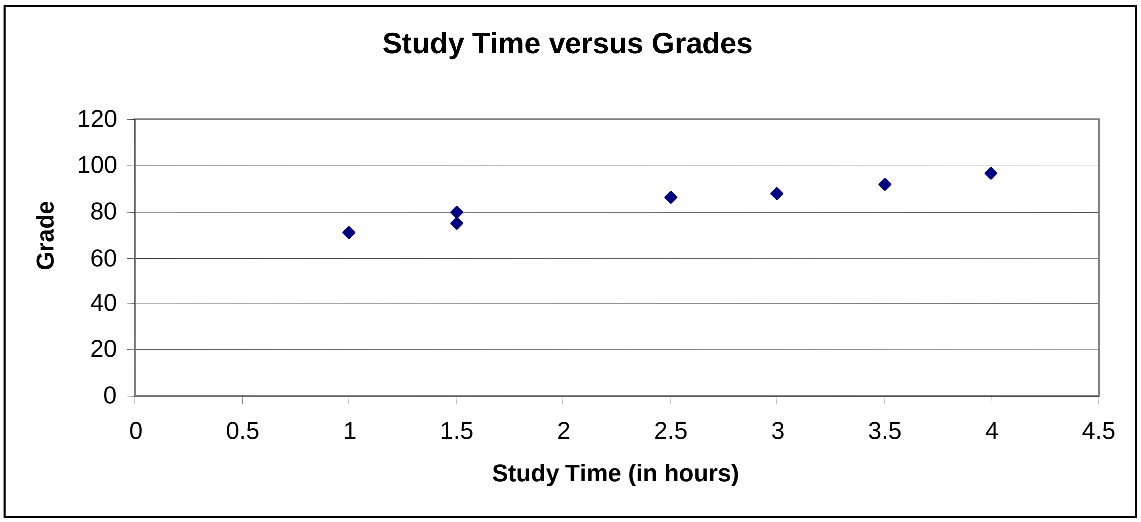


Figure 6

Answer: The dots on the scatterplot are close to forming a straight line. So, there is a strong positive correlation between the number of hours the student studied and her test grades. It seems that the more the student studied, the better the grades were.

MCAS Problems

2007

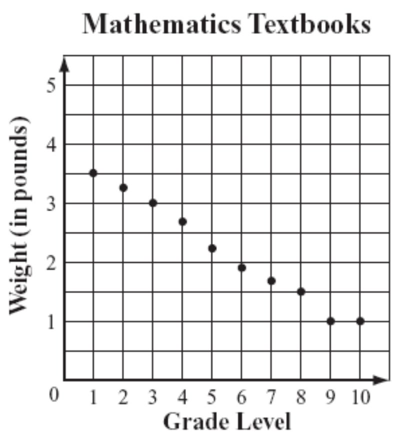
http://www.doe.mass.edu/mcas/2007/release/g8math.pdf

2007 Grade 8 #26

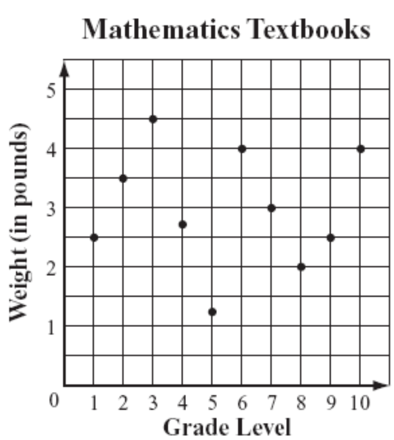
26 In David's school district, there is a positive correlation between the grade level and the weight of the mathematics textbook used by each grade.

Which of the following scatterplots best represents this correlation?

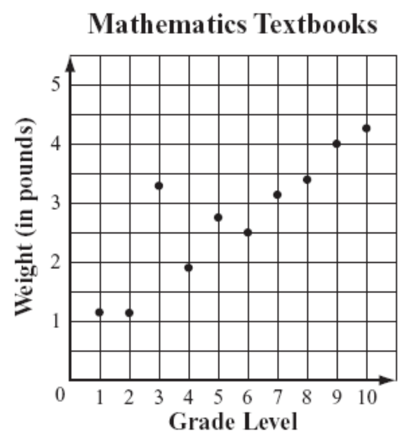
A.



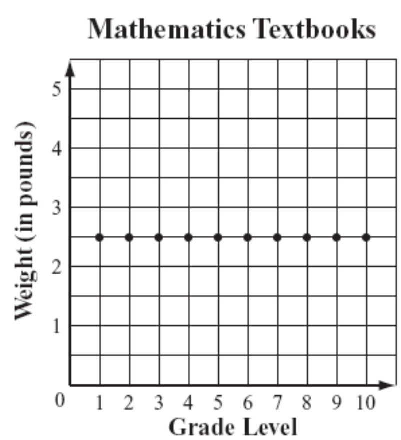
B.



C.



D.



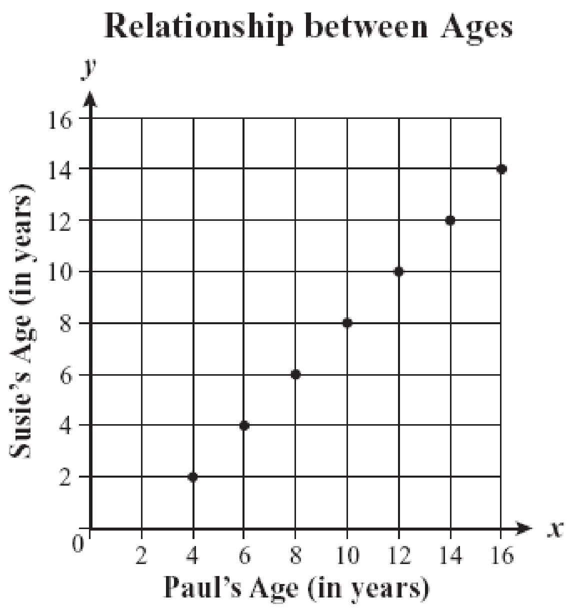
Answer: Although there is a single outlier for the data point at grade level 3, scatterplot C best shows a positive correlation.

2006

http://www.doe.mass.edu/mcas/2006/release/g6math.pdf

Grade 6 #3

3 The graph below represents the relationship between Paul's age and Susie's age.



Which of the following best describes the relationship between Paul's age and Susie's age for all the points shown on the graph?

A. Susie is twice as old as Paul.

B. Susie is 2 years older than Paul.

C. Susie is half as old as Paul.

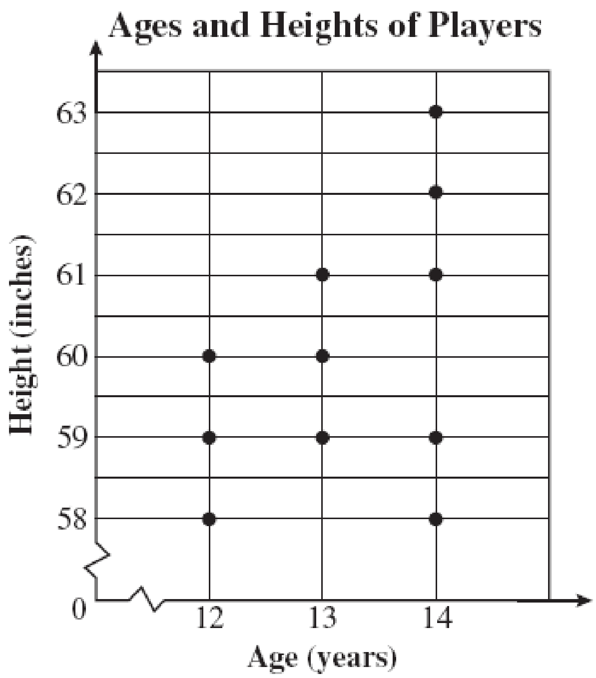
D. Susie is 2 years younger than Paul.

Answer: We are using a scatterplot in this example to look at the relationship between the variables x and y. For each of the data points (4,2), (6,4), and so on, Susie's Age which is the y value is 2 less than the x value. The correct answer is D.

2005

http://www.doe.mass.edu/mcas/2005/release/g8math.pdf
Grade 8 #11

11 The scatterplot below shows the ages and heights of 11 players on the school football team. Each dot represents one player.



What is the total number of 14-year-olds who are **more than** 60 inches tall?

A. 0

B. 2

C. 3

D. 5

Answer: Look first at the vertical line indicating 14-year-olds. Only 3 of them are more than 60 inches tall and the correct answer is C.

Box and Whisker Diagrams

A box and whisker diagram , sometimes referred to as a boxplot, is a graph of a data set that consists of a line extending from the minimum value to the maximum value, along with a box with lines drawn at the first quartile Q_1, the median, and the third quartile Q_3.

Just as the median divides sorted data into two equal parts, the quartiles Q_1, Q_2, Q_3 divide the sorted data into four equal parts. Q_1 separates the bottom 25% of the data from the top 75%, Q_2 is the median which separates the bottom 50% from the top 50%, and Q_3 separates the bottom 75% from the top 25%.

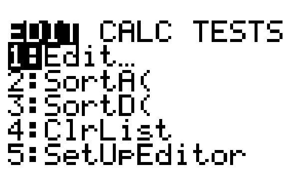
Example 1

Given the sorted data 1, 3, 6, 10, 15, 21, 28, 36, the median or Q_2 is 12.5 which is the average of the 4th and 5th vales, i.e. $\frac{10+15}{2}$. There are various ways to compute Q_1 and Q_3 which at times is confusing when using different types of technology. We will compute Q_1 as the median of the bottom half, i.e. the median of 1, 3, 6, 10 which results in Q_1=4.5. Likewise Q_3 is the median of the top half of the data, i.e. the median of 15, 21, 28, 36 which results in Q_3=24.5. These results are consistent with those generated by the TI-84.

Let's now draw a box and whisker diagram for the data in Example 1. We already know that Q_1=4.5, the median Q_2=12.5 and Q_3=24.5. The low or minimum value is 1 and the high or maximum value in the data set is 36.

Using a TI-84, enter Example 1 data in list 1 as follows:

1. Press the "STAT" key. The first entry "Edit" is already selected. Press the "ENTER" key.



2. Enter each score into L_1, pressing the down arrow after each score:

   1, 3, 6, 10, 15, 21, 28, 36

3. Then find the quartiles. Press the "STAT" key. Use the arrow to select "CALC" and press "ENTER" selecting "1-Var Stats". The first screen of statistics is displayed showing the mean or average.

4. Continue to press the down arrow and following screens will contain quantities that include the minimum, median and maximum. The output screens will tell us that for Example 1 data:

   a. The mean is 15

   b. Confirm the quartile measurements and the minimum and maximum values in the data set.

The TI-84 **does not require** that the initial data in L_1 be sorted. That is a nice feature. Now we can draw the box and whisker diagram:

1. Set equation Y_1 equal to L_1. Press "Y=" , then press "2ND" "STAT" to select LIST. L_1 is already selected. Press "ENTER".

2. We want to select the box and whisker option for the plot1. Go to "STAT PLOT" by pressing the "2ND" key followed by the "Y=" key. Press the "ENTER" key to select "Plot1".

3. Turn Plot1 On, for Type select the 4^{th} option (box and whisker), Xlist should be L_1 and Freq set to 1.

At this point, the Y_1 screen will have $Y_1 = L_1$ and Plot1 highlighted. Since we have already selected the box and whisker type of graph, select ZOOM->ZoomStat which is option 9, and then depress the GRAPH button. The output will appear as:

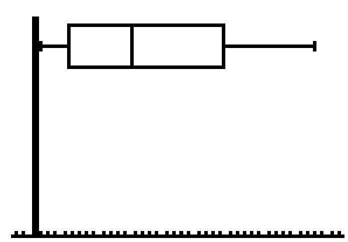


Figure 7

The two boxes in a box and whisker diagram show us how spread out the two middle quarters of the data are. The whiskers are the low and high vales, and they show us the spread of the bottom and top quarters.

Using the TRACE button, and repeated use of the left and right arrows, we can see all quartile measurements along with minimum and maximum values or whiskers. The Q_1 quartile measurement is shown in Figure 8:

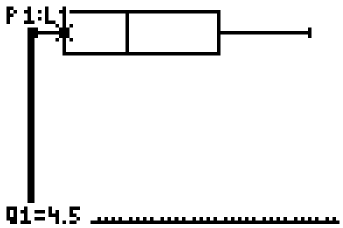


Figure 8

In Figure 7, Figure 8 and Figure 9, note the whiskers at the minimum and maximum values. Figure 9 displays the Q_3 quartile measurement. In the upper left hand corners of Figure 8 and Figure 9, P1:L1 refers the vertical line being the 1st percentile for the data list L1. In the next section, we define and discuss percentiles.

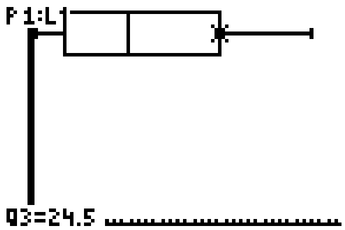


Figure 9

Examining any of the previous figures, we can see that there is a larger spread in the data from the median to the 3rd quartile than from the 1st quartile to the median. Also, the data spread from the 3rd quartile to the maximum is greater than the spread from the minimum to the 1st quartile. Obviously, data in the two higher quarters are more spread out than the two bottom quarters. Box and whisker diagrams are very useful in examining these types of spreads in data.

Without the use of the TI-84, we could also draw our own box and whisker diagram. Using Example 1 data, set up a horizontal axis to line up with the minimum and maximum values in the data set, i.e. from 1 to 36. Next draw a box with Q_1 at 4.5, the median Q_2 at 12.5 and Q_3 at 24.5. Put the whiskers at the minimum and maximum values and we get the following:

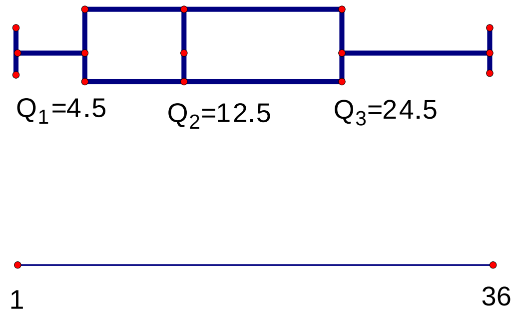


Figure 10

Setting up the appropriate scale for each of the quartile measurements in Figure 10 can be a bit demanding, so be sure to have an eraser handy. We used Geometer's Sketchpad to draw our box and whisker.

Deciles and Percentiles

Quartiles divide the data into fourths or quarters. Deciles group the data into tenths and there are nine values, denoted by $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$, which separate the sorted data into ten groups with **approximately** 10% of the values in each group. There are also 99 percentiles, denoted by $P_1, P_2, ..., P_{99}$, which separate the data into 100 groups with **about** 1% of the values in each group.

The first percentile, P_1, is the number that divides the bottom 1% of the data from the top 99%; the second percentile, P_2, is the number that divides the bottom 2% from the top 98%, and so forth. The second quartile, fifth decile, and fiftieth percentile of a data set are all the same and all equal the median; i.e. the median = $Q_2 = D_5 = P_{50}$.

The process of finding the percentile that corresponds to *x* is

$$\text{Percentile of value } x = \left(\text{number of values less than } x / \text{total number of values}\right) \times 100$$

Equating quartiles, deciles and percentiles, we can set up the table:

Table 4

Quartiles	Deciles
$Q_1 = P_{25}$	$D_1 = P_{10}$
$Q_2 = P_{50}$	$D_2 = P_{20}$
$Q_3 = P_{75}$	$D_3 = P_{30}$
	$D_i = P_{10i}$
	$D_9 = P_{90}$

Let's now do some computations to find deciles and percentiles with Example 1 data. As you will notice, there are a few issues to keep in mind.

Example 2

To find the 50th percentile or P_{50}, we want to multiply 0.50 (the percent) times the total number of values (n=8) in Example 1, i.e. P_{50} requires that we compute $0.50 * 8 = 4$. **Because this is an integer and we want the 50th percentile to be equal to the median, take the mean of the 4th and 5th values in the sorted data.** We then find that $P_{50} = (10+15)/2 = 12.5$. Note that $P_{50} = Q_2 = D_5$. In other words, the 50th percentile P_{50}

is the same as the median Q_2. From Table 4, we can see that 12.5 is also the 5th decile D_5.

Example 3

Find the value that separates the bottom 30% of the data from the top 70%. This requires that we find D_3. From Table 4, $D_3 = P_{30}$, and P_{30} requires that we compute 0.30*8=2.4. **Since this result is not an integer, always round up**. In this case round up to 3. The 3rd value in the data from Example 1 is 6 and $D_3 = P_{30} = 6$. Note that the decile and percentile computations are approximate, but certainly accurate enough to interpret large data sets.

As an exercise, show that P_1 in Figure 8 and Figure 9 is the 1st value in the data set for Example 1, i.e. P_1=1.
Answer: P_1 requires that we compute $0.01 \times 8 = 0.08$. Round this value up to 1 and the 1st value in the sorted data set is 1.

Example 4

Find the percentile corresponding to the value 28 in the Example 1 data. There are 6 values less than 28, so the

1. Percentile of 28= (number of values less than 28 / total number of values) * 100
2. Percentile of 28= (6 /8) * 100 = 75

When computing the percentile corresponding to a particular value in a sorted data set, always round the computed result. For example, if the percentile result were 62.5, we would round the result to 63 but if the result were 62.2, it would be rounded down to 62.

A summary of the issues to remember when performing decile and percentile computations are:

- When computing the percentile P_i**,** an integer result in the calculations requires us to take the mean of the ith and (i+1)st values.
- When finding a decile D_i**,** always round up if the interim result is not an integer.
- When computing the percentile corresponding to a particular value in a sorted data set, always round the computed result.

MCAS Problems

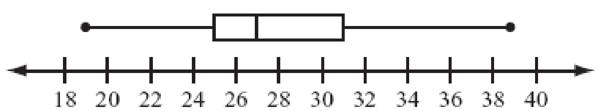
After each of the following MCAS problems, additional remarks are included for further study of the problems.

2007

http://www.doe.mass.edu/mcas/2007/release/g8math.pdf

2007 Grade 8 #33

33 Ms. Simmons made the box-and-whisker plot below to show some statistics about the ages of the students in her class at a community college.



Ages of Students (in years)

Which of the following best represents the median age of the students in her class?

A. 25

B. 27

C. 29

D. 31

Answer: Since the 2nd quartile Q_2 is at about 27, it best represents the median and the answer is B.

1. We could extend this problem and ask: “What values best represent the first and third quartiles?”
Answer: $Q_1 = 25$ and $Q_3 = 31$.
2. Another question that could be asked is: “Where is the data more spread out?”
Answer: There is a larger spread in the data from the median to the 3rd quartile than from the 1st quartile to the median. Also, the data spread from the 3rd quartile to the maximum (which is 39) is slightly larger than the spread from the minimum (which is 19) to the 1st quartile.

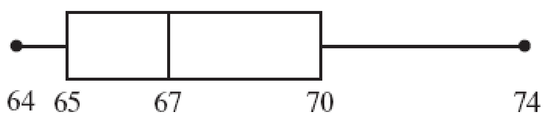
2004

http://www.doe.mass.edu/mcas/2004/release/g10math.pdf

2004 Grade 10 #30

30 The box-and-whisker plot shown below represents the heights, in inches, of the members of the Central High School girls' basketball team.

Girls' Heights in Inches



What is the median height of the members of the team?

A. 67 inches

B. 68 inches

C. 69 inches

D. 70 inches

Answer: The minimum value is 64, $Q_1 = 65, Q_2 = 67, Q_3 = 70$ and the maximum value is 74. The 2nd quartile Q_2 is the median and the correct answer is A.

Notice that the spread in this example from the median to maximum is much larger than the spread from the minimum to the median. In other words, the lower 50% of the heights are clustered from 64 to 67 inches, whereas the upper 50% of girls' heights extend from 67 to 74 inches.

2003

http://www.doe.mass.edu/mcas/2003/release/g10math.pdf

2003 Grade 10 #21

21 The highest possible score on a college admissions mathematics examination is 800. The stem-and-leaf plot shows the scores for a group of 20 students who were granted early admission to their chosen universities.

Stem	Leaf
68	0 5 5
70	0 0 5 5
71	5
73	0 0 5
74	0 5
75	0 5
78	0 0 5
80	0 0

Key
68 | 0 = 680

a. What is the median score for these 20 students? Show or explain how you obtained your answer.

b. What is the range of the scores for these 20 students? Show or explain how you obtained your answer.

c. What are the lower (first) quartile and the upper (third) quartile?

d. Make a box-and-whisker plot that displays the same data given in the stem-and-leaf plot above. Be sure to label the minimum, the lower quartile, the median, the upper quartile, and the maximum on your box-and-whisker plot.

Answer:

a) The sorted data is 680, 685, 685, 700, 700, 705, 705, 715, 730, 730, 735, 740, 745, 750, 755, 780, 780, 785, 800, 800. There are 20 values and the median is the average of the 10th and 11th data values, i.e. the median = $\frac{730+735}{2} = 732.5$.

b) The minimum value is 680 and the maximum value is 800. The range is 800 – 680 = 120.

c) The lower (first) quartile Q_1 is the median of the bottom half of the data. The median of the first 10 values is the average of the 5th and 6th values: Q_1 = $\frac{700+705}{2} = 702.5$. The third quartile Q_3 is the median of the top half of the data. Q_3 is the average of the 15th and 16th values and is equal to $\frac{755+780}{2} = 767.5$.

d) Using a TI screen capture and showing the lower quartile Q_1, we have

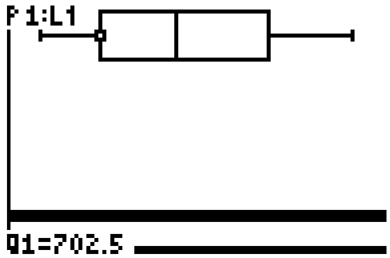


Figure 11

The left whisker or minimum is 680, and the median Q_2 is shown in Figure 12.

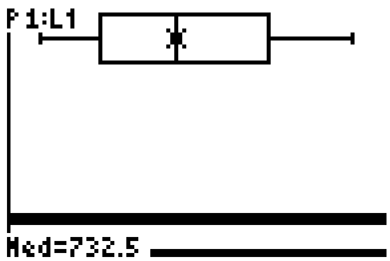


Figure 12

The upper quartile Q_3 which equals 767.5 is located at the right edge of the box. Finally, the right whisker or maximum is 800.

Since we have actual data in this problem, we can refer to the material on Deciles and Percentiles and ask the following:

1. Find the value that separates the bottom 40% of the scores from the top 60%. Answer: From Table 4 (page 58), $D_4 = P_{40}$ and P_{40} requires that since there are 20 data values, we compute 0.40*20=8. Since this is an integer, average the 8th and 9th values. Then, $D_4 = P_{40} = \frac{715+730}{2} = 722.5$, and the value 722.5 separates the bottom 40% from the top 60%.

2. Find the value that separates the bottom 62% of the scores from the top 38%. Answer: In this case, we want to evaluate P_{62}. This requires we compute 0.62*20=12.4. Since the result is not an integer, round up to 13. The 13th value in the data for this example is 745 and we can say that 62% of the scores are less than 745.

3. Find the percentile corresponding to the score 800. Answer: Corresponding to the score of 800, the
   - percentile= (number of values less than 800/total number of values) * 100
   - percentile= (18/20) * 100 = 90

   So for this example, a score of 800 puts the student in the 90th percentile.

Frequency Tables

A **frequency table** shows how often an item, a number, or a range of numbers occurs.

Here are two different ways (case 1 and case 2) to use a frequency table to show data. The one you use depends on the type of data you have to show.

Case 1 – A frequency table can show how often a particular item occurs. When you are counting responses, keep track of each response.

Example 1

A class participates in a survey in which they name their main source of news – TV, radio, Internet, newspapers, or magazines.

The answers to the survey are: TV, TV, newspaper, Internet, Internet, TV, magazines, Internet, Internet, TV, TV, radio, TV, newspaper, newspaper, Internet, radio, TV, magazines, TV, magazines, newspaper.

To make a frequency from the data, list each item in the data. Then count and record the number of times each item occurs. You may wish to use a tally to help you count.

Source	Tally	Frequency
TV	卌 | | |	8
Radio	| |	2
Internet	卌	5
Newspaper	| | | |	4
Magazines	| | |	3

Example 2

A teacher asks her class, "What's the main reason you use your computer: Internet, games, word processing, or something else?" The results are below.

Results: Internet, Internet, games, word processing, games, word processing, something else, word processing, Internet, games, games, Internet, games, Internet, word processing, something else, Internet, Internet.

To make a frequency table, first list each item in the data. Then count and record the number of times each item occurs.

Computer Use	Tally	Frequency
Internet	𝍸 𝍷𝍷	7
word processing	𝍷𝍷𝍷𝍷	4
games	𝍸	5
something else	𝍷𝍷	2

Example 3

A die was rolled 20 times. The results are shown below. Display the data in a frequency table.

5	2	5	4	1	6	5	2	5	1
3	6	1	3	4	5	3	5	3	4

~List the numbers on the cube in order.
~Use a tally mark for each result.
~Count the tally marks and record the frequency.

Number	Tally	Frequency
1	𝍷𝍷𝍷	3
2	𝍷𝍷	2
3	𝍷𝍷𝍷𝍷	4
4	𝍷𝍷𝍷	3
5	𝍸 𝍷	6
6	𝍷𝍷	2

Case 2 – Another way to make a frequency table is to group the data into intervals. This is called a **grouped frequency table**.

Example 4

Suppose the television show *Not You!* gets the following ratings over a 20-week period:

15.3, 12.2, 13.4, 15.6, 18.9, 20.2, 21.3, 22.4, 25.9, 20.1,
18.8, 17.4, 19.1, 17.2, 18.1, 19.3, 17.2, 14.3, 12.1, 15.3

To create a grouped frequency table:

1. Choose a scale that includes all the data (in this case 10-29.9).

2. Divide that scale into equal intervals (10-14.9, 15-19.9, etc.).

3. Count the number of data points in each interval. A tally can help you count the data.

Ratings for *Not You!*		
Interval	Tally	Frequency
10-14.9	||||	4
15-19.9	~~||||~~ ~~||||~~ |	11
20-24.9	||||	4
25-29.9	|	1

Example 5

Nolan Ryan holds the record for most career strikeouts by a major league pitcher. Ryan's season strikeout totals were: 6, 133, 92, 125, 137, 329, 383, 367, 186, 327, 341, 260, 223, 200, 140, 245, 183, 197, 209, 194, 270, 228, 301, 232, 203, 157, and 46.

To create a grouped frequency table:

1. Choose a scale that includes all the data (in this case, 6-383).

2. Divide that range into equal intervals (0-49, 50-99, 100-149, etc.).

3. Count the number of data points in each interval.

Nolan Ryan's Strikeouts		
Interval	Tally	Frequency
0-49	||	2
50-99	|	1
100-149	||||	4
150-199	~~||||~~	5
200-249	~~||||~~ ||	7
250-299	||	2
300-349	||||	4
350-399	||	2

MCAS Problems

2006

2006 Grade 10 #14

14 Cosmic Bowling Center has 100 bowling balls, and their weights range from 8 through 16 pounds. The frequency table below shows the number of balls by weight.

Number of Bowling Balls by Weight

Weight (pounds)	8	9	10	11	12	13	14	15	16
Number of Balls	13	4	9	3	10	2	18	17	24

What is the median weight per ball for the 100 bowling balls?

A. 11 pounds

B. 12 pounds

C. 13 pounds

D. 14 pounds

The median is the average of the weights of the 50^{th} and 51^{st} bowling balls. There are 41 balls in 8 pound through 13 pound intervals. The 14 pound interval has the 42^{nd} through the 59^{th} bowling balls. Since both the 50^{th} and 51^{st} balls are in this interval, the median is 14 pounds.

Answer: D.

2000

2000 Grade 8 #29

29. Chris selected 50 students at random and asked them who they want for class president. The results are shown in the table below.

Candidate	Frequency
Jessica	30
Jeremy	4
Monique	16

Which statement is true regarding the probability that at least 5 of the next 10 students interviewed will want Jeremy for president?

A. It is impossible.

✓ B. It is unlikely.

C. It is likely.

D. It is certain.

The fraction of the students poled who want Jeremy is $\frac{4}{50}$ or 0.08. The fraction not wanting Jeremy is $\frac{46}{50}$ or 0.92.

Answers A and D are not valid. An impossible outcome requires that the probability of selecting Jeremy is 0 (zero) for each student interviewed, and a certain outcome requires that the probability is 1 (one) for each student interviewed. The only remaining options are answers B and C. Based on the 50 student sample, the probability of an individual student selecting Jeremy is 0.08. If 5 of the next 10 students interviewed were to select Jeremy, the required computation to compute its probability would involve $(.08)^5$ which is a very small number. Likewise, the probability of 6 selecting Jeremy would involve $(.08)^6$ which is even smaller than the previous case. Continuing in this manner, the probability of at least 5 of the next 10 selecting Jeremy involves the sum of very small numbers, so its probability is <u>unlikely</u>.

Answer: B

Note: The probability of exactly 5 out of the next 10 students selecting Jeremy involves the binomial distribution and is $C(10,5)(.08)^5(.92)^5$. The term $C(10,5)$ is the number of combinations of 5 things out of 10. The probability of at least 5 of the next 10 selecting Jeremy is $\sum_{i=5}^{10} C(10,i)(.08)^i(.92)^{10-i}$. Using a TI-84 to evaluate this expression, the probability is 0.0005857, a rather unlikely event.

Histograms

Another way to describe the data in a frequency table is to draw a graph called a **histogram**. In a histogram, data are grouped into convenient intervals. Histograms are similar to bar graphs except that each bar on a histogram has the total frequency for an interval of data, while each bar on a bar graph has a value associated with each datum.

Example 1

The frequency table below consists of test scores with the number of students that received each score. The intervals in the histogram of these test scores are [60, 65), [65, 70), and so on. A test score of 75 is considered to be in the interval 75-80, while a score of 80 is in the interval 80-85. We have defined the intervals to be closed on the left and open on the right, but we could have made the intervals open on the left and closed on the right. Can you notice what effect this type of change would have on the histogram?

Thirty students received the following scores on a test:

77, 67, 91, 63, 81, 89, 73, 67, 85, 89, 89, 95, 83, 79, 89,

66, 83, 88, 97, 74, 77, 88, 89, 95, 79, 83, 95, 83, 98, 94

Each bar in the histogram below represents the number of scores for that particular interval of scores. Notice that 63 is the only score in the interval of 60-65 so that bar has a height of 1. Also, there are 8 scores in the interval of 85-90 so that bar has a height of 8.

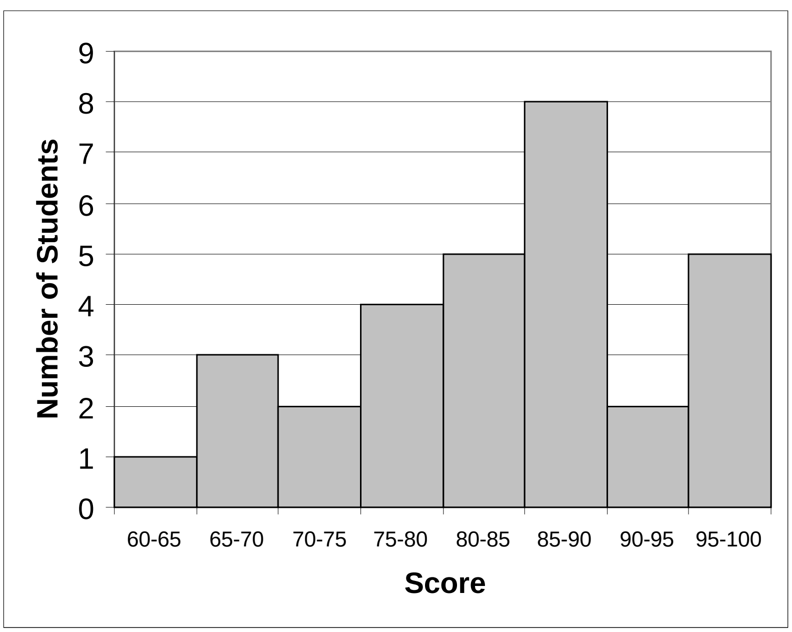


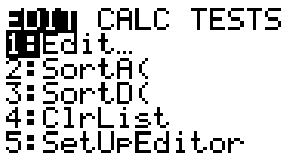
Table 5

Interval	Score	Number of Students
95-100	98	1
	97	1
	95	3
90-95	94	1
	91	1
85-90	89	5
	88	2
	85	1
80-85	83	4
	81	1
75-80	79	2
	77	2
70-75	74	1
	73	1
65-70	67	2
	66	1
60-65	63	1

Example 2

We can create the histogram for Example 1 on a TI-84 calculator. First, we must enter the scores and their frequencies into the lists L_1 and L_2 respectively:

1. Press the "STAT" key. The first entry "Edit" is already selected. Press the "ENTER" key.

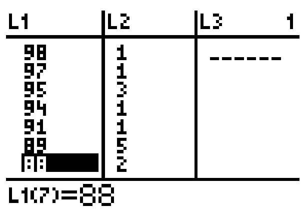


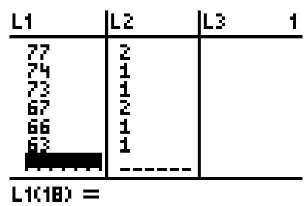
2. Enter each score into L_1, pressing the down arrow after each score:

   98, 97, 95, 94, 91, 89, 88, 85, 83, 81, 79, 77, 74, 73, 67, 66, 63

3. Enter the frequency for each score into L2, pressing the down arrow after each score. Make sure that each one is next to the correct score:

   1, 1, 3, 1, 1, 5, 2, 1, 4, 1, 2, 2, 1, 1, 2, 1, 1





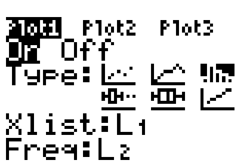
4. Now we can set up the plot for the histogram. Go to "STAT PLOT" by pressing the "2ND" key followed by the "Y=" key. Press "ENTER" to select "Plot1".

5. Select "ON" using the arrow key, then press "ENTER".

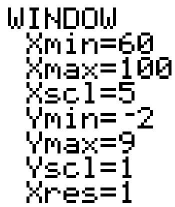
6. For "Type:" select the symbol that looks like a bar graph (third one).

7. For "Xlist:" input L_1 by pressing "2ND" "STAT" to select LIST. L_1 is already selected. Press "ENTER".

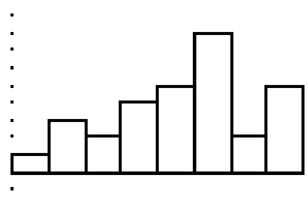
8. For "Freq:" input L_2 by pressing "2ND" "STAT" to select LIST. Select L_2 and press "ENTER". The settings should look like this:



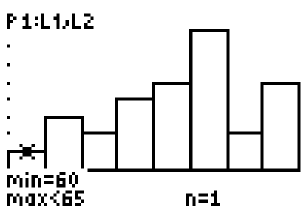
Now we can set up the formatting for the histogram. We want the X values to go from 60 to 100, the interval for X to be 5 and the maximum Y value to be 9. Press the "WINDOW" key and set each value to match the screen shot below:

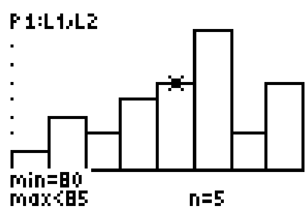


Now press the "GRAPH" key and the histogram will be created and should look like this:



Press the "TRACE" key and the cursor will appear at the top of the first bar. The interval is displayed, showing that min=60 and max=65. The number of scores is also displayed, showing that n=1. Move the cursor to the right using the arrow key, noting the change in the interval and the number of scores (n):



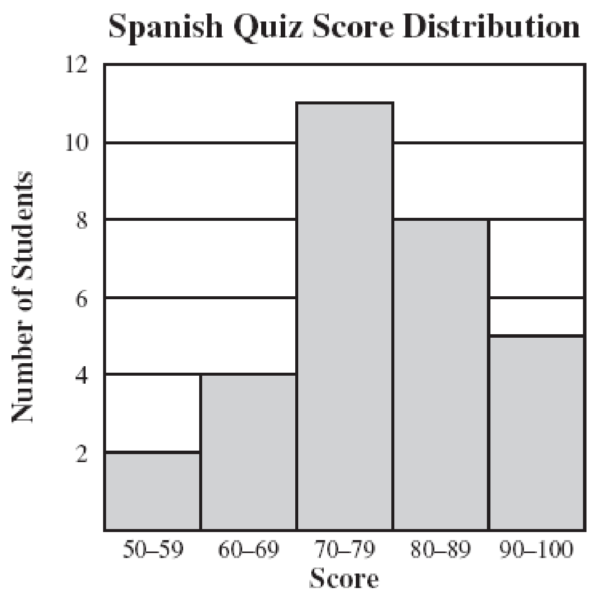


MCAS Problems

2006

2006 Grade 10 #19

19 Ms. Ruiz drew the histogram shown below on her board to display the score distribution for last week's Spanish quiz.



What fraction of the students received a score of 70 or more?

First we need to count the number of scores above 70. The number of students for the last three intervals is 24.

Interval	Number of Students
70-79	11
80-89	8
90-100	5
Total	24

Next we need to count the total number of students.

Interval	Number of Students
50-59	2
60-69	4
70-79	11
80-89	8
90-100	5
Total	30

We determine the fraction of students with score above 70 by placing 24 in the numerator and 30 in the denominator.

$$\frac{24}{30} = \frac{\cancel{2} \cdot 2 \cdot 2 \cdot \cancel{3}}{\cancel{2} \cdot \cancel{3} \cdot 5} = \frac{4}{5}$$

So the answer is $\frac{4}{5}$.

2004

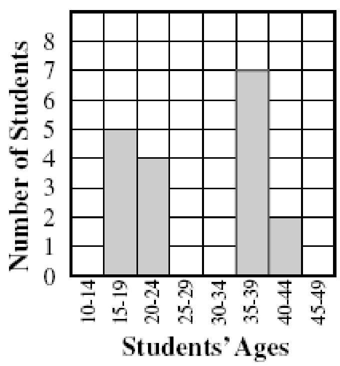
2004 Grade 10 #13

13 The chart below shows a random sample of students' ages at a community college.

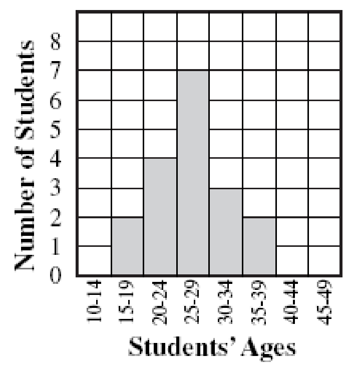
22	18	35	43	44	19
18	38	36	20	19	37
37	20	19	38	38	21

Administrators at the college constructed a histogram of the students' ages. Which of the following histograms **best** represents the distribution of students' ages?

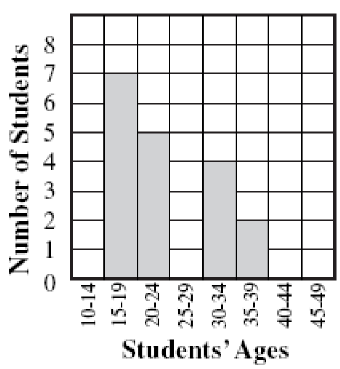
A.



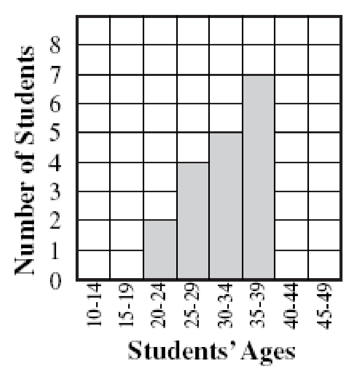
C.



B.



D.



Answer: A.

First we try to eliminate any wrong answers by comparing the data in the table to the first interval bar of each chart. Answers A, B and C have bars for the interval 15-19 but D does not. A quick glance at the table shows that there are ages listed between 15 and 19, so D is wrong.

A closer look at the table shows that in the interval 15-19 there are the ages 18, 18, 19, 19, and 19. Therefore the total number of students in this interval is 5. We see that answer A is the only histogram with 5 students in the interval of 15-19. With enough time, we can verify that answer A frequencies satisfy the rest of the students' ages.

Circle Graphs

A **circle graph** indicates a whole is broken into parts. Microsoft Excel facilitates creation of circle graphs with the creation of pie charts. Let's take a look at a brief example.

Example 1

We have data indicating the costs per month for various energy sources and want to create a pie chart indicating the percentage use of each energy source.

Table 6

Power	Cost/month
electricty	$100
gas	$125
propane	$50

Using Excel, highlight this data, under the Chart Wizard select the pie chart type, under the data labels option, select category name, values, and percentage. After creating the pie chart, save it as a new sheet. The completed pie chart will take on the following appearance:

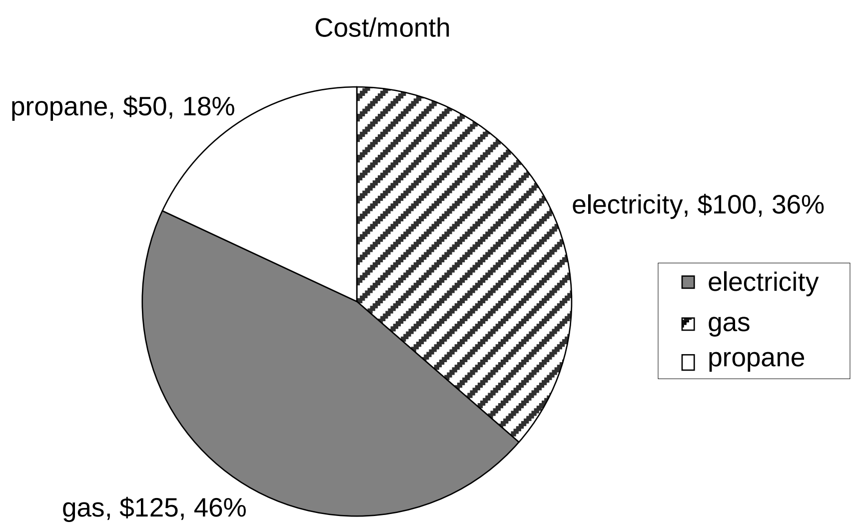


Figure 13

Mathematically, each central angle in Figure 13 can be computed as the appropriate percent times the total central angle which is 360°. The angle measurements are shown in Table 7.

Table 7

Power	Percent	Angle in degrees
Electricity	36	129.6
Gas	46	165.6
Propane	18	64.8

Without Excel, we can still create the circle graph manually. To compute the percentages as generated in Table 7, total the costs/month in Table 6 to get a total cost of $275 per month. The rounded percent of electricity cost in Table 7 is computed as a monthly cost divided by the total cost times 100%, i.e. $\frac{100}{275} \times 100\% = 36.34\%$ or 36% rounded for the electricity percent. Likewise, compute the percentages for gas and propane. Draw a circle with a compass and use a protractor to measure the computed angles.

Another technology based option to create the circle chart is with geometry software. The authors have used Geometers' Sketchpad. Knowing the results listed in Table 7, Sketchpad's output might look like:

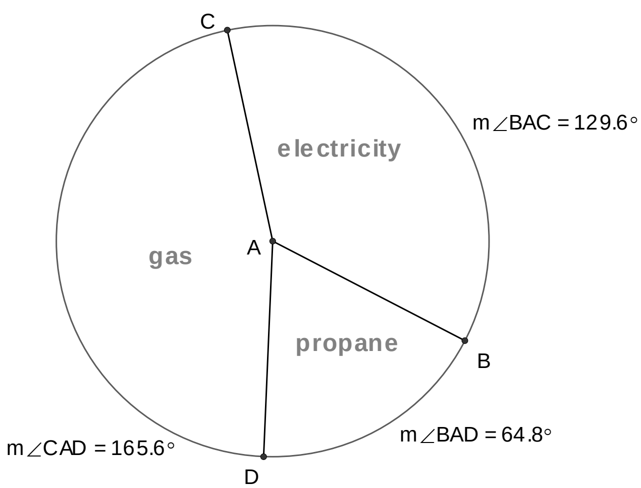


Figure 14

Example 2

Table 8 through Table 10 show various scenarios for the available data. Make a circle graph for each and compute the central angle for each data entry.

Case 1

Table 8

Farmland Usage in the United States 1987	
Use	% of U.S. Farmland
Cropland	46
Conservation Land	3
Woodland	3
Pasture	43
Other	5

An Excel generated chart is shown in Figure 15. Since cropland represents 46% of the farmland usage, its central angle is $\frac{46}{100} \times 360^{\circ} = 165.6^{\circ}$.

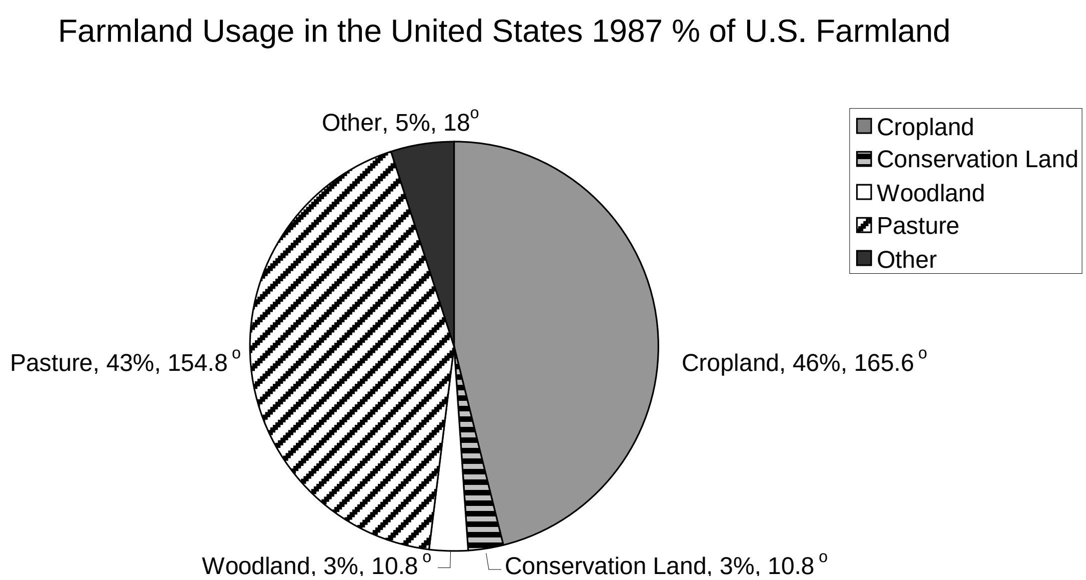


Figure 15

Case 2

Make a circle graph for Juan's weekly budget shown in Table 9 and compute the central angles for each budget item.

Table 9

Juan's Weekly Budget	
Recreation (r)	20%
Lunch (l)	25%
Clothes (c)	15%
Savings (s)	40%

A sample circle graph with central angles is shown in Figure 16. The central angle for Recreation is $20/100 \times 360^{\circ} = .2 \times 360^{\circ} = 72^{\circ}$. Likewise, compute the central angles for the other budget items.

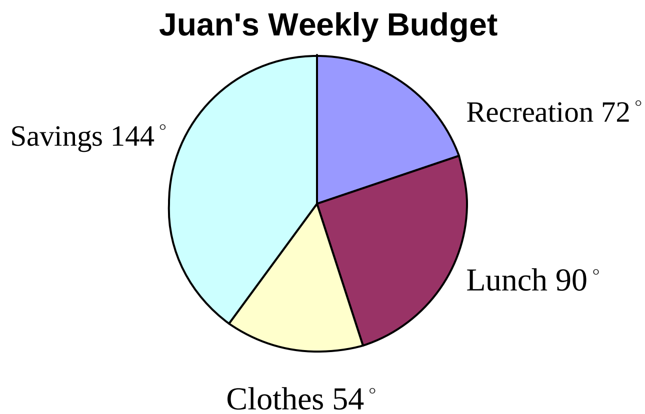


Figure 16

Case 3

Draw a circle graph of the data in Table 10.

Table 10

Site	Visits (millions)
Abraham Lincoln's Birthplace	0.3
Big South Fork	0.4
Cumberland Gap	1.3
Mammoth Caves	1.8

Follow the procedure outlined for Example 1. First add all entries to find that the total number of visits (in millions) is 0.3 + 0.4 + 1.3 + 1.8 = 3.8. Then compute the percentages for each entry and finally compute each central angle. Note that an alternate way to compute the central angles without computing the percentages first is to use proportions to find the measures of the central angles. So for Case 3:

$0.3/3.8 = a/360$	$0.4/3.8 = b/360$	$1.3/3.8 = c/360$	$1.8/3.8 = d/360$
$a \approx 28°$	$b \approx 38°$	$c \approx 123°$	$d \approx 171°$

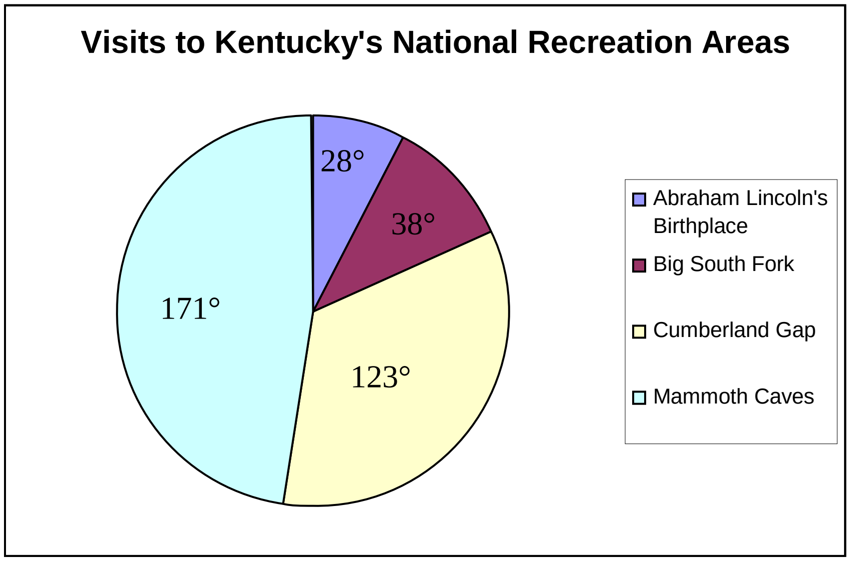

Visits to Kentucky's National Recreation Areas
28°
38°
171°
123°
Abraham Lincoln's Birthplace
Big South Fork
Cumberland Gap
Mammoth Caves

Figure 17

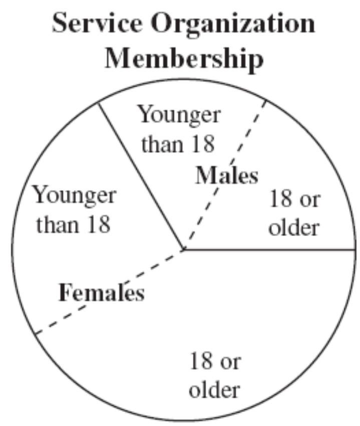
MCAS Problems

2005

http://www.doe.mass.edu/mcas/2005/release/g10math.pdf

2005 Grade 10 #18

18 The circle graph shown below represents the membership of a service organization. In this organization, $\frac{2}{3}$ of the members are female.



Approximately what fractional part of the total membership consists of males who are 18 or older?

Answer: For the females, the circle graph angle is $\frac{2}{3} \times 360^{\circ} = 240^{\circ}$. The males in the circle graph represent the remaining 120° which appears to be evenly split between the 18 or older group and those younger than 18. The angle represented by the 18 or older group is therefore 60° and the proportion of the total circle represented by the males 18 or older group is $\frac{60^{\circ}}{360^{\circ}} = \frac{1}{6}$.

2005 Grade 10 #23

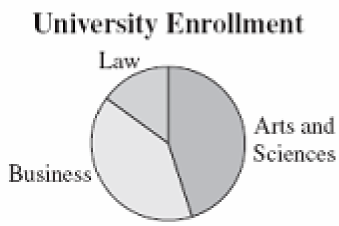
23 A local university is divided into three colleges. The table below shows the number of students enrolled in each college.

University Enrollment

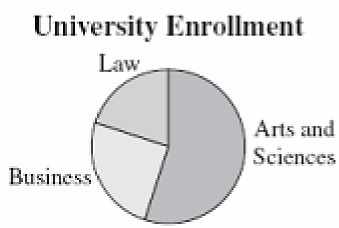
College	Number of Students
Arts and Sciences	8036
Business	2977
Law	1014

Which of the following circle graphs best represents the data in the table?

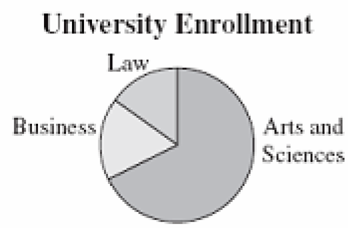
A.



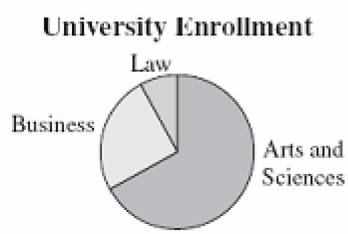
B.



C.



D.



Answer: There are 12027 total students in the university. Using the discussion in this section that appears in Case 3 and Figure 17, the proportions for each college are: Arts and Sciences - $\frac{8036}{12027} = \frac{a}{360}$ and $a \approx 241^\circ$, Business - $\frac{2977}{12027} = \frac{b}{360}$ and $b \approx 89^\circ$, Law - $\frac{1014}{12027} = \frac{c}{360}$ and $c \approx 30^\circ$. Circle graph D best represents the angle measurements.

2004

http://www.doe.mass.edu/mcas/2004/release/g8math.pdf

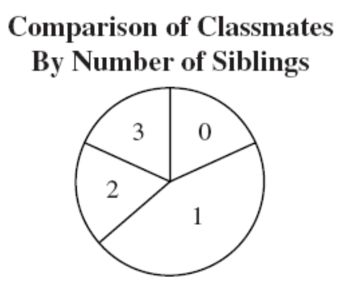
2004 Grade 8 #11

11 Stephanie conducted a survey to determine the number of siblings that each of her classmates has. The data appears in the table below.

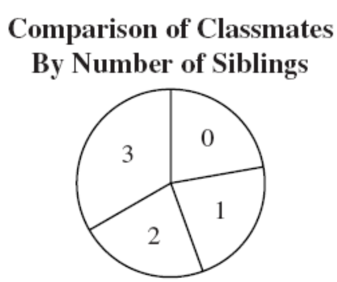
Number of Siblings	Classmates Having That Number of Siblings
0	𝍸 I
1	𝍸 III
2	𝍸 𝍸 I
3	𝍸

Which of the following circle graphs best represents the data in the table?

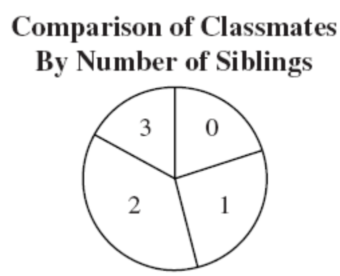
A.



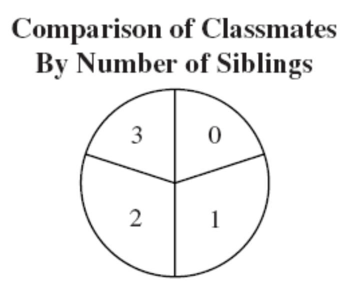
B.



C.



D.

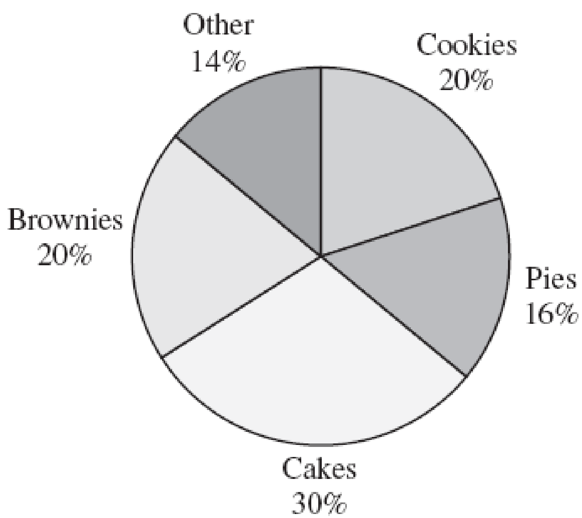


Answer: In the frequency table for this problem, there are 30 classmates in all. The percentages for number of siblings are: for 0 siblings, 6/30 implies 20%, 1 sibling implies 8/30 or 27% rounded, 2 siblings implies 11/30 or 37% rounded, and 3 siblings implies 5/30 or 17% rounded. Note that due to rounding the total is not 100%, but that is okay. The rounded central angles for these percentages are: 0 siblings – 20% (0.2×360) results in 72 degrees, 1 sibling – 27% yields 97 degrees rounded, 2 siblings – 37% yields 133 degrees rounded, and 3 siblings – 17% yields 61 degrees rounded. Again due to rounding, the total is not exactly 360 degrees, but for the sake of answering this problem, the estimates are adequate. Answer C is the correct choice since only in that case are the central angles for 1 and 2 siblings larger than 90 degrees with the angle for 2 siblings greater than the angle for 1 sibling.

2004 Grade 8 #24

24 The circle graph below shows the percentages of people who brought each type of baked good to sell at a recent bake sale.

Percentage of People Bringing Baked Goods



If 15 people brought cakes to sell, what is the total number of people who brought baked goods to sell at the bake sale?

A. 45

B. 50

C. 70

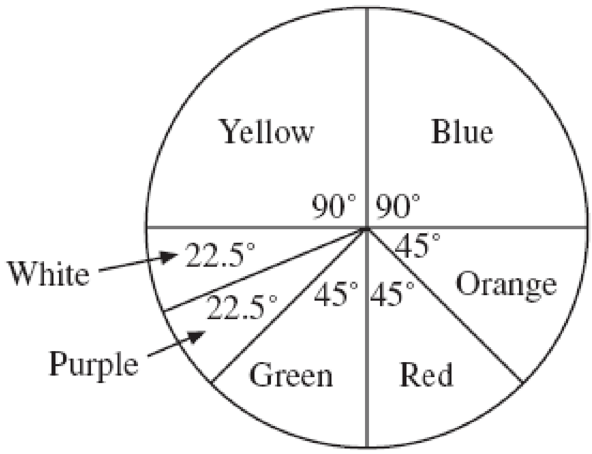
D. 75

Answer: 0.30x=15 where x is the total number of people. Solving, we get x=50 and the correct answer is B.

http://www.doe.mass.edu/mcas/2004/release/g10math.pdf

2004 Grade 10 #29

29 The circle graph below shows the colors of 160 marbles.



What is the total number of green marbles?

A. 8

B. 20

C. 40

D. 45

Answer: Using the discussion in this section that appears in Case 3 and Figure 17, the proportion of green marbles is $\frac{45^{\circ}}{360^{\circ}} = \frac{1}{8}$. We can form the equation $x = \frac{1}{8} * 160$ to find the total number x of green marbles and x=20. The correct answer is B.

Venn Diagrams

As an extension of the circle graphs, we can discuss intersecting circles. A **Venn diagram** consists of intersecting circles and shows relationships among sets of data or objects. Each circle is named for the data in it. Data that belong in more than one circle reside where the circles intersect or overlap.

Example 1

School coaches plan to send notices to all students playing a fall or winter sport. The data appears in Table 11. How many notices do they need to send?

Table 11

Season	**Students**
Fall	155
Winter	79
Both Fall and Winter	28

Creating a Venn diagram for this data, we find the situation shown in Figure 18.

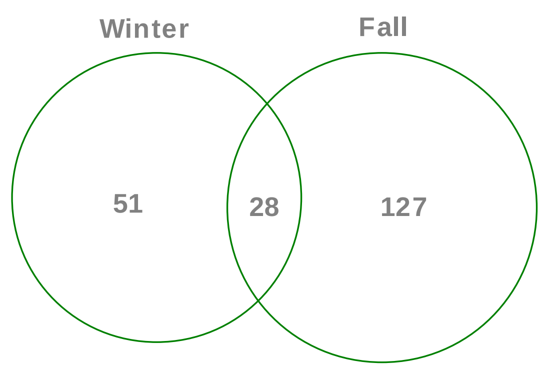


Figure 18

Enumerating the number of notices to be sent in each part of Figure 18, 51 play only a winter sport, 28 play both winter and fall sports, and 127 play only a fall sport. There are 206 notices that need to be sent out.

We can also use a Venn diagram to find the Greatest Common Factor (GCF) of two numbers.

Example 2

Find the GCF of 30 and 42.

The prime factors of each number can be written as 30 = 2 x 3 x 5 and 42 = 2 x 3 x 7.

Looking at a Venn diagram for this situation in Figure 19,

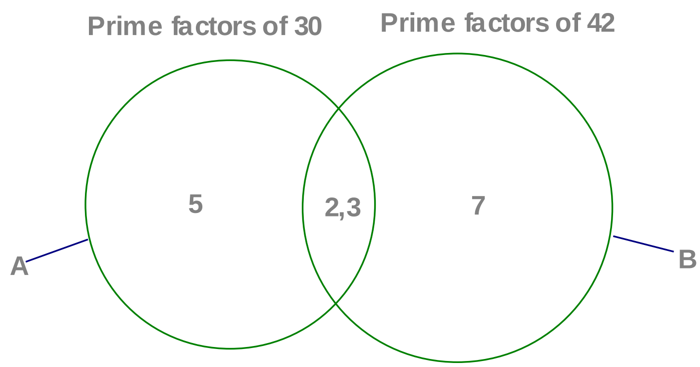


Figure 19

we can see that the prime factors 2 and 3 are common to both circles. In other words, the GCF is 2×3, or 6.

The **intersection** of sets A and B, written A∩B, is the set of elements that are in both A and B. In terms of Example 2, let's refer to set A as the prime factors of 30 and set B as prime factors of 42. Then A∩B = {2, 3} which means that A∩B is made up of the collection of elements 2 and 3.

The **union** of sets A and B, written as $A \cup B$, is the set of elements in either A or B (or in both). For this example, $A \cup B$ = {5, 2, 3, 7}.

The **complement** of set A, written as $\overline{A}$, is the collection of elements in $A \cup B$ but outside A.

Mathematically, $A \cap \overline{B}$ represents the set of elements in set A and not in set B. In Example 2, $A \cap \overline{B}$={5} and $\overline{A} \cap B$={7}.

Example 3

Let A = {1, 3, 5, 7, 9, 11} and B = {1, 4, 7, 10}. Find $A \cap B$, $A \cup B$, $A \cap \overline{B}$, and $\overline{A} \cap B$.

Answer: $A \cap B$={1, 7}; $A \cup B$={1, 3, 4, 5, 7, 9, 10, 11}; $A \cap \overline{B}$={3, 5, 9, 11}; $\overline{A} \cap B$={4, 10}.

Try to make a Venn diagram for Example 3.

There is a set that has no elements in it. It is called the empty set, or null set. The symbol { } or symbol Ø can be used to refer to this set. The empty set might refer to many things, including

- the set of whole numbers between 11 and 12;
- the set of U.S. presidents under 35 years of age;
- the set of all living dinosaurs.

Example 4

Let *E* = the set of even integers and O = the set of odd integers. Find E∩O.

Answer: *E* = {…, -6, -4, -2, 0, 2, 4, 6, …} and *O* = {…, -5, -3, -1, 1, 3, 5, …}
No integer is both odd and even, so the intersection $E \cap O$ = { }.

Rather than only consider two circles in the Venn diagrams, let's now look at three overlapping circles.

Example 5

There are 50 students in a survey of high school juniors, and most are enrolled in three courses: Biology, Math and English. Table 12 shows the number of students enrolled in each of the courses.

Table 12

Course	**Number of Students**
Biology	19
English	22
Math	24
Math and English	15
Math and Biology	11
Biology, English and Math	5

A Venn diagram for the data in Table 12 is shown in Figure 20.

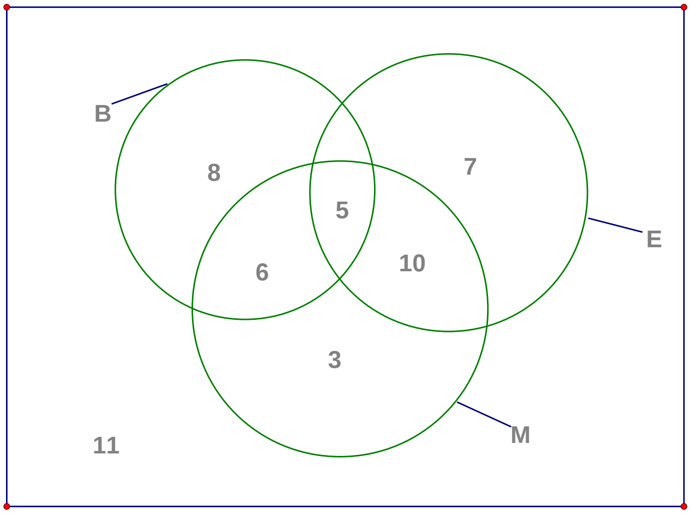


Figure 20

Confirm from Figure 20 that set B for Biology has 19 students, set E for English has 22 students, and set M for Math has 24 students. $M \cap E$ has 15 students, $M \cap B$ has 11 students and $B \cap E \cap M$ has 5 students. Since all Biology, English and Math criteria in Table 12 are satisfied, there aren't any students in $\overline{M} \cap B \cap E$, i.e. $\overline{M} \cap B \cap E$ ={ } or the null set. Finally, to satisfy the criteria that 50 students were surveyed, there are 11 students in $\overline{M} \cap \overline{B} \cap \overline{E}$.

In the Fractions document mentioned in the Acknowledgements and Preface page, we developed probability material as applications to fractions. As a refresher, the probability of an event occurring is the number of favorable outcomes divided by the total number of outcomes. Let's now introduce some probability questions related to Example 5 and Figure 20.

Example 6

Assume that the Venn diagram in Figure 20 is representative of an entire junior class at a high school.

1. What is the probability that a student selected at random is not taking any of the subjects Biology, English or Math?

   Answer: We want the probability of set $\overline{M} \cap \overline{B} \cap \overline{E}$ and denote this as P($\overline{M} \cap \overline{B} \cap \overline{E}$)=11/50 or .22.

2. What is the probability that a student selected at random is taking all three subjects?

   Answer: In this case, evaluate P($B \cap E \cap M$)=5/50 or 1/10.

3. What is the probability that a randomly selected student is taking at least two of the three courses?

Answer: In this situation, we have to consider:

$P[(B \cap E \cap M) \cup (B \cap \bar{E} \cap M) \cup (\bar{B} \cap E \cap M)] = \frac{5+6+10}{50} = \frac{21}{50}$ or .42.

Note that normally, when computing the probability to solve this problem, we would also have to add the $P\left(B \cap E \cap \overline{M}\right)$. However, we omitted it in this example since there aren't any students in that set.

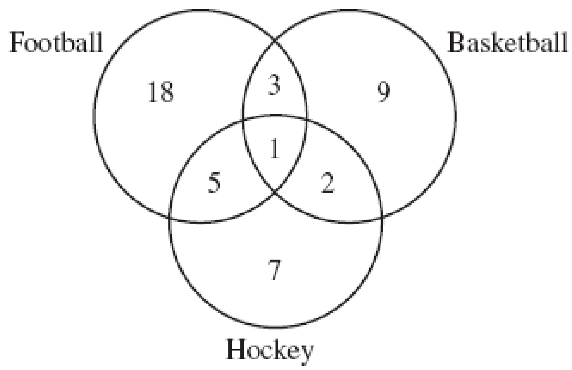
MCAS Problems

2005

http://www.doe.mass.edu/mcas/2005/release/g8math.pdf

2005 Grade 8 #18

18 Coach Wilson constructed a Venn diagram that shows the number of eighth-grade athletes who play football, basketball, and hockey.



Which phrase best identifies the number 5 shown in the diagram?

A. the total number of athletes who do not play all three sports

B. the total number of athletes who play both football and hockey, but not basketball

C. the total number of athletes who play either football or hockey, but not both

D. the total number of athletes who do not play basketball

Answer: B is the correct answer, i.e. the number 5 represents those students who play both football and hockey but do not play basketball. Using set notation with F representing football, B representing basketball and H representing hockey, the number 5 represents the set $F \cap H \cap \overline{B}$.

2004

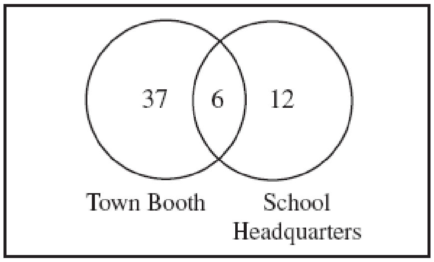
http://www.doe.mass.edu/mcas/2004/release/g8math.pdf

2004 Grade 8 #9

9. Last weekend, Lauren helped organize some students to participate in a fundraiser for charity. The students had a choice of working one shift at the information booth in town or one shift at the school headquarters. Students could also choose to work 2 shifts, one in town and one at school.

After the fundraiser, Lauren prepared a report for the school board. In her report, she drew the Venn diagram below to show where the students worked.

Students Working at the Fundraiser



a. Based on the Venn diagram, how many students worked shifts at the Town Booth?

b. Based on the Venn diagram, how many students participated in the fundraiser?

Answer: a) 43 worked at the town booth
b) 55 worked at the fundraiser

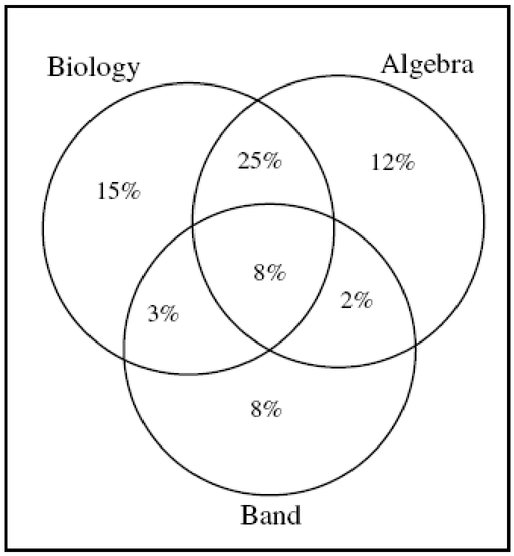
2003

http://www.doe.mass.edu/mcas/2003/release/g8math.pdf

2003 Grade 8 #9

9 At student registration, eighth-grade students selected the courses they would be taking next year as ninth graders. The counselor made the diagram below that shows a relationship among the percentages of students who chose to take Biology, Algebra, and/or Band.

Student Registration



a. According to the diagram, what percent of the eighth-grade students will be taking **all three** courses, Biology, Algebra, and Band, next year?

b. What percent of the eighth-grade students will be taking Algebra and Biology, but **not** Band, next year?

c. If 900 students signed up to take courses, how many will **not** be taking Biology, Algebra or Band? Show or explain your work.

Answer: a) 8% are in the intersection of all three circles

b) 25% are taking Biology and Algebra but not Band

c) In the union of all three circles, 73% of the students are accounted for. Therefore, 27% of the students will not be taking Biology, Algebra or Band. Of the 900 students who signed up to take courses, .27*900=243 will not take any of the three mentioned.

Regression

A least squares regression curve produces a model that minimizes the sum of the squares of the errors between the actual and modeled data. The mathematics behind the generation of the least squares models involves calculus and in particular partial derivatives. We will focus on applications of least squares to data sets and interpretation of the goodness of the models.

Using the data from Example 1 in the Scatterplots section (page 47), a linear least squares line generated on a TI-84 is shown in Figure 21:

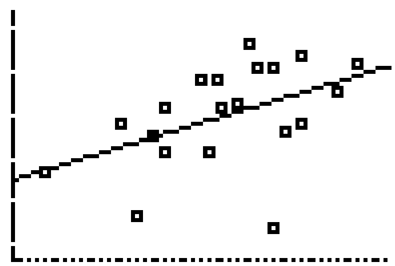


Figure 21

The generated model and correlation coefficient are displayed in a copy of the TI screen appearing in Figure 22.

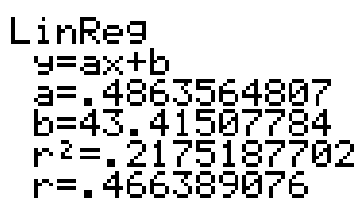


Figure 22

The procedure used to generate a linear least squares regression model on the TI is:
We will model the data with a linear regression model $y = ax + b$.

1. STAT, EDIT to create the data. Using Example 1 data described above, list 1, L_1, will contain the math averages and list 2, L_2, will contain the history averages.

2. 2nd QUIT, 2nd CATALOG, DiagnosticOn, enter (twice). This only has to be done the initial time a least squares regression is run on your TI-84 and its purpose is to enable output of the correlation coefficient statistic.

3. STAT, CALC, LinReg(ax+b)

4. LinReg(ax+b) L_1, L_2, y_1 enter. This statement generates a linear model using L_1 for the *x* data, L_2 for the y data, and automatically places the generated model in y_1 at the function screen so we can plot the generated curve. To generate y_1, press the VARS button, followed by Y-VARS, 1:Function and then 1: y_1.

5. 2nd STAT PLOT, option 1 , on. Be sure to select the first option for Type which is the scatterplot.

6. ZOOM, option 9. This option automatically set the x and y scales for the plot and will plot the actual data and generated model.

7. <u>turn off</u> plot 1 at y=screen or at STAT PLOT to enable further use of normal graphing capabilities on the TI.

Let's discuss the model generated in Figure 22 and displayed in Figure 21. Rounding to 2 decimal places, the slope for the linear model is a=.49 and the y intercept is b=43.42. The correlation coefficient is R=0.466 (printed on the TI screen as *r=.466…*). Since R is positive, it indicates a positive correlation in the data. However, R is not very close to the value 1 and this is due to the noise or outliers in the data. Statistically, R^2 represents the variance accounted for in the model and in this case $R^2 \approx 0.22$ or 22%. Generally, for scientific data, $R^2 \geq 90\%$ is considered acceptable whereas for social science data as we have in this example, $R^2 \geq 50\%$ would be good.

We can attempt to improve the R^2 in several ways. Let's first remove the 2 outliers at (62, 57) and (81, 54) since they are not in sync with the rest of the data, and repeat the linear least squares regression. In this case we get the model

```
LinReg
 y=ax+b
 a=.4720174575
 b=47.00758443
 r²=.3903293227
 r=.6247634134
```

Figure 23

The R^2 has improved to 39%. Note that R has also improved from 0.466 to 0.625. Now, with the outliers removed, let's try to see if a quadratic model will improve results further. On the TI, the only difference is to invoke QuadReg (option 5) rather than LinReg as done before. The quadratic model is a polynomial of degree 2 and the results are

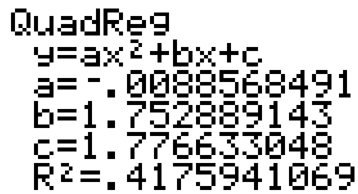


Figure 24

The R^2 has improved slightly to 42% and the graph of the quadratic model and actuals appears in Figure 25.

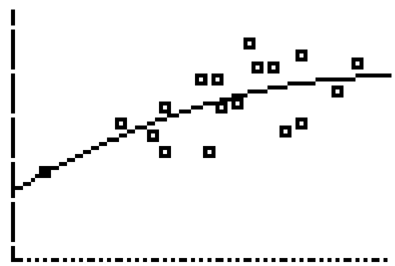


Figure 25

While the least squares regression models presented are accurate, one has to question the reliability of using mathematics averages (the independent variable) to predict history averages (the dependent variable). We leave that discussion to social scientists and want to emphasize that the data used was merely to implement the regression procedures.

MCAS Problems

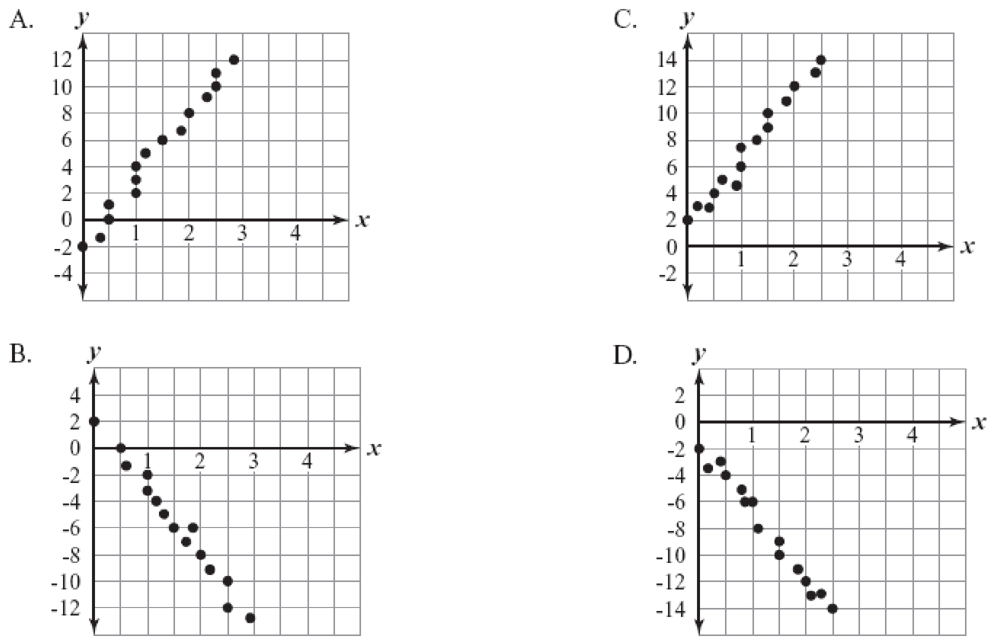
2007

http://www.doe.mass.edu/mcas/2007/release/g10math.pdf

2007 Grade 10 #10

10 Which of the following scatterplots is most likely to have a line of best fit represented by the equation below?

$$y = -5x + 2$$



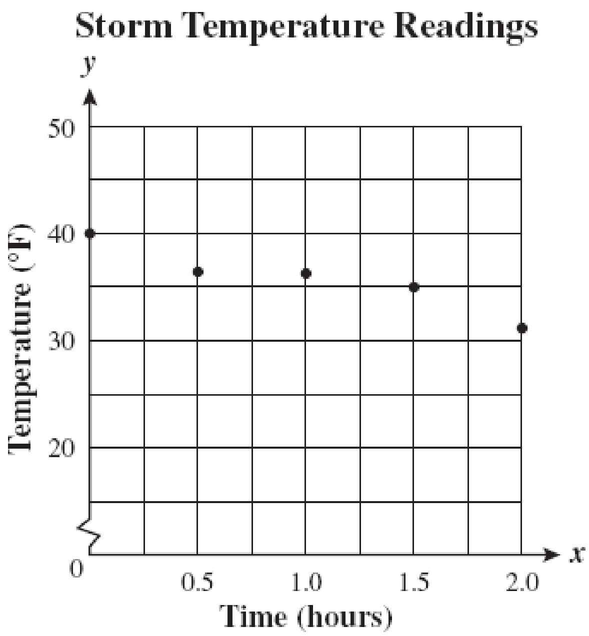
Answer: Since the y data is decreasing as x increases, the slope (and correlation) must be negative. Using a linear model, answer B appears to have a y intercept equal to 2 with a negative slope for the line of best fit.

2006

http://www.doe.mass.edu/mcas/2006/release/g10math.pdf

2006 Grade 10 #3

3 During the beginning of a recent storm, a weather broadcaster took temperature readings every half hour and plotted the data on the scatterplot below.



Which of the following most closely approximates the equation of the line of best fit for the data?

A. $y = -40x + 40$

B. $y = -3x + 40$

C. $y = 40x + 40$

D. $y = 3x + 40$

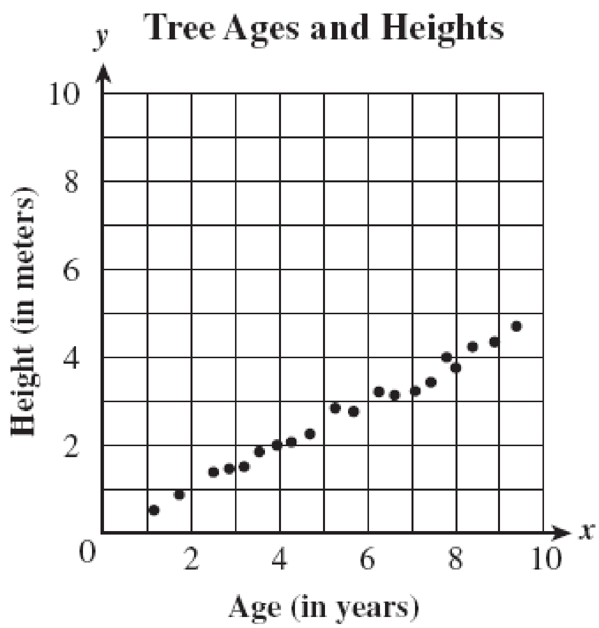
Answer: The y intercept for the actual data is 40. The slope of the linear trend is negative. That leaves us with answers A and B to choose from. Estimating the slope from the data while using the actual data (0, 40) and (1.5, 35), the slope for the best fit regression line is approximately $\frac{y_2 - y_1}{x_2 - x_1} = \frac{35-40}{1.5-0} = \frac{-5}{1.5} = \frac{-10}{3}$. Answer B, y= -3x+40 is the best selection.

2005

http://www.doe.mass.edu/mcas/2005/release/g10math.pdf

2005 Grade 10 #35

35 The scatterplot below shows the ages and heights of 20 trees on a tree farm.



If x = age in years and y = height in meters, which of the following equations best approximates the line of best fit for this scatterplot?

A. $y = \frac{1}{2}x$

B. $y = \frac{1}{2}x + 5$

C. $y = 2x$

D. $y = 2x + 5$

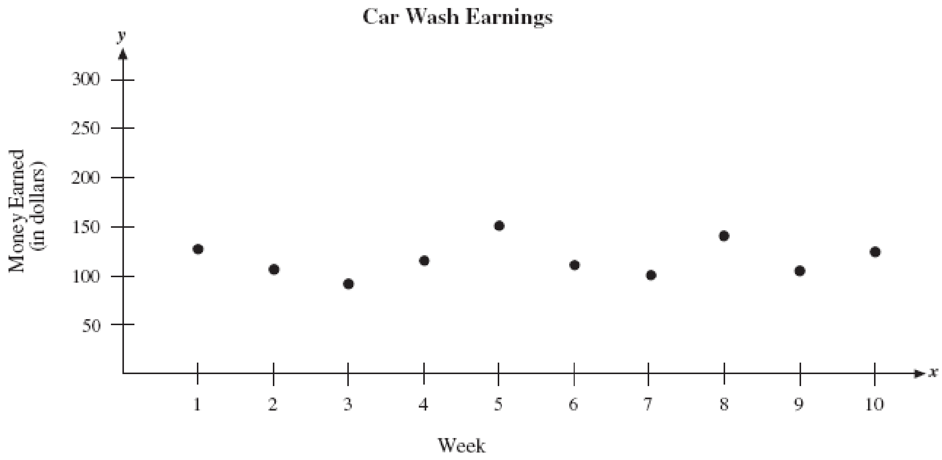
Answer: The y intercept from the actual data appears to be about y=0. The slope is positive. Using the points (1, .5) and (8, 3.8) as estimates of actual data points, the slope is approximately $\frac{3.8-0.5}{8-1} = \frac{3.3}{7} = 0.47$. The best approximation for the line of best fit is answer A, $y = \frac{1}{2}x$.

2003

http://www.doe.mass.edu/mcas/2003/release/g10math.pdf

2003 Grade 10 #30

30 The students at Albemarle High held a car wash each week for 10 weeks to earn money for the student council. The students made the scatter plot below to represent the amount of the money they earned each week.



Which of the following equations **best** represents the line of best fit for these data?

A. $y = 110$

B. $y = 110x$

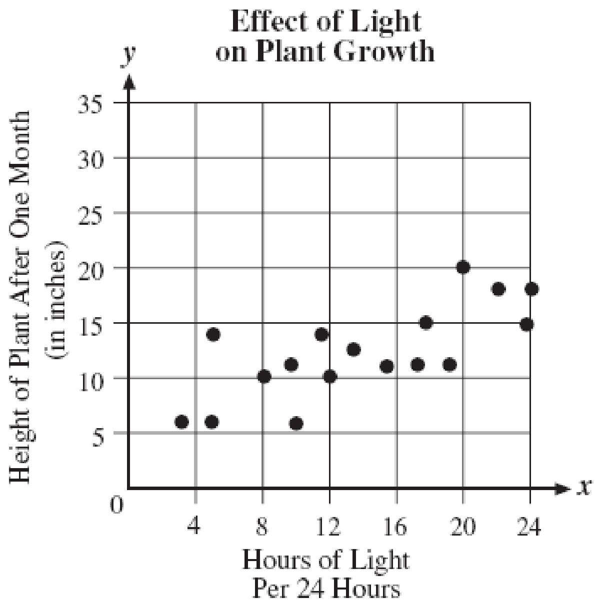
C. $y = x + 55$

D. $y = -x + 55$

Answer: Looking at all the data, although there are fluctuations from week to week, there does not appear to be an overall increase or decrease in the money earned per week. All the data is concentrated about the line y=110 meaning that the students earn about $110 per week. The answer is A.

2003 Grade 10 #39

39 Jenny studied the effect of light on plant growth. She graphed a scatterplot to represent her data.



Which of the following **best** represents the equation for the line of best fit for the data shown?

A. $y = -0.4x + 5$

B. $y = 0.4x + 5$

C. $y = -4x + 5$

D. $y = 4x + 5$

Answer: Using an approach similar to that used for the 2005 MCAS problem Grade 10 #35, The y intercept from the actual data appears to be about y=5. The slope is positive. Using the points (8, 10) and (24, 15), the slope for the best fit line is approximately $\frac{15-10}{24-8} = \frac{5}{16} = 0.31$. It appears that answer B, $y = 0.4x + 5$ is the best selection for the line of best fit.

References

The following resources have been used mainly for measures of central tendency, measures of variability, bar graphs, line plots, stem-and-leaf plots, frequency tables, histograms, circle graphs and Venn diagrams.

At the grade 6 level, the references included:

1. Cavanagh, Mary. Math to Know, Great Source Education Group, 2000.

2. Charles, Randall, Judith Branch-Boyd, Mark Illingworth, Darwin Mills, Andy Reeves. Mathematics: Course 1, Pearson Prentice Hall, 2004.

3. Larson, Roland, Laurie Boswell, and Lee Stiff. Passport to Mathematics: An Integrated Approach Book 1, D.C. Heath & Company, 1997.

4. Manfre, Edward, James Moser, Joanne Lobato, Lorna Morrow. Heath Mathematics Connections (green book), D.C. Heath & Company, 1996.

5. Manfre, Edward, James Moser, Joanne Lobato, Lorna Morrow. Heath Mathematics Connections (purple book), D.C. Heath & Company, 1996.

At the grade 8 level, the sources used were:

1. Brown, Richard, Mary Dolciani, Robert Sorgenfrey, and William Cole. Algebra: Structure and Method Book 1, McDougal Littell, 1997.

2. Cavanagh, Mary. Math on Call, Great Source Education Group, 1998.

3. Cavanagh, Mary. Algebra to Go, Great Source Education Group, 2000.

4. Charles, Randall, David Davison, Marsha Landau, Leah McCracken, and Linda Thompson. Prentice Hall Mathematics: Pre-Algebra, Pearson Prentice Hall, 2004.

5. McConnell, John, Susan Brown, Zalman Usiskin, Susan Eddins, Cathy Hynes Feldman, James Flanders, Margaret Hackworth, Daniel Hirschhorn, and Lydia Polonsky. Algebra (Teacher's Edition Part 1), Scott Foresman, 1996.

6. Saxon, John. Algebra 1: An Incremental Development, Saxon Publishers, 2nd edition, 1990.

For scatterplots, box and whisker plots, quartiles and deciles, regression material, and TI-84 calculation procedures, additional resources used were:

1. Morgan, Larry. Statistics Handbook for the TI-83, Texas Instruments Inc., 1997.

2. Triola, Mario F. Elementary Statistics Using Excel, Addison Wesley Longman Inc., 2001.

3. Weiss and Hassett, Introductory Statistics, Addison Wesley publishing Co., 2nd edition, 1987