CHAPTER 8
Statistics

8-1. Mean, Median, Mode, and Range

The **mean** of a data set is the sum of the values in the data set divided by the number of values in the data set. The mean is also called the arithmetic mean.

$$\text{Mean} = \frac{\text{sum of the values}}{\text{number of values}} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

The sum of the values $=$ mean $\times$ number of values .

Weighted average of two groups $= \dfrac{\{\text{sum of the values of group 1}\} + \{\text{sum of the values of group 2}\}}{\text{total number of values}}$

The median of a set of data is the middle number when the data are arranged in order. If there are two middle numbers, the median is the mean of the two numbers. In the set of data $\{3, 7, 8, 10, 14\}$, the median is 8 and in the set of data $\{3, 7, 10, 14\}$, the median is $\dfrac{7+10}{2}$, or 8.5.

The **mode** of a set of data is the number that appears most frequently. Some set of data have more than one mode, and others have no mode. In the set of data $\{3, 7, 8, 10, 14\}$, there is no mode since each number appears only once. In the set of data $\{3, 7, 7, 10, 14, 19, 19, 25\}$, the modes are 7 and 19, since both numbers appear twice.

The **range** in a set of data is the difference between the greatest value and the least value of the data.

Example 1 □ Find the mean, median, mode, and range of the data set $\{132,\ 149,\ 152,\ 164,\ 164,\ 175\}$.

Solution □ $\text{Mean} = \dfrac{\text{sum of the values}}{\text{number of values}} = \dfrac{132+149+152+164+164+175}{6} = \dfrac{936}{6} = 156$

$\text{Median} = \dfrac{152+164}{2} = 158$

$\text{Mode} = 164$

$\text{Range} = 175 - 132 = 43$

Example 2 □ In a geometry class of 15 boys and 12 girls, the average (arithmetic mean) test score of the class was 81. If the average score of the 15 boys was 83, what was the average score of the 12 girls?

Solution □ Let $x =$ the average score of 12 girls.

$81 = \dfrac{(83\times 15) + (x \times 12)}{15+12}$ Weighted average formula

$81 = \dfrac{1245 + 12x}{27}$ Simplify.

$81 \times 27 = 1245 + 12x$ Multiply each side by 27.

$2187 - 1245 = 12x$ Subtract 1245 from each side.

$942 = 12x$ Simplify.

$x = 78.5$ Divide each side by 12.

The average score of 12 girls is 78.5.

Exercise - Mean, Median, Mode, and Range

1

Test Scores	67	75	87	91
Number of Students	1	3	2	2

The test scores of 8 students are shown in the table above. Let m be the mean of the scores and M be the median of the score. What is the value of $M-m$?

A) -6

B) 0

C) 3

D) 6

2

The average (arithmetic mean) of five numbers $n,\ n-3,\ 2n+1,\ 3n-4$, and $5n+10$ is 8. Which of the following is true?

A) median $=5$, range $=18$

B) median $=5$, range $=25$

C) median $=7$, range $=18$

D) median $=7$, range $=25$

3

The average (arithmetic mean) of two numbers is $\frac{1}{2}x+1$. If one of the numbers is x, what is the other number?

A) $x+2$

B) $x-2$

C) -2

D) 2

4

The average (arithmetic mean) of a set of n numbers is 19. If the average of the 6 greatest numbers in the set is 29 and the average of the remaining numbers is 7, what is the value of n?

A) 9

B) 10

C) 11

D) 12

5

The average (arithmetic mean) of m, n, and -1 is 0. What is the value of $m+n$?

6

The average (arithmetic mean) test score for all the students in a class is 84. The average score of m boys in the class was 79, while that of n girls was 87. What is the ratio of m to n?

7

A student has an average (arithmetic mean) score of 86 points for 4 tests. What total score does this student need in the next two tests in order to have an average of 90 for all 6 tests?

8-2. Standard Deviation

The **variance**, denoted d^2, and the **standard deviation** of a data set $\{x_1, x_2, \cdots, x_n\}$ are defined as follows.

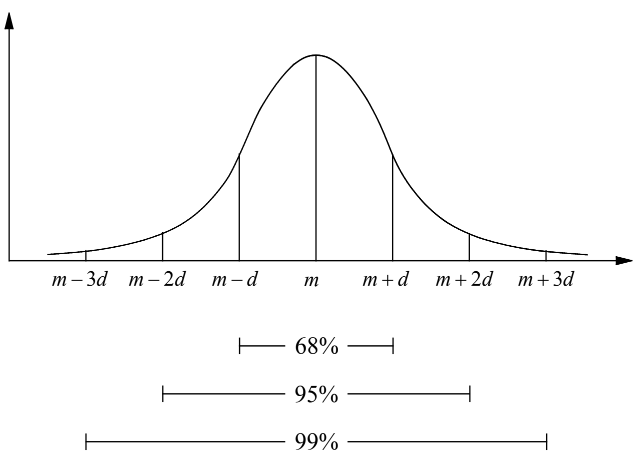
Variance $= d^2 = \dfrac{(x_1 - m)^2 + (x_2 - m)^2 + \cdots + (x_n - m)^2}{n}$, in which m is the mean of n numbers.

Standard deviation $= \sqrt{d^2} = \sqrt{\dfrac{(x_1 - m)^2 + (x_2 - m)^2 + \cdots + (x_n - m)^2}{n}}$

In a data set, if the same number k is added to each number, the mean and the median are increased by k, while the range and the standard deviation remain unchanged.
If each number in a data set is multiplied by the same number k, the mean, median, range, and standard deviation are all multiplied by k.
In general, *when the measures are clustered close to the mean the standard deviation is small, and when the measures are widely spread apart the standard deviation is relatively large*.

The graph of the normal distribution, called the **normal curve**, is a symmetrical bell shaped curve which shows the relationship between the standard deviation and the observations.

The Normal Curve



Normal distributions have these properties.

1. The mean m is at the center.
2. About 68% of the values are within one standard deviation from the mean.
3. About 95% of the values are within two standard deviations from the mean.
4. About 99% of the values are within three standard deviations from the mean.

Example 1 □ Find the variance and the standard deviation of the data set $\{3, 5, 12, 16, 19\}$.

Solutions □ Mean $= \dfrac{3+5+12+16+19}{5} = \dfrac{55}{5} = 11$

$$\text{Variance} = d^2 = \frac{(3-11)^2 + (5-11)^2 + (12-11)^2 + (16-11)^2 + (19-11)^2}{5} = 38$$

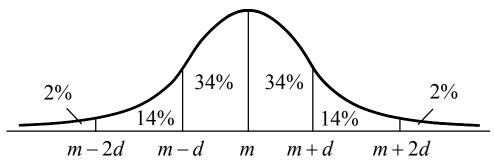
$$\text{Standard deviation} = \sqrt{d^2} = \sqrt{38} \approx 6.2$$

Example 2 □ On a test, 82 is the arithmetic mean and 6 is the standard deviation. What value is exactly two standard deviations more than the mean?

Solutions □ Two standard deviations more than the mean is $82 + 2(6) = 94$.

Exercise - Standard Deviation

Questions 1-3 refer to the following information.



The figure above shows a normal distribution with mean m and standard deviation d, including approximate percentages of the distribution corresponding to the regions shown. Suppose the SAT math scores of 1,200 students entering a certain university are normally distributed with a mean score of 600 and standard deviation of 60.

1

Approximately how many of the students have SAT scores between 660 and 720?

2

Approximately how many of the students have SAT scores less than 540?

3

Approximately how many of the students have SAT scores greater than 720?

Questions 4-6 refer to the following information.

Number of Children	0	1	2	3	4
Frequency	1	2	4	0	1

The table above shows the frequency distribution of the number of children in each of 8 families.

4

Let m be the mean of the data set above. What is the value of m?

5

Let d be the standard deviation of the data set above. What is the value of d?
(Round your answer to the nearest hundredth.)

6

Add 2 to each entry on the original list. Let m_a and d_a be the new mean and the new standard deviation of the data set. Which of the following is true?

A) $m_a = m + 2$ and $d_a = d + 2$

B) $m_a = m$ and $d_a = d + 2$

C) $m_a = m + 2$ and $d_a = d$

D) $m_a = m$ and $d_a = d$

7

Multiply each entry by 2 on the original list. Let m_p and d_p be the new mean and the new standard deviation of the data set. Which of the following is true?

A) $m_p = 2m$ and $d_p = 2d$

B) $m_p = m$ and $d_p = d$

C) $m_p = 2m$ and $d_p = d$

D) $m_p = m$ and $d_p = 2d$

8-3. Graphical Displays

The **frequency** of a particular category is the number of times the category appears in the data.
A **frequency distribution** is a table or graph that presents the categories along with their associated frequencies.
The **relative frequency** of a category is the associated frequency divided by the total number of data.
Relative frequencies may be expressed in terms of percents, fractions, or decimals. The sum of the relative frequencies in a relative frequency distribution is always 1(100%).

Note: In data interpretation questions, you need to *distinguish between the change in absolute numbers and the percent change* in the questions.

A **line graph** shows the change in a certain data over a period of time. The time is always on the horizontal axis, and the variable measured is always on the vertical axis.

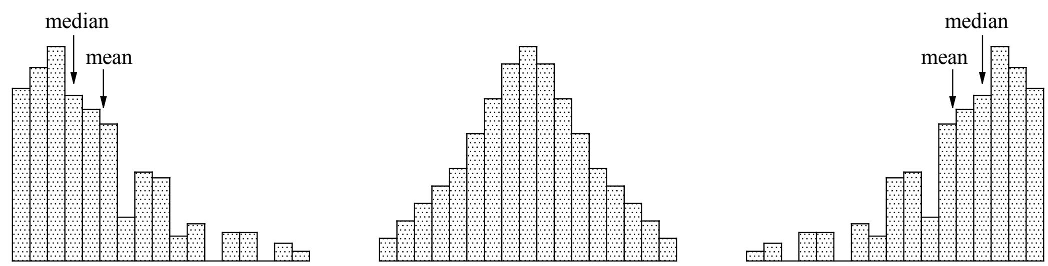
In a **bar graph**, rectangular bars represent the categories of data. The height of vertical bars or the length of a horizontal bars show the frequency or relative frequency.

A **histogram** is useful in displaying frequency distributions that are similar to bar graphs, but they have a number line for the horizontal axis. Also, there are no spaces between the bars.

Distributions are **skewed** if the majority of their values fall either to the left or to the right. In right-skewed data the mean is usually greater than the median, while in left-skewed data the mean is usually less than the median.

Outliers are a small group of values that are significantly smaller or larger than the other values in the data. When there are outliers in the data, the mean will be pulled in their direction, while the median remains the same. If outliers are significantly smaller than the other values, the mean will be smaller than the median. If outliers are significantly larger than the other values, the mean will be larger than the median

Patterns of Distribution

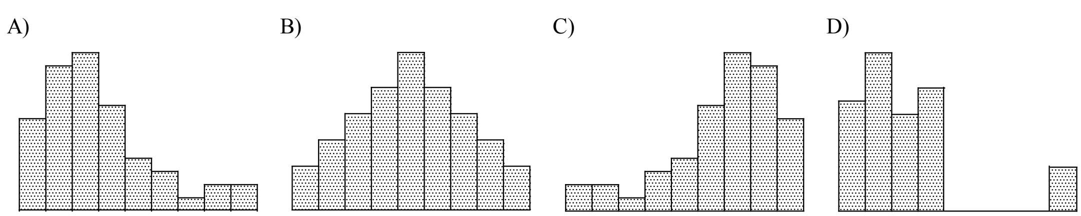


Distribution is skewed to the right.
Majority of the values fall to the left.
Median < Mean.

Distribution is symmetric.
Mean = Median.

Distribution is skewed to the left.
Majority of the values fall to the right.
Mean < Median.

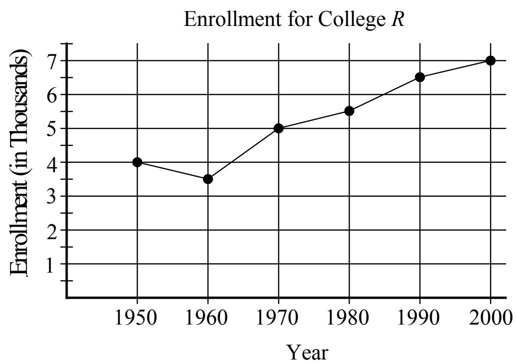
Example 1 □ In the graph below which data set will have a median greater than the mean?



Solutions □ Choice C is correct. When the distribution is skewed to the left, the data set will usually have a greater median than mean. Choice A is incorrect because the distribution is skewed to the right, meaning the mean is greater than the median. Choice B is incorrect because the distribution is symmetric, meaning the median is equal to the mean. Choice D is incorrect because the outlier on the far right pulls the mean to the right but leaves the median alone. So its mean is greater than the median.

Exercise - Graphical Displays

Questions 1-3 refer to the following information.



The line graph above shows the enrollment for College R between 1950 and 2000.

1

According to the graph above, College R showed the greatest change in enrollment between which two decades?

A) 1950 to 1960

B) 1960 to 1970

C) 1970 to 1980

D) 1980 to 1990

2

What is the average rate of increase in enrollment per decade between 1950 and 2000?

A) 500

B) 600

C) 750

D) 875

3

If enrollment increases by approximately the same percentage between 2000 and 2010 as it decreased between 1950 and 1960, what is the expected enrollment in 2010?

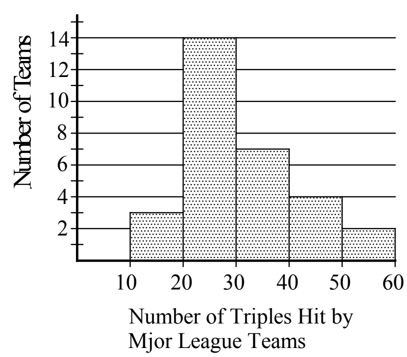
A) 7,250

B) 7,540

C) 7,650

D) 7,875

4



The histogram above shows the distribution of the number of triples hit by 30 major league baseball teams in a certain year. Which of the following could be the median number of triples represented in the histogram?

A) 19

B) 27

C) 32

D) 34

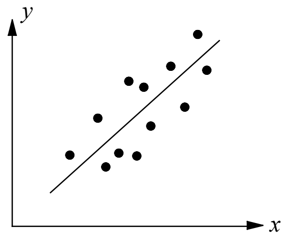
8-4. Scatter Plots and the Regression Lines

A **scatter plot** is a mathematical diagram represented by a set of dots that display the relationship between two numerical variables.
The x- coordinate of the dot, measured along the horizontal axis, gives the value of one variable and the y- coordinate of the dot, measured along the vertical axis, gives the value of the other variable.

When the set of dots are plotted, usually there is no single line that passes through all the data points, but we can approximate a linear relationship by finding a line that best fits the data. This line is called the **regression line** or the **best fit line**.

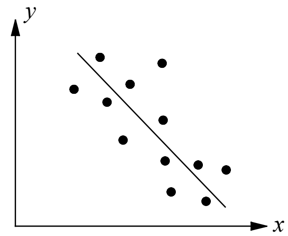
If the slope of a regression line is positive, then the two variables have a **positive association**.
If the slope of a regression line is negative, then the two variables have a **negative association**.
If the slope of a regression line cannot be determined, then the two variables have **no association**.

regression line with positive slope



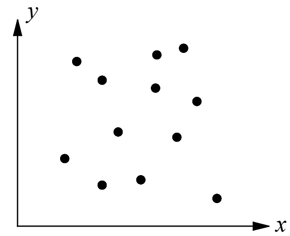
positive association

regression line with negative slope



negative association

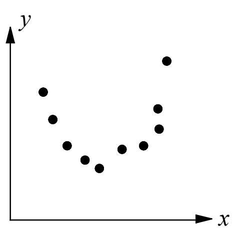
regression line cannot be determined



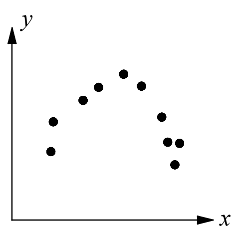
no association

Some scatter plots are not linear. There are quadratic and exponential models too.

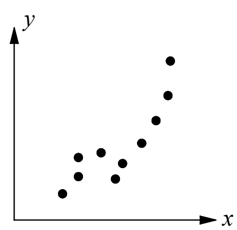
A scatter plot with a quadratic model shows that as x increases, y falls then rises.



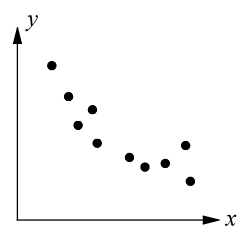
A scatter plot with a quadratic model shows that as x increases, y rises then falls.



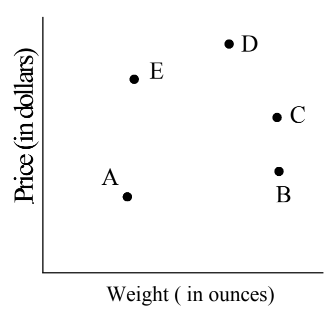
A scatter plot with an exponential model shows a positive association that is not linear.



A scatter plot with an exponential model shows a negative association that is not linear.



Example 1 □ The scatter plot at the right shows the prices and weights of various boxed products. Of the five labeled products, which has the best unit price?

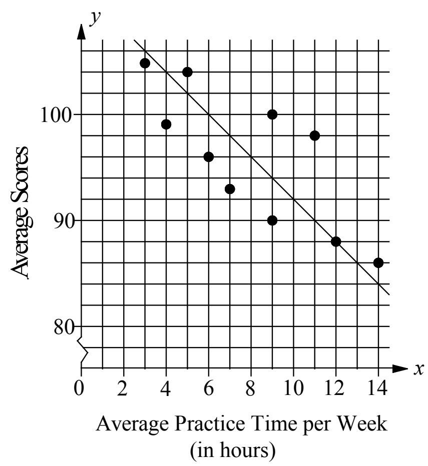


Solutions □ $\text{unit price} = \dfrac{\text{price}}{\text{number of units}}$

For each of the five points on the graph, product B has the lowest price to weight ratio. Therefore, product B is the one with best unit price.

Exercise - Scatter Plots and the Regression Lines

Questions 1 and 2 refer to the following information.



The scatter plot above shows the average scores of 10 golfers and their weekly practice times. The line of best fit is also shown.

1

What is the average score of the golfer that is farthest from the line of best fit?

A) 93

B) 96

C) 98

D) 99

2

There are two golfers whose average practice time is the same. What is the difference between their average scores?

A) 4

B) 6

C) 8

D) 10

3

What is the median score of the 10 golfers?

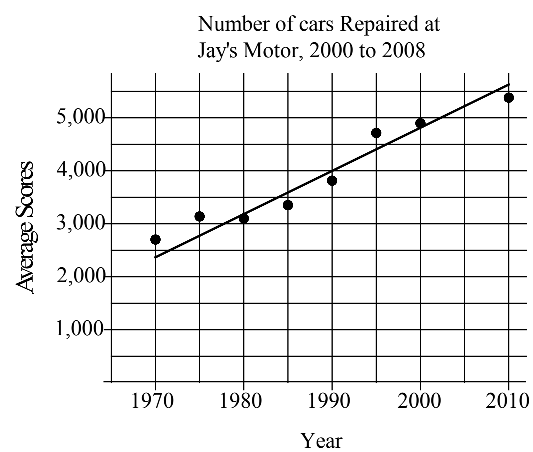
A) 96

B) 97

C) 98

D) 99

4



According to the line of best fit in the scatter plot above, which of the following best approximate the year in which the number of cars repaired by Jay's Motor was estimated to be 4,500?

A) 1996

B) 1998

C) 2000

D) 2002

8-5. Populations, Samples, and Random Selection

In statistics, the word **population** does not necessarily refer to people. It may, for example, refer to the number of smart phones sold or to the number of public high schools in California. The term population refers to the set of all individuals or objects under consideration. In the real world, studying the entire population is usually impractical. Only part of the whole population can be examined, and this part is called the **sample**.

When researchers make generalizations from the part to the whole, we say they make **inferences** from the sample to the population.

A sample that accurately reflects the entire population must satisfy the following two conditions.
1. *Large sample size.*
2. *The sample must be selected at random from the original population of interest.*

Large sample size leads to more reliable conclusions because larger samples experience less sampling variability. A random sample of size 300 from a population of size 100,000 is just as reliable as a random sample of size 300 from a population of size 1,000,000. What is important is the sample size, not the sample-to-population percentage or fraction. But when a selection procedure is biased, taking a large sample does not help.

Bias is the tendency to favor the selection of certain members of a population. When selecting a sample, it is crucial to use randomization. Randomly selecting samples from the original population of interest will most likely result in a decrease in the margin of error.

The **population parameter** is a measured characteristic about the population, calculated from the sample.

Population parameter = fraction of people in the random sample × total population

Example 1 □ An opinion survey about water use was conducted by DWR (Department of Water Resources) in two cities. City A has 120,000 residents and city B has 65,000 residents. Each survey was conducted with a sample of 400 residents from each city. The results of the survey will be used by the state government for future water regulations. Which of the following statements is true?

A) The opinions surveyed at City A are more accurate, since the larger city is more likely to have diversity of opinion.

B) The opinions surveyed at City B are more accurate, since it has a larger percentage of its people surveyed.

C) The opinions surveyed at City A are more accurate, since there is less sampling variability with a larger population.

D) Neither city is more likely to have a more accurately estimate.

Solutions □ Choice D is correct. A random sample of size 400 from a population of size 120,000 is just as reliable as a random sample of size 400 from a population of size 650,000.

Choice A is not correct. It is unclear whether the larger city has more diversity of opinion. Even if it does, how many of those diverse opinions are included in the sample is in question.

Choice B is not correct. In a survey, what is important is the sample size, not the percentage or fraction of the sample to the population.

Choice C is not correct. Larger *samples*, not larger populations, lead to more reliable conclusions because they experience less sampling variability.

Exercise - Populations, Samples, and Random Selection

Questions 1-3 refer to the following information.

	Voted for Candidate A	Voted for Candidate B	Voted for Other Candidates	Total
Ages 18 to 30	84	46	30	160
Ages 31 to 55	72	90	48	210
55 years or older	31	76	23	130
Total	187	212	101	500

A polling organization takes a random sample of 500 voters who voted for the mayoral election of a large western city. The organization gathered data right after the election, as shown in the table above.

1

According to the data above, what percent of people from ages 31 to 55 voted for candidate B ?

A) 38%

B) 43%

C) 45%

D) 48%

2

The total population of individuals in the city who voted for the election was about 450,000. What is the best estimate of the total number of votes for candidate B ?

A) 144,000

B) 168,000

C) 190,000

D) 210,000

3

According to the data above, how many times more likely is it for ages 18 to 30 year olds to vote for candidate A than it is for ages 55 years or older to vote for candidate A ?

A) 2.2

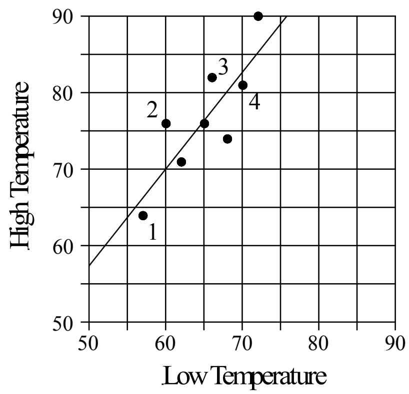
B) 2.4

C) 2.6

D) 2.8

Chapter 8 Practice Test

Questions 1-3 refer to the following information.



The graph above is a scatter plot with 8 points, each representing the low temperature and high temperature of 8 days in September in a certain city. Both the low temperatures and high temperatures are measured in degrees Fahrenheit. The line of best fit for the data is also shown.

1

Based on the line of best fit for the data shown, how many degrees does the high temperature increase when the low temperature increases by one degree?

A) 0.9

B) 1.3

C) 1.6

D) 1.8

2

What is the predicted high temperature of the day when the low temperature is 58?

A) 65

B) 68

C) 71

D) 74

3

Among the four days marked 1, 2, 3, and 4 in the scatter plot, on which day is the difference between the high temperature and the low temperature minimal?

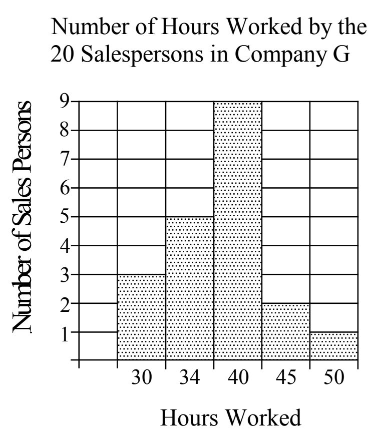
A) Day 1

B) Day 2

C) Day 3

D) Day 4

4



Based on the histogram above, what is the average number of hours worked by the 20 salespersons in Company G?

A) 36

B) 37

C) 38

D) 39

Questions 5 and 6 refer to the following information.

Frequency Distribution for List A

Number	0	4	5	6
Frequency	8	10	12	10

Frequency Distribution for List B

Number	7	10	11	15
Frequency	10	8	10	12

The table above shows the frequency distribution of two lists. List A and list B each contain 40 numbers.

5

What is the difference between the average of the numbers in list B and the average of the numbers in list A ?

A) 6.5

B) 7

C) 7.5

D) 8

6

List C contains 80 numbers: the 40 numbers in list A and the 40 numbers in list B. Let m be the average of 80 numbers in list C and M be the median of 80 numbers in list C. What is the value of $m - M$?

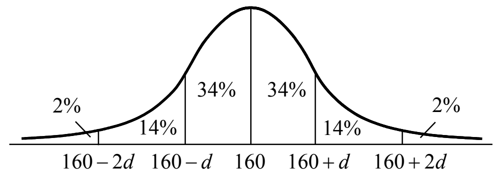
A) 1

B) 1.5

C) 2

D) 2.5

7



The figure above shows a standard normal distribution with mean of 160 and standard deviation d, including approximate percents of the distribution corresponding to the regions shown. If the value 148 is at the 12th percentile of the distribution, which of the following is the best estimate of the standard deviation d of the distribution?

A) 5

B) 10

C) 15

D) 20

8

The tables below give the distribution of ratings of two different laptops by 100 people each.

Ratings of Laptop A by 100 Reviewers

Ratings	5	4	3	2	1
Frequency	28	45	11	7	9

Ratings of Laptop B by 100 Reviewers

Ratings	5	4	3	2	1
Frequency	22	24	18	20	16

Which of the following is true about the data shown for the ratings of the two laptops?

A) The standard deviation of the ratings of laptop A is larger.

B) The standard deviation of the ratings of laptop B is larger.

C) The standard deviation of the two ratings are the same.

D) The standard deviation of the two ratings cannot be determined with the data provided.

Answer Key

Section 8-1

1. B 2. B 3. D 4. C 5. 1
6. $\frac{3}{5}$ 7. 196

Section 8-2

1. 168 2. 192 3. 24 4. 1.75 5. 1.09
6. C 7. A

Section 8-3

1. B 2. B 3. D 4. B

Section 8-4

1. C 2. D 3. B 4. A

Section 8-5

1. B 2. C 3. A

Chapter 8 Practice Test

1. B 2. B 3. A 4. C 5. B
6. A 7. B 8. B

Answers and Explanations

Section 8-1

1. B

$$m = \frac{67 + 75 \times 3 + 87 \times 2 + 91 \times 2}{8} = 81$$

To find the median of the test scores, put all the test scores in order, from least to greatest.

67, 75, 75, 75, 87, 87, 91, 91
There are two middle values, 75 and 87.
The median is the mean of these two values.

$$M = \frac{75 + 87}{2} = 81$$
$$M - m = 81 - 81 = 0$$

2. B

$$\text{Average} = \frac{n + n - 3 + 2n + 1 + 3n - 4 + 5n + 10}{5}$$
$$= \frac{12n + 4}{5}$$

Since the average of five numbers is 8,

$$\frac{12n + 4}{5} = 8.$$

Solving the equation for n yields $n = 3$.
Therefore, the five numbers are 3, 0, 7, 5, and 25.
Arranging the five numbers in order, we get
0, 3, 5, 7, and 25.

The median of this set of numbers is 5, and the range of this set of numbers is $25 - 0$, or 25.

3. D

Let the other number $= y$.

$\frac{x + y}{2} = \frac{1}{2}x + 1$	Avg. of two numbers is $\frac{1}{2}x + 1$.
$2(\frac{x + y}{2}) = 2(\frac{1}{2}x + 1)$	Multiply each side by 2.
$x + y = x + 2$	Distributive property
$y = 2$	Subtract each side by x.

The other number is 2.

4. C

If the average of a set of n numbers is 19, then the sum of n numbers is $19n$.
If the average of the 6 greatest numbers in the set is 29, the sum of those 6 numbers is 6×29, or 174. There are $n - 6$ remaining numbers, with sum $19n - 174$ and average is 7.

Therefore, $\frac{19n - 174}{n - 6} = 7$.

$$19n - 174 = 7(n - 6)$$
$$19n - 174 = 7n - 42$$
$$12n = 132$$
$$n = 11$$

5. 1

The average of m, n, and -1 is 0.

$$\frac{m + n + (-1)}{3} = 0$$
$$m + n + (-1) = 0$$
$$m + n = 1$$

6. $\frac{3}{5}$

Weighted average of the two groups

$$= \frac{\begin{Bmatrix}\text{sum of the values} \\ \text{of group 1}\end{Bmatrix} + \begin{Bmatrix}\text{sum of the values} \\ \text{of group 2}\end{Bmatrix}}{\text{total number of values}}$$

$$84 = \frac{79m + 87n}{m+n} \Rightarrow 84(m+n) = 79m + 87n$$

$$\Rightarrow 84m + 84n = 79m + 87n \Rightarrow 5m = 3n$$

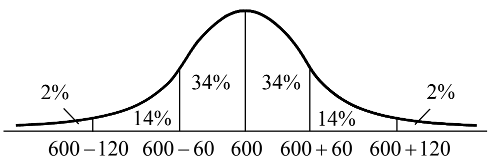
$$\Rightarrow m = \frac{3}{5}n \Rightarrow \frac{m}{n} = \frac{3}{5}$$

7. 196

The student's total score for his 4 tests is 86×4, or 344. In order to have an average of 90 for all 6 tests, the student needs 90×6, or 540, points total. So the total score needed on the next two tests is $540 - 344$, or 196.

Section 8-2

1. 168



Fourteen percent of the students have SAT scores between 660 and 720.
14% of 1,200 is 1200×0.14, or 168.

2. 192

Sixteen percent of the students have SAT scores less than 540.
16% of 1,200 is 1200×0.16, or 192.

3. 24

Two percent of the students have SAT scores greater than 720.
2% of 1,200 is 1200×0.02, or 24.

4. 1.75

First, put all the data in order.
0, 1, 1, 2, 2, 2, 2, 4

$$m = \frac{0+1+1+2+2+2+2+4}{8} = \frac{14}{8} = 1.75$$

5. 1.09

$$d^2 = [(0-1.75)^2 + (1-1.75)^2 + (1-1.75)^2 + (2-1.75)^2 + (2-1.75)^2 + (2-1.75)^2 + (2-1.75)^2 + (4-1.75)^2] \div 8 = 1.1875$$

$$d = \sqrt{1.1875} \approx 1.09$$

6. C

Add 2 to each entry on the original list. The new list is 2, 3, 3, 4, 4, 4, 4, 6.

$$m_a = \frac{2+3+3+4+4+4+4+6}{8} = \frac{30}{8} = 3.75$$

$$d^2 = [(2-3.75)^2 + (3-3.75)^2 + (3-3.75)^2 + (4-3.75)^2 + (4-3.75)^2 + (4-3.75)^2 + (4-3.75)^2 + (6-3.75)^2] \div 8 = 1.1875$$

$$d = \sqrt{1.1875} \approx 1.09$$

The mean is increased by 2, but the standard deviation is unchanged.

Choice C is correct.

7. A

Multiply each entry by 2 on the original list. New list is 0, 2, 2, 4, 4, 4, 4, 8.

$$m_p = \frac{0+2+2+4+4+4+4+8}{8} = \frac{28}{8} = 3.5$$

$$d^2 = [(0-3.5)^2 + (2-3.5)^2 + (2-3.5)^2 + (4-3.5)^2 + (4-3.5)^2 + (4-3.5)^2 + (4-3.5)^2 + (8-3.5)^2] \div 8 = 4.75$$

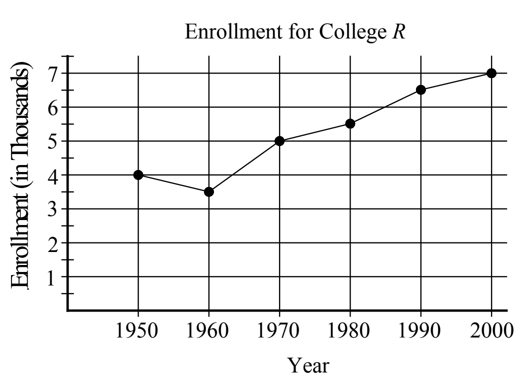
$$d = \sqrt{4.75} \approx 2.18$$

The mean and standard deviation are multiplied by 2.

Choice A is correct.

Section 8-3

1. B



Year	Changes in enrollment	
1950 – 1960	500	decrease
1960 – 1970	1,500	increase
1970 – 1980	500	increase
1980 – 1990	1,000	increase

The greatest change in enrollment occurred between 1960 and 1970.

2. B

Average rate of increase in enrollment

$= \dfrac{\text{number increased}}{\text{change in decades}} = \dfrac{7000-4000}{5}$

$= \dfrac{3000}{5} = 600$

3. D

Percent of decrease in enrollment between 1950 and $1960 = \dfrac{\text{number decreased}}{\text{original number}} = \dfrac{500}{4000} = 0.125$

The percent decrease was 12.5%, so the expected increase is 12.5%.

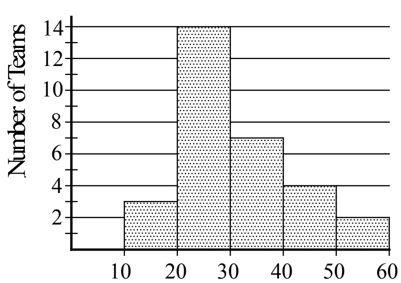
Let x = expected number of increase.
Expected percent increase

$= \dfrac{\text{number of increase}}{\text{original number}} = \dfrac{x}{7000} = 0.125$

$x = 7000 \times 0.125 = 875$

Therefore, the expected enrollment in 2010 is $7,000 + 875$, or 7,875.

4. B



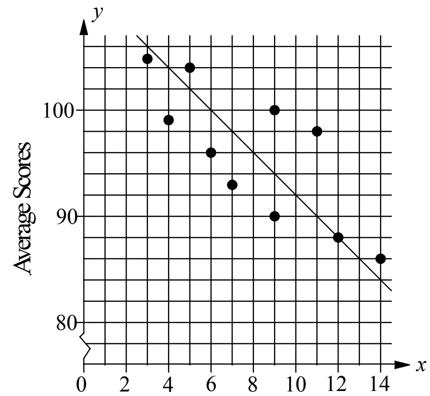
Number of Triples Hit by Mjor League Teams

The median of a data set is the middle value when the data are arranged in order. Since there are 30 teams the middle value is the average of 15th and 16th value. Since there are 3 teams who hit less than 20 triples and 13 teams who hit more than 30 triples, the median number should be between 20 and 30. Therefore, of the choices given, only 27 could be the median number of triples hit by 30 teams.

Choice B is correct.

Section 8-4

1. C



The average score of the golfer that is farthest from the line of best fit is located at $(11, 98)$.
The golfer's average score is 98.

Choice C is correct.

2. D

There are two golfers whose average practice time is 9. The difference between their average scores is $100 - 90$, or 10.

Choice D is correct.

3. B

The list of the 10 golfers' average scores listed from highest to lowest is 105, 104, 100, 99, 98, 96, 93, 90, 88, and 86.
There are two middle values, 98 and 96.
The median is the average of these two numbers.

$\text{Median} = \dfrac{98 + 96}{2} = 97$

4. A

According to the graph, the horizontal line that represents 4,500 cars repaired by Jay's Motor intersects the line of best fit at a point where the horizontal coordinate is between 1995 and 2000, and closer to 1995 than 2000. Therefore, of the choices given, 1996 best approximates the year in which the number of cars repaired by Jay's Motor was estimated to be 4,500.

Section 8-5

1. B

The total number of people who are between age 31 and 60 is 210. From that age group, 90 people voted for candidate B. Thus the percent of those who voted for candidate B is $\frac{90}{210} \approx 0.428 \approx 43\%$.

Choice B is correct.

2. C

The best estimate of the total number of votes can be obtained by multiplying the fraction of people who voted for candidate B and the total population of voters. $\frac{212}{500} \times 450{,}000 = 190{,}800$.

Therefore, of the choices given, 190,000 is the best estimation.

3. A

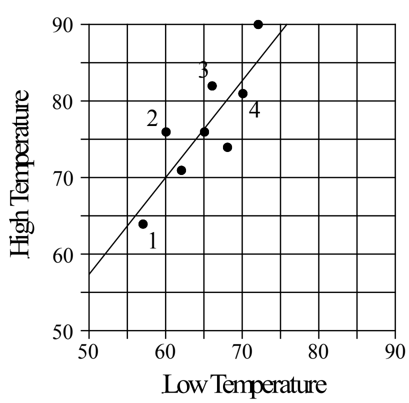
According to the data, 84 out of 160 people whose ages are between 18 and 30 voted for candidate A and 31 out of 130 people who are 55 years or older, voted for candidate A.

The ratio is $(\frac{84}{160}) \div (\frac{31}{130}) \approx 2.2$.

2.2 times more likely.

Chapter 8 Practice Test

1. B



The slope of the line of best fit is the value of the average increase in high temperature when the low temperature increases by one degree.

Using approximate values found along the line of best fit $(60,70)$ and $(76,90)$, the approximate slope can be calculated as $\frac{90-70}{76-60} = 1.25$.

Of the choices given, 1.3 is the best estimation. Therefore, the high temperature increases 1.3 degrees when the low temperature increases by one degree.

2. B

When the low temperature is 58, the graph shows that the high temperature is between 65 and 70, but closer to 70. Of the choices given, 68 is the best estimation.

3. A

In Day 1, the approximate high temperature is 64 and the approximate low temperature is 57. The difference is $64-57$, or 7 degrees.

In Day 2, the approximate high temperature is 76 and the approximate low temperature is 60. The difference is $76-60$, or 16 degrees.

In Day 3, the approximate high temperature is 82 and the approximate low temperature is 67. The difference is $82-67$, or 15 degrees.

In Day 4, the approximate high temperature is 81 and the approximate low temperature is 70. The difference is $81-70$, or 11 degrees.

The difference between the high and the low temperature was minimum on Day 1.

4. C

3 people worked for 30 hours.
5 people worked for 34 hours.
9 people worked for 40 hours.
2 people worked for 45 hours.
1 person worked for 50 hours.

Average number of hours worked

$$= \frac{\text{total number of hours}}{\text{total number of people}}$$

$$= \frac{3 \cdot 30 + 5 \cdot 34 + 9 \cdot 40 + 2 \cdot 45 + 1 \cdot 50}{20}$$

$$= \frac{760}{20} = 38$$

5 . B

Frequency Distribution for List *A*

Number	0	4	5	6
Frequency	8	10	12	10

Frequency Distribution for List *B*

Number	7	10	11	15
Frequency	10	8	10	12

Average of the numbers in List B

$= \frac{10 \times 7 + 8 \times 10 + 10 \times 11 + 12 \times 15}{40} = \frac{440}{40} = 11$

Average of the numbers in List A

$= \frac{8 \times 0 + 10 \times 4 + 12 \times 5 + 10 \times 6}{40} = \frac{160}{40} = 4$

Therefore, the difference between the average of the two lists is $11 - 4$, or 7.

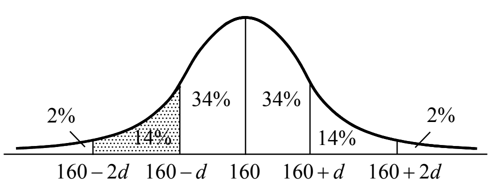
6. A

Because the lists A and B each contain 40 numbers, the average of the numbers in list C is the average of the individual averages of the numbers in lists A and B. Thus the average of the numbers in list C is $\frac{4+11}{2}$, or 7.5.

If you look at the numbers in the two lists, you will see that the 40 numbers in list A are all less than or equal to 6, and the 40 numbers in list B are all greater than or equal to 7. Thus the two middle numbers in list C are 6 and 7, and the average of these numbers is $\frac{6+7}{2}$, or 6.5.

Therefore, $m = 7.5$ and $M = 6.5$, and $m - M = 7.5 - 6.5 = 1$.

7. B



If the value 148 is at the 12th percentile of the distribution, the value must be in the shaded region which is in between $160 - d$ and $160 - 2d$.

If $d = 5$, $160 - d = 160 - 5 = 155$ and $160 - 2d = 160 - 2 \cdot 5 = 150$, which does not include 148.

If $d = 10$, $160 - d = 160 - 10 = 150$ and $160 - 2d = 160 - 2 \cdot 10 = 140$, which includes 148.

If $d = 15$, $160 - d = 160 - 15 = 145$ and $160 - 2d = 160 - 2 \cdot 15 = 130$, which does not include 148.

If $d = 20$, $160 - d = 160 - 20 = 140$ and $160 - 2d = 160 - 2 \cdot 20 = 120$, which does not include 148.

Choice B is correct.

8. B

Ratings of Laptop A by 100 Reviewers

Ratings	5	4	3	2	1
Frequency	28	45	11	7	9

Ratings of Laptop B by 100 Reviewers

Ratings	5	4	3	2	1
Frequency	22	24	18	20	16

The standard deviation is a measure of how far the data set values are from the mean. In the data set for laptop A, the large majority of the data are in two of the five possible values, which are the two values closest to the mean. In the data set for laptop B, the data are more spread out, thus by observation, the data for laptop B have a larger standard deviation.

Choice B is correct.