

Assignments for Prob/Stat/Discrete Unit 4B Chapter 2: Descriptive Statistics

Day	Date	Assignment (Due the next class meeting)
		4.3 Worksheet
		4.4 Worksheet
		4.5 Worksheet
		Unit 4B Practice Test
		Unit 4 Test

NOTE: Be prepared for daily quizzes.

HW reminders:

- If you cannot solve a problem, get help **before** the assignment is due.
- Help is available before school, during lunch, or during IC
- For extra practice, visit www.interactmath.com
- Click ENTER, then scroll down to Larson, Elementary Statistics 4th edition. Pick the assignment you need extra practice with. You can get immediate feedback and hints.

Don't forget that you can get 24-hour math help from www.smarthinking.com

Section 4.3: Measures of Central Tendencies

Objectives:

1. Can you determine the mean, median, and mode of a population and of a sample?
2. Can you determine the weighted mean of a data set and the mean of a frequency distribution?
3. Can you describe the shape of a distribution as symmetric, uniform, or skewed and compare the mean and median for each?

Measures of Central Tendency:

- A value that represents a typical, or central, entry of a _____.
- Most common measures of central tendency:
 - _____
 - _____
 - _____

Mean (_____)

- The _____ of all the data entries divided by the number of _____.
- Sigma notation: $\Sigma x =$ _____ all of the data entries (x) in the data set.
- Population mean: Using every entry from the entire population. $\mu = \frac{\Sigma x}{N}$
- Sample mean: Using data collected from a sample. $\bar{x} = \frac{\Sigma x}{n}$

Median: The value that lies in the _____ of the data when the data set is **ordered**.

- Measures the center of an ordered data set by dividing it into two _____ parts.
- If the data set has an
 - odd number of entries: median is the middle data entry.
 - even number of entries: median is the _____ of the two middle data entries.

Mode: The data entry that occurs with the _____ frequency.

- If no entry is repeated the data set has _____.
- If two entries occur with the same greatest frequency, each entry is a mode (_____).

Example 1: The prices (in dollars) for a sample of roundtrip flights from Chicago, Illinois to Cancun, Mexico are listed. Find the mean, median, and mode price of the flights.

872 432 397 427 388 782 397



Example 2: At a political debate a sample of audience members was asked to name the political party to which they belong. Their responses are shown in the table. What is the mode of the responses?

Political Party	Frequency, f
Democrat	34
Republican	56
Other	21
Did not respond	9



How could the mean of the above data be found? This is called a _____.

Weighted Mean

- The mean of a data set whose entries have varying _____.
- $\bar{x} = \frac{\sum(x \cdot w)}{\sum w}$ where w is the weight of each entry x .

Example 4: You are taking a class in which your grade is determined from four sources: 65% from your test scores, 10% from your assignments, 20% from your final exam, and 5% from your binder grade. Your scores are 86 (test average), 96 (assignment average), 82 (final exam), and 98 (binder). What is the weighted mean of your scores? If the minimum average for an A is 90, did you get an A?

Comparing the Mean, Median, and Mode

- All three measures describe a typical entry of a data set.
- Advantage of using the mean:
 - The mean is a reliable measure because it takes into account _____ of a data set.
- Disadvantage of using the mean:
 - Greatly affected by _____ (a data entry that is far removed from the other entries in the data set). Do **NOT** use the mean as a “typical entry” if there exists one or more extreme outliers!

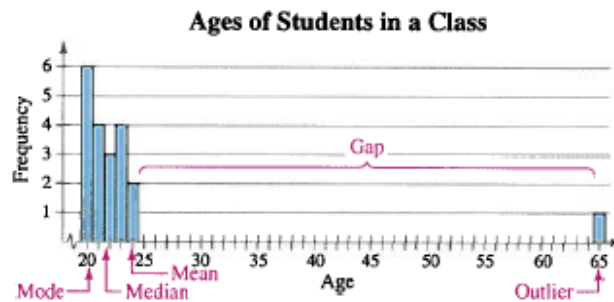
Example 5: Find the mean, median, and mode of the sample ages of a college class shown. Which measure of central tendency best describes a typical entry of this data set? Are there any outliers?

Ages in a class						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Things to consider...

- 1) Is there an outlier? If so, do not use the _____.
- 2) Is the mode a “typical central” entry?
- 3) The median is a good choice if the mean cannot be used.

Sometimes a graphical approach can help you decide which entry is best to use.

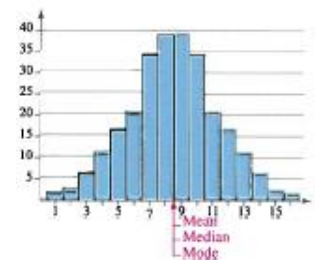


Shapes of Distributions

- Symmetric
- Uniform
- Skewed Left
- Skewed Right

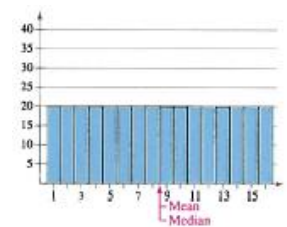
Symmetric Distribution

- A _____ line can be drawn through the middle of a graph of the distribution and the resulting halves are approximately _____ images. The mean, median, and mode are the _____ values.



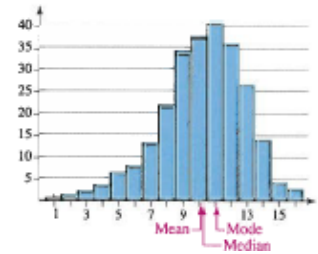
Uniform Distribution (_____)

- _____ entries or classes in the distribution have equal or **approximately** equal frequencies.
- Symmetric (and so the mean, median, and mode are the same values.)



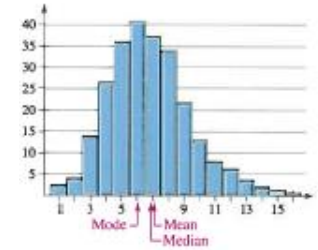
Skewed Left Distribution (negatively skewed)

- The “_____” of the graph elongates more to the left.
- The _____ is to the left of the median.



Skewed Right Distribution (positively skewed)

- The “_____” of the graph elongates more to the right.
- The _____ is to the right of the median.



Using a calculator to find Mean and Median ☺

Step 1: Input your data.

Button	Comments
STAT	
EDIT	(or hit ENTER)
Arrow up to L1 and hit CLEAR	To erase any previous entries
Type in each data entry, and hit ENTER after each one.	Note: The data does NOT have to be ordered.
Make sure you don't forget to clear out L1... there could be “hidden” entries that impact your calculations!	

Step 2: Perform calculations with your data.

Button	Comments
	Note: L1 must be entered!!!
STAT	
CALCULATE	Use the right arrow button to select this.
ENTER	To select 1-Var Stats (one variable statistics)
ENTER	To activate the command. L1 is the default... you can pick another list by typing in a new one (hit 2 nd first.)
You can easily see the mean, and if you scroll down, you will find the median (M.) Many other statistical values are also given that we will discuss this semester.	

Example 6: Use your calculator to find the mean and median of the following list of monthly fees, in dollars, for car insurance for teenagers:

247, 130, 290, 150, 160, 221, 205, 187, 199, 220, 263, 144, 172, 261, 220, 201, 189, 240, 156

4.4 Notes: Measures of Variation

Objectives:

4. Can you determine the range of a data set?
5. Can you determine the variance and standard deviation of a population and of a sample?
6. Can you use the Empirical Rule to interpret standard deviation?

Range: How _____ is the data?

- ❖ The _____ between the maximum and minimum data entries in the set.
- ❖ The data must be _____.
- ❖ Range = Max. data entry – Min. data entry

Example 1: A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries.

Starting salaries (1000s of dollars)
41 38 39 45 47 41 44 41 37 42

Deviation: How far is an entry from the _____?

- The difference between the data entry, x , and the _____ of the data set.
- _____ data set:
 - Deviation of $x = x - \mu$
- _____ data set:
 - Deviation of $x = x - \bar{x}$

(Keep in Mind, we will use the Calculator to find Standard deviation, this just lets you know how)

Example 2: Use the starting salaries from example 1 to find the deviations for each entry.

Step 1: Find the sample mean $\bar{x}$.

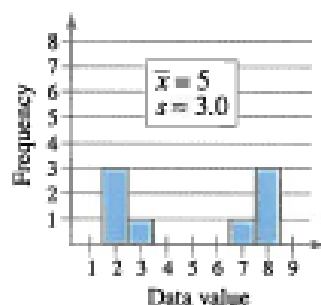
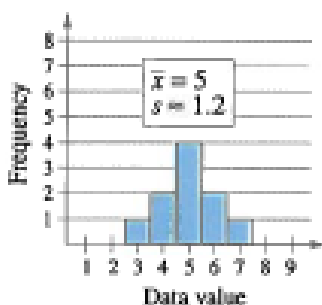
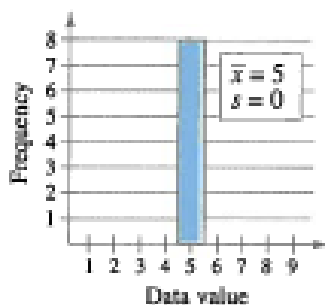
Step 2: Find the deviations: How far is each entry from the mean?

Salary (\$1000s), x	Deviation: $x - \mu$	=
41	$41 - 41.5 =$	
38	$38 - 41.5 =$	
39	$39 - 41.5 =$	
45	$45 - 41.5 =$	3.5
47	$47 - 41.5 =$	5.5
41	$41 - 41.5 =$	-0.5
44	$44 - 41.5 =$	2.5
41	$41 - 41.5 =$	-0.5
37	$37 - 41.5 =$	-4.5
42	$42 - 41.5 =$	0.5

$$\sum x = \quad \quad \quad \sum (x - \bar{x}) =$$

Standard Deviation

- A measure of the typical amount an entry deviates from the _____.
- The “average” of the deviations (kind of.)
- The more the entries are _____, the greater the standard deviation.
- Sample standard deviation uses the symbol _____.
- Population standard deviation uses the symbol _____.



Variance

- Another measure of spread.
- It is the value of the standard deviation _____.
- Not as useful to our purposes as standard deviation, because it is not in the same units as our data list (units _____ instead.)
- Used at a higher level in more advanced statistics courses.
- Sample variance uses the symbol _____.
- Population variance uses the symbol _____.

Formulas for Standard Deviation and Variance

	Standard Deviation	Variance
Population	$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x - \mu)^2}{N}}$ Step 1: Find the mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the number of entries. Step 6: Square root this value (the variance.)	$\sigma^2 = \frac{\sum(x - \mu)^2}{N}$ Step 1: Find the mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the number of entries.
Sample	$s = \sqrt{s^2} = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$ Step 1: Find the sample mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the sample size minus 1. Step 6: Square root this value (the variance.)	$s^2 = \frac{\sum(x - \bar{x})^2}{n-1}$ Step 1: Find the sample mean. Step 2: Find each deviation from the mean. Step 3: Square each deviation. Why? Step 4: Find the sum of those squares. Step 5: Divide by the sample size minus 1.

(Keep in Mind, we will use the Calculator to find Standard deviation, this just lets you know how)

Example 3: A corporation hired 10 graduates. The starting salaries for each graduate are shown. Find the population variance and standard deviation of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

Step 1: Recall $\mu = 41.5$

Step 2: We already found the deviations are as shown.

Step 3: Find the squares of the deviations.

Step 4: Find the sum of the squares of the deviations.

Step 5: Divide by the number of entries.

Salary, x	Deviation: $x - \mu$	Squares: $(x - \mu)^2$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = \underline{\hspace{1cm}}$
38	$38 - 41.5 = -3.5$	$(-3.5)^2 = \underline{\hspace{1cm}}$
39	$39 - 41.5 = -2.5$	$(-2.5)^2 = 6.25$
45	$45 - 41.5 = 3.5$	$(3.5)^2 = 12.25$
47	$47 - 41.5 = 5.5$	$(5.5)^2 = 30.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
44	$44 - 41.5 = 2.5$	$(2.5)^2 = 6.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
37	$37 - 41.5 = -4.5$	$(-4.5)^2 = 20.25$
42	$42 - 41.5 = 0.5$	$(0.5)^2 = 0.25$
$\Sigma(x - \mu) = 0$		$\Sigma(x - \mu)^2 =$

(Keep in Mind, we will use the Calculator to find Standard deviation, this just lets you know how)

Example 4: The starting salaries are for the Chicago branches of a corporation. The corporation has several other branches, and you plan to use the starting salaries of the Chicago branches to estimate the starting salaries for the larger population. Find the **sample** standard deviation of the starting salaries.

Starting salaries (1000s of dollars)

41 38 39 45 47 41 44 41 37 42

How is this problem different than Example 3? How can we adjust our results to find the sample standard deviation rather than the population standard deviation?

Using the TI-83 to find standard deviation and variance:

Step 1: Input your data in a list (STAT, EDIT...)

Step 2: Perform calculations (STAT, CALCULATION, 1-VAR STAT, ENTER)

Example 5: Sample office rental rates (in dollars per square foot per year) for Miami's central business district are shown in the table. Use a calculator to find the mean rental rate, the standard deviation, and the variance. (*Adapted from: Cushman & Wakefield Inc.*)

Note: What type of standard deviation and variance do we want?

Office Rental Rates		
35.00	33.50	37.00
23.75	26.50	31.25
36.50	40.00	32.00
39.25	37.50	34.75
37.75	37.25	36.75
27.00	35.75	26.00
37.00	29.00	40.50
24.50	33.00	38.00

Working With Bell-Shaped Distributions:

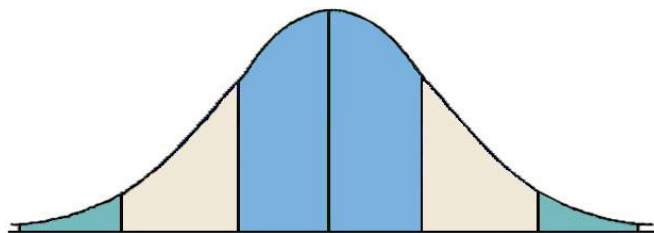
- Often described as _____ distributions.
- Symmetrical.
- Area under curve = 1.

Use the information above to predict the shaded area in the normal curves given to you in the 2.4 Activity.

Creating a Diagram Using Mean and Standard Deviation:

For data with a (_____) bell-shaped distribution, the standard deviation can be used to separate the data into sections.

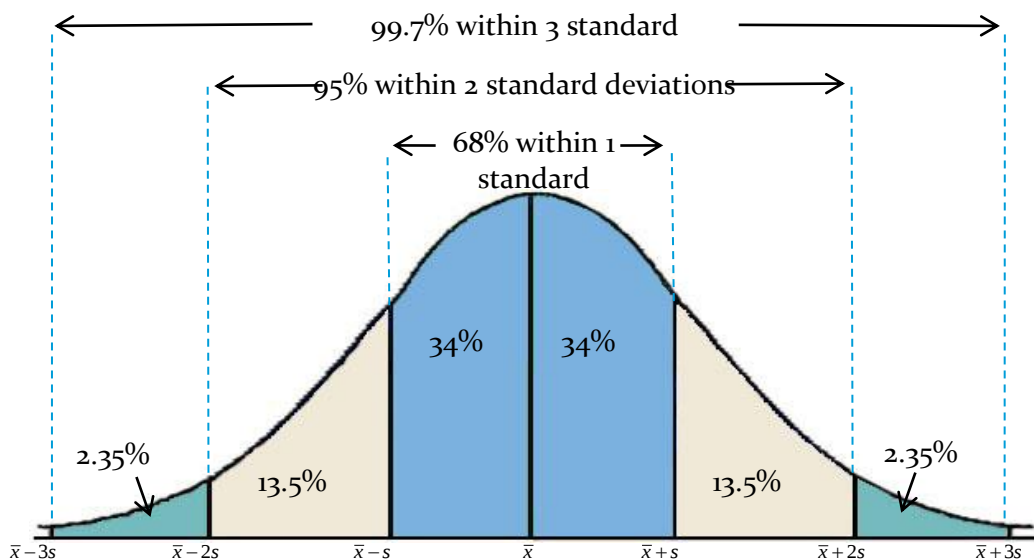
Example 2: A math test has a normal distribution, and the mean is 72% with a standard deviation of 6. Draw the distribution.



Empirical Rule (68 – 95 – 99.7 Rule)

For data with _____ distributions, the standard deviation has the following characteristics:

- About 68% of the data lie within _____ standard deviation of the mean.
- About 95% of the data lie within _____ standard deviations of the mean.
- About 99.7% of the data lie within _____ standard deviations of the mean.



Example 2: In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20-29) was 64 inches, with a sample standard deviation of 2.71 inches. Estimate the percent of the women whose heights are between 64 inches and 69.42 inches.

Step 1: Draw a diagram.

Step 2: Use the Empirical Rule
(68 – 95 – 99.7 Rule)

Example 3: IQ scores for adults have a bell-shaped distribution with a mean of 90 and a standard deviation of 12. Use the Empirical Rule to find the percentage of adults with scores between 78 and 102.

Step 1: Draw a diagram.

Step 2: Use the Empirical Rule.

Unusual Scores: Any score that is more than _____ standard deviations away from the mean (above OR below) is considered to be unusual and is an _____.

Example 4: Use the normal distribution created in #3 to identify which of the following scores, if any, are unusual and would be considered outliers: 100, 60, 70, 117

4.5 Notes: Measures of Position

Objectives:

8. Can you determine the quartiles of a data set?
9. Can you determine the interquartile range of a data set?
10. Can you create a box-and-whisker plot?
11. Can you interpret other fractiles such as percentiles?
12. Can you determine and interpret a standardized score (z-score)?

Quartiles

- Approximately divide an ordered data set into _____ equal parts.
- 1st quartile, Q_1 : About _____ of the data fall on or below Q_1 .
- 2nd quartile, Q_2 : About _____ of the data fall on or below Q_2 (median).
- 3rd quartile, Q_3 : About _____ of the data fall on or below Q_3 .

Example 1: The test scores of 15 employees enrolled in a CPR training course are listed. Find the first, second, and third quartiles of the test scores.

13 9 18 15 14 21 7 10 11 20 5 18 37 16 17

Step 1: Order the data.

Step 2: Find the median.

Step 3: Find the quartiles
(the medians of each half.)

Example 2: Using the data from Example 1, answer the questions below.

- a) About what % of employees score below 10?
- b) About what % of employees score below 18?
- c) About what % of employees score above 15?

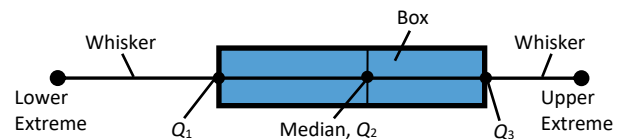
Interquartile Range (___)

- The _____ between the _____ and first quartiles.
- $IQR = Q_3 - Q_1$

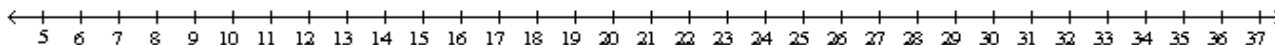
Example 3: Find the IQR for the data in Example 1.

Box-and-whisker plot

- Divides data set into quartiles.
- Requires (____-number summary):
 1. Lower extreme (minimum value)
 2. Q_1 (lower quartile)
 3. Q_2 (median)
 4. Q_3 (upper quartile)
 5. Upper extreme (maximum value)



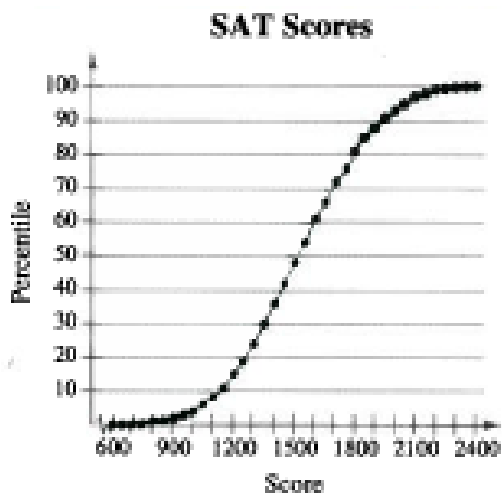
Example 4: Draw a box-and-whisker plot from the data in Example 1. Recall that Min = 5 Q1 = 10 Q2 = 15 Q3 = 18 Max = 37



Fractiles are numbers that partition (divide) an _____ data set into _____ parts. The chart below describes different types of fractiles.

Fractiles	Summary	Symbols
Quartiles	Divides data into ___ equal parts	$Q1, Q2, Q3$
Deciles	Divides data into ____ equal parts	$D1, D2, D3, \dots, D9$
Percentiles	Divides data into _____ equal parts	$P1, P2, P3, \dots, P99$

Example 5: The ogive shown represents the cumulative frequency distribution for SAT test scores of college-bound students in a recent year. (Source: College Board Online)



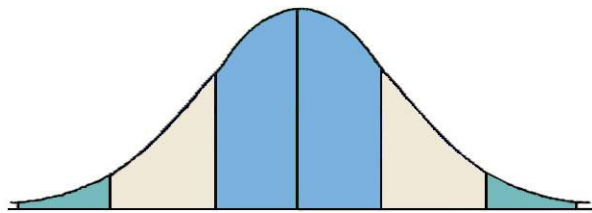
a) What test score represents the 72nd percentile (P_{72})? How should you interpret this?

b) What percentile is a score of 1800? Interpret this score.

Example 6: Sara is taller than 28 of the students in her class. There are 32 students in her class. What is her percentile ranking?

Standard Normal Distribution

- Mean = _____
- Standard Deviation = _____

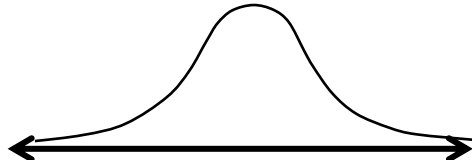


Standardized Score (z-score)

- ▶ Represents the number of _____ a given value x falls from the mean μ .
- ▶ Used to transform any score to fit a **Standard Normal Distribution** (mean = _____ and standard deviation = _____.)
- ▶ An **unusual score** (or an _____) is more than 2 standard deviations above or below the mean.
- ▶
$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

Example 1: In 2007, Forest Whitaker won the Best Actor Oscar at age 45 for his role in the movie *The Last King of Scotland*. Helen Mirren won the Best Actress Oscar at age 61 for her role in *The Queen*. The mean age of all best actor winners is 43.7, with a standard deviation of 8.8. The mean age of all best actress winners is 36, with a standard deviation of 11.5.

- a) Find Forest Whitaker's z-score for best actor winners.



- b) Find Helen Mirren's z-score for best actress winners.



- c) Compare your results. Whose age was a more unusual one?

Example 2: The weights of 19 high school basketball players have a bell-shaped distribution, with a mean of 180 pounds and a standard deviation of 15 pounds. Use standardized scores (z-scores) to determine if the weights of the following basketball players are unusual (outliers.)

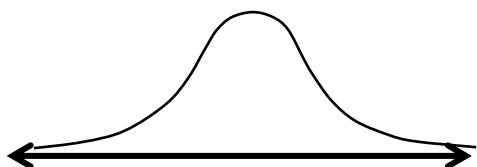
A) 140 pounds



B) 195 pounds



C) 180 pounds



Outliers... There are two main methods for determining whether or not a score is an outlier.

1) Find its z-score. If it is more than ____ standard deviations from the mean, that score is an outlier.

OR

2) Find the Interquartile Range (IQR). If the score is more than $1.5(\text{IQR})$ units higher than Q_3 or more than $1.5(\text{IQR})$ units lower than Q_1 , then it is an outlier.

Example 3: A collection of data has $Q_1 = 49$, $Q_2 = 53$, and $Q_3 = 57$. Are the following values outliers?

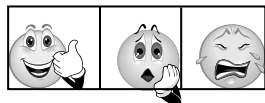
a) 40

b) 30

c) 70

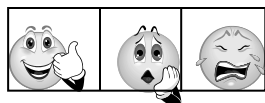
Prob/Stat/Discrete
Unit 4B Objectives

Name _____



Objective #1: Can you determine the mean, median, and mode of a population and of a sample?

The times (in seconds) for a sample of seven sports cars to go from 0 to 60 miles per hour: 3.7 4.0 4.8 4.8 4.8 5.1



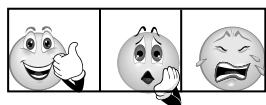
Objective #2: Can you determine the weighted mean of a data set?

a) The average starting salaries (by degree attained) for 25 employees at a company are given. What is the mean starting salary for these employees?

8 with MBAs: \$45,500

17 with BAs in business: \$32,000

b) A science class has the following percentages for grading categories: Tests 40%, Labs 20%, Final 25%, Homework 15%. Terry has a 70% average on tests, 90% on labs, 85% on the final, and 94% on homework. What is her weighted mean?

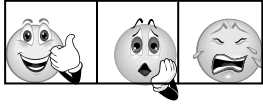


Objective #3: Can you describe the shape of a distribution as symmetric, uniform, skewed right, or skewed left?

A) Each data class has approximately the same frequency.

B) Median is greater than the mean.

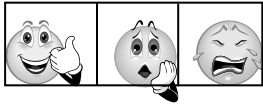
C) Mean is greater than the median.



Objective #4: Can you determine the range of a data set?

Weekly salaries (in dollars) for a sample of registered nurses are listed.

774 446 1019 795 908 667 444 960



Objective #5: Can you use variance and standard deviation?

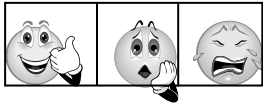
a) Find the variance and standard deviation of the follow sample data:

17 8 13 18 15 9 10 11 6

b) What does the standard deviation tells about a data set?

c) The standard deviation of a data set is 32 inches. What is the variance?

d) The variance of a data set is 4.5 m^2 . What is the standard deviation?



Objective #6: Can you use the Empirical Rule to interpret standard deviation?

a) Draw the following normal curve: mean of 50, standard deviation of 6.

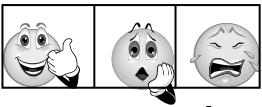
For b - d) Using the graph from 6a...

b) What percentage of the data is between 44 and 56?

c) Between 44 and 62?

d) Less than 38?

e) The mean value of land and buildings per acre from a sample of farms is \$2400, with a standard deviation of \$450. Between what two values do about 95% of the data lie? (Assume the data set has a bell-shaped distribution.)

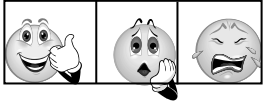


Objective #8: Can you determine the quartiles of a data set?

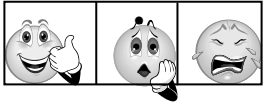
The number of wins for each Major League Baseball team in 2006 are listed.

97	87	86	70	61	96	95	90	78	62
93	89	80	78	97	85	79	78	71	83
82	80	75	67	66	88	88	76	76	76

Find the quartiles of the data set.



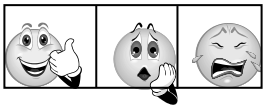
Objective #9: Can you determine the interquartile range of a data set from Objective #8?



Objective #10: Can you create a box-and-whisker plot using the information from Objectives 8 and 9?



- b) What percentage of the data is lower than 76?
- c) What percentage of the data is lower than 88?
- d) What percentage of the data is between 76 and 88?

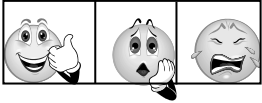


Objective #11: Can you interpret other fractiles such as percentiles?

a) Sam took the ACT, and so did 321 other students at his school. 199 students had scores equal to Sam's or less than Sam's score. What is Sam's percentile?

b) Freddy's weight when he was 7 years old is at the 40th percentile. What does this mean?

c) In 2007 there were 768 “oldies” radio stations in the United States. If one station finds that 84 stations have a larger daily audiences than it has, what percentile does the station come closest to in daily audience rankings?



Objective #12: Can you determine and interpret the standard score (z-score)?

The weights of 19 high school football players have a bell-shaped distribution, with a mean of 186 pounds and a standard deviation of 18 pounds. Find the z-scores for the following player weights.

A) 178 pounds

B) 249 pounds

c) Draw a diagram showing the location of each weight on the normal curve.

d) Which of these scores, if any, is unusual?