Name: ____

Foundations of Common Core Math 3

Unit 1 - Statistics

Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write.

- H. G. Wells

WAKE COUNTY PUBLIC SCHOOL SYSTEM

APEX HIGH SCHOOL
1501 LAURA DUNCAN ROAD
APEX, NC 27502

Foundations of Common Core Math 3

Unit 1 Statistics

Day	Date	Homework
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Foundations of Common Core Math 3
Unit 1: One-Variable Data
PRE-ASSESSMENT

The purpose of this assessment is to see what you already know about one-variable statistics. Please answer each question to the best of your ability.

1) What is the difference between a measure of center and a measure of spread?

2) What is an outlier? Does it affect the mean of a sample? Does it affect the median? Give specific examples to support your answer.

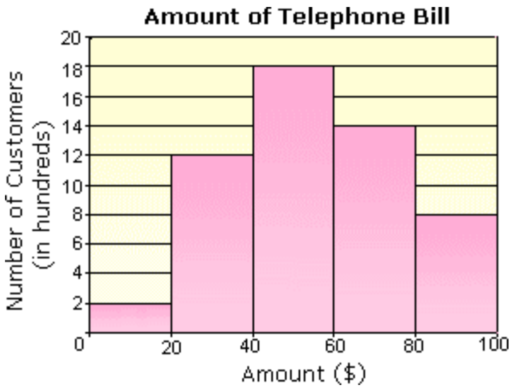
3) The following is a sample set of quiz grades for the first quiz in Mr. Smith's class: 72, 90, 85, 75, 87, 100, 30, 86, 82, 100, 75

Find the mean and the median of the grades. Which measure best represents the data? Why? Describe the distribution of grades in context.

4) The following is the sample set of quiz grades for the second quiz in Mr. Jones' class: 78, 86, 70, 39, 93, 78, 82, 79, 78, 85

Find the median and interquartile range for this set of data. How do the two sets of quiz grades compare?

5) The histogram below shows the graphs of telephone bills of customers in a certain city. How many customers had bills of at least $40?



6) In your own words, explain what a standard deviation represents. Can a standard deviation be negative? Can it be zero? Explain your answers thoroughly – you may give examples to support your answer.

7) If a sample is randomly selected, it means that (choose one)
a) no thought is given to the items being selected.
b) every item has the same chance of being selected.
c) it doesn't matter which items are selected.
d) the size of the sample can be different every time.

8) If a question is biased, it means that (choose one)
a) it may have two answers.
b) only certain people have to answer it.
c) the answer may or may not be right.
d) the wording may make people answer a certain way.

9) A student in AP Statistics is interested in finding out how many students believe that the dress code is reasonable. He asks a sample of senior girls the following question: "Do you agree that the dress code is old-fashioned and that it's hard to buy clothes that fit the current dress code?" Is this a fair question? Explain your answer.

Foundations of Common Core Math 3 – Unit #1 Modeling with Statistics

Topics in this unit:

- **Measures of Center**
  - find median and mean of given data set
  - compare the values of median and mean and determine the best measure of center
  - provide information about a data set given its histogram model
  - determine the shape of a data set given its histogram model (skewness)
- **Using Normal Distributions**
  - find standard deviation of given data set
  - when is a distribution approximately normal
  - how to compare distributions using the standard normal curve
  - find probabilities of events that are normally distributed
- **Sampling and Study Design**
  - how to select samples
  - why sampling choices influence the bias present in the sample

By the end of this unit students will be able to:

- Find measures of center by hand and using a calculator.
- Find standard deviation using a calculator.
- Read a given histogram and provide information about the data, specifically its distribution.
- Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use a table to estimate areas under the normal curve.
- Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
- Use data from a sample survey to estimate a population mean or proportion.
- Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
- Use data from a randomized experiment to compare two treatments.
- Evaluate reports based on data.

By the end of this unit, students will be able to answer the following questions:

- What does it mean when data follows a normal distribution?
- How does standardizing data help our understanding of a distribution?
- What are the advantages and disadvantages of observational studies and experiments?
- Is anything in life truly random?
- How can we analyze data to make inferences and/or predictions, based on surveys, experiments, probability and observational studies?

VOCABULARY

- **Mean** is the same as finding the average of a data set. Mean is often used to measure center in business, engineering, and computer science. Mean is most useful when the data set does NOT have any outliers.
- **Median** is the middle value of a data set. Extreme values (outliers) do not affect the median, so it is a better measurement of center when outliers are involved.
- **Standard deviation** is a numerical value used to indicate how widely individuals in a group vary. If individual observations vary greatly from the group mean, the standard deviation is big; and vice versa.
- **Skewness** is a measure of the degree of asymmetry of a distribution. If the left tail (tail at small end of the distribution) is more pronounced than the right tail (tail at the large end of the distribution), the function is said to be skewed left. If the reverse is true, it is skewed right. If the two are equal, it is said to be symmetrical and is representative of a normal distribution.
- **A normal distribution** is a symmetric, bell shaped curve with a single peak. Approximately 68% of observations fall within one standard deviation of the mean, 95% fall within two standard deviations of the mean, 99.7% fall within three standard deviations of the mean. (This is called **The Empirical rule**.)
- **A standard normal curve** is a normal curve where the mean is zero and the standard deviation is one.
- **A z-score** represents a value's position on a standard normal curve. It is used to compare values from two different data sets.
- **Population** is the entire group of individuals that we want information about.
- **Census** is a complete count of the population.
- **Sample** is the part of the population that we actually examine in order to gather information.
- **Experimental design** is the method to gather data.
  - **Surveys**: most often involve the use of a questionnaire to measure the characteristics and/or attitudes of people.
  - **Observational studies**: individuals are observed and certain outcomes are measured, but no attempt is made to affect the outcome.
  - **Experiments**: treatments are imposed prior to observation.
- **Sampling design** is the method used to choose a sample from the population.
  - **Simple random sample** (SRS): all individuals in the population have the same probability of being selected and all groups of the sample size have the same probability of being selected
  - **Stratified random sample**: the researcher divides the entire target population into different subgroups, or strata, and randomly selects the subjects proportionally from the different strata.
  - **Cluster sample**: the entire population is divided into groups, or clusters, and a random sample of these clusters are selected. All individuals in the selected clusters are included in the sample.
  - **Systematic random sample**: the researcher selects a number at random, n, and then selects every nth individual for the study.
  - **Convenience sample**: subjects are taken from a group that is conveniently accessible to a researcher.
  - **Voluntary Response sample:** subjects select themselves, for example an on-line survey.
- **A parameter** is a value that represents a population**.** It is typically represented by Greek letters
- **A statistic** is a value that represents a sample. It is not represented by Greek letters
- **Simulation** is a way to model random events.
- **Random number generators** (e.g. on graphing calculators) and **random digit tables** can be used to randomize selection for simulations.

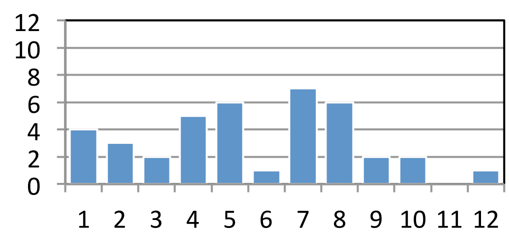
Warm Ups and Exit Ticket Questions

1) Tell whether the answers to the questions are quantitative or categorical data. If the data is quantitative, indicate the unit of measurement.

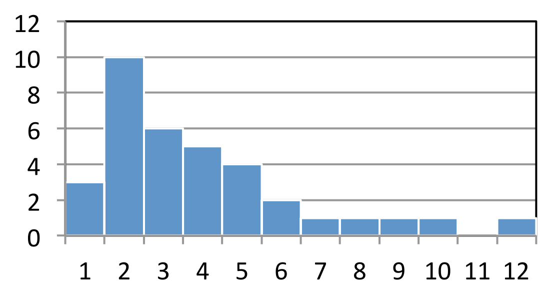
a) What is your height in centimeters?
b) What is your favorite band?
c) What would you like to do after you graduate from high school?
d) How much time do you spend doing homework each day?
e) On a scale of 1 to 5, with 5 being excellent and 1 being poor, how would you rate the food in the cafeteria?

2) Use the graphs below. The graphs show information about a class of students.

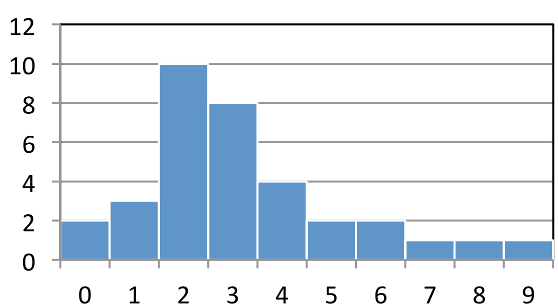
Graph A



Graph B



Graph C



a) Which graph might show the number of children in the student's families? Explain.

b) Which graph might show the birth months of the students? Explain.

c) Which graph might show the number of toppings students like on their pizzas?

d) Give a possible title, a label for the vertical axis, and a label for the horizontal axis for each graph, based on your answers to questions a-c.

3) Annette has a hamster that is three years old. She wonders if her hamster is old compared to other hamsters. On the internet, she finds out that the median age for a hamster is 2 ½ years.

a) What does the median tell Annette about the life span of a hamster?

b) How would knowing how the data vary from the least value to the greatest value help Annette predict the life span of her hamster?

4) Ms. Jackson's class also collected data on the number of pets each student has at their home. Use this data to answer the following questions.

a) Make a histogram of the data. Include the graph or a sketch of the graph from the calculator.

b) Find the median.

c) Describe the distribution (shape, center, spread, outliers).

d) Do you think the students live in a city, the suburbs, or the country? Explain.

Number of Pets

Number	Frequency
0	2
1	2
2	5
3	4
4	1
5	2
6	3
7	0
8	1
9	1

5) Using salary information from the Wacky Widget Company, answer the questions below.

Job	**Number of Employees**	**Annual Salary**
President	1	$200,000
Vice President	1	$50,000
Supervisor	2	$25,000
Sales Representative	4	$21,000
Warehouse Worker	2	$15,000
Custodian	2	$15,000
Clerical Worker	3	$12,000

a) Find the mean salary.

b) Find the median salary.

c) Why do you suppose a recent newspaper headline reporting "Average worker at Wacky Widget making $32,000" upset most employees?

d) Create a newspaper headline that you feel is more appropriate.

6) Is it possible for a data set to have a standard deviation of 0? If so, give an example. If not, tell why not.

7) For the following data set, find the mean, median, and standard deviation:

4, 6, 6, 8, 10, 12, 16, 18

8) Give an example of a data set that is not normally distributed. Draw a curve that represents your data set.

9) A visitor from the star Alpha Centauri has selected you to provide her with information about our solar system. She asks about the length of a typical day in our solar system. Study the following table.

Planet	Approximate Length of a Day in Earth Hours
Mercury	1416
Venus	5832
Earth	24
Mars	24.5
Jupiter	10
Saturn	11
Uranus	22
Neptune	16

a) Compute the mean length of a day in our solar system in hours.
b) How many Earth days is this?
c) Find the median length of a day in our solar system.

10) The miles per gallon for city travel of ten cars and ten SUVs are given below:

Cars	MPG	SUVs	MPG
Geo Metro	46	Jeep Grand Cherokee	14
Honda Civic CX	42	Ford Explorer	21
Hyundai Excel GS	29	Chevy Silverado	21
Mazda 323	29	Toyota Tacoma	22
Plymouth Sundance	26	Nissan Frontier	16
Saturn SL	28	Chevy Suburban	12
Eagle Summit	31	GMC Yukon	12
Nissan Sentra E	29	Ford F150	13
Ford Festiva GL	35	Jeep Wrangler	18

Write a few sentences comparing the two types of automobiles. Make predictions about the shape, center, spread, and any outliers of each type of automobile.

11) Which of the following data sets will have the largest standard deviation? Which will have the smallest standard deviation?

A. 3, 6, 9, 12, 15

B. 5, 7, 9, 11, 13

C. 1, 5, 9, 13, 17

12) List a reasonable unit for the following measurements.

a) height of a person

b) height of a building

c) distance from your house to school

d) length of time to brush your teeth

e) length of time to drive across the state

13) **Read the problem and pick out the important information. Tell which information is not needed in order to solve the problem.**

It takes Alex 30 minutes to walk to school. His house is 2 miles from the school. Just before Alex left his house to walk to school, he counted the money he had in his pocket. He had $2.10 in dimes. After walking for 20 minutes Alex discovered he had a small hole in his pocket. Alex counted his money and estimated that he was losing one dime every 4 minutes through the hole. How much money did Alex have after 20 minutes?

14) For the following scenarios determine which one should have a symmetrical distribution, which one should be skewed left, and which one should be skewed right.

A) Incomes for all professional athletes

B) Test scores on an easy test

C) Heights of people

15) You are planning to take on a part time job as a waiter at a local restaurant. During your interview, the boss told you that their best waitress, Jenny, made an average of $70 a night in tips last week. However, when you asked Jenny about this, she said that she made an average of only $50 per night last week. She provides you with a copy of her nightly tip amounts from last week (see below).

Day	Tip Amount
Sunday	$50
Monday	$45
Wednesday	$48
Friday	$125
Saturday	$85

A) Calculate the mean tip amount.

Calculate the median tip amount.

B) Which value is Jenny's boss using to describe the average tip? Why do you think he chose this value?

C) Which value is Jenny using? Why do you think she chose this value?

Exploring Distributions

Adapted from Core Plus Course 1: Unit 2, Lesson 1

"The statistical approach to problem solving includes refining the question you want to answer, designing a study, collecting the data, analyzing the data collected, and reporting your conclusions in the context of the original question."

Investigation 1

Exploring Distributions Activity

How many pennies can you stack using your dominant hand?

How many pennies can you stack using your non-dominant hand?

Here are the rules to answer these questions:

1) You can touch pennies **only** with the one hand you are using.
2) You have to **place the pennies one at a time** on the stack **without touching any of pennies already on the stack.**
3) Once you let go of a penny it **cannot** be moved.
4) When a penny or pennies fall off the stack you are done.
5) Your score is the number of pennies you had stacked before a penny or pennies fell.

Each member of the group should perform this experiment once using their **dominant hand**, **and** then once using their **non-dominant hand**.

When your group has finished, write your data on the blackboard.

Use the data from the entire class to do the following:

1. Organize the data from least to greatest for both dominant and non-dominant hands. You may want to use a stem-and-leaf plot.

2. Measures of center:

- The **median** is the midpoint of an ordered list of data – at least half the values are at or below the median and at least half are above it. When there are an odd number of values, the median is the one in the middle. When there is an even number of values, the median is the average of the two in the middle.
- The **mean**, or arithmetic average, is the sum of the values divided by the number of values.
- The **mode** is the item of data that appears more frequently than any other in the set. Data with 2 (3 or more) modes are called bimodal (trimodal or multimodal). Data sets have no mode when each item of the set has equal frequency.

Find the median, mean, and mode for each set of data (dominant hand, and non-dominant hand).

3. When describing a distribution it is important to include information about its shape, its center, and its spread.

a) Compare the shapes of the 2 distributions (skewed left, skewed right, symmetrical). Discuss the similarities and differences between the 2 distributions.
b) Are there any numbers that fall outside the overall pattern of either distribution?
c) Compare the centers of the two distributions.

4. Assume the maximum value of the dominant data set was replaced with 500 pennies. Will that have a greater effect on the mean, median or mode? Defend your answer.

5. A histogram of the data will be is clearer if nearby values are grouped together in the frequency distribution. To do this divide up the range of possible values into equal intervals. Each interval is called a **class**, and the **class interval** is range of each class. The **class limits** are the upper and lower values in each interval. For example the table below divides the data into classes with a class interval of 5. The class limits are 0, 5, 10, 15, 20, 25, 30, …..

# of Pennies Stacked	**Frequency**
0-5	
5-10	
10-15	
15-20	
20-25	
etc…….	

If n = # of pennies:

The class 0-5 counts the values of n such that $0 \le n < 5$

The class 5-10 counts the values of n such that $5 \le n < 10$

The class 10-15 counts the values of n such that $10 \le n < 15$

etc……

Make 2 **frequency distribution** tables, one for the dominant hand data, and one for non-dominant hand data. Each table will have 2 columns: (1st column: **# of pennies stacked**, 2nd column: the **frequency** (i.e. the number of times this value occurred)). You can use a class interval of 5 (as in the table above), or choose another class interval if you think it more appropriate.

6. Make a histogram for each of your tables from #5 above. Be sure to label the axes.

7. Do either of the have outliers? Defend your answer.

Exploring Distributions Activity Name: _______________________

Investigation 1 (Answer Sheet to turn in.)

Let “D” represent Dominant Hand and “N” represent Non-Dominant Hand

1. Data (from least to greatest)

   (D)

   (N)

2. Mean (D)- Mean (N)-

   Median (D)- Median (N)-

   Mode (D)- Mode (N)-

3. a) Shape (D)-

   Shape (N)-

   b) Are there any numbers that fall outside the overall pattern of either distribution?

   c) Compare the centers

4. Assume the maximum value of the dominant data set was replaced with 500 pennies. Will that have a greater effect on the mean, median or mode? Defend your answer.

5. Complete the frequency distributions which divide the data into classes. You can use a class interval of 5 (as the example in the investigation, or choose another class interval if you think it more appropriate.

Dominant Hand	
# of Pennies Stacked	**Frequency**

Non-Dominant Hand	
# of Pennies Stacked	**Frequency**

6. Make a histogram for each of your tables from #6 above. Be sure to label the axes.

Dominant Hand

Non-Dominant Hand

7. Do either of the have outliers? Defend your answer.

Measuring Variability

Adapted from Core Plus Course 1: Unit 2, Lesson 2 (Student Edition and Teacher Edition)

"The observation that no two snowflakes are alike is somewhat amazing. But in fact, there is variability in nearly everything. When a car part is manufactured, each part will differ slightly from the others. If many people measure the length of a room to the nearest millimeter, there will be many slightly different measurements. If you conduct the same experiment several times, you will get slightly different results. Because variability is everywhere, it is important to understand how variability can be measured and interpreted.

Variability is present in all processes and in all physical things. Measures of variability tell us how spread out a distribution is. Examples of people for whom the variability in a distribution is important are doctors who must decide if a test result is abnormal, engineers who need to know how far off measurements are likely to be, educators who need to know how many items to put on a standardized test to get a precise measure of a student's achievement, and machine operators who must know how much variability is normal in the output of their machine."

Investigation 1

Median vs. Mean

Discuss the following with your partner or group. Write your answers on your own paper. Be prepared to share your answers with the class.

The heights of Washington High School's basketball players are:
5 ft 9in, 5 ft 4in, 5 ft 7 in, 5ft 6 in, 5 ft 5 in, 5 ft 3 in, and 5 ft 7 in.

A student transfers to Washington High and joins the basketball team. Her height is 6 ft 10in.

1) What is the mean height of the team before the new player transfers in? What is the median height?

2) What is the mean height after the new player transfers? What is the median height?

3) What effect does her height have on the team's height distribution and stats (center and spread)?

4) How many players are taller than the new mean team height? How many players are taller than the new median team height?

5) Which measure of center more accurately describes the team's typical height? Explain.

How do I know which measure of central tendency to use?

MEAN

Use the mean to describe the middle of a set of data that *does not* have an outlier.

Advantages:
- Most popular measure in fields such as business, engineering and computer science.
- It is unique - there is only one answer.
- Useful when comparing sets of data.

Disadvantages:
- Affected by extreme values (outliers)

MEDIAN

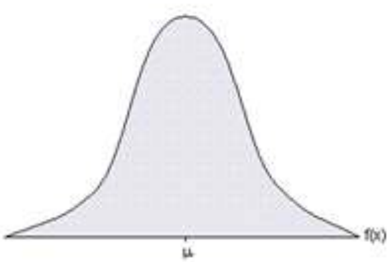
Use the median to describe the middle of a set of data that *does* have an outlier.

Advantages:
- Extreme values (outliers) do not affect the median as strongly as they do the mean.
- Useful when comparing sets of data.
- It is unique - there is only one answer.

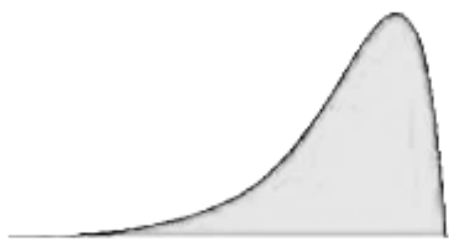
Disadvantages:
- Not as popular as mean.

http://regentsprep.org/REgents/math/ALGEBRA/AD2/measure.htm

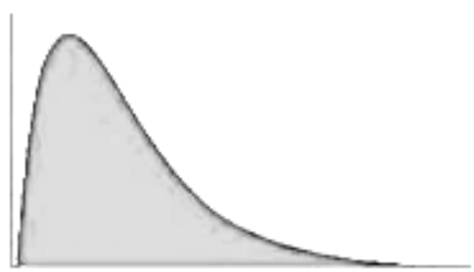
Mound-shaped and symmetrical (Normal)



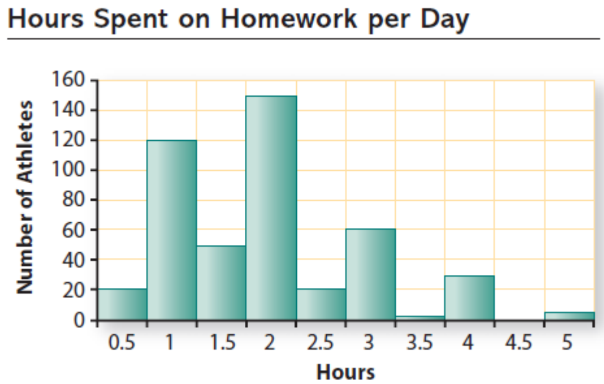
Skewed Left



Skewed Right



1. The histogram displays the results of a survey filled out by 460 varsity athletes in a football, women's basketball, and men's basketball from schools around Detroit Michigan. These results were reported in a school newspaper.



a. What is the shape of this graph?

b. Discuss the information contained in the graph and the paragraph. There are a few unusual features of this distribution. What are they and why do you think they occurred?

c. Estimate the median, lower quartile, and upper quartile. Remember you are to estimate not calculate them.

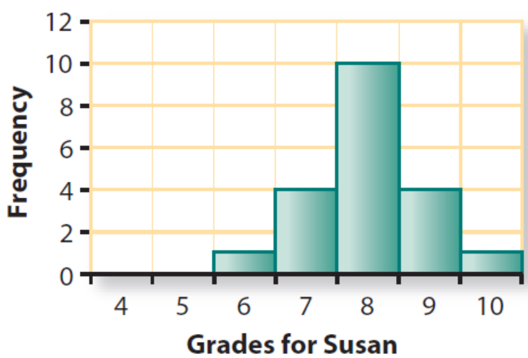
2. The math homework grades for two ninth-grade students at Lakeview High School are given.

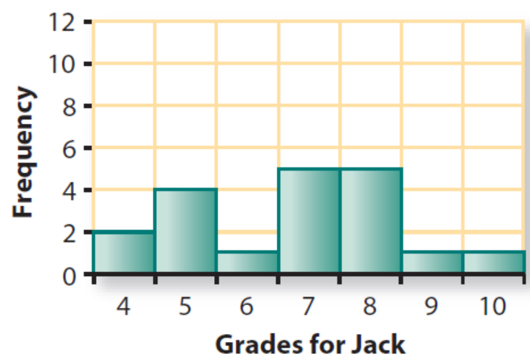
Susan's Homework Grades

8, 8, 7, 9, 7, 8, 8, 6, 8, 7,
8, 8, 8, 7, 8, 8, 10, 9, 9, 9

Jack's Homework Grades

10, 7, 7, 9, 5, 8, 7, 4, 7,
5, 8, 8, 8, 4, 5, 6, 5, 8, 7





a. What is the highest grade Susan's and Jack's teacher assigns for homework? The lowest?

b. Just by viewing the graphs, which of the students has greater variability in his/her grades? Be detailed in how you determined your answer.

c. Put the 20 grades for Susan in an ordered list and find the:

i. median

ii. Lower and upper quartiles. (Remember find the quartiles by finding the medians for the lower and upper halves).

iii. Mark these measures on your ordered list.

iv. Find the mean.

d. Put the 19 grades for Jack in an ordered list and find the:

i. median

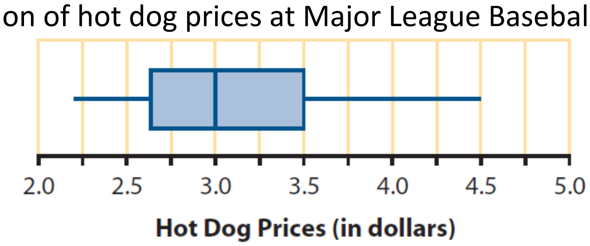
ii. Lower and upper quartiles. (Remember find the quartiles by finding the medians for the lower and upper halves; leave out the median because there is an odd number of values).

iii. Mark these measures on your ordered list.

iv. Find the mean.

e. For which student are the lower and upper quartiles farther apart? What does this tell you about the variability of the grades of the two students?

3. The following box plot shows the distribution of hot dog prices at Major League Baseball parks.



Source: www.teammarketing.com/fci.cfm?page=fci_mlb2004.cfm

a. Is this distribution skewed to the left, skewed to the right, or is it symmetric? Explain your reasoning.

b. Estimate the five-number summary. What does each value tell you about hot dog prices at Major League Baseball parks?

4. Box plots are most useful when the distribution is skewed or has outliers or if you want to compare two or more distributions. The math homework grades for five ninth-grade students at Lakeview High, Maria (M), Tran (T), Gia (G), Jack (J), and Susan (S), are show with corresponding box plots.

Maria's Grades

8, 9, 6, 7, 9, 8, 8, 6, 9, 9, 8, 7, 8, 7, 9, 9, 7, 7, 8, 9

Tran's Grades

9, 8, 6, 9, 7, 9, 8, 4, 8, 5, 9, 9, 9, 6, 4, 6, 5, 8, 8, 8

Gia's Grades

8, 9, 9, 9, 6, 9, 8, 6, 8, 6, 8, 8, 8, 6, 6, 6, 3, 8, 8, 9

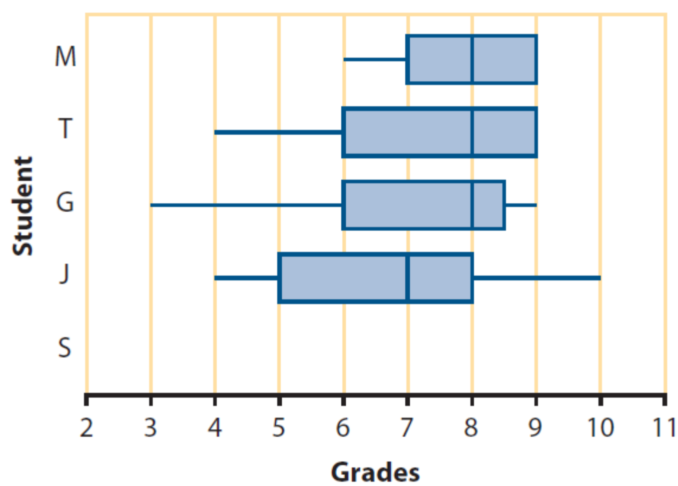
Jack's Grades

10, 7, 7, 9, 5, 8, 7, 4, 7, 5, 8, 8, 8, 4, 5, 6, 5, 8, 7

Susan's Grades

8, 8, 7, 9, 7, 8, 8, 6, 8, 7, 8, 8, 8, 7, 8, 8, 10, 9, 9, 9

Math Homework Grades



a. On the diagram above, plot Susan's homework grades.

b. Why do the plots for Maria (M) and Tran (T) have no whisker at the upper end?

c. Why is the lower whisker on Gia's (G) box plot so long? Does this mean there are more grades for Gia in that whisker than in the shorter whisker?

d. Which distribution is the most symmetric? Which distribution(s) are skewed left?

e. Use the box plots to determine which of the five students has the lowest median grade.

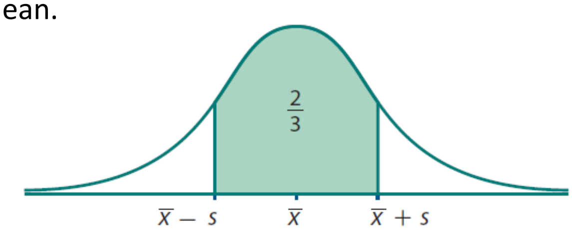
f. Based on the box plots, which of the five students seems to have the best homework record?

Standard Deviation

Interquartile Range (IQR) is a very useful measure of spread for data that is skewed or has outliers. For data which has a normal distribution, described as symmetric, mound-shaped, without outliers, a different measure of spread called the standard deviation is typically used.

Standard Deviation (***s***) is based on the sum of the squared deviations from the mean. In a normal distribution, about 68% (two-thirds) of the value fall within one standard deviation to the left and one standard deviation to the right of the mean.

(Note: $\bar{x}$ represents mean and *s* represents Standard Deviation)



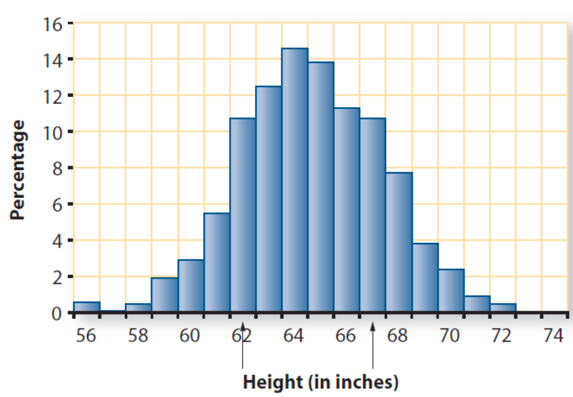
1. On each of the following distributions, the arrows enclose the middle two-thirds of the values. For each distribution:

i. Estimate the mean ($\bar{x}$).

ii. Estimate the distance from the mean to one of the two arrows. This distance is roughly the standard deviation.

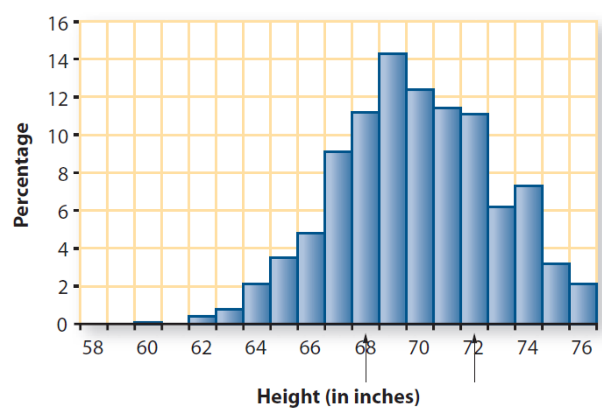
a. Heights of a large sample of young adult women in the United States

Heights of Young Adult Women



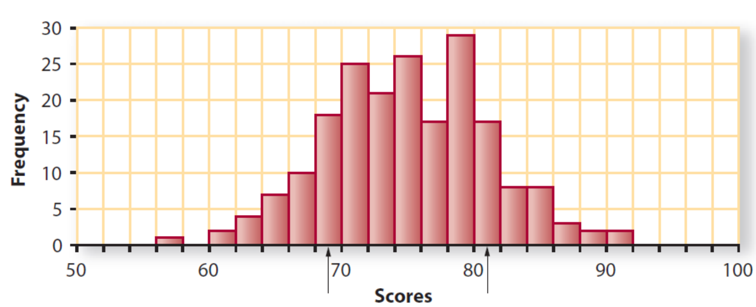
b. Heights of a large sample of young adult men in the United States

Heights of Young Adult Men



c. Achievement test scores for all ninth graders in one high school

Achievement Test Scores



2. Another measure of where a value, x, lies in a distribution is its deviation from the mean.

deviation from the mean equals the value minus the mean ($x - \bar{x}$)

In 2003, LeBron James was the first-round draft pick and the NBA Rookie of the Year. The following table gives the number of points he scored in the seven games he played in the first month of his freshman season at St. Vincent-St. Mary High School in Akron, OH. That season he led his high school team to a perfect 27-0 record and the Division III state title.

Points Scored by LeBron James in His First Month

Date	Opponent	Total Points	Deviation
Dec. 3	Cuyahoga Falls	15	
Dec. 4	Cleveland Central Catholic	21	
Dec. 7	Garfield	11	
Dec. 17	Benedictine	27	
Dec. 18	Detroit Redford	18	
Dec. 28	Mansfield Temple Christian	20	
Dec. 30	Mapleton	21	

Source: www.cleveland.com/hssports/lebron/agate.ssf?/hssports/lebron/lebron_stats.html

a. Find the mean number of points scored.

i. For each game, find the deviation from the mean. (above)

ii. For which game(s) is James' total points farthest from the mean?

iii. For which game(s) is James' total points closest to the mean?

b. For which game would you say that he has the most "typical" deviation?

c. In James' rookie season in the NBA, he averaged 20.9 points per game.

i. The highest number of points he scored is a game that season was 20.1 points above his average. How many points did he score that game?

ii. In his first NBA game, he scored 25 points. What was the deviation from his season average that year?

iii. In one game that season. LeBron had a deviation from his average of -12.9 points. How many points did he score that night?

3. Investigation: Consider the following test scores:

Student	Test 1	Test 2	Test 3	Test 4
Johnny	65	82	93	100
Will	82	86	89	83
Anna	80	99	73	88

Who is the best student?

How do you know?

Thinking about the Situation

Discuss the following with your partner or group. Write your answers on your own paper. Be prepared to share your answers with the class.

What is the mean test score for each student?

Based on the mean, who is the best student?

If asked to select one student, who would you pick as the best student? Explain.

Deviation from the Mean

Discuss the following with your partner or group. Write your answers on your own paper. Be prepared to share your answers with the class.

Usually we calculate the mean, or average, test score to describe how a student is doing. Johnny, Will, and Anna all have the same average. However, these three students do not seem to be "equal" in their test performance. We need more information than just the typical test score to describe how they are doing. One thing we can look at is how consistent each student is with their test performance. Does each student tend to do about the same on each test, or does it vary a lot from test to test? Measures of spread will give us that information. In statistics, deviation is the amount that a single data value differs from the mean.

1) Complete the table below by finding the deviation from the mean for each test score for each student.

	Score x	Mean $\bar{x}$	Deviation from the Mean $x - \bar{x}$
Johnny			
Test 1			
Test 2			
Test 3			
Test 4			

	Score x	Mean $\bar{x}$	Deviation from the Mean $x - \bar{x}$
Will			
Test 1			
Test 2			
Test 3			
Test 4			

	Score x	Mean $\bar{x}$	Deviation from the Mean $x - \bar{x}$
Anna			
Test 1			
Test 2			
Test 3			
Test 4			

2) What is the sum of the deviations the mean?

How does this relate to the mean being the balance point for a set of data?

Calculating the Standard Deviation

Below is the formula for calculating the standard deviation. It looks pretty complicated, doesn't it? Let's break it down step by step so we can see how it is finding the "average deviation from the mean."

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}}$$

Step 1: In the table below, record the data values (test scores) in the second column labeled "Value." The x is used to denote a value from the data set. *The first value is written for you.*
Step 2: Find the mean of the test scores (we did this in the "Thinking About the Situation") and record at the bottom of the second column next to the symbol μ. Mu (pronounced "mew") is the lowercase Greek letter that later became our letter "m." μ is another symbol that we use for mean (in addition to $\bar{x}$).
Step 3: In the third column, find the deviation from the mean for each test score by taking each test score and subtracting the mean. *The first difference has been done for you.*
Step 4: Add the values in the third column to find the sum of the deviations from the mean. If you have done everything correctly so far, the sum should be zero. The capital Greek letter Σ, called sigma, is a symbol that is used to indicate the sum.
Step 5: Square each deviation to make it positive and record these values in the last column of the table. *The first value is done for you.*
Step 6: Find the sum of the squared deviations by adding up the values in the fourth column and putting the sum at the bottom of the column. This is the sum of the squared deviations from the mean.
Step 7: Find the average of the squared deviations from the mean by dividing the sum of column four by the number of data values (the number of test scores).
Step 8: "Un-do" the squaring by taking the square root. Now you have found the standard deviation! The symbol for standard deviation is the lower-case letter sigma, σ.

Let's start with Johnny's data.

Johnny's Data

Test	Value (x)	Deviations from the Mean Value – Mean $(x-\mu)$	Squared Deviations from the Mean (Value – Mean)^2 $(x-\mu)^2$
1	65	65-85 = -20	$(-20)^2 = 400$
2			
3			
4			
	Mean $\mu =$	Sum $\sum(x-\mu) =$	Sum $\sum(x-\mu)^2 =$

$$\text{average of squared deviations} = \frac{\text{sum of squared deviations}}{\text{number of data values}} = \frac{\sum(x-\mu)^2}{n} =$$

$$\sigma = \text{square root of}\left(\frac{\text{sum of squared deviations}}{\text{number of data values}}\right) = \sqrt{\frac{\sum(x-\mu)^2}{n}} =$$

Now repeat the process with Will and Anna's data.

Will's Data

Test	Value (x)	Deviations from the Mean Value – Mean $(x-\mu)$	Squared Deviations from the Mean (Value – Mean)^2 $(x-\mu)^2$
1			
2			
3			
4			
	Mean $\mu =$	Sum $\sum(x-\mu) =$	Sum $\sum(x-\mu)^2 =$

$$\text{average of squared deviations} = \frac{\text{sum of squared deviations}}{\text{number of data values}} = \frac{\sum(x-\mu)^2}{n} =$$

$$\sigma = \text{square root of } \left(\frac{\text{sum of squared deviations}}{\text{number of data values}}\right) = \sqrt{\frac{\sum(x-\mu)^2}{n}} =$$

Anna's Data

Test	Value (x)	Deviations from the Mean Value – Mean $(x-\mu)$	Squared Deviations from the Mean (Value – Mean)^2 $(x-\mu)^2$
1			
2			
3			
4			
	Mean $\mu =$	Sum $\sum(x-\mu) =$	Sum $\sum(x-\mu)^2 =$

$$\text{average of squared deviations} = \frac{\text{sum of squared deviations}}{\text{number of data values}} = \frac{\sum(x-\mu)^2}{n} =$$

$$\sigma = \text{square root of } \left(\frac{\text{sum of squared deviations}}{\text{number of data values}}\right) = \sqrt{\frac{\sum(x-\mu)^2}{n}} =$$

Discussion Questions:

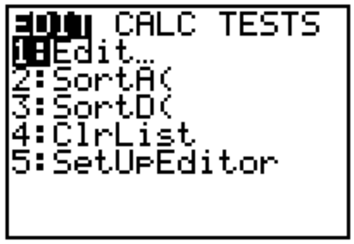
1) Why is the sum of the third column always equal to zero?
2) Translate into words: $\sum(x-\mu)^2$.
3) Interpret Anna's standard deviation in context.
4) Who is the best student? How do you know?

Finding the mean, median, and standard deviation on the calculator

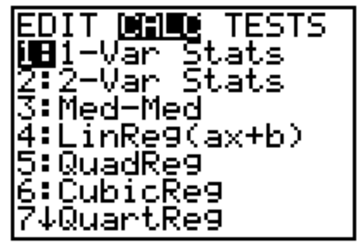
Enter the following data into L_1 in the calculator.

X
28
48
53
25
38
23
49
32

Step 1: The data should already be entered as lists in the calculator. Press STAT.



Step 2: Press right arrow button so that CALC is highlighted and the CALC menu appears (as shown). 1: 1-VAR STATS should already be highlighted.



Step 3: Press ENTER. 1-VAR STATS will appear on the Home Screen.

Step 4: Now enter the list names.

For this example, the data are stored in L1.

Notice above "1" on the number keypad is "L1" in yellow. Press the yellow 2ND button, then "1" on the keypad. L1 should appear on the screen.

Measuring Variability Practice

1) Superbowl XLIII featured 2 of the NFL's most unknown offensive linemen. The data sets give the name of the players and their weights (lbs).

Cardinals	
Mike Gandy	316
Reggie Wells	308
Lyle Sendlein	300
Duece Luti	332
Levi Brown	322

Steelers	
Max Starks	345
Chris Kemoeata	344
Justin Hartwig	312
Darnell Stapelton	305
Willie Colon	315

a. Find the mean and median weights of both the Steelers' and Cardinals' offensive lines.

b. Given the Cardinals' standard deviation is 12.4, the Stealers' standard deviation is 18.9, and the information you found in problem a, compare the Steelers' and the Cardinals' offensive lines. How are they different? How are they alike?

c. Assume that the Cardinals' offensive linemen each put on 15 pounds.

i. Calculate the mean and median for this new group of data.

ii. Which statistical values changed compared to the original group of Cardinals' linemen? Which stayed the same? Why do you think this happened?

2) Consider the heights of the people in the following two groups:

i. the members of the Charlotte Bobcats

ii. the adults living in Charlotte

a. Which group would you expect to have the larger mean height? Explain your reasoning.

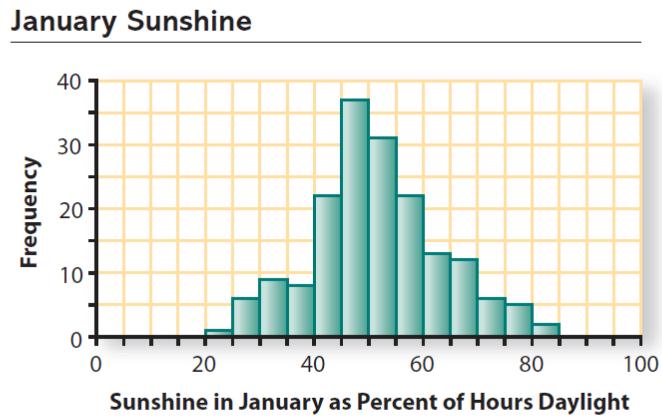
b. Which group would you expect to have the larger standard deviation? Explain.

3) Without calculating, match the sets of values from Column A with the same standard deviation in Column B.

	Column A
Q	1, 2, 3, 4, 5
R	2, 4, 6, 8, 10
S	2, 2, 2, 2, 2
T	2, 6, 6, 6, 10
U	2, 2, 6, 10, 10

	Column B
V	10, 10, 10, 10, 10
W	4, 6, 8, 10, 12
X	4, 5, 6, 7, 8
Y	16, 16, 20, 24, 24
Z	4, 8, 8, 8, 12

4) Use the following data on U.S. weather to check your understanding of the standard deviation. The histogram below shows the percentage of time that sunshine reaches the surface of Earth in January at 174 different major weather-observing stations in all 50 states, Puerto Rico, and the Pacific Islands. The two stations with the highest percentages are Tucson and Yuma, AZ. The station with the lowest percentage is Quillayute, WA.



Source: www.ncdc.noaa.gov/oa/climate/online/ccd/avgsun.html

a. Estimate the mean and standard deviation of these percentages. Include the units of measurement.

b. About how many standard deviations from the mean are Tucson and Yuma?

c. Use your calculator to check your estimates of mean and standard deviation.

Normal Distribution

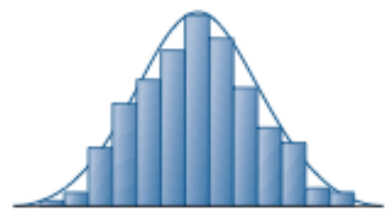
Adapted from Core Plus Course 3: Unit 4, Lesson 1

Taking political polls and manufacturing car parts do not seem similar at first glance. However, both involve processes that have variation in the outcomes. When Gallup takes a poll, different samples of voters would give slightly different estimates of the president's popularity. When Ford Motor Company builds a car; body panels will vary slightly in their dimensions even if the same machine makes them.

In this lesson, you will explore connections between a normal distribution and its mean and standard deviation and how those ideas can be used in modeling the variability in common situations.

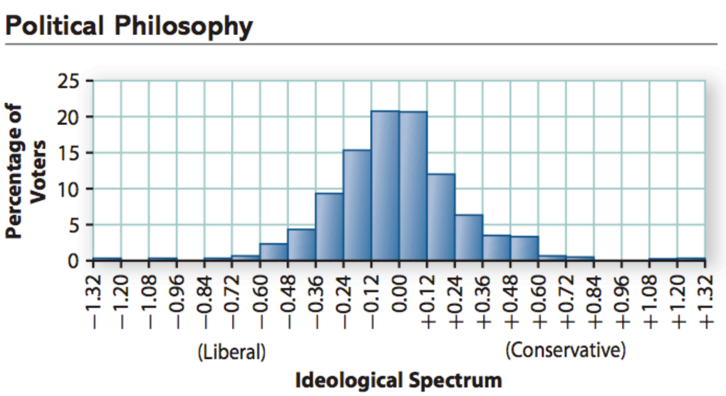
Investigation 1

Many measurements, such as human heights or the lengths or weights of supposedly identical object produced by machines, are **approximately normally distributed**. Their histograms are "bell-shaped," with the data clustered symmetrically about the mean and tapering off gradually on both ends. When measurements can be modeled by a distribution that is approximately normal in shape, the mean and standard deviation often are used to summarize the distribution's center and variability.



1. You can estimate the mean of a distribution by finding the "balance point" of the histogram. You can estimate the standard deviation of a normal distribution by finding the distance to the right of the mean and to the left of the mean that encloses 68% (about 2/3) of the values.

The histogram below shows the political points of view of a sample of 1,271 voters in the U.S. The voters were asked a series of questions to determine their political philosophy and then were rated on a scale from liberal to conservative. Estimate the mean and standard deviation of this approximately normal distribution.



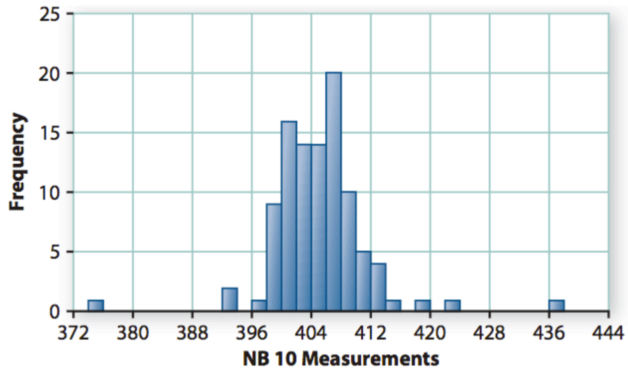
Source: Romer, Thomas, and Howard Rosenthal. 1984. Voting models and empirical evidence. *American Scientist*, 72: 465–473.

Mean ____________________

Std. Dev. __________________

In a science class, you may have weighed something by balancing it on a scale against a standard weight. To be sure a standard weight is reasonably accurate, its manufacturer can have it weighed at the National Institute of Standards and Technology (NIST). The accuracy of the weighing procedure at the NIST is itself checked about once a week by weighing a known 10-gram weight, NB 10. The histogram which follows is based on 100 consecutive measurements of the weight of NB 10 using the same apparatus and procedure. Estimate the mean and standard deviation of this distribution.

NB 10 Weight



Source: Freedman, David, *et al.* *Statistics, 3rd edition.* New York: W. W. Norton & Co, 1998.

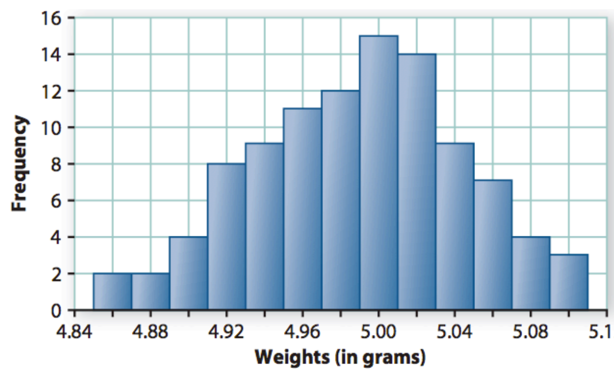
Mean ___________________

Std. Dev. _________________

2. The data below and in the accompanying histogram give the weights, to the nearest hundredth of a gram, of a sample of 100 nickels.

Nickel Weights (in grams)

4.87	4.92	4.95	4.97	4.98	5.00	5.01	5.03	5.04	5.07
4.87	4.92	4.95	4.97	4.98	5.00	5.01	5.03	5.04	5.07
4.88	4.93	4.95	4.97	4.99	5.00	5.01	5.03	5.04	5.07
4.89	4.93	4.95	4.97	4.99	5.00	5.02	5.03	5.05	5.08
4.90	4.93	4.95	4.97	4.99	5.00	5.02	5.03	5.05	5.08
4.90	4.93	4.96	4.97	4.99	5.01	5.02	5.03	5.05	5.09
4.91	4.94	4.96	4.98	4.99	5.01	5.02	5.03	5.06	5.09
4.91	4.94	4.96	4.98	4.99	5.01	5.02	5.04	5.06	5.10
4.92	4.94	4.96	4.98	5.00	5.01	5.02	5.04	5.06	5.11
4.92	4.94	4.96	4.98	5.00	5.01	5.02	5.04	5.06	5.11



The mean weight of this sample is 4.9941 grams. Find the median weight from the table above. How does it compare to the mean weight?

The standard deviation of this distribution is 0.0551 grams. On the histogram above mark the point on the horizontal axis that is the mean, then mark the points that are one, two, and three standard deviations above and below the mean.

What percentages of the weights in the table are within?

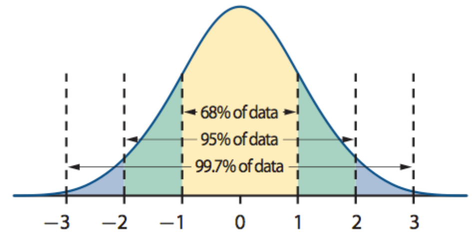
1 std dev of the mean: ________________ 2 std dev of the mean: ________________

3 std dev of the mean: ________________

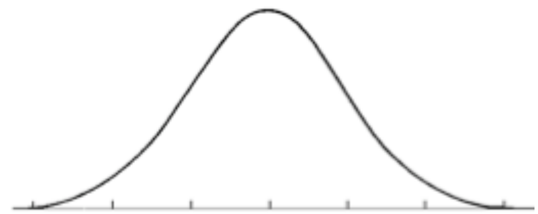
In questions 1 and 2, you looked at *approximately normal distributions* of real data that were samples taken from a larger population. The symbol for the mean of a sample is $\bar{x}$. The symbol for the standard deviation of a sample is s. For the next questions, you will think about theoretical populations that have a *perfectly normal distribution*. The symbol for the mean of a population is μ, "mu". The symbol for the standard deviation of a population is σ, "sigma".

Remember: The *population* includes all objects of interest whereas the *sample* is only a portion of the *population*.

All normal distributions have the same overall shape, differing only in their mean and standard deviation. Some look tall and skinny. Others look more spread out. All normal distributions, however, have certain characteristics in common. They are symmetric about the mean, 68% of the values lie within 1 standard deviation of the mean, 95% of the values lie within 2 standard deviations of the mean, and 99.7% of the values lie within 3 standard deviations of the mean.



3. Suppose that the distribution of the weights of newly minted coins is a normal distribution with mean $\mu = 5$ grams and standard deviation $\sigma = 0.10$ grams. Label the points on the distribution that are the mean; and one, two, and three standard deviations above and below the mean.

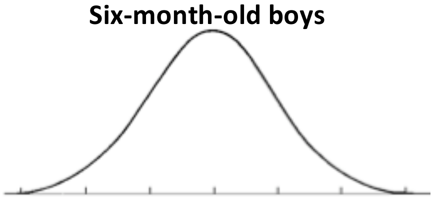


Between what two values do the middle 68%, 95%, and 99.7% of the weights of coins lie?

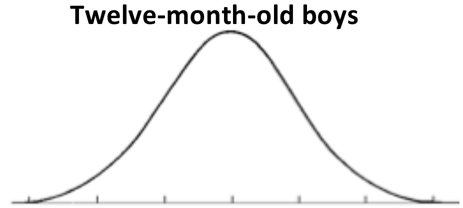
68%: ________________________ 95%: ________________________ 99.7%: ________________________

4. The weights of babies of a given age and gender are approximately normally distributed. This fact allows a doctor to use a baby's weight to find the weight percentile to which the baby belongs. (Recall that a value x in a distribution lies at, say, the 27th percentile if 27% of the values in the distribution are less than or equal to x.) The weight of six-month-old boys has a mean $\mu = 17.25$ pounds and a standard deviation $\sigma = 2.0$ pounds. The weight of twelve-month-old boys has a mean $\mu = 22.50$ pounds and a standard deviation $\sigma = 2.2$ pounds. Label the points on each distribution that are the mean; and one, two, and three standard deviations above and below the mean.

Six-month-old boys



Twelve-month-old boys



About what % of six-month-old boys weigh between 15.25 and 19.25 pounds? ____________________

About what % of twelve-month-old boys weigh more than 26.9 pounds? ___________________________

A twelve-month-old boy who weighs 24.7 pounds is at what percentile for weight? _________________

A six-month-old boy who weights 21.25 pounds is at what percentile for weight? ____________________

☑ Check Your Understanding:

Scores on the verbal section of the SAT are approximately normally distributed with a mean $\mu = 502$ and standard deviation $\sigma = 113$.

What percentage of students score between 389 and 615 ? ____________ over 615? ______________

If your score 389, what is your percentile? ______________________

Investigation 2

Often, you are interested in comparing values from two different distributions. For example: Is Sophie or her brother Pablo taller compared to others of their gender? Was your SAT or your ACT score better? Answers to questions like these require the comparison of values from two different normal distributions.

1. This table gives information about the heights of young Americans aged 18 to 24. Each distribution is approximately normal.

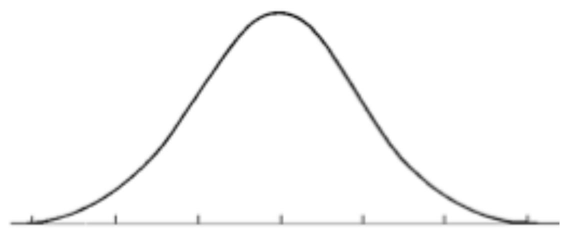
Heights of American Young Adults (in inches)

	Men	Women
Mean μ	68.5	65.5
Standard Deviation σ	2.7	2.5

Sketch the two distributions. Include a scale on the horizontal axis.

Men

Women



Alexis is 3 standard deviations above the average height. How tall is she? ___________________

Marvin is 2.1 standard deviations below the average height. How tall is he? ___________________

Miguel is 74" tall. How many standard deviations above average height is he? ___________________

Jackie is 62" tall. How many standard deviations below average height is she? ___________________

Steve is 71" tall. Mary is 68" tall. Who is relatively taller for his or her gender? ___________________
Explain your reasoning.

____

____

2. The **standardized value** tells how many standard deviations a given value lies from the mean of a distribution. For example: Alexis is 3 standard deviations above average height so her standardized height is 3, Marvin is 2.1 standard deviations below average height so his standardized height is -2.1.

The symbol for **standardized value** is z. Another name for standardized value is "**z-score**".

Note: a z-score is positive (+) for values above the mean, and negative (-) for values below the mean.

Calculate the z-scores for Miguel's and Jackie's heights in Problem 1. Think about whether or not your calculations give z-scores for Miguel's and Jackie's heights that have the correct sign (+/-).

Calculation of z-score for Miguel's height: ___________________________________

Calculation of z-score for Jackie's height: ___________________________________

Using μ for the population mean, and σ for the population standard deviation, write a formula for computing the z-score of a data value x.

$$z =$$

3. Now consider how z-scores can help you make comparisons. (Refer to table in Problem 1).

Find the z-score for the height of a young woman who is 5 feet tall. __________________

Find the z-score for the height of a young man who is 5 feet 2 inches tall. __________________

Is the young woman or the young man shorter relative to his or her own gender? __________________
Explain your reasoning.

____

____

4. In an experiment about the effects of mental stress, subjects' systolic blood pressure and heart rate were measured before and after doing a stressful mental task. Their systolic blood pressure increased an average of 22.4 mm Hg (millimeters of Mercury) with a standard deviation of 2. Their heart rates increased an average of 7.6 beats per minute with a standard deviation of 0.7. Each distribution was approximately normal.

Suppose that after completing a task, Mark's blood pressure increased by 25 mm Hg and his heart rate increased by 9 beats per minutes. On which measure did he increate the most, relative to the other participants? Explain your reasoning.

____

____

☑ Check Your Understanding:

Kevin earned a grade of 50 on a normally distributed test with mean 45 and standard deviation 10. On another normally distributed test with mean 70 and standard deviation 15, he earned a 78. On which of the two tests did he do better, relative to the others who took the tests? Explain your reasoning.

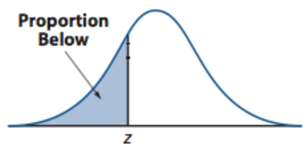
____

____

Investigation 3

In Investigation 1 you learned that all normal distributions are symmetric about the mean with 68% of the values within 1 standard deviation of the mean, 95% of the values within 2 standard deviations of the mean, and 99.7% of the values within 3 standard deviations of the mean. In Investigation 2 you learned how to calculate a **z-score** which tells how many standard deviations a given value lies from the mean of a distribution. In this investigation you will use z-scores to find the location of any value in a distribution that is approximately normal.

The following **z-table** gives the proportion (percentage) of values in a normal distribution that are less than the given z-score.



Proportion of Values in a Normal Distribution that Lie Below a Standardized Value *z*

z	Proportion Below	*z*	Proportion Below	*z*	Proportion Below	*z*	Proportion Below
−3.5	0.0002	−1.7	0.0446	0.1	0.5398	1.9	0.9713
−3.4	0.0003	−1.6	0.0548	0.2	0.5793	2.0	0.9772
−3.3	0.0005	−1.5	0.0668	0.3	0.6179	2.1	0.9821
−3.2	0.0007	−1.4	0.0808	0.4	0.6554	2.2	0.9861
−3.1	0.0010	−1.3	0.0968	0.5	0.6915	2.3	0.9893
−3.0	0.0013	−1.2	0.1151	0.6	0.7257	2.4	0.9918
−2.9	0.0019	−1.1	0.1357	0.7	0.7580	2.5	0.9938
−2.8	0.0026	−1.0	0.1587	0.8	0.7881	2.6	0.9953
−2.7	0.0035	−0.9	0.1841	0.9	0.8159	2.7	0.9965
−2.6	0.0047	−0.8	0.2119	1.0	0.8413	2.8	0.9974
−2.5	0.0062	−0.7	0.2420	1.1	0.8643	2.9	0.9981
−2.4	0.0082	−0.6	0.2743	1.2	0.8849	3.0	0.9987
−2.3	0.0107	−0.5	0.3085	1.3	0.9032	3.1	0.9990
−2.2	0.0139	−0.4	0.3446	1.4	0.9192	3.2	0.9993
−2.1	0.0179	−0.3	0.3821	1.5	0.9332	3.3	0.9995
−2.0	0.0228	−0.2	0.4207	1.6	0.9452	3.4	0.9997
−1.9	0.0287	−0.1	0.4602	1.7	0.9554	3.5	0.9998
−1.8	0.0359	0.0	0.5000	1.8	0.9641		

Examples using the z-table:

- 6.68% of the data in a normal distribution lie below a data value with a z-score of -1.5
- 93.32% of the data in a normal distribution lie below a data value with a z-score of 1.5

Draw a sketch to illustrate the area of interest, then use the z-table to answer the question.

1. If a value from a normal distribution is 1.3 standard deviations above the mean (i.e. z-score = 1.3), what percentage of the values are:

a. below it?

b. above it?

% below ____________

% above _____________

2. For a normal distribution what percentage of the values are within ±1.3 standard deviation of the mean (i.e. between z-scores of -1.3 and +1.3)?

% between z-scores of -1.3 and +1.3 ______________

3. For a normal distribution what percentage of the value are within 1.3 standard deviations above the mean and 0.7 standard deviations below the mean (i.e. between z-scores of -0.7 and +1.3)?

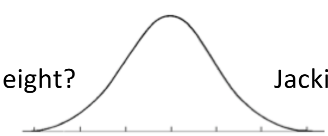
% between z-scores of -0.7 and +1.3 ______________

4. The heights of American young men aged 18 to 24 are approximately normally distributed with a mean $\mu = 68.5$ and standard deviation $\sigma = 2.7$. The heights of American young women aged 18 to 24 are approximately normally distributed with a mean $\mu = 65.5$ and standard deviation $\sigma = 2.5$.

Miguel is 74" tall. What is percentile for height (i.e. what % of young men are the same height or shorter than Miguel?)?

Miguel's percentile for height ______________

Jackie is 62" tall. What is her percentile for height?



Jackie's percentile for height _________

Abby is 5 feet 8 inches tall. What percentage of young women are between Jackie and Abby in height?

% of young women between Jackie & Abby in height _____________

5. You can also use a z-table to calculate a data value given the percentile.

Gabriel is at the 90th percentile for height. Approximately what is her height? ____________________

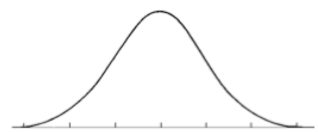
Yvette is at the 31st percentile for height. Approximately what is her height? _____________________

☑ Check Your Understanding:

In 2006-2007 all 11th grade students in Pennsylvania were tested in math using the Pennsylvania System of School Assessment (PSSA). The scores were approximately normally distributed with a mean $\mu = 1330$ and standard deviation $\sigma = 253$.

What % of 11th graders scored :

above 1500? _________ below 1000? _________ between 1200 and 1600? _________



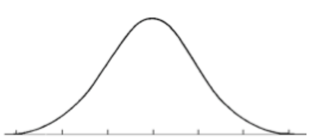
What PSSA score would be at the 50th percentile? ________________

Joe's PSSA score was at the 76th percentile. What was his score? ________________

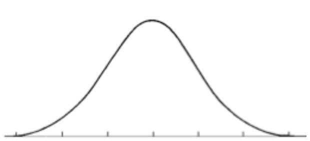
The IQ scores on the Standford-Binet intelligence test are approximately normally distributed with a mean mean $\mu = 100$ and standard deviation $\sigma = 15$.

What % of people have an IQ:

greater than 75? _________ less than 125? _________ between 120 and 150? _________



What % of people have an IQ greater than 140 or less than 80? ____________________



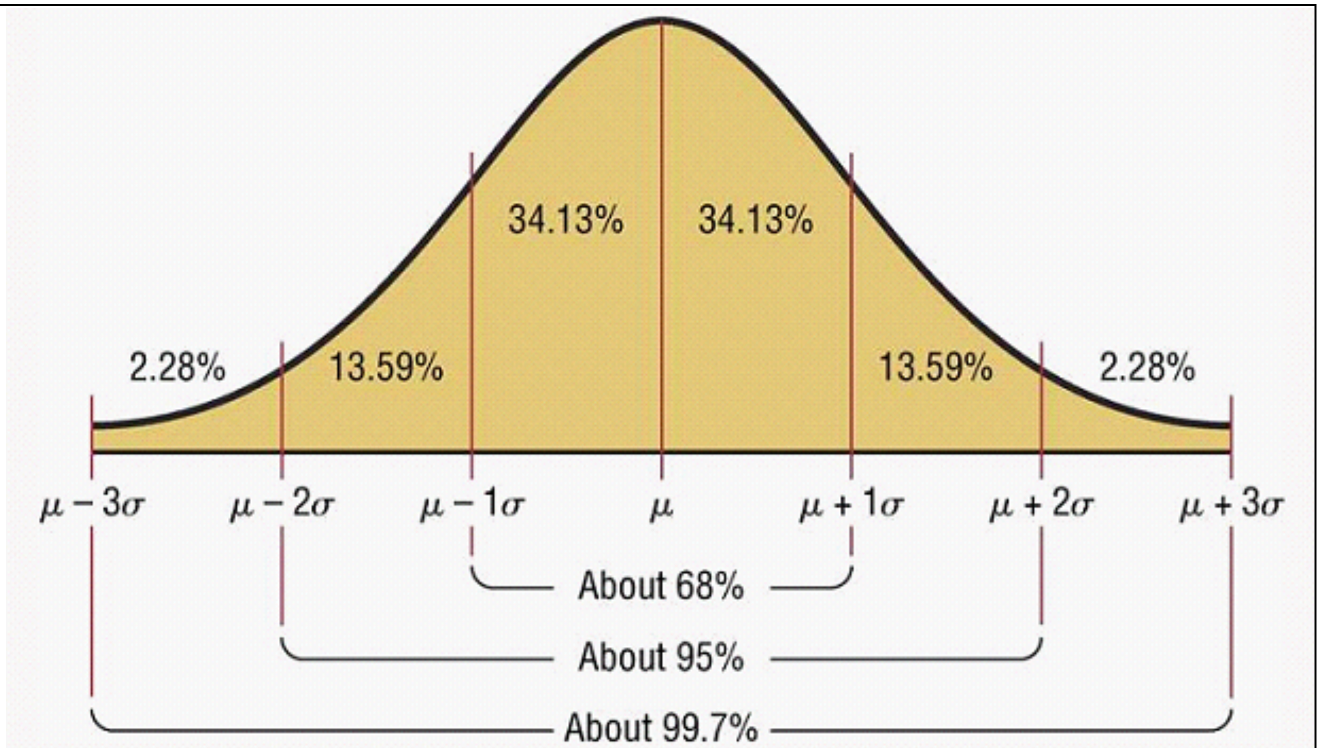
What IQ score would be at the 90th percentile? ________________

What IQ score would be at the 25th percentile? ________________

Normal Probability Distribution

A normal distribution is a continuous, symmetric, bell-shaped distribution of a variable. The properties of a normal distribution, are:

1. A normal distribution curve is bell-shaped
2. The mean, median, and mode are equal and are located at the center of the distribution
3. A normal distribution curve is unimodal (i.e., it has only one mode)
4. The curve is symmetric about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.
5. The curve is continuous, that is, there are no gaps or holes. For each value of X, there is a corresponding value of Y.
6. The curve never touches the x axis. Theoretically, no matter how far in either direction the curve extends, it never meets the x axis – but it gets increasingly closer
7. The total area under a normal distribution curve is equal to 1.00, or 100%



The Empirical Rule states that:

The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%.

Using the Empirical Rule

For each problem set, label the normal curve with the appropriate values, and use the curve to answer the questions.

1. The mean score on the midterm was an 82 with a standard deviation of 5. Find the probability that a randomly selected person:

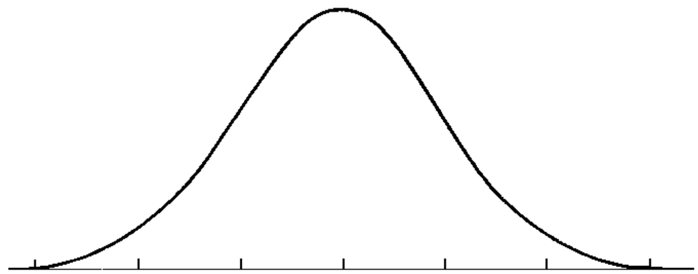
   a. scored between 77 and 87

   b. scored between 82 and 87

   c. scored between 72 and 87

   d. scored higher than 92

   e. scored less than 77



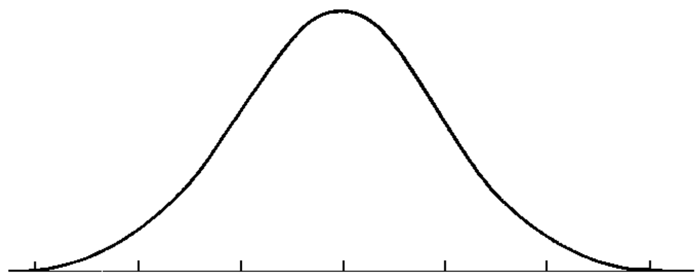
2. The mean SAT score is 490 with a standard deviation of 100. Find the probability that a randomly selected student:

   a. scored between 390 and 590

   b. scored above 790

   c. scored less than 490

   d. scored between 290 and 490



3. The mean weight of college football players is 200 pounds with a standard deviation of 30. Find the probability that a randomly selected player:

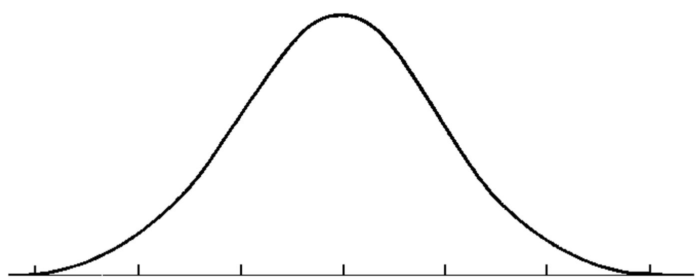
   a. weighs between 170 and 260

   b. weighs less than 170

   c. weighs over 290

   d. weighs less than 140

   e. weighs between 140 and 230



4. The average life of a car tire is 28,000 miles with a standard deviation of 3000. Find the probability that a randomly selected tire will have a life of:

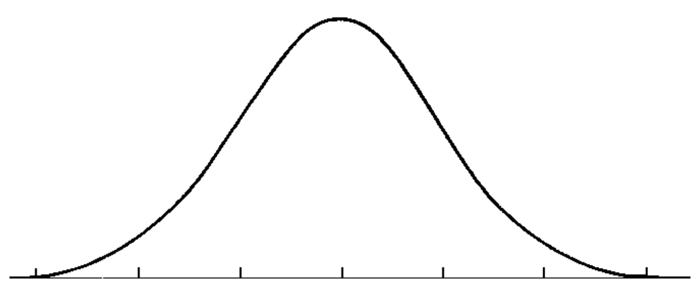
   a. between 19,000 and 37,000 miles

   b. less than 25,000 miles

   c. between 31,000 and 37,000 miles

   d. over 22,000 miles

   e. below 31,000 miles



Working with Z Scores

1. The mean score on the SAT (math & verbal) is 1500, with a standard deviation of 240. The ACT, a different college entrance examination, has a mean score of 21 with a standard deviation of 6.

(a) If Bobby scored 1740 on the SAT, how many points above the SAT mean did he score?

(b) If Kathy scored 30 on the ACT, how many points above the ACT mean did she score?

(c) Is it sensible to conclude that because Bobby's difference is bigger that he outperformed Kathy on the admissions test? Explain.

(d) Determine how many standard deviations above the mean Bobby scored by dividing your answer to part (a) by the standard deviation of the SAT scores.

(e) Determine how many standard deviations above the mean Kathy scored by dividing your answer to part (b) by the standard deviation of the ACT scores.

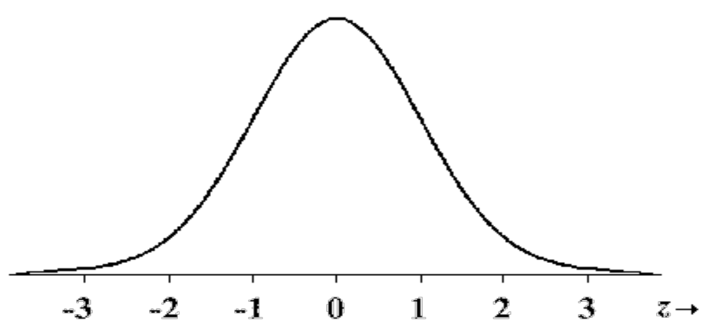
The values you just calculated are called z-scores. A z-score is an indication of how many standard deviations above or below the mean a given data point is located. Z-scores are used to convert data from different scales to a common scale so a more accurate comparison can be made between them.

Z-scores are locations on the Standard Normal Curve, which has a mean of 0 and a standard deviation of 1.

To find a z-score, use the formula:

$$z = \frac{x - \mu}{\sigma}$$

where x is the given value, μ is the mean, and σ is the standard deviation.



Now mark the values that you found in parts (d) and (e) on the standard normal curve above.

(f) Who had the higher z-score on their admissions test?

(g) Who performed better on his or her admissions test compared to his or her peers?

(h) Calculate the z-score for Peter who scored 1380 on the SAT. Calculate the z-score for Kelly who scored 15 on the ACT.

(i) Does Peter or Kelly have the higher z-score?

(j) What does it mean to have a negative z-score? Can you think of situations in which it would be beneficial to have a negative z-score?

2. The weight of an average 3 month child is 12.5 pounds with a standard deviation of 1.5 pounds. Benjamin is a healthy 3 month old child who weighs 13.9 pounds.

(a) Determine the z-score for Benjamin's weight at 3 months.

(b) Interpret what the z-score means (in context) in a sentence.

(c) The weight of an average 6 month old is 17.25 pounds with a standard deviation of 2.0 pounds. If Benjamin had the same z-score at 6 months as he did at 3 months, determine how much a 6 month old Benjamin would weigh.

3. The height of women aged 20 to 29 are approximately normal with mean 64 inches and standard deviation 2.7 inches. Men the same age have mean height 69.3 inches with standard deviation 2.8 inches. John and his sister June both play basketball for N. C. State University. John is 81 inches tall; June is 74 inches tall.

(a) Compared to their respective peers, who is the tallest?

(b) How tall would a woman be who has a z-score of 1.5?

(c) If a man has a z-score of -0.5 and a woman has a z-score of 1.2, which is tallest?

4. The following table shows the scores of Toni on six different scales of an aptitude test. Also shown are the means and standard deviations of these scales.

Test	Mean	Standard Deviation	Score	Z-Score
Clerical Ability	50	15	41	
Logical Reasoning	40	4	47	
Mechanical Ability	120	25	100	
Numerical Reasoning	100	10	105	
Spatial Relations	70	20	90	
Verbal Fluency	60	6	70	

(a) Calculate the z-scores for each.

(b) On which test did Toni score the highest? On which did Toni score the lowest?

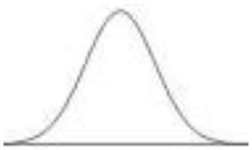
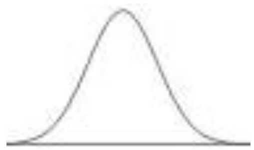
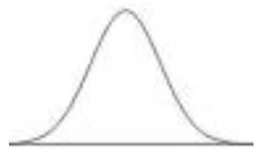
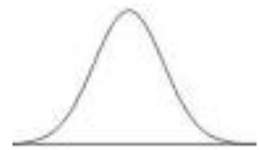
Normal Distribution Practice 1

For all questions, assume that the distribution is normal.
Shade the appropriate area of the normal curve, then solve using the z-table.

1. A survey found that mean length of time that Americans keep their cars is 5.3 years with a standard deviation of 1.2 years. If a person decides to purchase a new car, find the probability that he or she has owned the old car for

a) less than 2.5 years b) between 3 and 6 years c) more than 7 years

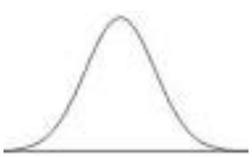
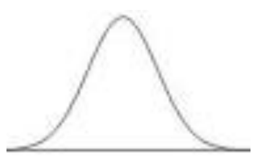
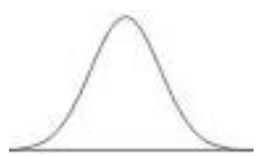
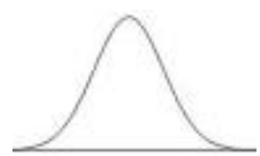
d) The length of time John keeps his car is in the 90^{th} percentile. Determine how long John keeps his car.

a)  b)  c)  d) 

2. The average waiting time at Walgreen's drive-through window is 7.6 minutes, with a standard deviation of 2.6 minutes. When a customer arrives at Walgreen's, find the probability that he will have to wait

a) between 4 and 6 minutes b) less than 3 minutes c) more than 8 minutes

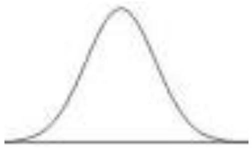
d) Only 8% of customers have to wait longer than Mrs. Sickalot. Determine how long Mrs. Sickalot has to wait.

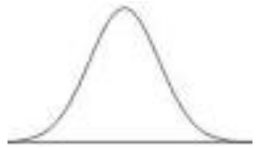
a)  b)  c)  d) 

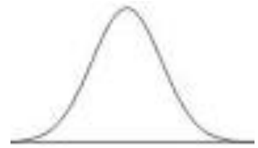
3. The scores on an Algebra II test have a mean of 76.4 and a standard deviation of 11.4. Find the probability that a student will score

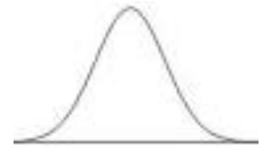
a) above 78   b) below 60   c) between 80 and 85

d) Mr. Reeves scales his tests so that only 5% of students can receive an A. What is the minimum score Andrea can make on this test and still get an A?

a) 

b) 

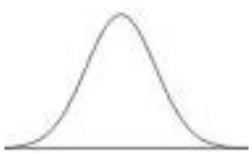
c) 

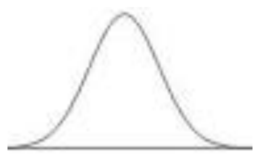
d) 

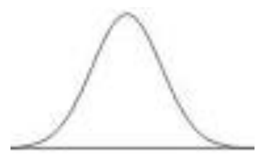
4. The average life of automobile tires is 30,000 miles with a standard deviation of 2000 miles. If a tire is selected and tested, find the probability that it will have a lifetime

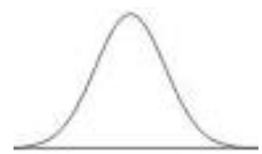
a) between 25,000 and 28,000 miles   b) between 27,000 and 32,000 miles

c) over 35,000 miles   d) The tire company will replace tires whose tread life falls in the lowest 150% of all tires of this model. What is the lifetime of a tire that qualifies for replacement?

a) 

b) 

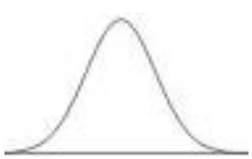
c) 

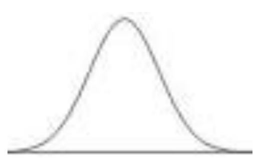
d) 

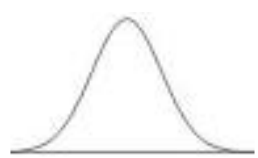
5. The mean height of an American man is 69” with a standard deviation of 2.4”. If a man is selected at random, find the probability that he will be

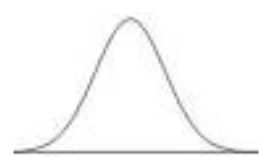
a) between 68” and 71” tall   b) shorter than 67”

c) taller than 72”   d) If Jose is in the 75^{th} percentile, how tall is he?

a) 

b) 

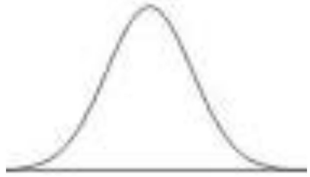
c) 

d) 

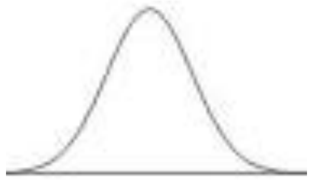
Normal Distribution Practice 2

For all questions, assume that the distribution is normal.
Shade the appropriate area of the normal curve, then solve using the z-table.

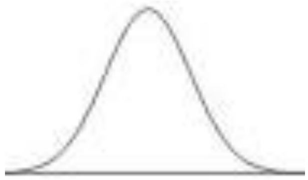
1. The percentage impurity of a chemical can be modelled by a normal distribution with a mean of 5.8 and a standard deviation of 0.5. Obtain the probability that a sample of the chemical has percentage impurity between 5 and 6.



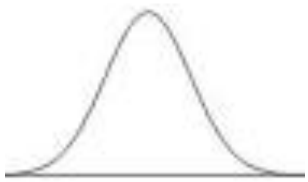
2. Melons sold on a market stall have weights that are normally distributed with a mean of 2.18 kg and a standard deviation of 0.25 kg. For a melon chosen at random, find the probability that its weight lies between 2 kg and 2.5 kg.



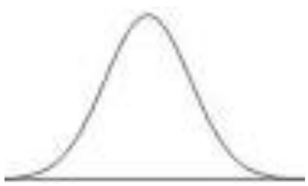
3. A teacher travels from home to work by car each weekday by one of two routes, *X* or *Y*. For route *X*, her journey times are normally distributed with a mean of 30.4 minutes and a standard deviation of 3.6 minutes. Calculate the probability that her journey time on a particular day takes between 25 minutes and 35 minutes.



4. Soup is sold in tins which are filled by a machine. The actual weight of soup delivered to a tin by the filling machine is always normally distributed about the mean weight with a standard deviation of 8g. The machine is set originally to deliver a mean weight of 810g.
   (a) Determine the probability that the weight of soup in a tin, selected at random, is less than 800g.



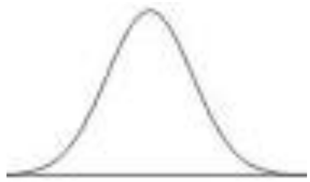
   (b) Determine the probability that the weight of soup in a tin, selected at random, is between 795 g and 820 g.



5. The weight, X grams, of a particular variety of orange is normally distributed with mean 205 and standard deviation 25. A wholesaler decides to grade such oranges by weight. He decides that the smallest 30 percent should be graded as small, the largest 20 percent graded as large, and the remainder graded as medium.

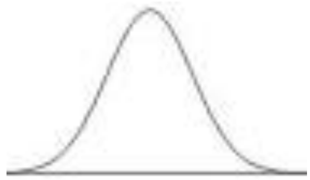
Determine, to one decimal place, the maximum weight of an orange graded as:

(i) small

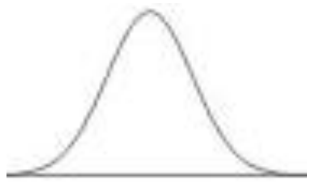


(ii) medium.

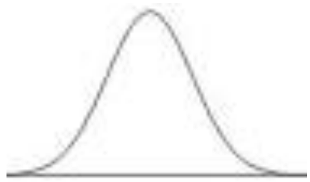
6. Jars of bolognese sauce, sold by a supermarket, are stated to have contents of weight 500 g. The weights, in grams, of the actual contents of jars in a large batch are normally distributed with mean 506 and standard deviation 5. Find the weight which is exceeded by the contents of 99.9% of the jars in this batch.



7. The distance, in kilometers, travelled to work by the employees of a city council may be modelled by a normal distribution with mean 7.5 and standard deviation 2.5. Find d such that 10% of the council's employees travel less than d kilometers to work.



8. An airline operates a service between Manchester and Paris. The flight time may be modelled by a normal distribution with mean 85 minutes and standard deviation 8 minutes. In order to gain publicity for the service, the airline decides to refund fares when a flight time exceeds q minutes. Find the value of q such that the probability of fares being refunded on a particular flight is 0.001.



Experimental Design and Sampling Design

Experimental Design: methods to gather data

- **Surveys:**

- **Observational studies:**

- **Experiments:**

Identify the experimental design:

a. A random sample of 30 digital cameras is selected and divided into 2 groups. One group is given a brand-name battery, while the other is given a generic battery. Pictures are taken under identical conditions and the battery life of the two groups is compared.

b. Using SAT scores to predict a student's college success has created some controversy. A study was conducted to examine the SAT and college GPA information of 105 students who graduated from a state university with a B.S. in computer science.

c. Before viewing a lecture, students were given a summary of the instructors' prior teaching evaluations. These evaluations were rated either: Charismatic instructor or Punitive instructor. All the students then watched the same twenty-minute lecture given by the exact same lecturer. Following the lecture, students answered three questions about the leadership qualities of the lecturer.

d. How much television is too much for children? Parents answered a questionnaire about their children's age, characteristic behavior, and television viewing habits.

e. A study investigated the impact of participating in a school garden program. Subjects were 320 sixth-grade students at 3 schools. At 2 of the schools, garden-based learning activities were incorporated into the regular science class. The 3rd school did not include a garden program as part of its science class. Two questionnaires – Garden Vegetable Frequency Questionnaire and a taste test – were given to assess the outcome.

Population: The entire group of individuals that we want information about.
Census: A complete count of the population.
Sample: A part of the population that we actually examine in order to gather information.
Use the sample to generalize to population .

Why would we not use a census all the time?

1. Not accurate- Look at the U.S. census: it has a huge amount of error, and it takes a long time to compile the data making the data obsolete by the time we get it!
2. Very expensive- taking a census of any population takes time is VERY costly to do!
3. Perhaps impossible- Suppose you wanted to know the average weight of the white-tail deer population in Texas, would it be feasible to do a census?
4. If using destructive sampling, you would destroy population
   Examples: breaking strength of soda bottles , lifetime of flashlight batteries

Sampling Design: methods to choose the sample from the population

- **Simple Random Sample:**

- **Stratified Random Sample:**

- **Cluster Sample**:

- **Systematic random sample**:

- **Convenience sample:**

- **Voluntary Response sample:**

Identify the sampling design:

a. The Educational Testing Service (ETS) needed a sample of colleges. ETS first divided all colleges into groups of similar types (small public, small private, etc.) Then they randomly selected 3 colleges from each group.

b. A county commissioner wants to survey people in her district to determine their opinions on a particular law up for adoption. She decides to randomly select blocks in her district and then survey all who live on those blocks.

c. A local restaurant manager wants to survey customers about the service they receive. Each night the manager randomly chooses a number between 1 & 10. He then gives a survey to that customer, and to every 10th customer after them, to fill it out before they leave.

d. A professor at UNC is interested in studying drinking behaviors among college students. The professor decides to stand at the main entrance to the Student Union and surveys the students who enter.

e. A building contractor has a chance to buy an odd lot of 5000 used bricks at an auction. She is interested in determining the proportion of bricks in the lot that are cracked and therefore unusable for her current project, but she does not have enough time to inspect all 5000 bricks. Instead, she randomly selects a portion of the stack and checks the 100 bricks in that portion to determine whether each is cracked.

f. A professor at a NCSU is interested in studying drinking behaviors among college students. The professor teaches a Sociology 101 class to mostly college freshmen and decides to use his or her class as the study sample. He or she passes out surveys during class for the students to complete and hand in.

Observational Study or Experiment?

For each situation, determine whether the research conducted is an *observational study* or an *experiment.* Explain your reasoning.

1. In an attempt to study the health effects of air pollution, a group of researchers selected 6 cities in very different environments; some from an urban setting (e.g. greater Boston), some from a heavy industrial setting (e.g. eastern Ohio), some from a rural setting (e.g. Wisconsin). Altogether they selected 8000 subjects from the 6 cities, and followed their health for the next 20 years. At this time their health prognoses were compared with measurements of air pollution in the 6 cities.

2. Among a group of women aged 65 and older who were tracked for several years, those who had a vitamin B_{12} deficiency were twice as likely to suffer severe depression as those who did not.

3. Forty volunteers suffering from insomnia were divided into two groups. The first group was assigned to a special no-desserts diet while the other continued desserts as usual. Half of the people in these groups were randomly assigned to an exercise program, while the others did not exercise. Those who ate no desserts and engaged in exercise showed the most improvement.

4. A study in California showed that students who study a musical instrument have higher GPAs than students who do not, 3.59 to 2.91. Of the music students, 16% had all A's, compared with only 5% among the students who did not study a musical instrument.

5. Scientists at a major pharmaceutical firm investigated the effectiveness of an herbal compound to treat the common cold. They exposed each subject to a cold virus, and then gave him or her either the herbal compound or a sugar solution known to have no effect. Several days later, they assessed the patient's condition, using a cold severity scale of 0 to 5.

6. In 2001, a report in the *Journal of the American Cancer Institute* indicated that women who work nights have a 60% greater risk of developing breast cancer. Researchers based these findings on the work histories of 763 women with breast cancer and 741 women without the disease.

7. To research the effects of dietary patterns on blood pressure in 459 subjects, subjects were randomly assigned to three groups and had their meals prepared by dietitians. Those who were fed a diet low in fat and cholesterol lowered their systolic blood pressure by an average of 6.7 points when compared with subjects fed a control diet.

8. Some people who race greyhounds give the dogs large doses of vitamin C in the belief that the dogs will run faster. Investigators at the University of Florida tried three different diets in random order on each of five racing greyhounds. They were surprised to find that when the dogs ate high amounts of vitamin C, they ran more slowly.

Sampling Methods

Identify the sampling method: *simple random, cluster, stratified, convenience, voluntary response, or systematic.*

1. Every fifth person boarding a plane is searched thoroughly.
2. At a local community College, five math classes are randomly selected out of 20 and all of the students from each class are interviewed.
3. A researcher randomly selects and interviews fifty male and fifty female teachers.
4, A researcher for an airline interviews all of the passengers on five randomly selected flights.
5. Based on 12,500 responses from 42,000 surveys sent to its alumni, a major university estimated that the annual salary of its alumni was 92,500.
6. A community college student interviews the first 100 students to enter the building to determine the percentage of students that own a car.
7. A market researcher randomly selects 200 drivers under 35 years of age and 100 drivers over 35 years of age.
8. All of the teachers from 85 randomly selected nation's middle schools were interviewed.
9. To avoid working late, the quality control manager inspects the last 10 items produced that day.
10. The names of 70 contestants are written on 70 cards, The cards are placed in a bag, and three names are picked from the bag.
11. 32 sophomores, 35 juniors and 49 seniors are randomly selected from 230 sophomores, 280 juniors, 577 seniors at a certain high school.
12. To ensure customer satisfaction, every 35th phone call received by customer service will be monitored.
13. Calling randomly generated telephone numbers, a study asked 855 U.S. adults which medical conditions could be prevented by their diet.
14. A pregnancy study in Chicago, randomly selected 25 communities from the metropolitan area, then interviewed all pregnant women in these communities.

Are these samples representative? Explain your thinking.

15. To determine the percentage of teenage girls with long hair, *Teen* magazine published a mail-in questionnaire. Of the 500 respondents, 85% had hair shoulder length or longer (*USA Today*, July 1, 1985).
16. A college psychology professor needs subjects for a research project to determine which colors average American adults find restful. From the list of all 743 students taking introductory psychology at her school, she selects 25 students using a random number table.
17. To evaluate the reliability of cars owned by its subscribers, *Consumer Reports* magazine publishes a yearly list of automobiles and their frequency-of-repair records. The magazine collects the information by mailing a questionnaire to subscribers and tabulating the results from those who return it.
18. Oranges from an orchard need to be samples to see if they are sweet enough for juice. The orchard has 25,000 orange trees. Each tree has at least 500 oranges. Claire decides to randomly choose 800 trees and test one orange from each tree.