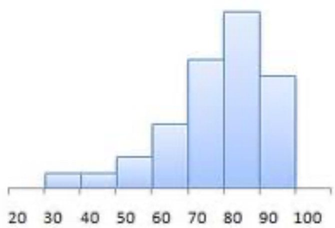
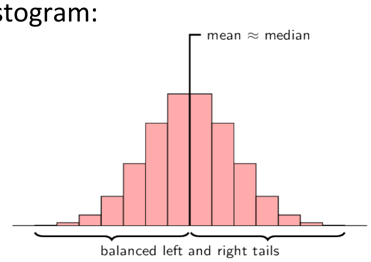
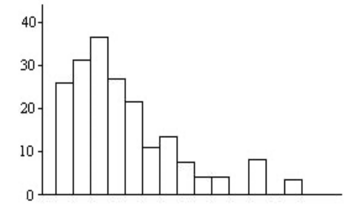
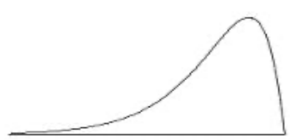
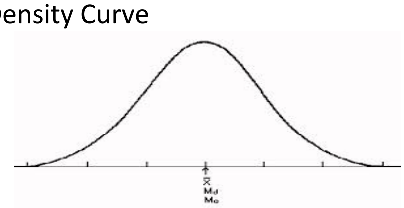
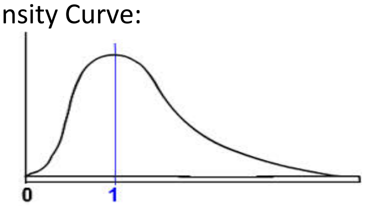
SM3 HW 14.2 Normal Distributions

OBJECTIVES:

- Summarize, represent, and interpret data on a single count or measurement variable
- 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 calculators, spreadsheets, and tables to estimate areas under the normal curve.

VOCABULARY:

- **Symmetric Distributions** are ones where one half of the distribution's graph is a mirror image of the other half
- **Skew** is when one half of the distribution is larger than the other half [it's the TAIL not the BUMP]
    - A distribution is called **"Left Skewed"** if the larger **tail** is on the left
    - A distribution is called **"Right Skewed"** if the larger **tail** is on the right

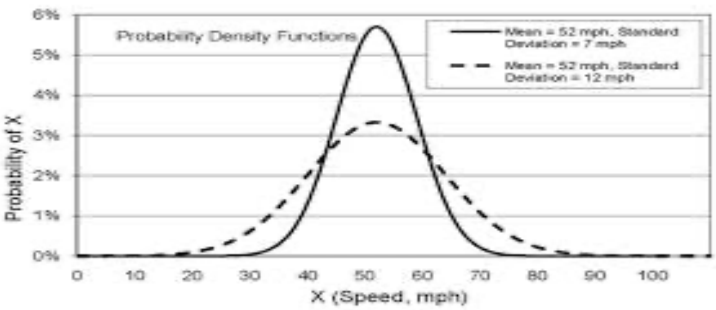
<table>
<tr><th>LEFT SKEWED DISTRIBUTIONS</th><th>SYMMETRIC DISTRIBUTION</th><th>RIGHT SKEWED DISTRIBUTIONS</th></tr>
<tr><td>Histogram:
</td><td>Histogram:
</td><td>Histogram:
</td></tr>
<tr><td>Density Curve:
</td><td>Density Curve
</td><td>Density Curve:
</td></tr>
</table>

- A **Histogram** separates the data into intervals of equal width called "classes" and then counts how many observations fell within each interval. The height of the bar for that internal indicates how many observations where in that "class". Classes must be defined so that no single observation can be included in more than one class.
- **Density Curves** are idealized curves to show the overall pattern and shape of a distribution
    - by definition, the **area under a density** curve is 1.
- **Standard Deviation** is a statistics calculated to describe the spread of a distribution using the formula:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}} \; \textit{this formula is for a sample of data,}$$

$$x_1 = \textit{first data value} \qquad \bar{x} = \textit{mean of the data set} \qquad n = \textit{number of values in the data set}$$

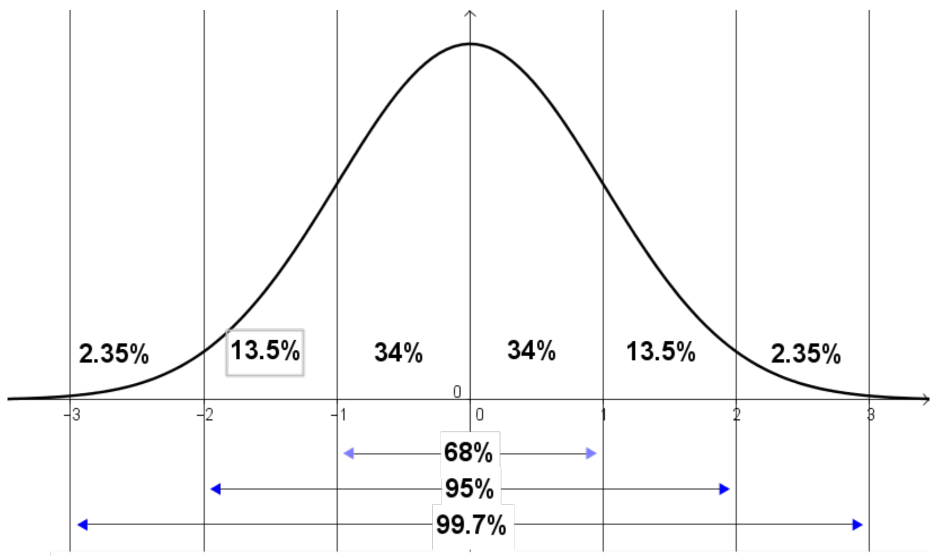
    - The symbol σ "sigma" is used to represent the standard deviation of the whole population's
    - The standard deviation can be thought of as the "average" distance of each data point from the mean. Larger standard deviations mean the data has a more spread out distribution. A smaller standard deviation means is more tightly grouped around the mean and less spread out.



The solid taller curve has a standard deviation of 7, the dashed shorter curve has a standard deviation of 12.

They both have the same mean and an area of 1 under the curve

- A **Normal Distribution** is a probability distribution that has ALL of the following specific characteristics:
  - The distribution is "bell" shaped
  - it has one peak (called unimodal)
  - it is symmetric with the left half being a mirror image of the right half
  - there are no outliers
  - Points of inflection on the normal curve lie one standard deviation away from the mean
  - **FOLLOWS THE 68-95-99.7 RULE (EMPIRICAL RULE)**
    - If you go one standard deviation above and below the mean it will contain 68% of the data
    - If you go two standard deviations above and below the mean it will contain 95% of the data
    - If you go three standard deviations above and below the mean it contains 99.7%% of the data



  - **PLEASE BE AWARE THAT A CURVE OR DATA SET CAN HAVE ALL OF THE FIRST FIVE CHARACTERISTICS BUT IF IT DOESN'T FOLLOW 68-95-99.7 RULE THEN IT CAN NOT BE A NORMAL DISTRIBUTION.**
  - Many variables have outcomes that behave like normal distributions but aren't exactly the values we expect or may have some slightly skewed distributions but they are similar enough that we call them "approximately normal" and still use normal calculations to find probabilities of events.
  - If we are told that a variable follows a normal distribution then we can use the probabilities of the Empirical Rule to help us answer questions about the variable.

Example 1: ACT test scores are approximately normally distributed. One year the scores had a mean $\mu = 21$ and a standard deviation $\sigma = 5$

1. What interval of scores would contain the middle 95% of ACT scores for that year?
   The middle 95% of the scores lie within 2 standard deviations of the mean, so we would have:
   $\mu - 2\sigma$ *on the lower end and* $\mu + 2\sigma$ *on the upper end. Giving us 21-2(5)=11 and 21+2(5)=31 so the interval containing the middle 95% of the scores is from 11 to 31 or (11, 31)*

2. What percentage of ACT scores is less than 26?

26 is one standard deviation (5) above the mean so there is 50% (the scores below the mean) plus 34% (the scores from the mean to one standard deviation above the mean) for a total of 84%

3. What percentage of ACT scores lie between 16 and 31?
   16 is one standard deviation below the mean and 31 is two standard deviations above the mean so we have 34% to the left and 34% +13.5% to the right. 34%+34%+13.5%=82.5%

What if I asked you what percentage of scores were above a 27 which is the score needed to earn a scholarship from the University I am applying to? 27 doesn't lie on one of my standard deviation intervals so we need to use a different method to calculate the probability or percentage:

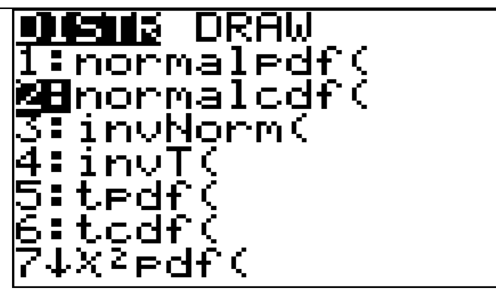
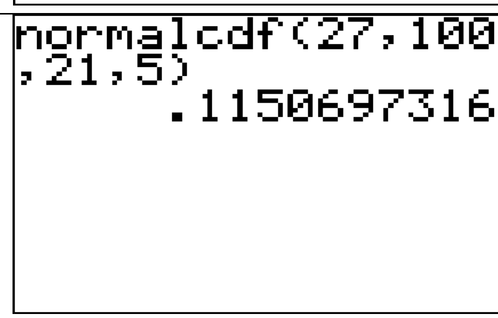
USING THE CALCULATOR TO FIND NORMAL PROBABILITIES:

The calculator has to be one with statistics functions on it. The TI-83's and TI-84's in your class room have stats capabilities and are easy to use.

IF A CALCULATOR IS NOT AVAILABLE, THIS WEBSITE HAS A NICE ONLINE CALCULATOR TO FIND NORMAL PROBABILITIES:

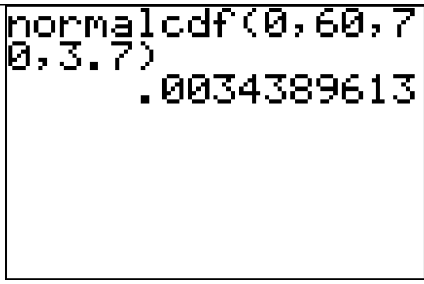
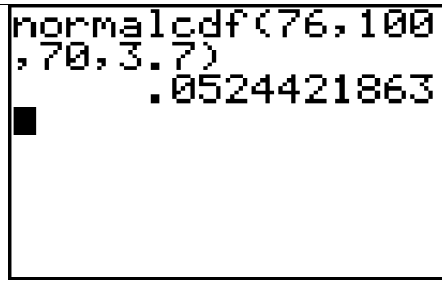
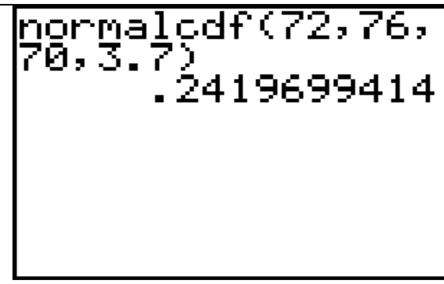
http://www.mathportal.org/calculators/statistics-calculator/normal-distribution-calculator.php

TI-84/83:

The distribution features are found by pushing the 2nd button then VARS button. A menu like the one at the right should appear. Option 2, **normalcdf(** is the feature that you want to use. This feature is the normal cumulative distribution function. It will calculate the percentage of data that fall between two numbers. Select option 2 by using your arrow keys to arrow down to 2 and pushing ENTER .	
The values required for this function are **Normalcdf(lowest value, highest value, mean, st. dev.)** The comma key is located right about the number 7 key. So to answer the question of what percentage of the scores were above a 27 we would enter the following into the calculator: **normalcdf(27, 100, 21, 5)** the 100 entered for the highest value could have been any number as long as it was more than 3 standard deviations away from the mean so we don't cut off any of the tail. We can see from the screen shot to the right that about 11.5% of the scores are higher than a score of 27	

EXAMPLE 2: If the distribution of the heights of males aged 20 – 29 in the United States is approximately normal with a mean of 70 inches with a standard deviation of 3.7 inches. Using the Empirical Rule or a calculator find the following percentages:

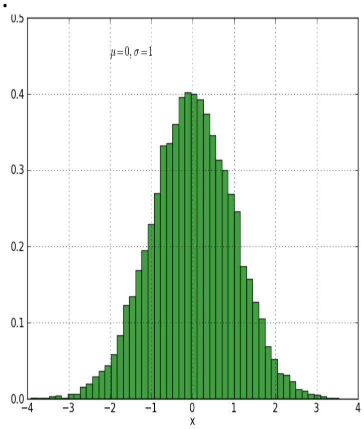
1. What percentage of U.S. men in that age group are under 5 ft tall?
2. To play basketball at Dunk University you must be over 6 ft. 4 in. tall, what percentage of these men would qualify to play?
3. Calvin Klunk Fashion looks for male models between 6 ft tall and 6 ft. 4 in. tall what percentage of U.S. men could model for the company?

		
ABOUT 0.344% First 0 was a number well lower than 3 standard deviations from the mean of 70.	**ABOUT 5.24%** 100 is a big number well above 3 standard deviations above the mean of 70. [6’4” = 76 in.]	**ABOUT 24.2%** **[**6ft=72 in][6’4”=76in] the problem gave us both an upper and lower limit

5.3 EXERCISES **NAME_______________________________________**

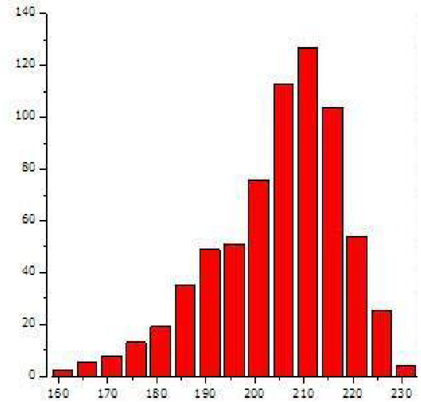
For 1-9 indicate if the distribution could be approximately normal. If not, indicate which characteristic is not met.

1.



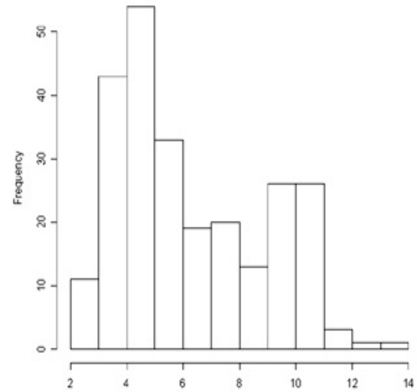
Normal | Not Normal
Why is it not normal?

2.



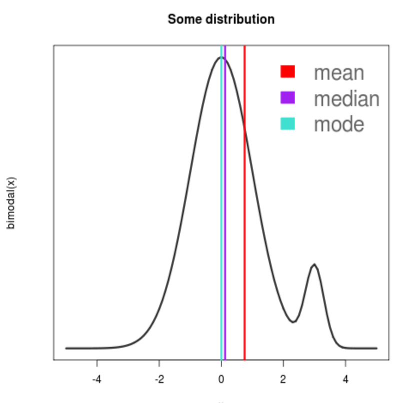
Normal | Not Normal
Why is it not normal?

3.



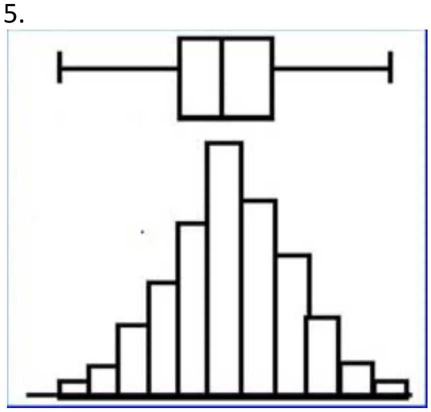
Normal | Not Normal
Why is it not normal?

4.



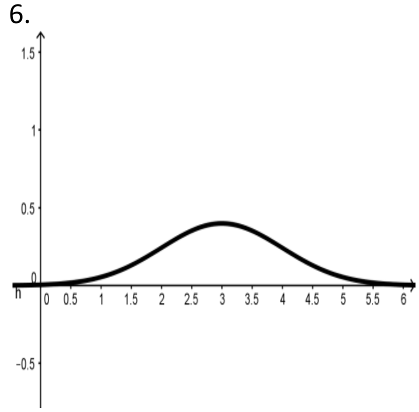
Normal | Not Normal
Why is it not normal?

5.



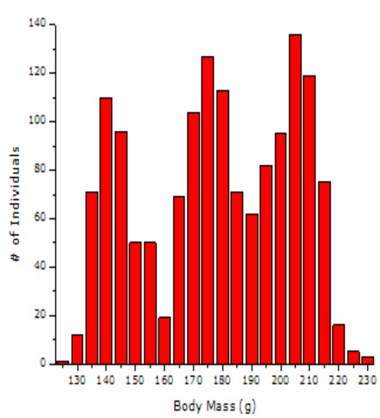
Normal | Not Normal
Why is it not normal?

6.



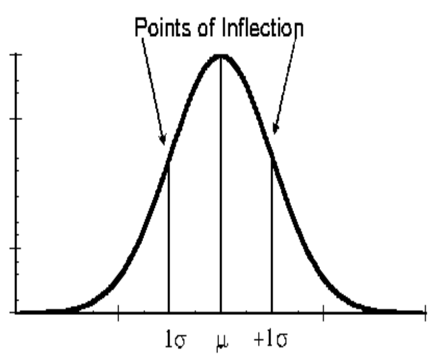
Normal | Not Normal
Why is it not normal?

7.



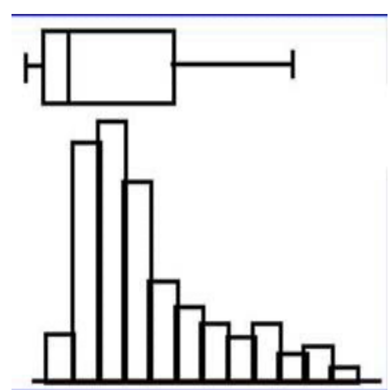
Normal | Not Normal
Why is it not normal?

8.



Normal | Not Normal
Why is it not normal?

9.



Normal | Not Normal
Why is it not normal?

10. Use the 68%-95%-99.7% Rule (Empirical Rule) to see if the following data set could be approximately Normal:

5.5	5.6	4.9	5.1	5.3	5.6	5.4	5.3	5.6	5.7
5.6	5.5	5.6	5.3	5.4	5.3	5.8	5.1	5.3	5.4
5.4	5.5	5.6	5.3	5.5	5.3	5.8	5.7	5.9	5.2

Using the calculator you find that the mean=5.45 and the standard deviation=0.23

a. What is the interval of values that are 1 σ away from the mean?
   2σ from the mean?
   3σ from the mean?

b. What % of the values lies within one standard deviation of the mean?

c. What % of the values lies within two standard deviations of the mean?

d. What % of the values lies within three standard deviations of the mean?

e. Does the data fit the Empirical Rule?

f. Would you say that the data set is Approximately Normal or not Normal?

Find the following solutions using the 68%-95%-99.7% Rule or a calculator.
IF A CALCULATOR IS NOT AVAILABLE, THIS WEBSITE HAS A CALCULATOR TO FIND NORMAL PROBABILITIES:
http://www.mathportal.org/calculators/statistics-calculator/normal-distribution-calculator.php

11. The mathematics portion of the SAT is known to be Normal and has a mean score of 500 and a standard deviation of 100.

a. What is the interval that contains the middle 99.7% of scores?

b. What percentage of SAT scores is greater than 600?

c. What percentage of SAT scores is between 300 and 700?

12. Americans consume 16.5 pounds of ice cream per year with a standard deviation of 3.25 pounds. If the distribution of consumed ice cream is Normal:

a. What is the interval that contains 68% of the pounds consumed each year?

b. What percentage of pounds consumed is less than 10 pounds?

c. What percentage of pounds consumed is between 5 pounds and 11pounds?

13. The height of a NBA basketball players is Normally distributed with mean = 79 inches & st. dev. = 3.89 inches.

a. What is the interval that contains 95% of the heights?

b. What percentage of the heights is greater than 81 inches?

c. What percentage of the heights is between 73 inches and 77 inches?

14. This year's ACT Scores are normally distributed and had a mean $\mu = 19$ and a standard deviation $\sigma = 5.4$ Joey knows that Yale University usually only accepts students that performed in the top 1% on the ACT test. Joey has great grades and got a 34 on his ACT test, will Joey qualify for acceptance to Yale University this year?

15. A bag of Lay's Potato Chips have weights that are normally distributed with a mean of 9.12 ounces and a standard deviation of 0.05 ounces. If the print the weight on the front of the bag as being 9 ounces, what percentage of the bags of chips would actually be under weight? Why don't they print the weight as 9.12 ounces?