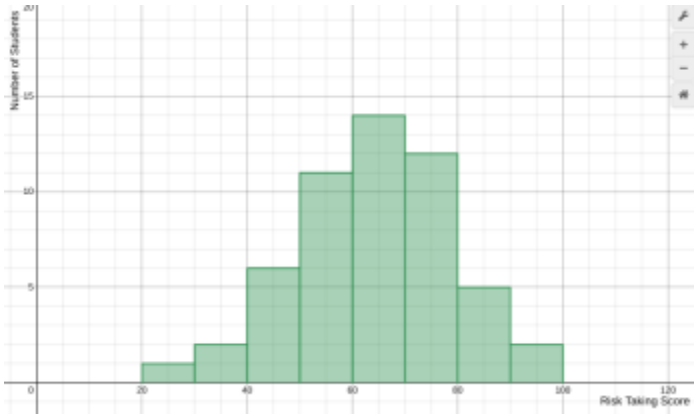
Data Analysis Lesson #7: Normal Distribution Part 1 Name _____

1. **Risk Taking Results:** The results of last year's students Risk Taking Survey are shown in the table below. The score was created by combining the responses of the 5 questions and scaling this so that is has a maximum of 100 and a minimum of 0 (meaning 100 = Maximum Risk Taker and 0 = Minimum Risk Taker)

Risk Taking Score	Number of Students	Relative Frequency
$0 \leq x \leq 10$	0	0.00
$10 < x \leq 20$	0	0.00
$20 < x \leq 30$	1	0.019
$30 < x \leq 40$	2	0.038
$40 < x \leq 50$	6	0.113
$50 < x \leq 60$	11	0.208
$60 < x \leq 70$	14	0.264
$70 < x \leq 80$	12	0.226
$80 < x \leq 90$	5	0.094
$90 < x \leq 100$	2	0.038
Total	53	1.000

The RELATIVE FREQUENCY HISTOGRAM of the Risk Taking Scores is shown here:



a. What does the relative frequency of 0.113 mean for a risk taking score $40 < x \leq 50$?
Response:

b. Use the table to determine the percentage of students that had a Risk Taking Score:
More than 60.
Response:
Between 40 and 70.
Response:

c. The MEAN Risk Taking Scores is 62.9 and the Standard Deviation is 15.4. On the image mark the MEAN and mark ONE STANDARD DEVIATION above and below the mean.
   i. Approximately what percent of the values in this data set are within one standard deviation of the mean (i.e., between 62.9 - 15.4 = 47.5 and 62.9 + 15.4 = 78.3)? Show your work.
   Response:

   ii. Using a similar strategy, approximately what percent of the values in this data set are within two standard deviations of the mean(i.e., between 62.9 - 2*15.4 = 32.1 and 62.9 + 2*15.4 = 93.7)? Show your work.
   Response:

d. On the image, draw a smooth curve that comes reasonably close to passing through the MIDPOINTS of the tops of the bars in the histogram. Does this curve look approximately bell shaped and symmetric?
Response:

e. **Based on your work above**, how would you answer the question, "How Risk Taking are these students?" Justify your answer.
Response:

2. **The Normal Distribution**
   a. Watch: The Normal Distribution (also linked on Canvas). Describe what is meant by the 68-95-99.7 Rule.
   Response:

   b. Examine your answers to questions 1ci and 1cii above. Does the Risk Taking data follow the 68-95 part of the 68-95-99.7 Rule? Explain your answer.
   Response:

   c. Now you know: A NORMAL CURVE is a smooth curve that is **symmetric and bell shaped**. Data distributions that have these properties are called approximately NORMAL.

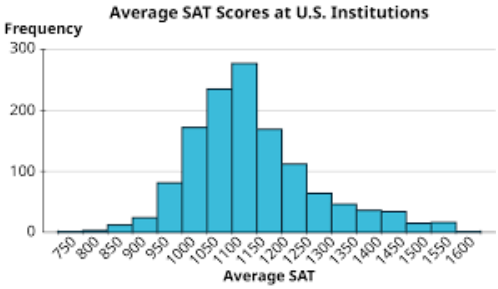
   Using Risk Taking Scores as an example, what does it mean about the class's Risk Taking preferences that these scores are approximately NORMAL (i.e., symmetric and bell shaped)? Be specific.
   Response:

Turn in when finished and start Part 2

Data Analysis Lesson #7: Normal Distribution Part 2 *Name* _____

What data would we expect to be Normal? Recall that data that follows a NORMAL DISTRIBUTION is bell shaped. That means that most of the scores are clustered around the mean and the frequency of scores decreases as you move further away from the mean in either direction.

1. Here is a sample of SAT scores for students attending U.S. colleges and universities.



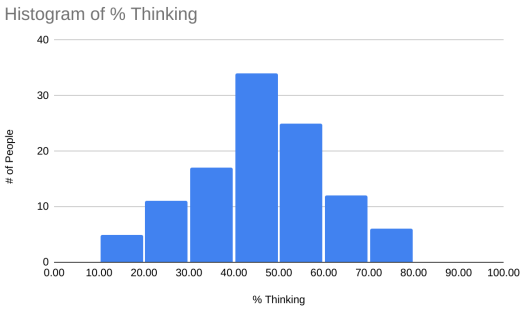
   a. Describe how this approximately fits a Normal distribution's bell curve.
   Response:

   b. Why does it make logical sense that SAT scores would follow this pattern?
   Response:

   c. One feature of a histogram that follows the normal distribution is that the MEAN (or average) is equivalent (or very close to) the MEDIAN. Why would this feature most likely create a bell curve?
   Response:

   d. Notice that the data from this histogram is sampled from a population of students attending college or university. Would you expect this data to be approximately Normal if it was a sample of everyone who took the SAT? Why or why not?
   Response:

2. This histogram shows the % Thinking (opposite of Feeling) for a sample of Lincoln High School students.



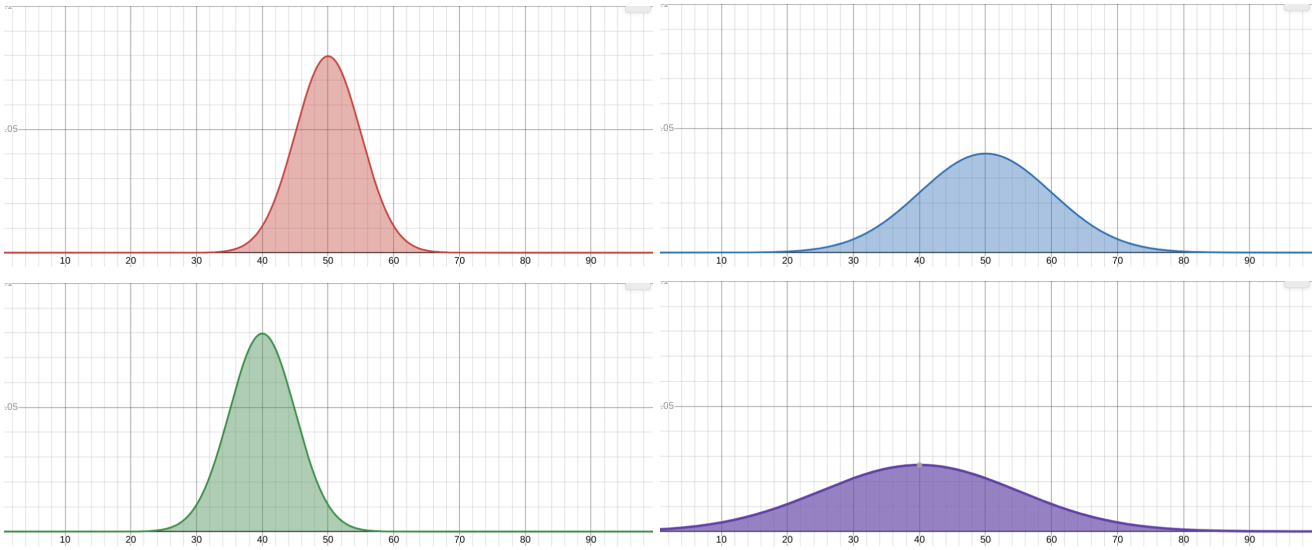
   a. Does this sample appear to follow an approximately Normal distribution? Explain why or why not using some of the ideas from the previous question.
   Response:

   b. Do you think that %Thinking scores logically should follow a normal distribution? Explain why or why not.

Response:

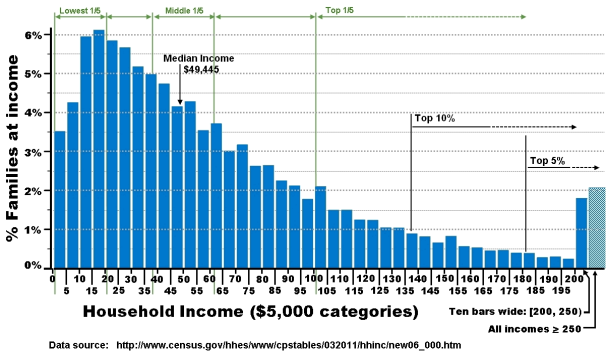
3. Next lesson we will learn more about the Normal Distribution and how to use it to calculate %s that we have only approximated up to now. What we will learn is that you can provide a close approximation to a Normal Bell Curve if you know the MEAN and STANDARD DEVIATION of a sample (and assume your sample is representative of the population).

   Examine the Normal curves below and match the curves with the appropriate Mean and Standard Deviation pairs. Explain your reasoning.



A: Mean = 40, SD = 15, B: Mean = 50, SD = 5, C: Mean = 40, SD = 10, D: Mean = 50, SD = 10
Response:

4. Some data does not fit the features of a Normal distribution. For example, the histogram below shows household income in the U.S.
   a. Why would we NOT expect household income to follow a Normal distribution?
      Response:



   b. In economics, the GINI COEFFICIENT is used to measure the amount of income inequality in a country. So that a GINI COEFFICIENT = 0 means PERFECT EQUALITY and a GINI COEFFICIENT = 100 means PERFECT INEQUALITY. The GINI COEFFICIENT for the U.S. is 42, for example.
      If a country's household income perfectly follows the Normal Distribution, do you think the GINI COEFFICIENT for that country would be greater or smaller than the U.S.? Explain your thinking.
      Response: