Section 3.4 – Normal Distribution

MDM4U
Jensen

Part 1: Dice Rolling Activity

a) Roll two 6-sided number cubes 18 times. Record a tally mark next to the appropriate number after each roll. After rolling the cube 18 times, determine the frequency for each number by counting the tally marks.

Individual Data		
Sum of the Numbers Rolled	Tally Marks	Frequency
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		

b) Create a histogram or bar graph for your Individual Data for Activity 2.

c) Determine the value of each of the following for your data set.

Mean =
Median =
Mode =

Range =
Interquartile Range =
Standard Deviation =

Record the combined class data for Activity 2 in the table below.

Sum of the Numbers Rolled	Frequency
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

d) Create a histogram or bar graph for the Class Data for Activity 2.

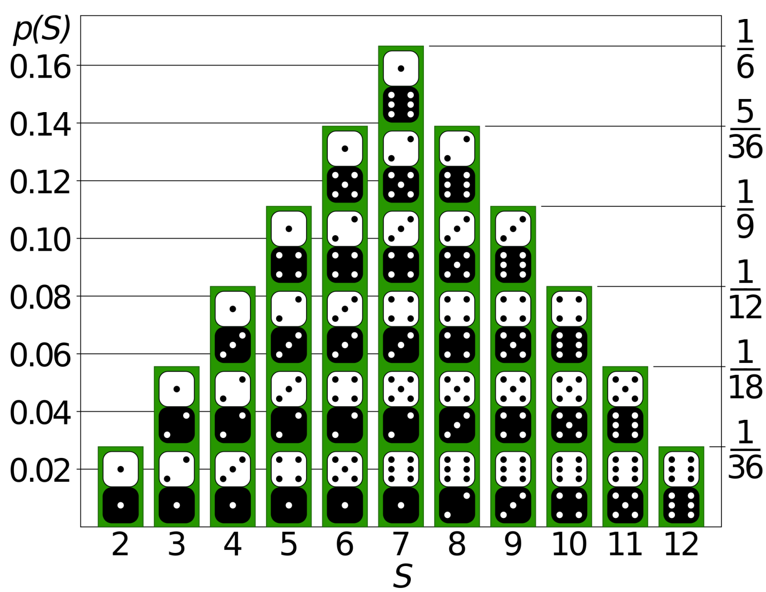
e) Determine the value of each of the following for the class data set.

Mean =
Median =
Mode =

Range =
Interquartile Range =
Standard Deviation =

f) Describe the frequency distribution of class set of data. What do you notice about shape, measures of central tendency, etc.?

Summary of activity: If you were to draw a smooth curve through the tops of the rectangles on the distribution of the class data, it would approximately form what is called a normal curve. Shown to the right is the theoretical probability distribution for rolling two dice. The more trials we complete, the closer our distribution would look to the actual probability distribution shown:



Part 2: Properties of a Normal Distribution

If you are to measure many similar things which have differences caused by a random variation, those results will typically be distributed symmetrically and unimodally about the mean. Statisticians observe this mound shaped curve so often that its mathematical model is known as the normal distribution. Distributions that are close to normal include: scores on tests taken by many people, repeated measurements of the same quantity (heights, weights), characteristics of biological populations (yields of corn), and chance outcomes (dice rolling example).

The normal curve was first used in the 1700's by French mathematicians and early 1800's by German mathematician and physicist Karl Gauss. The curve is known as the Gaussian distribution and is also sometimes called a bell curve.

General Properties of all Normal Distributions:

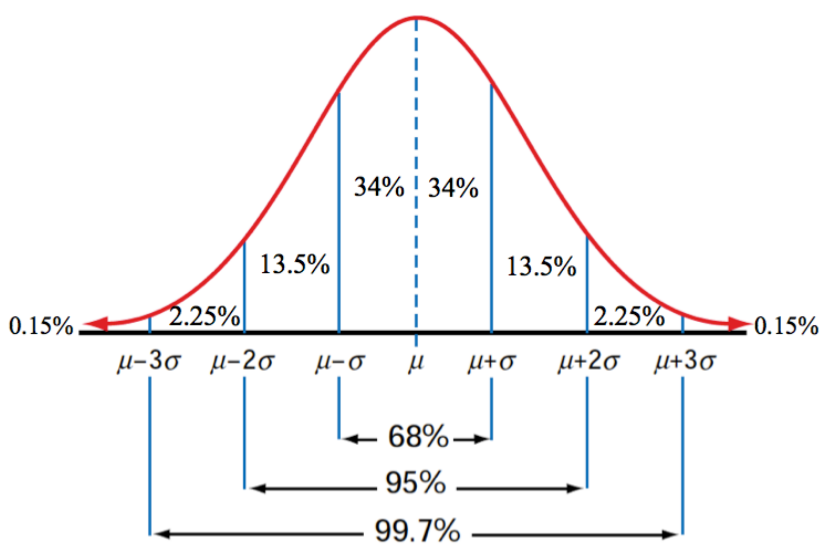
- It is unimodal and symmetrical; the mean, median and mode are equal and fall at the line of symmetry
- It is shaped like a bell, peaking in the middle and sloping down toward the sides. It approaches zero at the extremes

The Empirical Rule:

Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean. We'll soon show how to find these numbers precisely; but one simple rule (the Empirical Rule) is usually all we need.

In any normal distribution…

- Approximately 68% of the observations fall within 1 standard deviation of the mean
- Approximately 95% of the observations fall within 2 standard deviations of the mean
- Approximately 99.7% of the observations fall within 3 standard deviations of the mean



Notation Used:

Any particular normal distribution is completely described by two numbers: its mean μ and standard deviation σ

$$\boldsymbol{X \sim N(\bar{\mu},\ \sigma^2)}$$

X **- Data**

N **– Normal distribution**

$\boldsymbol{\mu}$ **– Population mean**

$\boldsymbol{\sigma}$ **– Population Standard Deviation**

Example:

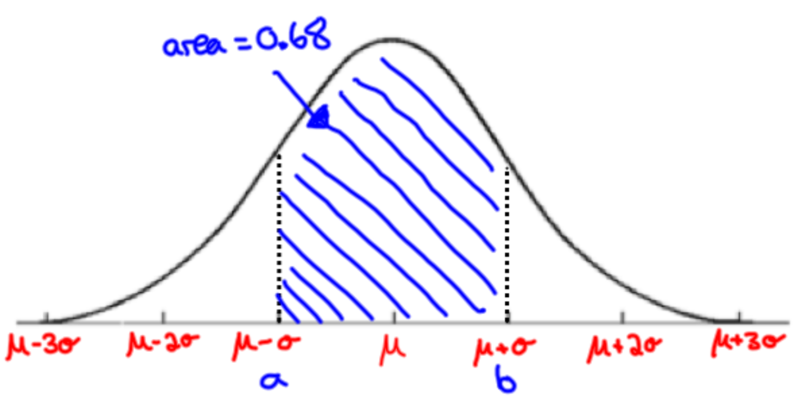
$X \sim N(50, 5^2)$

This normal distribution would have a mean of 50, a standard deviation of 5 and a variance of 25.

Part 3: Area Under a Normal Curve

The area under every normal curve equals 1. In any normal distribution, the percent of the data that lies between two specific values, *a* and *b*, is the area under the normal curve between endpoints *a* and *b*

Example:



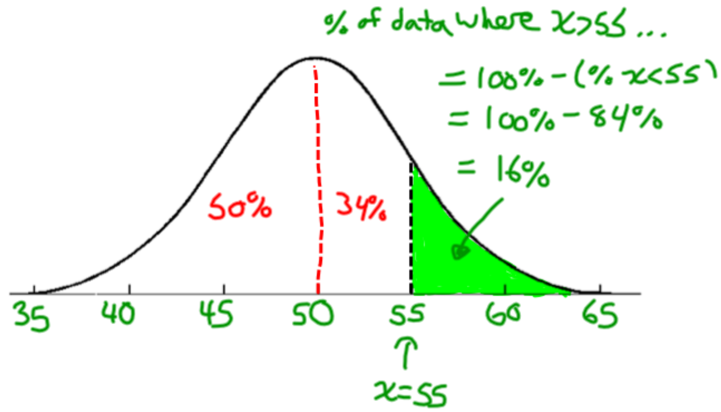
For ALL normal distribution problems you should sketch, label, shade the region of the curve, and answer the question.

Example 1: If $X \sim N(50, 5^2)$, shade in the area of the given interval on a normal curve and find the area of the shaded region using the Empirical Rule.

a) $x > 55$

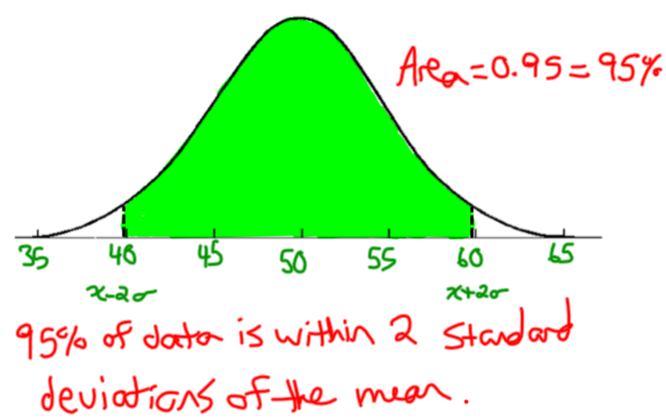
The shaded region on the diagram shows the area under the curve greater than 55. Using the empirical rule, we can see that approximately 16% of the data is greater than 55.

Area under curve = 100% - 50% - 34%
= 16%



b) $40 < x < 60$

Area under curve = 95%



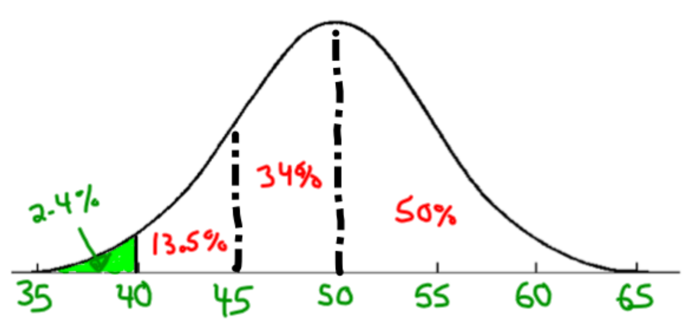
c) $x < 40$

Method 1:

Area under curve = 2.25% + 0.15%
= 2.4%

Method 2:

Area under curve = 100% - (50% + 34% + 13.5%)
= 2.5%

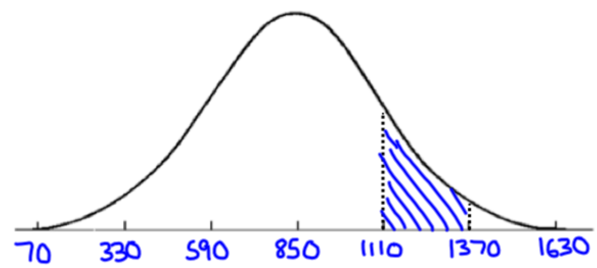


Note: different methods get slightly different answers because we are using the Empricial Rule which has rounded values.

Example 2: A real estate agent, working entirely on commission, makes an average of $850 with a standard deviation of $260 weekly selling property in the city. If we assume this distribution is roughly normal, what is the probability that a real estate agent will make between $1110 and $1,370 selling property in the city?

Area under curve = 13.5%

There is approximately a 13.5% chance that the agent makes between $1110 and $1370 in a week.



Example 3: Julie is an engineer who designs roller coasters. She wants to develop a ride that 95% of the population can ride. The average adult in North America has a mass of 71.8 kg, with a standard deviation of 13.6 kg.

a) Describe this information in the normal distribution notation

$X \sim N(71.8, 13.6^2)$

b) If she wanted to provide for 95% of the general population, what range of masses should she anticipate?

95% of the population would be within two standard deviations of the mean.

71.8 – 2(13.6) = 44.6
71.8 + 2(13.6) = 99

Therefore she should anticipate a range of 44.6 – 99 kg