Algebra II/Trig

Module 4, Unit 9 Lesson 9 – Normal Distributions and the Empirical Rule

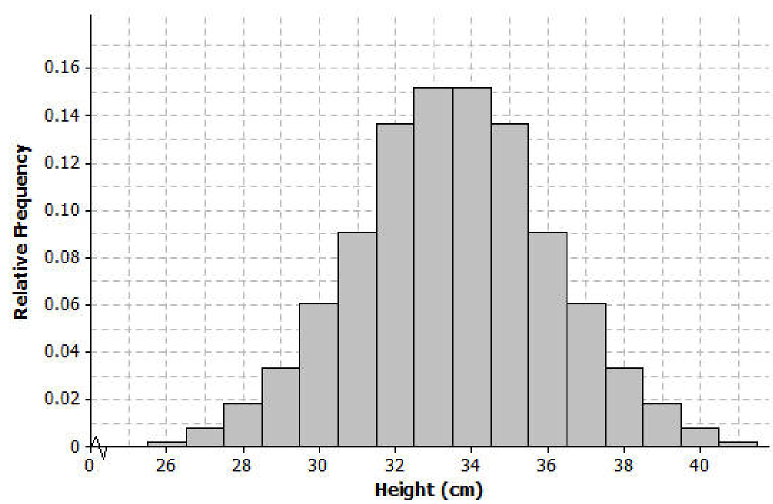
Example 1: Heights of Dinosaurs and the Normal Curve

A paleontologist studies prehistoric life and sometimes works with dinosaur fossils. The table below shows the distribution of heights (rounded to the nearest inch) of 660 procompsognathids, otherwise known as compys.

The heights were determined by studying the fossil remains of the compys.

Height (cm)	Number of Compys	Relative Frequency
26	1	0.002
27	5	0.008
28	12	0.018
29	22	0.033
30	40	0.061
31	60	0.091
32	90	0.136
33	100	0.152
34	100	0.152
35	90	0.136
36	60	0.091
37	40	0.061
38	22	0.033
39	12	0.018
40	5	0.008
41	1	0.002
Total	660	1.000

The following is a relative frequency histogram of the compy heights:



1. What does the relative frequency of 0.136 represent for the height of 32 cm?

2. What is the width of each bar? What does the height of the bar represent?

3. What is the area of the bar that represents the relative frequency for compys with a height of 32 cm?

4. The mean of the distribution of compy heights is 33.5 cm, and the standard deviation is 2.56 cm. Interpret the mean and standard deviation in this context.

5. Mark the mean on the graph, and mark *one* deviation above and below the mean.

   a. Approximately what percent of the values in this data set are within one standard deviation of the mean (i.e., between $33.5\text{ cm} - 2.56\text{ cm} = 30.94\text{ cm}$ and $33.5\text{ cm} + 2.56\text{ cm} = 36.06\text{ cm}$)?

   b. **A**pproximately what percent of the values in this data set are within *two* standard deviations of the mean?

6. Draw a smooth curve that comes reasonably close to passing through the midpoints of the tops of the bars in the histogram. Describe the shape of the distribution.

7. Shade the area of the histogram that represents the proportion of heights that are within one standard deviation of the mean.

Example 2: Gas Mileage and the Normal Distribution

A **normal curve** is a smooth curve that is symmetric and bell shaped. Data distributions that are mound shaped are often modeled using a normal curve, and we say that such a distribution is **approximately normal**. One example of a distribution that is approximately normal is the distribution of compy heights from Example 1. Another example is below. A salesman kept track of the gas mileage for his car over a 25-week span.

The mileages (miles per gallon rounded to the nearest whole number) were

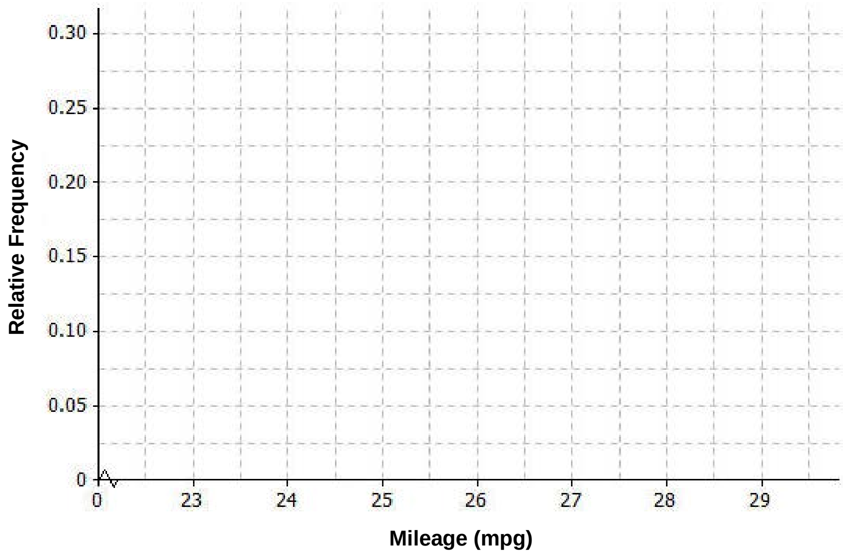
$23, 27, 27, 28, 25, 26, 25, 29, 26, 27, 24, 26, 26, 24, 27, 25, 28, 25, 26, 25, 29, 26, 27, 24, 26.$

a. Use Desmos to find the mean and standard deviation of the mileage data.

b. Calculate the relative frequency of each of the mileage values.

Mileage (mpg)	Frequency	Relative Frequency
23		
24		
25		
26	7	
27		
28		
29		
Total	25	

c. Construct a relative frequency histogram using the scale below.

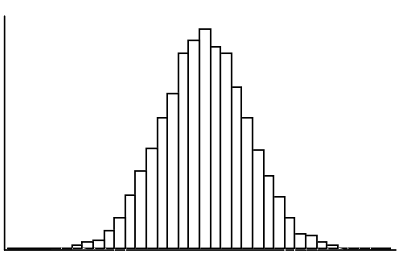


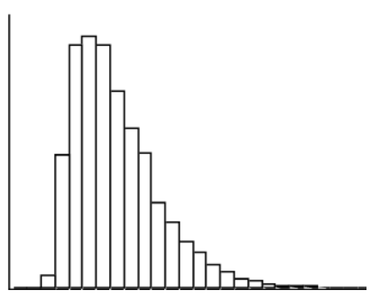
d. Describe the shape of the mileage distribution. Draw a smooth curve that comes reasonably close to passing through the midpoints of the tops of the bars in the histogram. Is this approximately a normal curve?

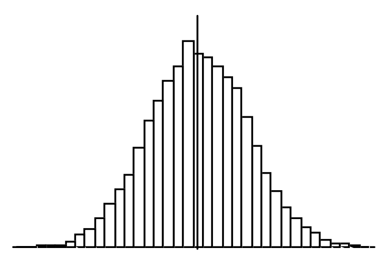
e. Mark the mean on the histogram. Mark one standard deviation to the left and right of the mean. Shade the area of the histogram that represents the proportion of mileages that are within one standard deviation of the mean. Find the proportion of the data within one standard deviation of the mean.

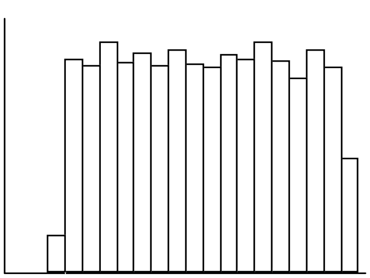
Example 3: Identifying Approximately Normal Distributions

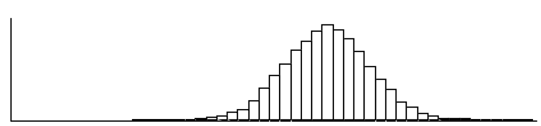
For each of the following distributions, indicate if it is approximately normal, skewed, or neither.











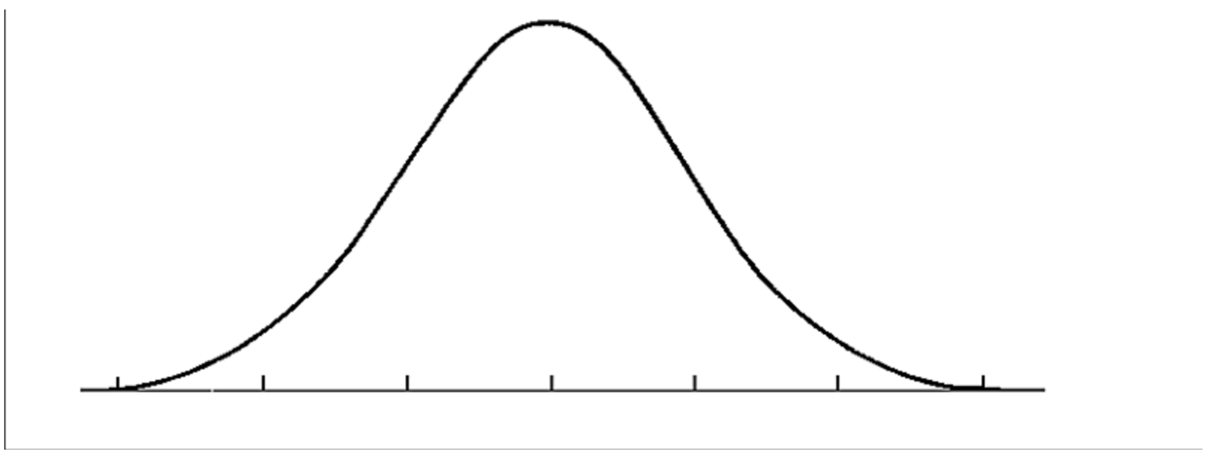
In previous examples we identified what *approximately* normal distributions look like, and started investigating how they behave. Specifically, we estimated the proportion of values within one and two standard deviations of the mean. Because those distributions were *approximately* normal, the answers varied a bit among the examples. Nevertheless, if a distribution is approximately normal, we can use a perfect normal curve to model the situation and give reasonable estimates about percentages and probabilities.

Normal curves are special, as the mean and standard deviation provide a complete description of the distribution. Remember that in such distributions the mean and median are approximately ____________.

Sketching a Normal Curve.
To determine a scale when drawing a normal curve, it is important to note that the mean corresponds to the ________________ of the curve. We then count by the standard deviation. If asked to sketch and scale a curve, you should go out 3 standard deviations in either direction from the mean.

Example 4: Scaling

The distribution of team batting average for the St. Louis Cardinals for the 50 years from 1964 to 2013 can be considered to be approximately normal with $\mu = 0.2637$ and $\sigma = 0.0096$. Use this information to mark the normal curve below.

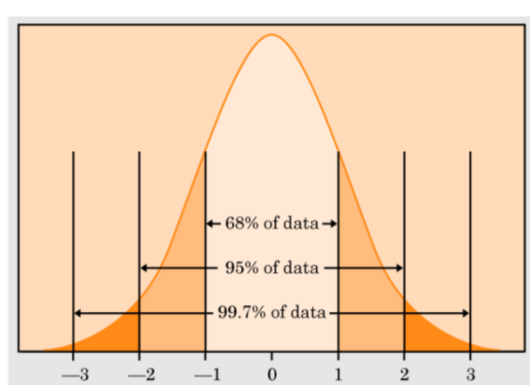


The Empirical Rule
As mentioned, normal distributions are completely described by the mean and standard deviation. The Empirical (aka 68-95-99.7) rule reinforces this fact. This rule states that, in a normal distribution:

- Approximately ____________% of the data lies within __________ standard deviation of the mean.
- Approximately ___________ % of the data lies within ___________ standard deviations of the mean.
- Approximately ___________ % of the data lies within ___________ standard deviations of the mean.

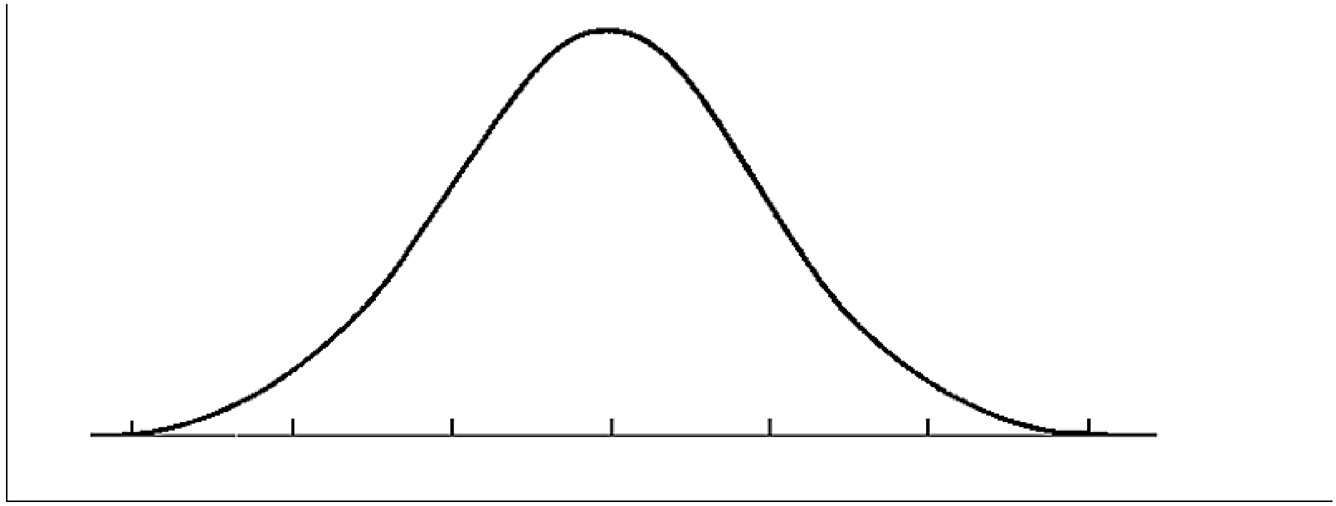
This powerful fact is illustrated in the diagram below.

Let's use this information to mark the curve you scaled above.



Example 5: Estimating using the Empirical Rule

500 freshmen at Schaumburg High School took an algebra test. The scores were distributed normally with a mean of 75 and a standard deviation of 7. Label the mean and three standard deviations from the mean.



Use the Empirical Rule to mark the percentages on the distribution above.

Use the information above to answer the questions below.

a) What percentage of scores are between scores 61 and 82?

b) What percentage of scores are between scores 75 and 82?

c) What percentage of scores are between scores 61 and 89?

d) What percentage of scores is less than a score of 61?

e) What percentage of scores is greater than a score of 96?

f) Approximately how many algebra students scored between 61 and 89?

g) Approximately how many algebra students scored between 68 and 82?

h) Approximately how many algebra students scored between 61 and 75?

i) Approximately how many algebra students scored between 89 and 96?

j) Approximately how many algebra students scored higher than 89?

Example 6: More Empirical Rule

The incubation time for Rhode Island Red chicks is approximately normally distributed with mean 21 days and standard deviation approximately 1 day. If 1000 eggs are being incubated, how many chicks do we expect will hatch in:

a. 19 to 23 days?

b. 21 days or fewer?

c. 19 days or fewer

d. 18 to 24 days?

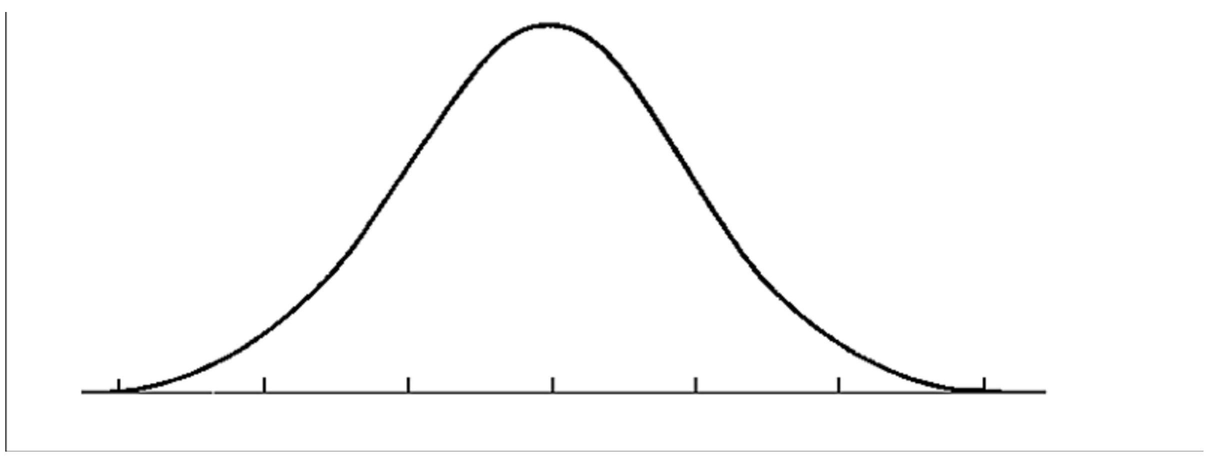
e. 22 days or more?

f. 18 to 21 days?

Practice: The Empirical Rule

Given an approximately normal distribution with a mean of 159 and a standard deviation of 17,

a) Label the curve below with the appropriate values and percentages.



b) What percent of values are within the interval (142, 176)?

c) What percent of values are within the interval (125, 193)?

d) What interval contains 99.7% of all values?

e) What percent of values are above 176?

f) What percent of values are below 125?