Lesson 9: Using a Curve to Model a Data Distribution

Classwork

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 or "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	1.00

Exercises 1–8

The following is a relative frequency histogram of the compy heights.
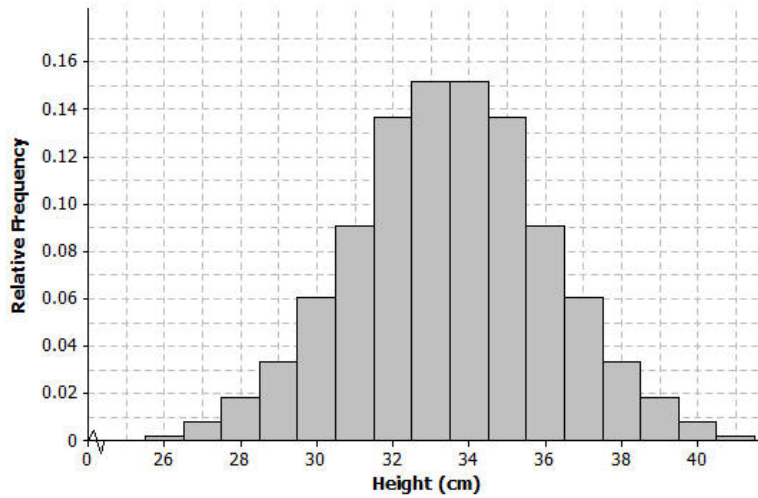
1. What does the relative frequency of 0.136 mean for the height of 32 cm?

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 <u>one</u> 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 $33.5 - 2.56 = 30.94$ cm and $33.5 + 2.56 = 36.06$ cm.)

   b. Approximately what percent of the values in this data set are within <u>two</u> 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 under the curve that represents the proportion of heights that are within one standard deviation of the mean.

8. Based on our analysis, how would you answer the question, "How tall was a compy?"

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. Distributions that are approximately normal occur in many different settings. For example, 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.

Exercise

9. Consider the following:

    a. Use technology to find the mean and standard deviation of the mileage data. How did you use technology to assist you?

    b. Calculate the relative frequency of each of the mileage values. For example, the mileage of 26 mpg has a frequency of 7. To find the relative frequency, divide 7 by 25, the total number of mileages recorded. Complete the following table.

| Mileage	Frequency	Relative Frequency
23		
24		
25		
26	7	
27		
28		
29		
Total	25	

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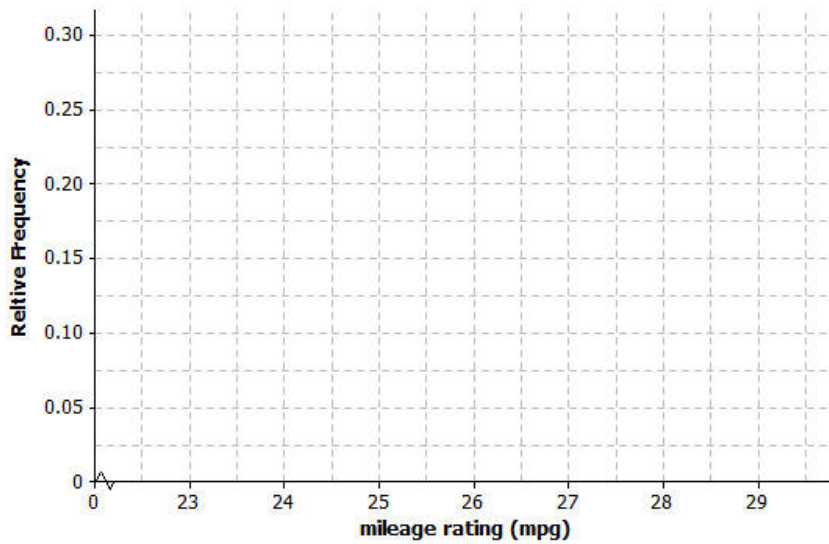
d.   Describe the shape of the mileage distribution. Draw a smooth curve that comes reasonably close to passing through the midpoints of the tops of the bars in the histogram. Is this approximately a normal curve?

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

Lesson Summary

- A normal curve is symmetric and bell shaped. The mean of a normal distribution is located in the center of the distribution. Areas under a normal curve can be used to estimate the proportion of the data values that fall within a given interval.
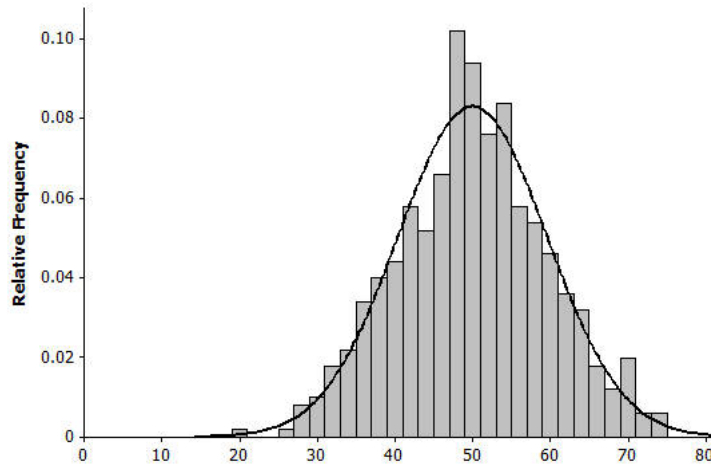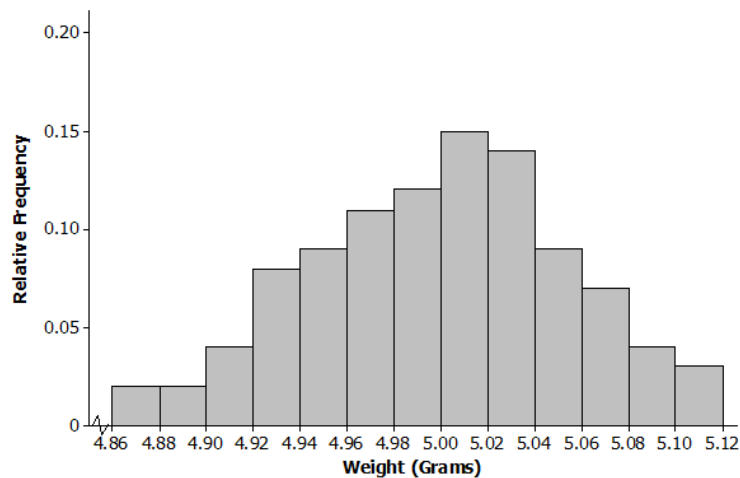

- When a distribution is skewed, it is not appropriate to model the data distribution with a normal curve.
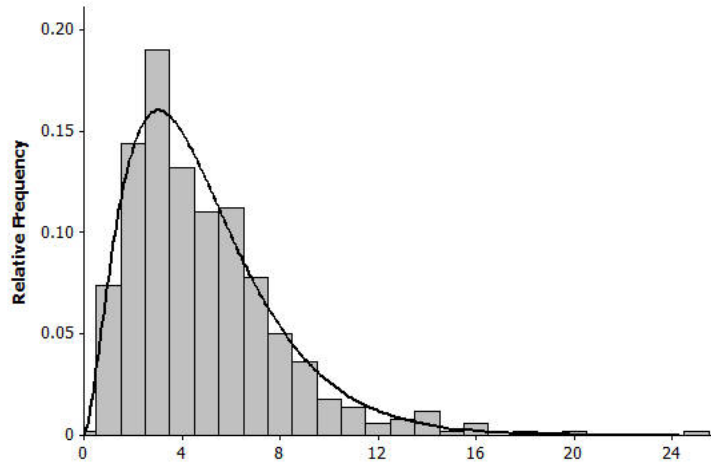
Problem Set

1. Periodically the U.S. Mint checks the weight of newly minted nickels. Below is a histogram of the weights (in grams) of a random sample of 100 new nickels.
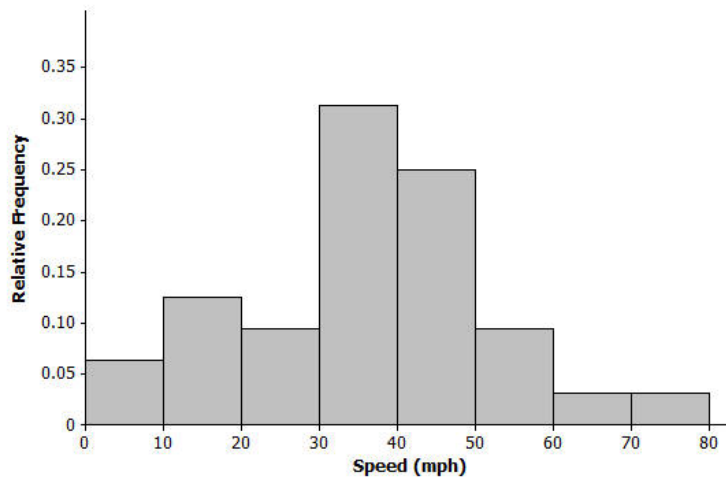

a. The mean and standard deviation of the distribution of nickel weights are 5.00 grams and 0.06 grams, respectively. Mark the mean on the histogram. Mark one standard deviation above the mean and one standard deviation below the mean.

b. Describe the shape of the 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?

c. Shade the area under the curve that represents the proportion of data within one standard deviation above and below the mean. Find the proportion of the data within one standard deviation above and below the mean.

2. Below is a relative frequency histogram of the gross (in millions of dollars) for the all-time top-grossing American movies (as of the end of 2012). Gross is the total amount of money made before subtracting out expenses, like advertising costs and actors' salaries.
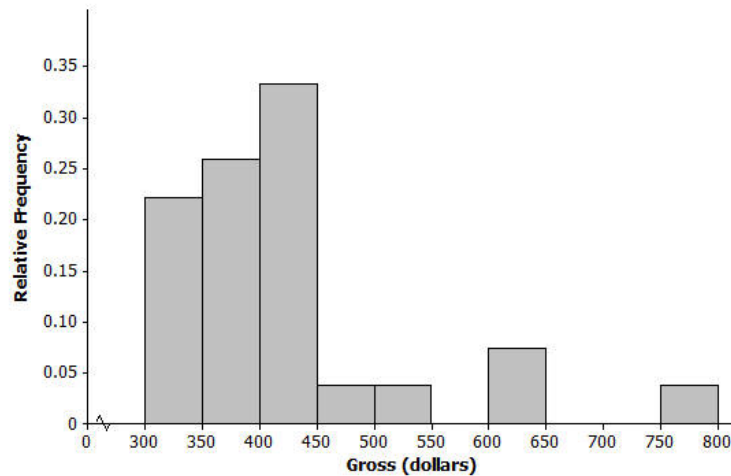
a. Describe the shape of the distribution of all-time top-grossing movies. Would a normal curve be the best curve to model this distribution? Explain your answer.

b. Which of the following is a reasonable estimate for the mean of the distribution? Explain your choice.

   i.   325 million

   ii. 375 million

   iii. 425 million

c. Which of the following is a reasonable estimate for the sample standard deviation? Explain your choice.

   i.   50 million

   ii. 100 million

   iii. 200 million

3. Below is a histogram of the top speed of different types of animals.

a. Describe the shape of the top speed distribution.

b. Estimate the mean and standard deviation of this distribution. Describe how you made your estimate.

c. Draw a smooth curve that is approximately a normal curve. The actual mean and standard deviation of this data set are 34.1 and 15.3. Shade the area under the curve that represents the proportion of the data values that are within one standard deviation of the mean.