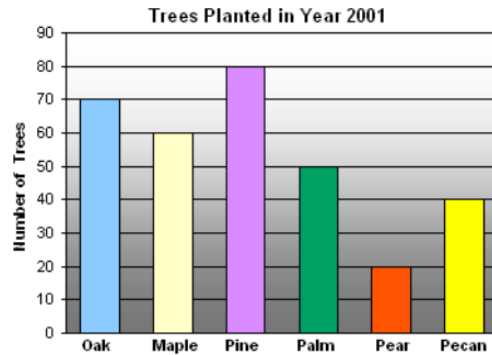




Note # 2: Describing our Data Using Graphs and Numbers

Problem 1. The following graph shows the different types of trees planted in Spring Hill City in 2001.

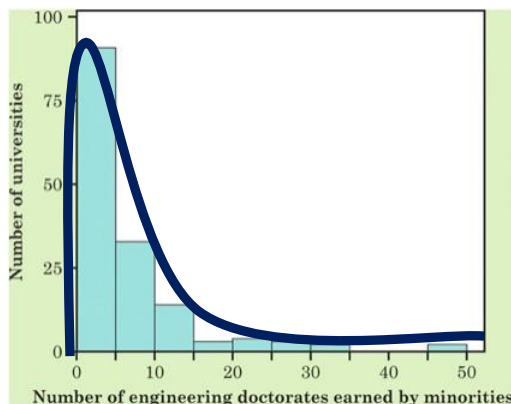


- What type of graph is this?
- What is the mode?
- Approximately how many Pear trees were planted?
- Approximately how many total trees were planted?

Solution:

- What type of graph is this? **Bar graph / Bar chart.**
- What is the mode? **Pine trees.**
- Approximately how many Pear trees were planted? **20 pear trees.**
- Approximately how many total trees were planted? **$70+60+80+50+20+40 = 320$ trees.**

Problem 2. The figure below is a histogram of the number of minority students (black, Hispanic, Native American) who earned doctorate degrees in engineering from each of 152 universities in the years 2000 through 2002.



- About how many universities had between 0 and 10 engineering doctorates earned by minorities?
- How would you describe the shape of this histogram?
- Based only on this histogram, does it appear as if there are any possible outliers?

Solution:

- About how many universities had between 0 and 10 engineering doctorates earned by minorities?

0 - 5 → about 90
 5 - 10 → about 30
 → Around 120 universities.

- How would you describe the shape of this histogram?

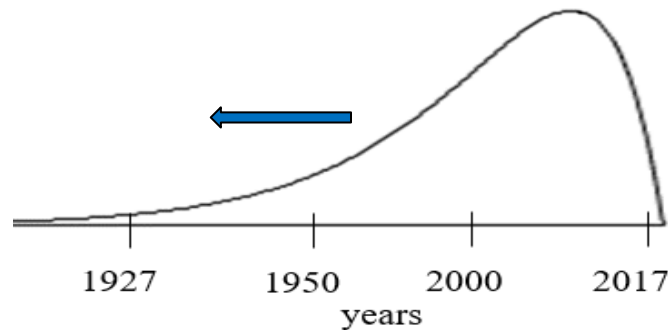
Modality - unimodal
 Skewness - skewed to right

- Based only on this histogram, does it appear as if there are any possible outliers? **Yes,** between 45 - 50 engineering doctorates.

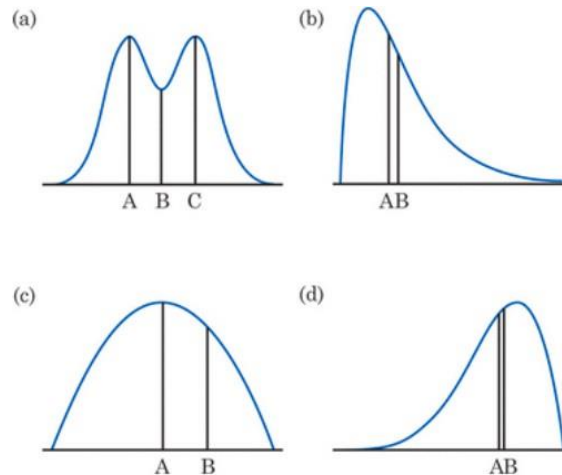
Problem 3. Kyla has a coin collection that she started with her grandfather. She specifically collects quarters. Most of her quarters are fairly new (made between 2000 and 2017) but she has quite a few quarters from between 1950 and 2000. She even has a few quarters made as early as 1927. If you were to describe the distribution of the years in which Kyla's quarters were made, would you expect this distribution to be left-skewed, right-skewed, or symmetric?

Solution:

Left-skewed



Problem 4. The following diagram shows four different density curves, each with 2-3 points marked. For each curve, identify which points represent the mean and the median (some points may not be used).



Solution:

- a. Symmetric, mean = B, median = B.
- b. Right - skewed, mean = B, median = A.
- c. Symmetric, mean = A, median = A.
- d. left - skewed, mean = A, median = B.



Problem 5. Lauren’s professor gives 15 point quizzes most weeks, for a total of 12 quizzes over the course of the semester. Lauren’s scores are: 15, 11, 12, 13, 9, 14, 14, 12, 10, 9, 15, and 2. Are any of her scores potentially outliers?

Solution:

To be able to find potential outliers, we need to find the five summary points (step1) first, then check for outliers (step2).

Step1: Find the five summary points (Min, Q_1 , M, Q_3 , and Max): first, we need to sort this data.

Sorted data set: 2 9 9 10 11 12 12 13 14 14 15 15

- i. Min = 2
- ii. $Q_1 = 9.5$ [avg. (9,10) = 9.5]
- iii. M = 12 [avg. (12,12) =12]
- iv. $Q_3 = 14$ [avg. (14,14) =14]
- v. Max = 15

Step2: Find potential outlier: To find potential outlier, we need to calculate IQR.

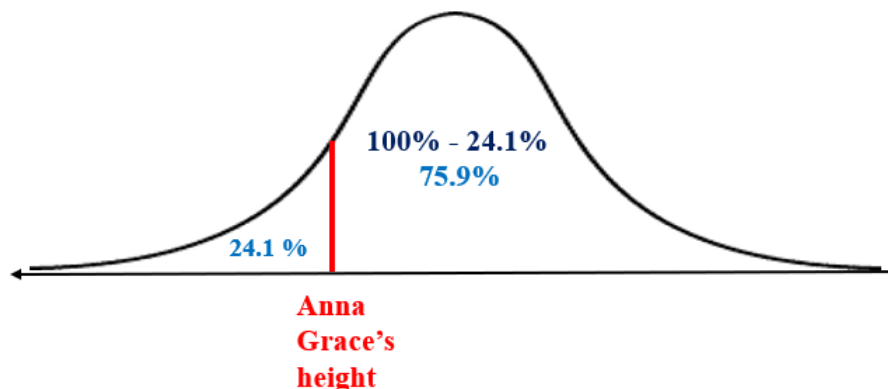
$$\therefore IQR = Q_3 - Q_1 = 14 - 9.5 = 4.5$$

- Smaller than Q_1 : $Q_1 - 1.5 \times IQR \rightarrow 9.5 - 1.5 \times 4.5 = 9.5 - 6.75 = 2.75$ (yes, min = 2)
- Bigger than Q_3 : $Q_3 + 1.5 \times IQR \rightarrow 14 + 1.5 \times 4.5 = 14 + 6.75 = 20.75$ (no, max = 15)

Are any of her scores potentially outliers? Yes.

Problem 6. Anna Grace is 2 years old and is 33 inches tall. Her doctor tells her parents that Anna Grace is in the 24.1st percentile for height. What does this mean?

Solution:





Problem 7. Suppose that John's course has three midterms. His scores are: 82, 97, and 85.

- What is the mean?
- What is the variance?
- What is the standard deviation?
- What would John's midterm scores need to look like in order for the standard deviation to be zero?

Solution:

- What is the mean? $\bar{x} = \frac{\sum x_i}{n} = \frac{82+97+85}{3} = \frac{264}{3} = 88$
- What is the variance? $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(82-88)^2 + (97-88)^2 + (85-88)^2}{3-1} = \frac{36+81+9}{2} = \frac{126}{2} = 63 \text{ points}^2$
- What is the standard deviation? $\because s = \sqrt{s^2} \rightarrow s = \sqrt{63} = 7.937253933 \approx 7.937 \text{ points}$
- All scores exactly the same.



Problem 8. For the following data sets, describe how the mean, median, IQR and standard deviation would change between the two data sets.

a. Data Set 1: 3, 5, 6, 7, 8; **Data Set 2:** 3, 5, 6, 7, 20

b. Data Set 1: 3, 3, 3, 3, 3, 3, 3; **Data Set 2:** 3, 3, 3, 3, 3, 3, 15

Solution:

a.

Data Set 1: 3, 5, 6, 7, 8

$$\bar{x} = 5.8, M = 6$$

$$Stdev. = 1.92, IQR = 3.5$$

Data Set 2: 3, 5, 6, 7, 20

$$\bar{x} = 8.2, M = 6$$

$$Stdev. = 6.76, IQR = 9.5$$

b.

Data Set 1: 3, 3, 3, 3, 3, 3, 3

$$\bar{x} = 3, M = 3$$

$$Stdev. = 0, IQR = 0$$

Data Set 2: 3, 3, 3, 3, 3, 3, 15

$$\bar{x} = 4.71, M = 3$$

$$Stdev. = 4.54, IQR = 0$$