

## 10.4 Measures of Central Tendency and Variation

Mode-->The number that occurs most frequently; there can be more than one mode ; if each number appears equally often, then there is no mode at all.  
(mode is great for shoe size, ice cream flavor, etc.)

Median-->The middle number when the numbers are all listed in increasing (or decreasing) order. If there are an even number of numbers, then take the arithmetic average (a.k.a. Mean) of the two middle numbers.

Mean-->(a.k.a. Arithmetic average) The number you get when you add all the numbers together and divide by the number of numbers in the list. This is typically what people mean when they simply say "average." Notation for mean:  $\bar{x}$ .

Box & Whisker Plot-->A type of graph that helps view a list of numbers quickly to see their spread.

Given a list of numbers, this is how to create a box and whisker plot:

1. Put the numbers in decreasing order.
2. Find the following information.
  1. (a) The lowest number in the list.
  2. (b) The lower quartile, which is the median of the lowest half of the numbers.
  3. (c) The median of the list of numbers.
  4. (d) The upper quartile, which is the median of the largest half of the numbers.
  5. (e) The highest number in the list.
3. Plot these 5 numbers on a number line.
4. Make a box from the lower quartile to the upper quartile, and indicate the median in that box with a vertical line segment where the median occurs.
5. Connect the lowest number to the lower quartile with a line segment and connect the upper quartile to the highest number with a line segment.

The box represents where 50% of the list of numbers lie. The whiskers represent about 25% of the other numbers.

IQR (Interquartile Range)-->The difference between the upper and lower quartile. For example, if the upper quartile is 93 and the lower quartile is 76, then the IQR is  $93 - 76 = 17$ .

Outlier-->Any value of data that lies more than 1.5 IQR units below the lower quartile OR more than 1.5 IQR units above the upper quartile.

nth Percentile-->A number is in the nth percentile of some data if it is greater than or equal to n% of the data.

Since the range of data doesn't give enough information about the distribution of the data, we need to look at other attributes of the data.

Variance-->The mean of the squared differences between each number and the mean of the list of numbers. For a list of numbers,

$$x_1, x_2, x_3, \dots, x_{n-1}, x_n, \text{ the variance} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} .$$

Standard Deviation-->The square root of the variance, usually denoted  $\sigma$ , i.e.  $\sigma = \sqrt{\text{variance}}$

Z-score-->For a particular number  $x$ , compared to a list of data values, the z-score is calculated with the following formula.

$$z = \frac{x - \bar{x}}{\sigma}$$

Where  $\sigma$  = standard deviation of the list of data,

and  $\bar{x}$  = the mean of that data set.

The z-score indicates how many standard deviations  $x$  is away from the mean.

$z > 0 \implies x$  is above the mean, i.e.  $x > \bar{x}$  .  
 $z < 0 \implies x$  is below the mean, i.e.  $x < \bar{x}$  .  
 $z = 0 \implies x$  is exactly the mean, i.e.  $x = \bar{x}$  .

Distribution-->Graph of the data values (on horizontal axis) vs. relative frequency of each number in the data (on vertical axis), on a Cartesian coordinate system.

Relative Frequency-->If we have  $n$  numbers in our data list, then the relative frequency of any  $x$  in the list is  
 (the number of times  $x$  occurs in the list)/  $n$ .

Bell-shaped Distribution-->A smooth, continuous distribution whose shape resembles a bell shape. (It's like a continuous histogram, or a "histogram-gone-smooth.")

1. The median will occur at the  $x$ -value where the vertical line cuts the region under the curve into two equal-area regions.
2. The mode will occur at the  $x$ -value where the highest point on the curve exists.
3. The mean will occur at the  $x$ -value where the distribution would be balanced (like a teeter-totter).

Normal Distribution-->A special bell-shaped distribution that's symmetric and the mean=mode=median.

Shape is determined by the mean and standard deviation (the larger the standard deviation, the flatter the normal curve will be).

For a normal distribution,

~68% of the data is in the interval  $[\bar{x}-\sigma, \bar{x}+\sigma]$

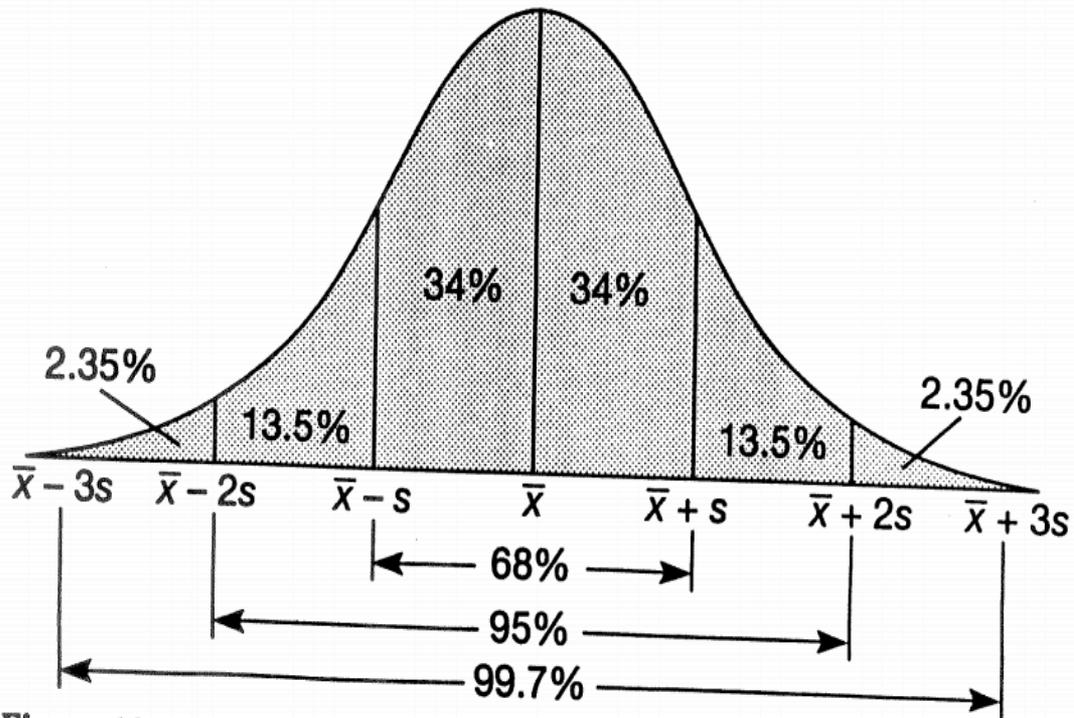
~95% of the data is in the interval  $[\bar{x}-2\sigma, \bar{x}+2\sigma]$

~99.7% of the data is in the interval  $[\bar{x}-3\sigma, \bar{x}+3\sigma]$

Mean Absolute Deviation (MAD)

Another way to measure the spread of the data—it's basically the average distance each point is away from the mean.

$$\text{MAD} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$



**Figure 10.31**

Ex 1 For 4, 6, 6, 6, 9, 11, 13, 13, 15, calculate the mean, mode, median, MAD and standard deviation.

Ex 2 Given these test scores:

	Test 1	Test 2
Andy	45	40
mean	30	25
s.d.	10	15

On which test did Andy do better?

Ex 3 If the average woman weighs 140 pounds and a weight of 162 is in the 84th percentile, what is the standard deviation?

Ex 4 Make a box and whisker plot for the following heights of children (in cm).

120, 121, 121, 124, 126, 128, 132, 134,  
140, 142, 147, 150, 152, 160

Ex 5 The average score on a test for 58 students was 87.3. Twelve more students took the test. The average for the 12 students was 90.7. What was the average for all the students?

Ex 6 Answer these conceptual questions.

(a) Could the mean of a set of scores ever be equal to the greatest or least score?

(b) When the data values are all the same, the mean, median and mode are the same. Can you create an example when the mean = median = mode, but the data values are not all the same?

(c) Suppose the average number of children per family for the employees at a university in a certain city is 2.59. Could this be a mean, median or mode?

(d) To report the average salary for Americans, would it be better to report the mean or median?

Ex 7 When a standardized test is scored, there was a mean of 400 and a standard deviation of 80. Suppose 5,000 students took the test and their scores had a normal distribution.

(a) How many scored between 320 and 480?

(b) How many scored between 240 and 560?

(c) How many scored between 160 and 640?

(d) How many scored above 640?

(e) What score is the 16th percentile? 84th percentile?

(f) If a student's score ranked 90th on this exam, what was the student's percentile rank?

(g) If Fred's score was in the 60th percentile, what was his ranking on this test?