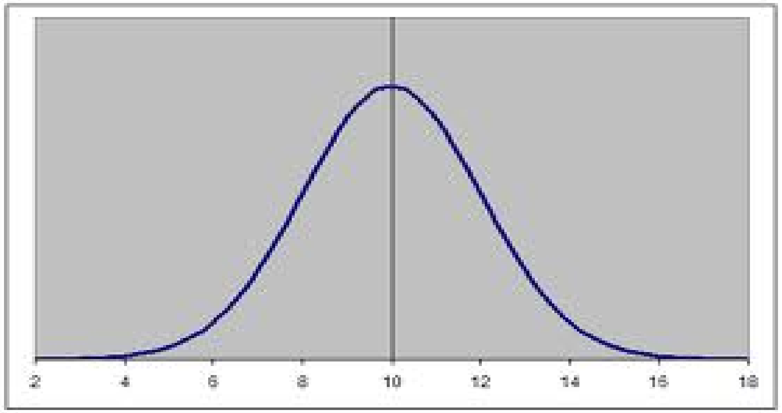
Section 6C

The Normal Distribution

The **normal distribution** is a symmetric, bell-shaped distribution with a single peak.

Its peak corresponds to the mean, median, and mode of the distribution.



When describing the shape of a normal distribution, the mean and the standard deviation should be stated.

Conditions for a Normal Distribution

A data set that satisfies the following four criteria is likely to have a nearly normal distribution:

1. Most data values are clustered near the mean.

   This gives the distribution a well-defined single peak.

2. Data values are spread evenly around the mean.

   This makes the distribution symmetric.

3. Larger deviations from the mean become increasingly rare.

   This produces the tapering tails of the distribution.

4. Individual data values result from a combination of many different factors.

The 68 – 95 – 99.7 Rule

For **any** normal distribution,

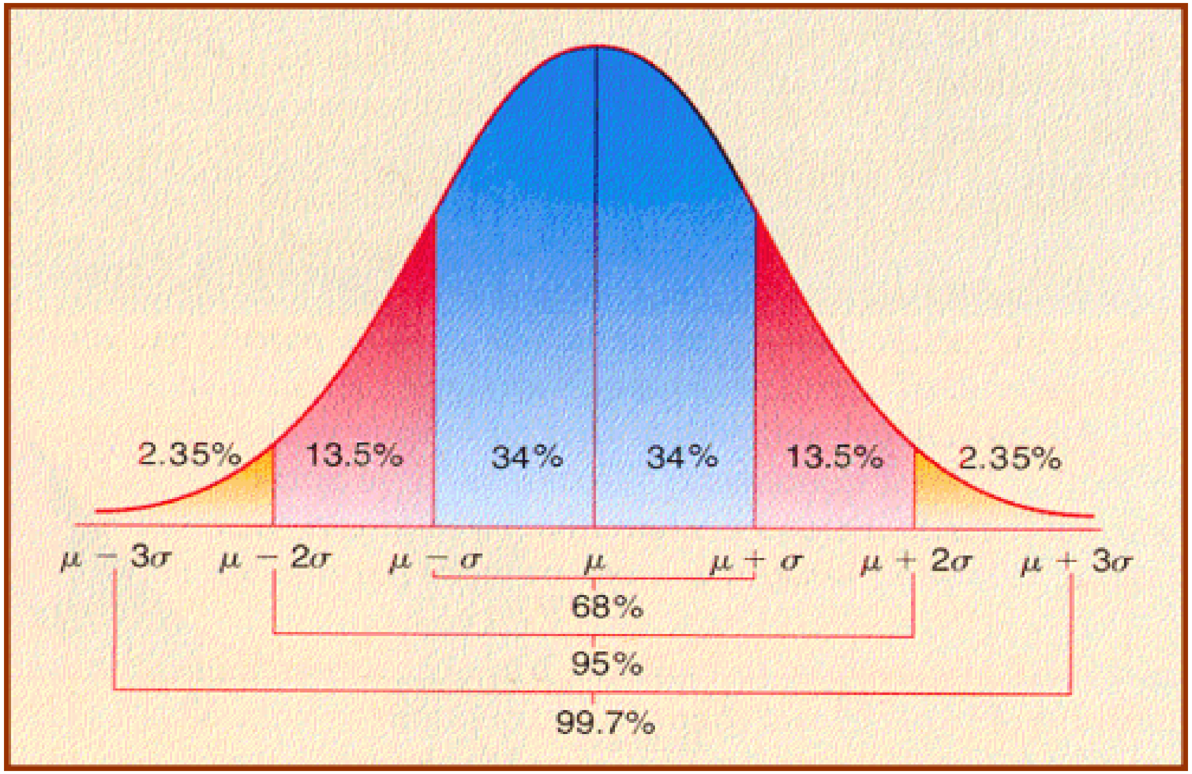
- about 68% of the data points fall within 1 standard deviation of the mean.

- about 95% of the data points fall within 2 standard deviations of the mean.

- about 99.7% of the data points fall within 3 standard deviations of the mean.

The 68 – 95 – 99.7 Rule

The standard deviation is abbreviated by the Greek letter σ (sigma).



A set of test scores is normally distributed with a mean of 100 and a standard deviation of 20. Use the 68 – 95 – 99.7 rule to find the percentage of scores in each of the following categories.

a. Less than 100

Since 100 is the mean, 50% of the data values are less than 100.

b. Less than 120

120 is one standard deviation above the mean, so there are 50% + 34% = 84% of the data values less than 120.

c. Less than 140

140 is two standard deviations above the mean, so there are 50% + 34% + 13.5% = 97.5% of the data values less than 140.

d. Between 60 and 140

60 is 2 standard deviations below the mean and 140 is 2 standard deviations above the mean. There is 95% of the data values between 60 and 140.

The 68 – 95 – 99.7 rule applies only to values that are exactly 1, 2, or 3 standard deviations from the mean.

We can generalize the rule if we know how many standard deviations from the mean a particular data value lies.

The number of standard deviations a data value lies above or below the mean is called its **standard score** (or z-score).

$$z = \text{standard score} = \frac{\text{data value} - \text{mean}}{\text{standard deviation}}$$

The standard score is positive for data values above the mean and negative for data values below the mean.

The scores were normally distributed with a mean of 67 and a standard deviation of 8. What is the standard score for an exam score of 88?

$$z = \frac{88 - 67}{8}$$

$$z = \frac{21}{8}$$

$$z = 2.6$$

Standard Scores & Percentiles

The **n^{th} percentile** of a data set is the smallest value in the set with the property that n% of the data values are *less than or equal to it*.

A data value that lies between two percentiles is said to be in the lower percentile.

We can use a standard score table (Table 6.3) to find percentiles from standard scores. The table gives the percentage of values in the distribution *less than or equal to* that value.

The body mass indices of American men between ages 30 and 50 are normally distributed with mean 26.2 and a standard deviation of 4.7. Determine the standard score and percentile of a BMI of 22.

$$z = \frac{22 - 26.2}{4.7}$$

$$z = \frac{-4.2}{4.7}$$

$$z = -0.9$$

From Table 6.3, the z-score of – 0.9 corresponds to the **18.41**