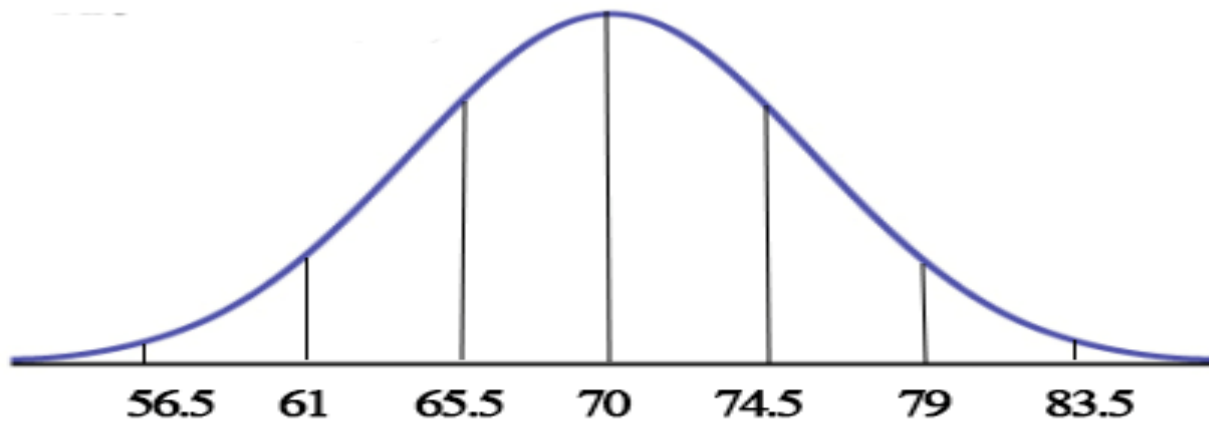
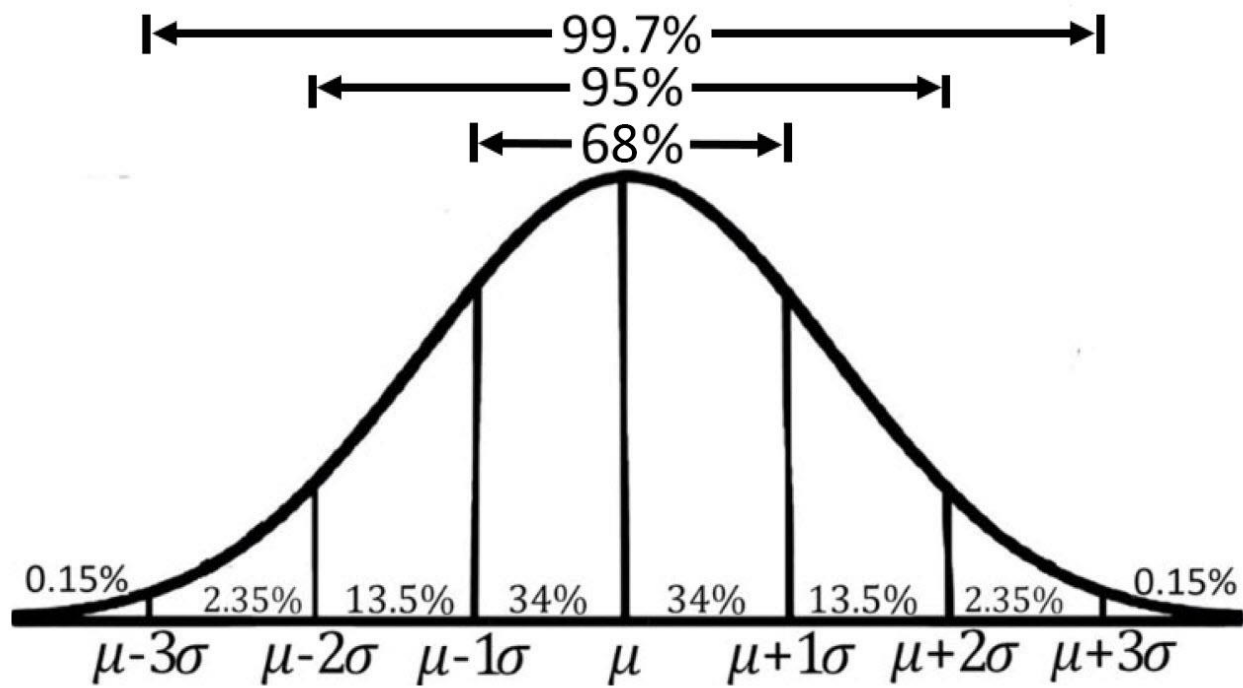
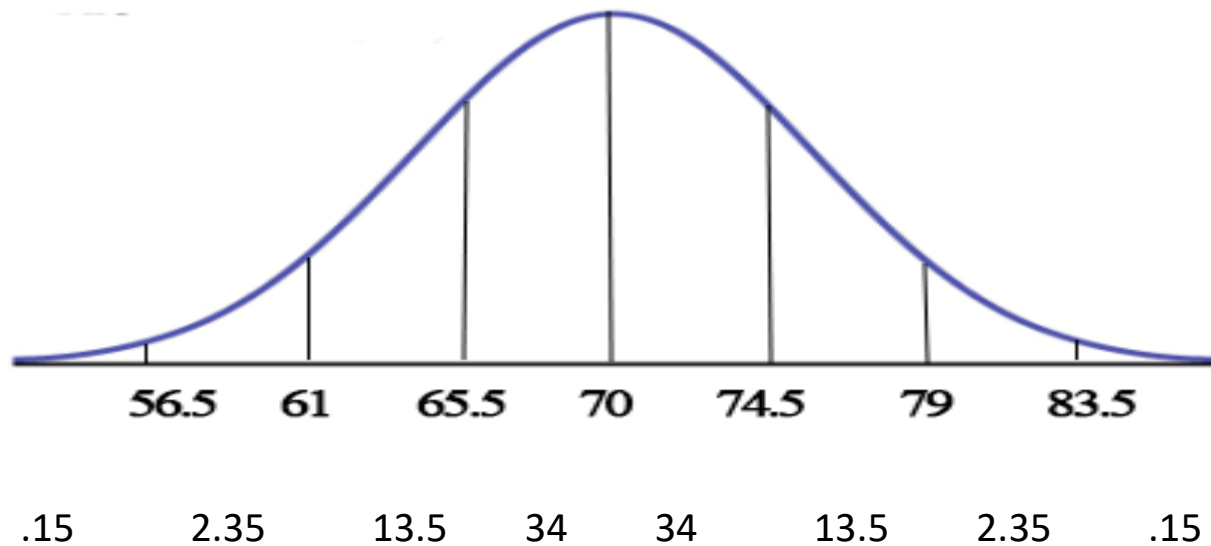


# Empirical Rule



What is the mean ?

What is the standard deviation ?



- 1) What percent of the data is between 65.5 and 74.5?
- 2) What percent of the data is between 61 and 74.5?
- 3) What percent of the data is less than 61?
- 4) What percent of the data is greater than 74.5?
- 5) What percent of the data greater than 79 and less than 56.5?
- 6) What percent of the data is greater than 82?
- 7) What percent of the data is between 68 and 73.5?
- 9) Find  $Q_1$
- 10) What value is at the 65<sup>th</sup> percentile?

## IQR and Outliers

IQR (Interquartile Range)  $Q_3 - Q_1$  Should be used for skewed data.

Using the IQR in order to determine if a data set has outliers. **(1.5 IQR rule)**

**Example 1 Use the IQR rule to determine if the data set below has outliers.**

**$S = \{12, 16, 22, 5, 18.5, 31\}$**

**$Q_1 = 13 \quad Q_3 = 21.12 \quad IQR = 21.12 - 13 = 8.12$**

**Lower Bound =  $Q_1 - 1.5(IQR) = 13 - 1.5(8.12) = .82$**

**Upper Bound =  $Q_3 + 1.5(IQR) = 21.12 + 1.5(8.12) = 33.3$**

**Is there a data value lower than .82? no !!**

**Is there a data value higher than 33.3? no !!**

**Data set S does not have outliers**

**Example 2 Use the IQR rule to determine if the data set below has outliers.**

**$K = \{4, 13, 15, 17, 20, 22, 25, 31, 38, 64\}$**

**$Q_1 = 15.5 \quad Q_3 = 29.5 \quad IQR = 29.5 - 15.5 = 14$**

**Lower Bound =  $Q_1 - 1.5(IQR) = 15.5 - 1.5(14) = -5.5$**

**Upper Bound =  $Q_3 + 1.5(IQR) = 29.5 + 1.5(14) = 50.5$**

**Is there a data value lower than -5.5? no !!**

**Is there a data value higher than 50.5? yes !!**

**Since  $64 > 50.5$ , 64 is an outlier**

Z score (How many standard deviations is a data set value from the mean?)

Example 3 A data set has an observation value of 46, a mean of 43, and a standard deviation of 4. How many standard deviations is the value 46 from the mean of 43 ?

$Z = (\text{observation} - \text{mean}) / \text{standard deviation}$

$Z = (46 - 43) / 4 = .75$

46 is .75 standard deviations above the mean.

Assuming a normal distribution, a Z score can also be used to determine if the observation value is an outlier.

If the Z score is greater than +3 or less than -3, then the associated observation value is an outlier.

Since .75 is not greater than +3 or less than -3, 46 is not an outlier for the data set.

#### Example 4

The Biology test scores from Mr. Arnold's class and Mrs Adam's class are approximately normal.

Mr. Arnold's test data: mean = 72, sd = 3.5     Mrs Adam's test data: mean = 74.5, sd = 4

John is in Mr. Arnold's class, his test score was 76; Ann is in Mrs Adam's class, her test score was 78. Which student performed better ?

Calculate the z score for John:  $(76 - 72) / 3.5 = \mathbf{1.145}$

Calculate the z score for Ann:  $(78 - 74.5) / 4 = \mathbf{.875}$

#### Conclusion:

The two students are in different classes and taught by different teachers. Since John has a higher z score, he performed better than Ann with respect to mean and standard deviation given for both testing environments.

TRUE or FALSE

- 1) The empirical rule only applies to data that is normal (bell shaped) not skewed data
- 2) If a data distribution is skewed right, the mean is less than the median
- 3) A z score cannot be 0.
- 4) The median and  $Q_2$  are the same.
- 5) The minimum value of a data set is always an outlier.
- 6) The range of a data set is found by subtracting  $Q_1$  from  $Q_3$
- 7) The mode of a data set is the value with the greatest frequency
- 8)  $Q_2$  is the same as the 50<sup>th</sup> percentile
- 9) The standard deviation of a data set is always a positive value
- 10) Possible z scores for a data set is a **discrete** set.

CIRCLE THE CORRECT ANSWER

- 11) If a data set has outliers the *mean median* should be used to summarize the data.
- 12) The statistical summary that **does not** belong to the standard five number summary for a data set is *maximum minimum  $Q_1$   $Q_3$  Median Mode*
- 13) The observation that falls in the middle of an ordered data set is the *Mean Mode Median*
- 14) A **standard normal distribution** has a mean of *0 1 -1* and a standard deviation of *0 1 -1*
- 15) The summary statistic that averages the squared deviations about the mean is the  
*standard deviation IQR mode range variance*

