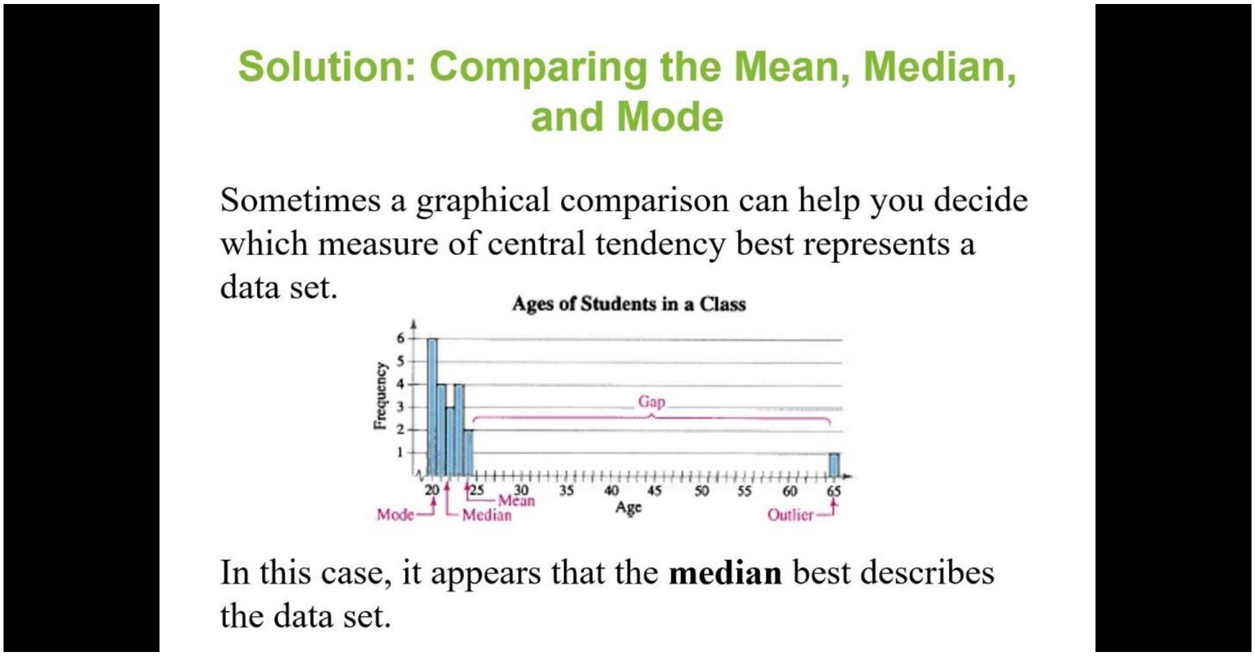
SPECIAL NOTE FOR MATH AND SCIENCE SUPPORT COURSE

MANY OF THE SLIDES IN THIS WEEK THREE SLIDE DECK **HAVE AS MUCH OR MORE TO DO WITH WEEK 5 OF THE COURSE** AS THEY HAVE TO DO WITH WEEK 3 OF THE COURSE

SO YES PLEASE STUDY THESE SLIDES CAREFULLY IN WEEK 3 BUT **PLEASE ALSO REVISIT THESE SLIDES HERE AGAIN IN WEEK 5 TOO**

THANKS FRIENDS AND VERY BEST WISHES !!!



Hi Friends and Greetings !!

As you have been finding out in your Weeks 1-3 Knewton Homework assignments, when there is a suspected outlier in the sample data, generally we shift towards preferring to use the median as a good measure of so-called central tendency.

On the other hand, if we don't have any suspected outliers in the sample data, then we very often stick with preferring to use the mean as a good measure of central tendency.

When we say that we start to talk about and learn about distributions using shape, center, and spread, then what central tendency is - it is the fancy phrase or expression for "center."

And the fancy phrase or expression for "spread" is "dispersion."

Thanks Friends and I hope that you can conclude the Weeks 1-3 Knewton Homework assignments as quickly as possible and then try to dive into the Week 4 Knewton Homework assignments as quickly as is practical too !

Good Luck and Best Wishes Friends and THANK YOU for your hard work and attention !!

:)

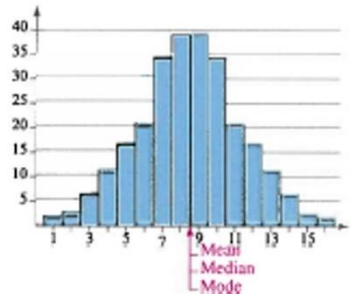
SHAPES OF DISTRIBUTIONS

The most common shapes of distributions that we encounter and learn about in this course are symmetrical / bell shaped, skewed left, and skewed right. But there are many other shapes of distributions that you might come across in the online textbook readings such as uniform, bi modal, multi modal, and a few other possible shapes too. Thanks Friends and best wishes !! Remember that the direction of the skewness, if there is any noticeable or substantial skewness present at all, is "in the direction of the longer tail" so to speak. Good Luck Friends !!   :)

The Shape of Distributions

Symmetric Distribution

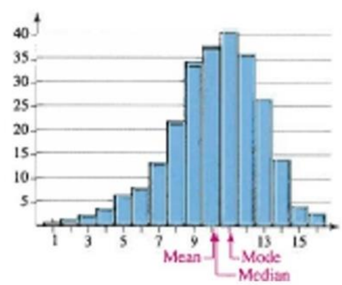
- A vertical line can be drawn through the middle of a graph of the distribution and the resulting halves are approximately mirror images.



The Shape of Distributions

Skewed Left Distribution (negatively skewed)

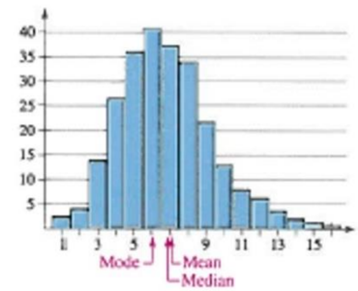
- The “tail” of the graph elongates more to the left.
- The mean is to the left of the median.



The Shape of Distributions

Skewed Right Distribution (positively skewed)

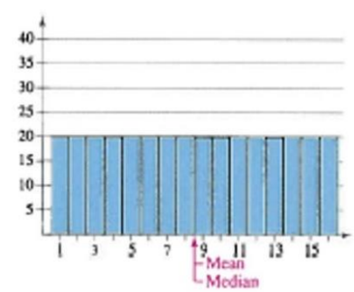
- The "tail" of the graph elongates more to the right.
- The mean is to the right of the median.



The Shape of Distributions

Uniform Distribution (rectangular)

- All entries or classes in the distribution have equal or approximately equal frequencies.
- Symmetric.



THE EMPIRICAL RULE

The Empirical Rule is also called the 68 95 99.7 Rule. You can sort of see why that would be the case by reviewing the slides below here.

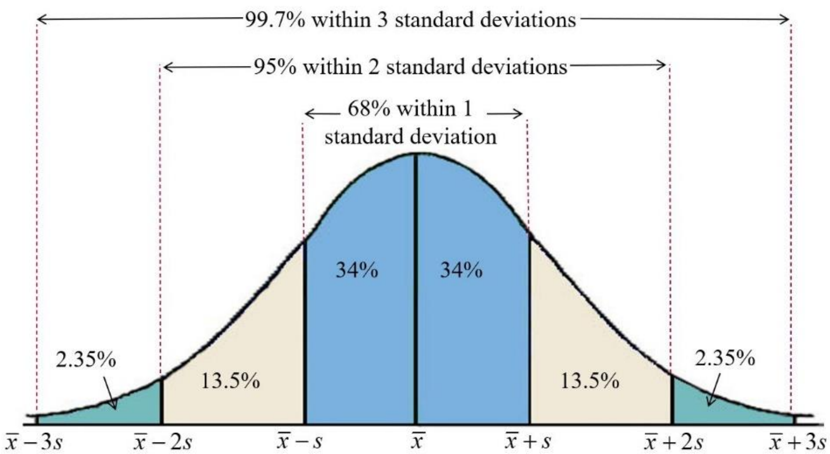
One of the most important things to keep in mind about the Empirical Rule is that it DOES NOT apply to any data set in the universe - it only properly applies to certain data sets that satisfy certain conditions or criteria. But when the Empirical Rule does properly apply to a particular data set, then the information on the following slides can be used as a nice aid and tool while working out the solutions and answers to certain Empirical Rule "related" problems and exercises. Thanks Friends and Best Wishes !!   :)

Interpreting Standard Deviation: Empirical Rule (68 – 95 – 99.7 Rule)

For data with a (symmetric) bell-shaped distribution, the standard deviation has the following characteristics:

- About **68%** of the data lie within one standard deviation of the mean.
- About **95%** of the data lie within two standard deviations of the mean.
- About **99.7%** of the data lie within three standard deviations of the mean.

Interpreting Standard Deviation: Empirical Rule (68 – 95 – 99.7 Rule)

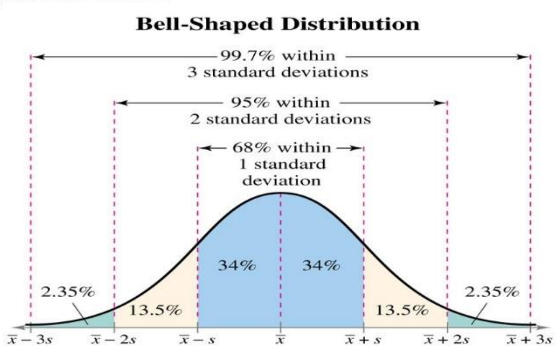


Example: Using the Empirical Rule

In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20-29) was 64.3 inches, with a sample standard deviation of 2.62 inches. Estimate the percent of the women whose heights are between 59.06 inches and 64.3 inches.

Solution: Using the Empirical Rule

- Because the distribution is bell-shaped, you can use the Empirical Rule.



34% + 13.5% = 47.5% of women are between 59.06 and 64.3 inches tall.

The way the answer was worked out for the example problem / exercise here was as follows:

We need the numerical labels for the seven "notches" or "tick marks" on the horizontal axis.

The middle of these seven notches or tick marks is x = 64.3 (which is the mean here)

Then we do these calculations to find the other six numerical labels for the tick marks:

64.3 - 2.62 - 2.62 - 2.62 = 56.44

64.3 - 2.62 - 2.62 = 59.06

64.3 - 2.62 = 61.68

64.3 + 2.62 = 66.92

64.3 + 2.62 + 2.62 = 69.54

64.3 + 2.62 + 2.62 + 2.62 = 72.16

So that the labels (and these are x values and NOT z values) for the seven tick marks from left to right in the picture then are:

56.44   59.06   61.68   64.3   66.92   69.54   72.16

And so when answering the question having to do with the percent of women whose heights are between 59.06 inches and 64.3 inches, what we see from the picture in the last slide above is that 13.5% occurs between x = 59.06 and x = 61.68 and that another 34% occurs between x = 61.68 and x = 64.3, and so that therefore the total percentage occurring between x = 59.06 and x = 64.3 is 13.5% + 34% = 47.5%

Thanks Friends and Be Well this Week 3 !!

:)

Z-SCORES, UNUSUAL OBSERVATIONS, AND VERY UNUSUAL OBSERVATIONS

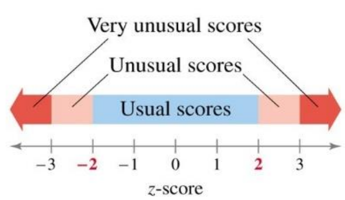
We explore potential / possible outliers in a data set using a concept involving the so-called fences in connection with a box and whiskers plot during Week 3. Concepts such as "unusual observations" and "very unusual observations" are separate from the concept of potential outlier and so I wanted to provide these slides here to try to reinforce what the situation and case is for how we identify and designate "unusual observations" and "very unusual observations" in a set of sample data. We will study and learn about z-scores in more detail during the upcoming Week 5 of the course. But we can please start to get acquainted with them here though. Thanks Friends and Enjoy Week 3 !! :)

The Standard Score

Standard Score (*z*-score)

- Represents the number of standard deviations a given value x falls from the mean μ.
- $z = \dfrac{\text{value - mean}}{\text{standard deviation}} = \dfrac{x-\mu}{\sigma}$

Solution: Comparing *z*-Scores from Different Data Sets



Both z-scores fall between –2 and 2, so neither score would be considered unusual. Compared with other Best Supporting Actor winners, Heath Ledger was relatively younger, whereas the age of Penelope Cruz was only slightly lower than the average age of other Best Supporting Actress winners.

Please ignore the written commentary at the bottom of this slide here. My reason for sharing this slide with you was just so that you could get a look at the graphic with the blue and pink and red areas / regions. Thanks Friends !! :)