Name ______________________________

6A: Characterizing Data

<u>Class Prep Assignment</u> <u>Due at the beginning of next class</u>

We Describe Four Characteristics of Data: <u>Shape</u>, <u>Center and Spread</u>, and <u>Outliers</u>

Example. The grades on the third exam for a MTH 95 class were as follows:
82 74 67 81 49 84 52 91 66 75 96 73 71 78 49 86 85 62 58

a) Make a histogram of the data to determine its shape.

Shape of the Histogram:

Unimodal

Bimodal

Multimodal

Symmetric

Skewed to the Left
(Mean less than median)

Skewed to the Right
(Mean greater than median)

Measures of Center or Average

Mean:

Median:

odd number of values:
even number of values:

Mode:

b) Arrange the grades above in order:

c) Find the mean

d) Find the median

e) Find the mode(s), if any

 Name ________________________________

6B: Measures of Variation

Class Prep Assignment | Due at the beginning of next class

Measures of Spread

Range:

Interquartile Range (IQR):

Standard Deviation:

Five-Number Summary and Boxplot: Minimum, Q_1, Median, Q_3, Maximum

Continuing with the test scores in order, find the following:

49, 49, 52, 58, 62, 66, 67, 71, 73, 74, 75, 78, 81, 82, 84, 85, 86, 91, 96

f) Five-number summary:

g) Range:

h) Interquartile Range (IQR):

i) Draw and label the boxplot:

Outliers

j) Are there any outliers in this data?

Which Measures to Use?
If the data is symmetric, use the mean and standard deviation
If the data is skewed, use the median and the IQR

 Name ____________________________

6B: Standard Deviation

Standard Deviation The "average deviation from the mean." Can be approximated by the Range÷4 if the data is evenly spread without outliers.

$$s = \sqrt{\frac{\sum (x - mean)^2}{n-1}}$$

49, 49, 52, 58, 62, 66, 67, 71, 73, 74, 75, 78, 81, 82, 84, 85, 86, 91, 96

Data	Deviation from Mean	Squared Deviation
49		
49		
52		
58		
62		
66		
67		
71		
73		
74		
75		
78		
81		
82		
84		
85		
86		
91		
96		
Sum of the squared deviations:		

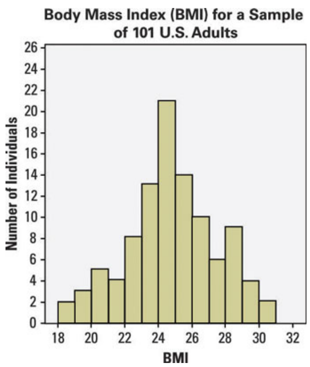
$$s = \sqrt{\frac{\sum (x - mean)^2}{n-1}} =$$

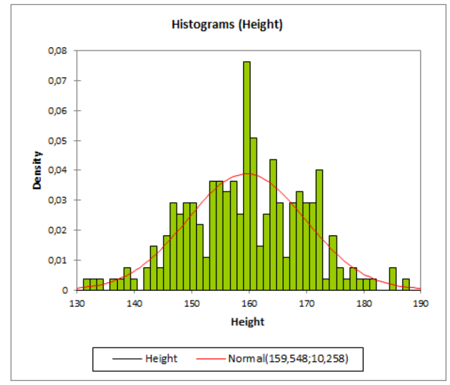
Standard Deviation Approximation: Range/4. How do they compare in this case?

6C: The Normal Distribution

The Normal Distribution – The bell-shaped curve

Used when the data is unimodal and approximately symmetric (mean = median)

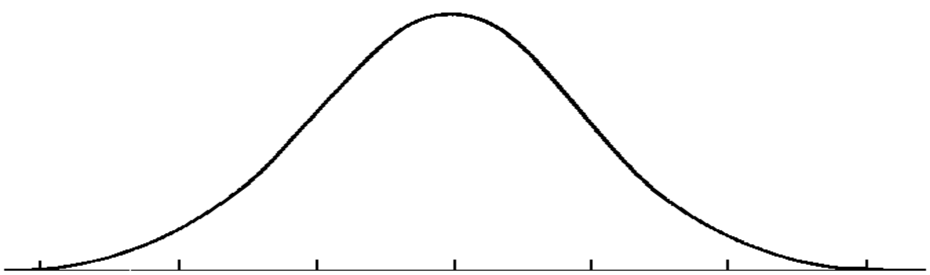




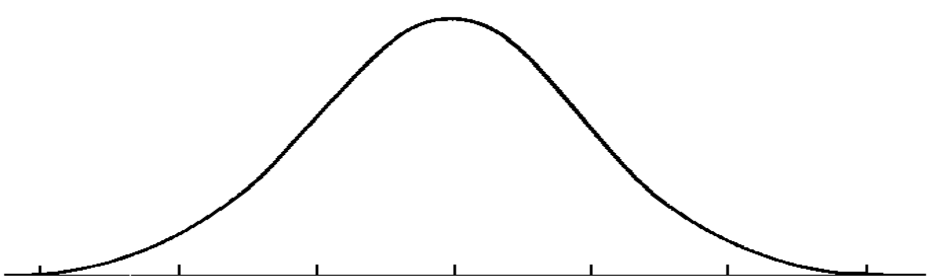
Sources: http://www.dummies.com/education/math/statistics/interpreting-histograms/
https://learnandteachstatistics.wordpress.com/2012/11/12/beware-of-excel-histograms/
https://www.xlstat.com/en/solutions/features/histograms

How to label a Normal Distribution - The standard deviation is the scale

Example. Heights of 10-year-olds of all genders closely follow a normal distribution with a mean of 55 inches and a standard deviation of 6 inches. Label the normal curve.



The 68-95-99.7 Rule for a Normal Distribution (Empirical Rule)



 Name ______________________________

Calculating Probabilities with the Empirical Rule

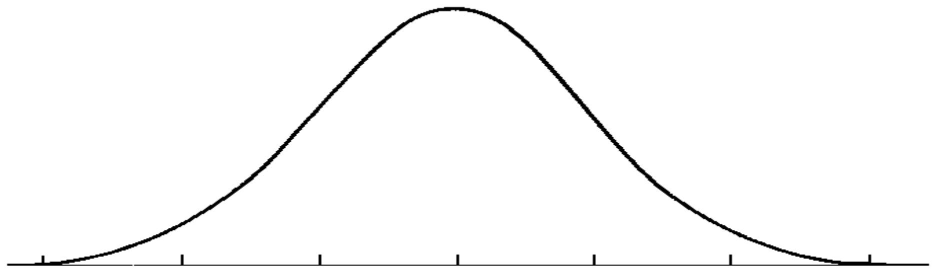
Example Continued. Find the probability that a randomly selected 10-year-old is:

a. between 49 and 61 inches

b. between 55 and 61 inches

c. greater than 61 inches

d. 37 inches or less



Z-Scores (Standard Scores)

The number of standard deviations that a value is away from the mean.

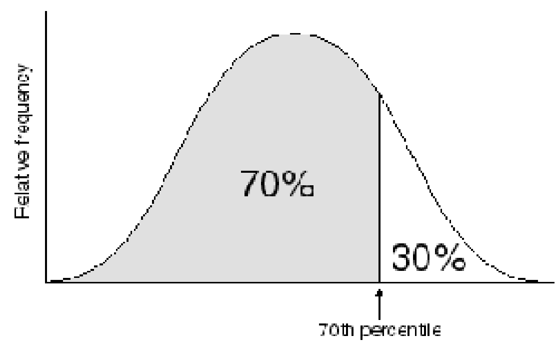
$$Z = \frac{x - mean}{s}$$

e. Find the Z-score for a 10-year-old height of 52 inches and give its interpretation.

Percentile

The x^{th} percentile is the value that x% of the data values are below.

f. A 10-year-old is in the 86^{th} percentile in height. This means the child is taller than ______ % of 10-year-olds.



Use the Empirical Rule to Find Percentiles

Find the corresponding percentiles for the Z-scores.

Z-score	Percentile
-3	
-2	
-1	
0	
1	
2	
3