Math-3
Lesson 7-1

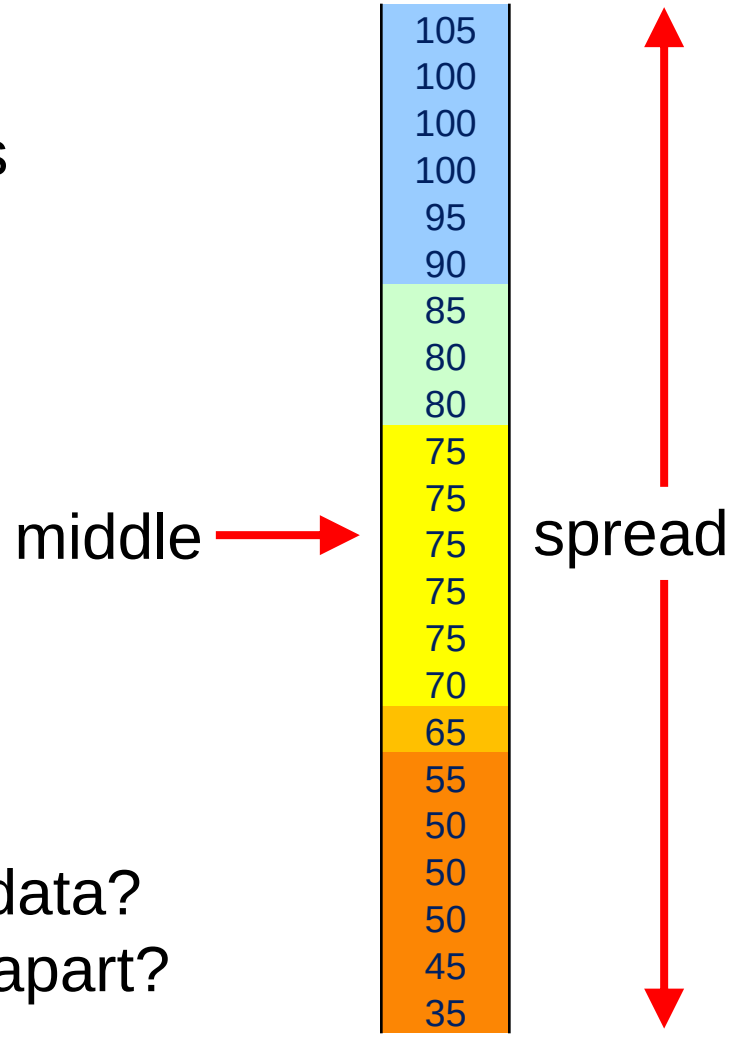
Review of Statistics
And
“Standard Deviation”

Statistics: The numerical values used to summarize and compare sets of data.

Data: A collection of measurements of a certain type:

Examples:

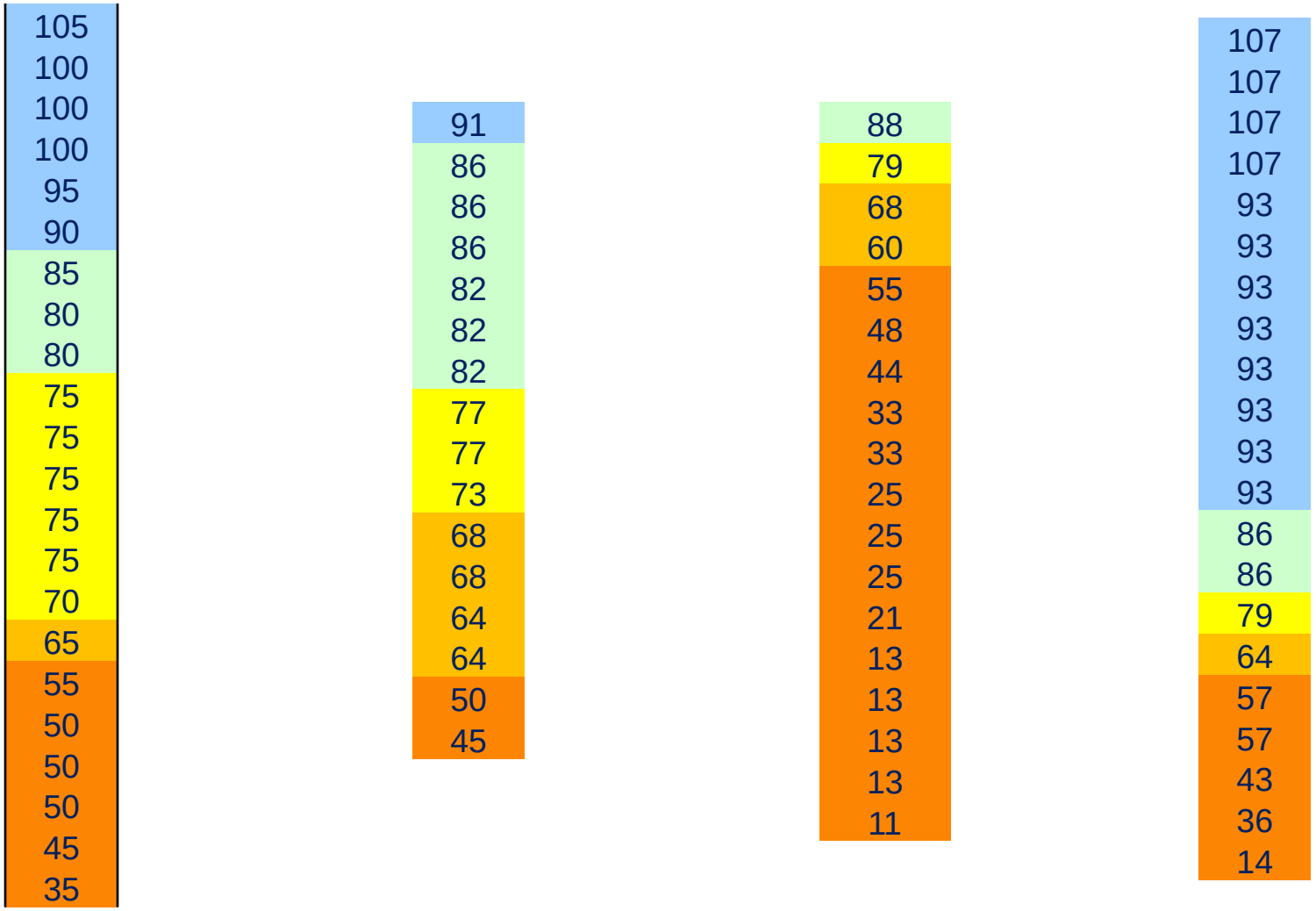
- Income
- Age
- Unit 6 Test Scores



Categories of Statistics:

- where is the middle of the data?
- how far is the data spread apart?

How can we find the “middle” of the data?



Measure of Central Tendency: a number used to represent the "center" or "middle" of the data set.

Mean is a measure of central tendency.

Mean: what you would normally call the "average".

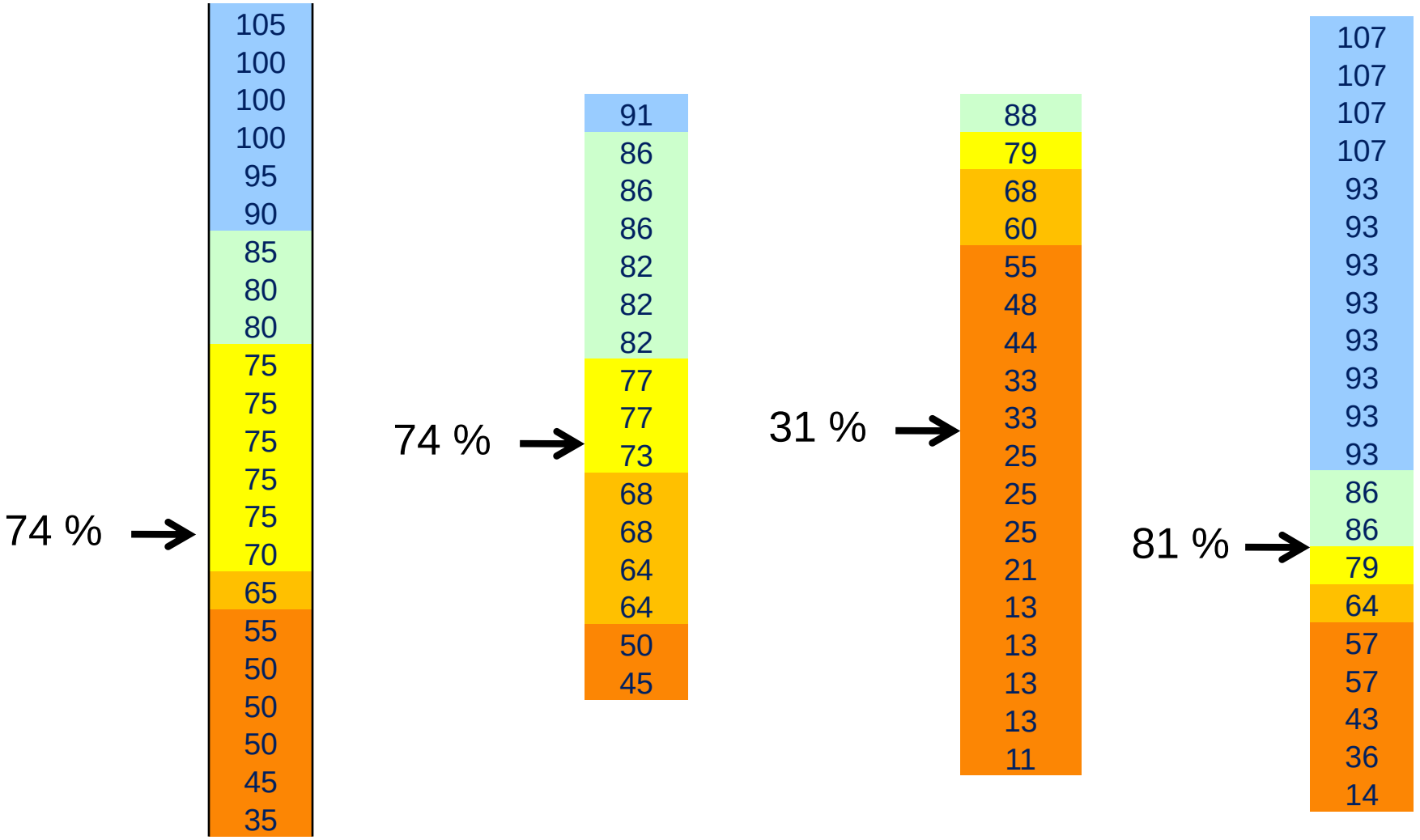
Add all the data together then divide by the number of data points.

$$Mean = \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

"x-bar"

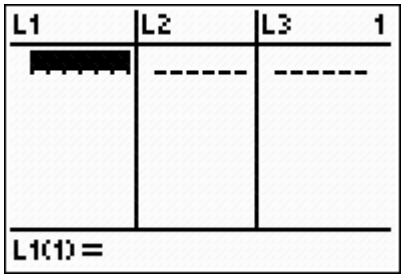
Grades for 4 different tests.

Mean (average) grade for each:

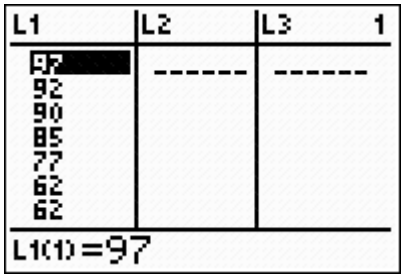


Let's use the calculator

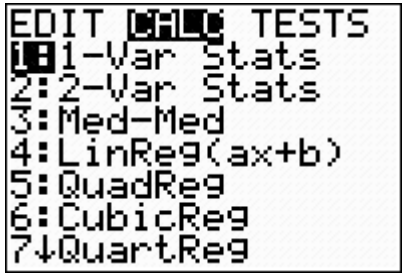
Enter the data into a list
→ "Stat" then "edit" option.



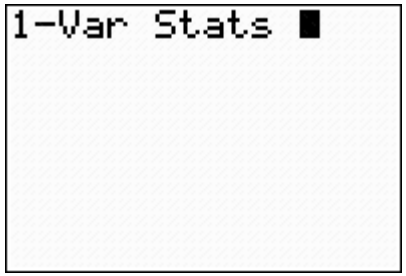
Enter the data into list "L1".



"stat" then scroll over to "calc"



Option 1: 1-var stats then "enter"

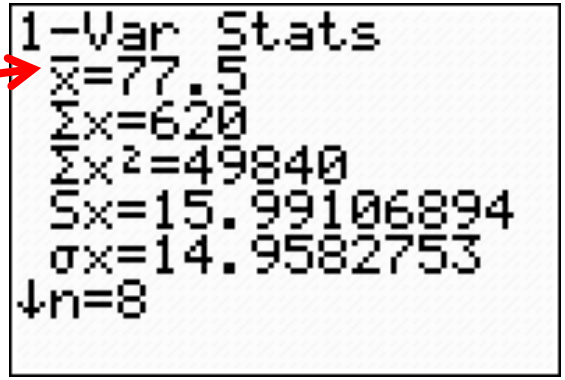


Find the mean of the following data.

{ 2, 3, 5, 7, 9, 11}

$$\bar{x} = \frac{2+3+5+7+9+11}{6} = 6.17$$

Mean:



Median: the number that is the middle number of the data set.

median: half of the data points are above this value and half are below this value.

Odd number of data: 4, 6, 8, 10, 12

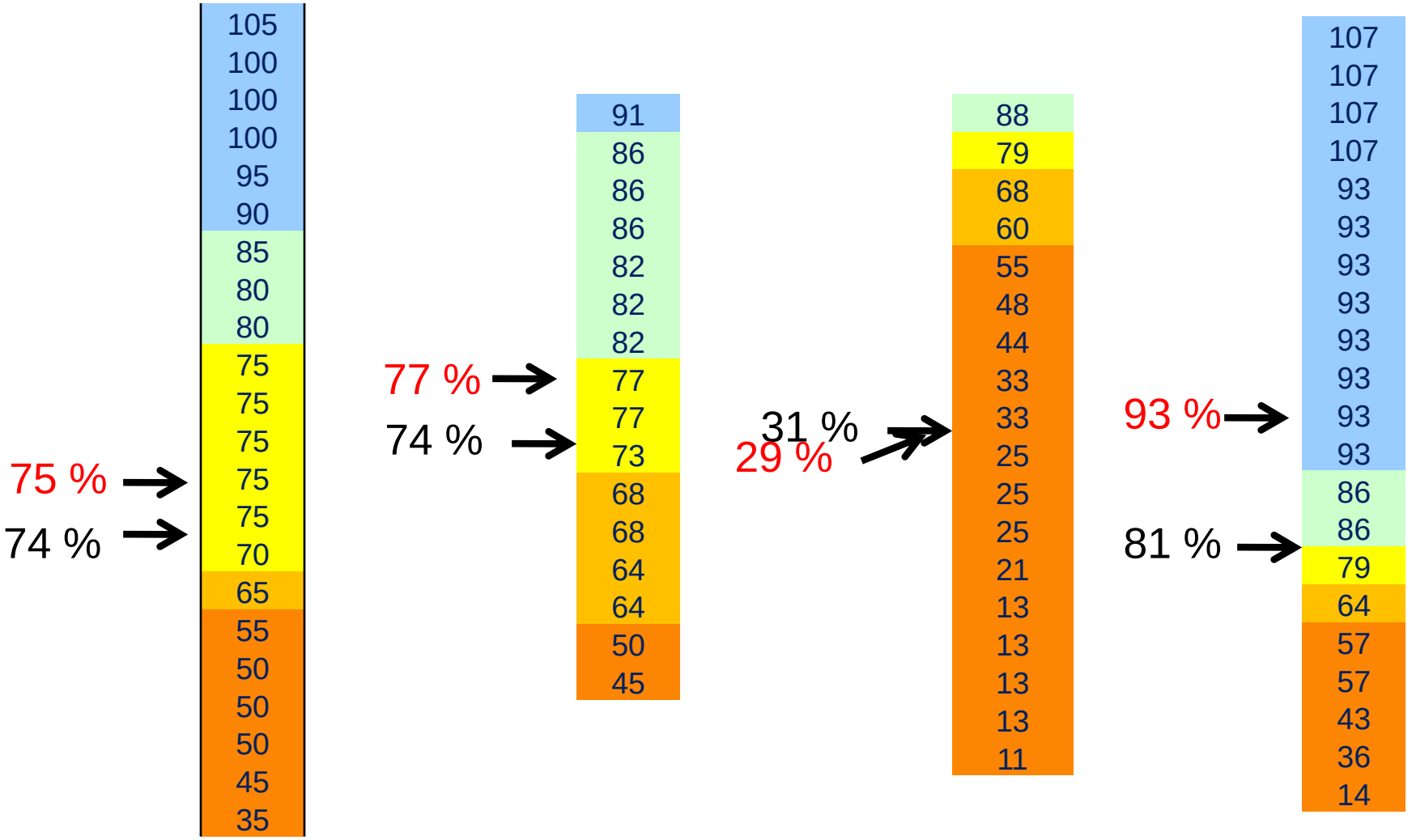
Even number of data: 3, 4, 6, 8, 10, 12

For an even number of data, take the mean of the numbers above and below the middle position.

→ median = 7

Average grade for each:

Median grade for each:



Mode: the number in the data set that occurs most frequently.

Data set: 1, 2, 4, 4, 6, 8, 8, 8, 10, 12

Frequency of occurrence: 4 (occurs 2 times),
8 (occurs 3 times),
all the rest (occur only once)

Mode = 8

Data set: 1, 2, 4, 4, 6, 8, 8, 10, 12

Frequency of occurrence: 4 (occurs 2 times),
8 (occurs 2 times),
all the rest (occur only once)

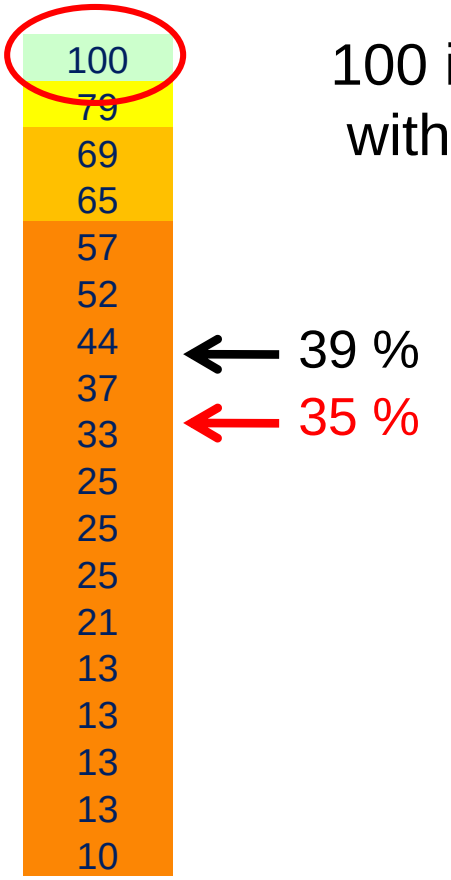
Mode = 4 and 8

"Bi-modal"

Vocabulary

Outlier: A data point that is much greater or much lower than most of the other data points.

Outliers tend to give misleading impression about a data set.



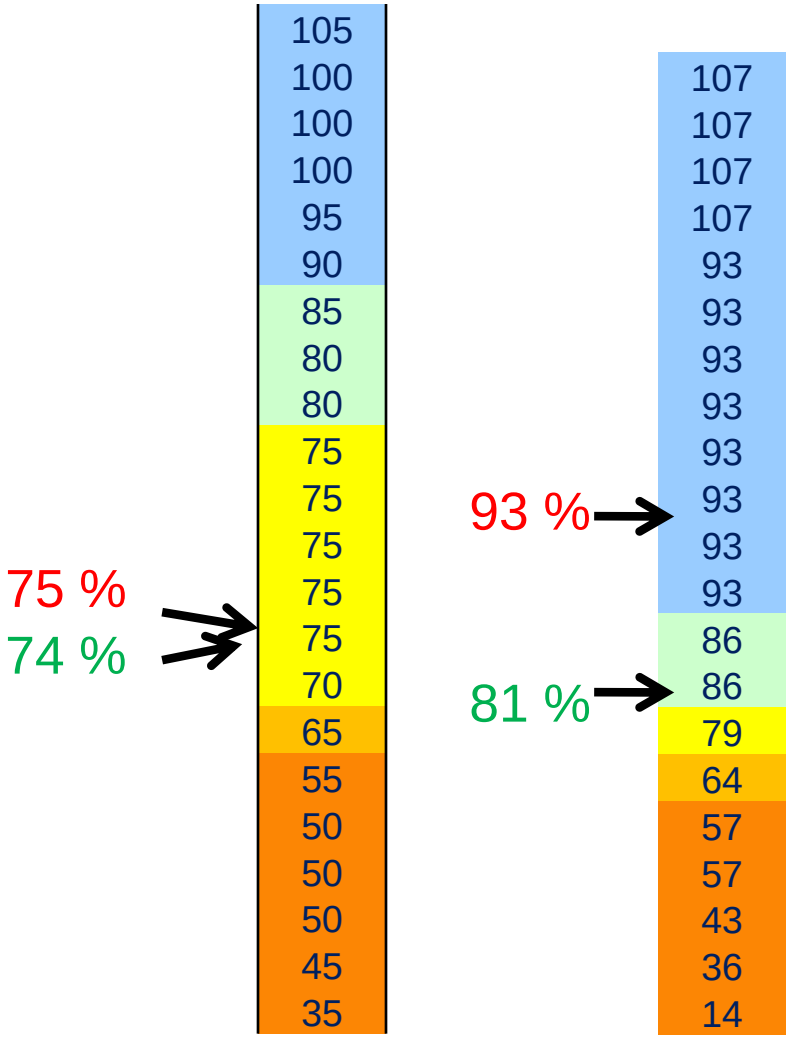
100 is 21 points above 79. All other points are within 10 points of the adjacent data point.

If we "throw out" the outlier (it not being a representative grade for this test/group) see how the mean is effected.

Average grade:

Median grade:

When you compare the mean with the median, you can see if the data is “skewed”



The mean is very sensitive to outliers (as it factors in their magnitude),

while the median is resistant to outliers."

Measure of Central Tendency: a statistic used to represent the “center” or “middle” of the data set.

mean — the average of the data measurements.

the difference between the greatest and least data point.

median — the middle number in the data set.

mode — the data point that occurs most frequently in the data set.

Measure of spread

Range: the difference between the greatest and least data point.

$Range = (105 - 35) = 70$

105
100
100
100
95
90
85
80
80
75
75
75
75
75
70
65
55
50
50
50
45
35

$Range = (91 - 45) = 46$

91
86
86
86
82
82
82
77
77
73
68
68
64
64
50
45

$Range = 77$

88
79
68
60
55
48
44
33
33
25
25
25
21
13
13
13
13
11

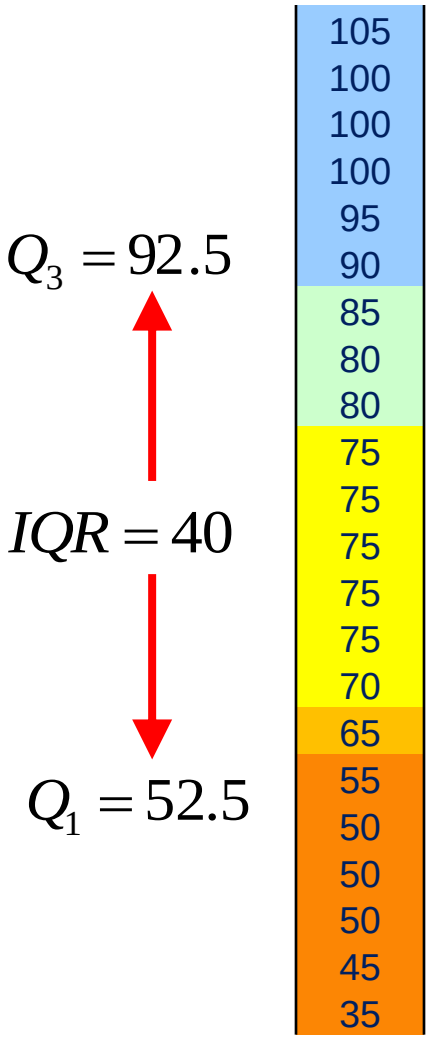
$Range = 93$

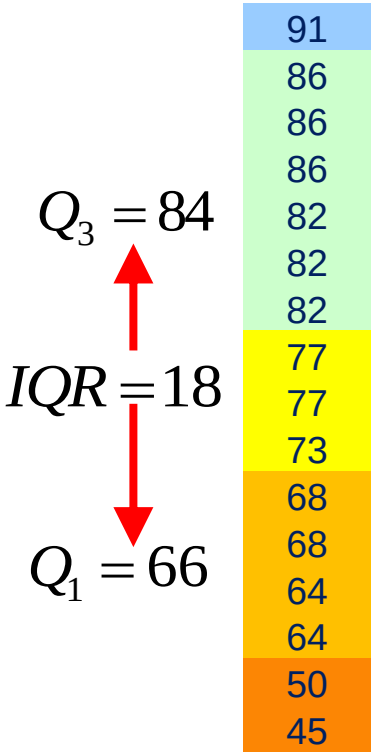
107
107
107
107
93
93
93
93
93
93
93
93
86
86
79
64
57
57
43
36
14

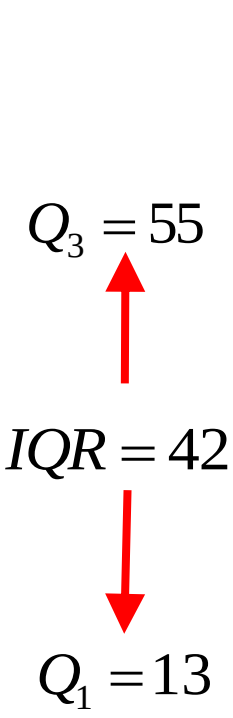
Q1: The median of the bottom ½ of the data

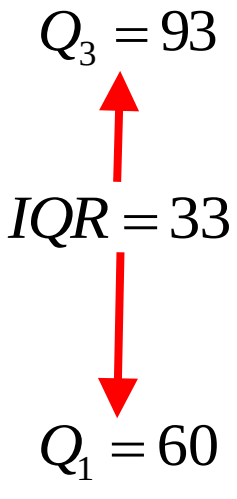
Q3: The median of the top ½ of the data

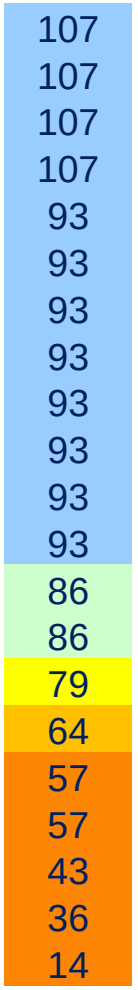
Inter-quartile Range: Q3 – Q1



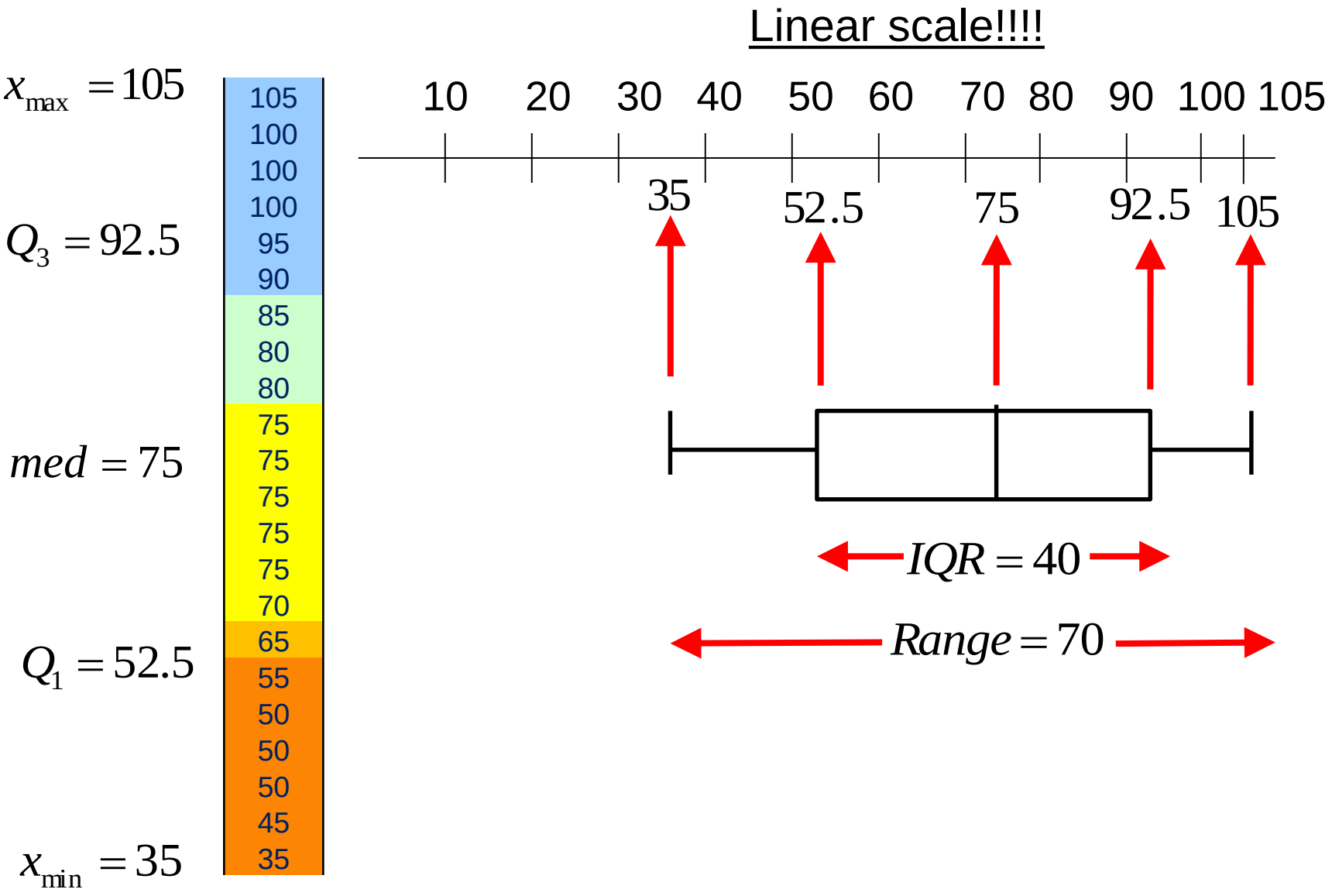






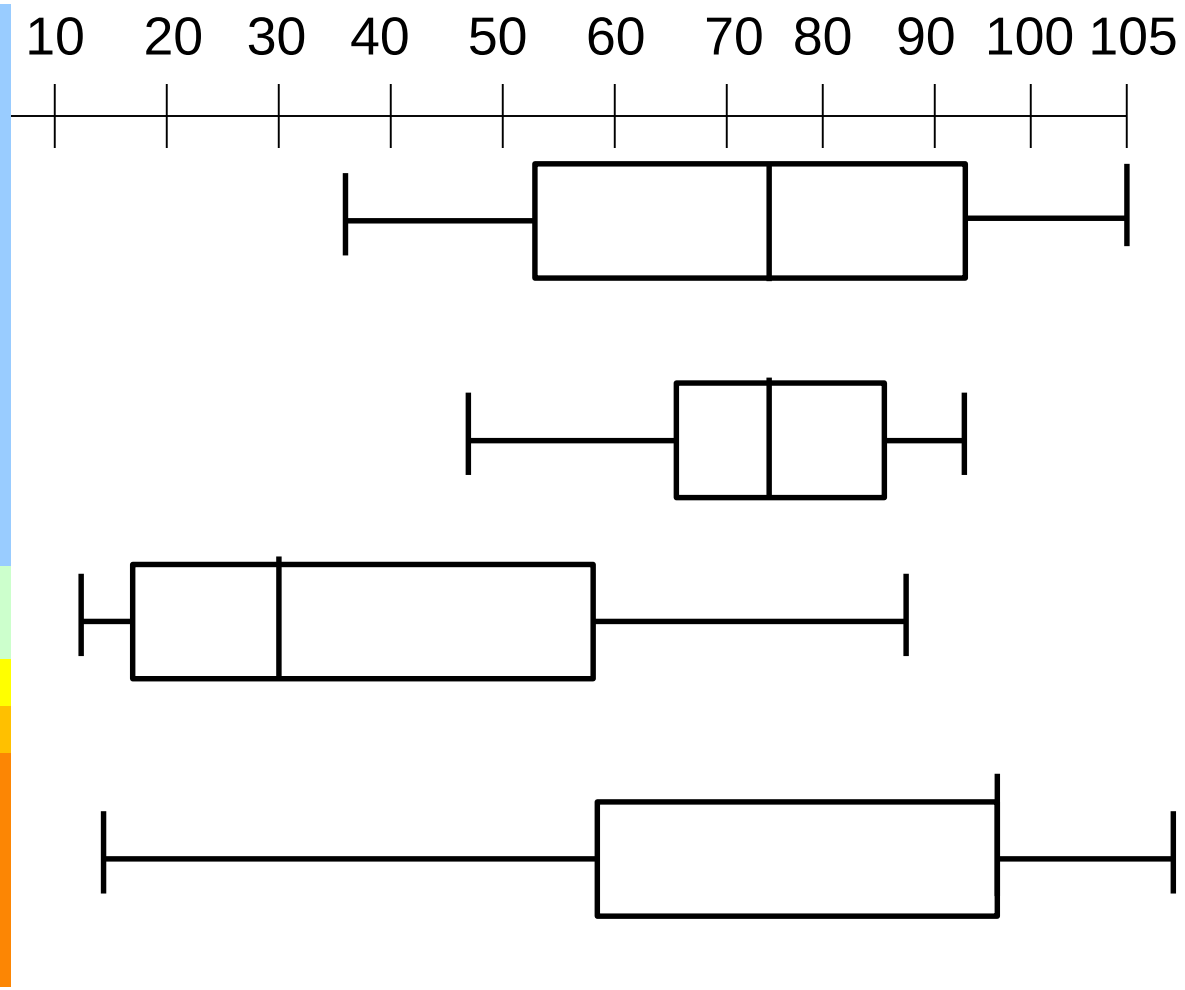


Box and Whisker Plot: a graphical representation of Min data point, Q1, median, Q3, max data point.

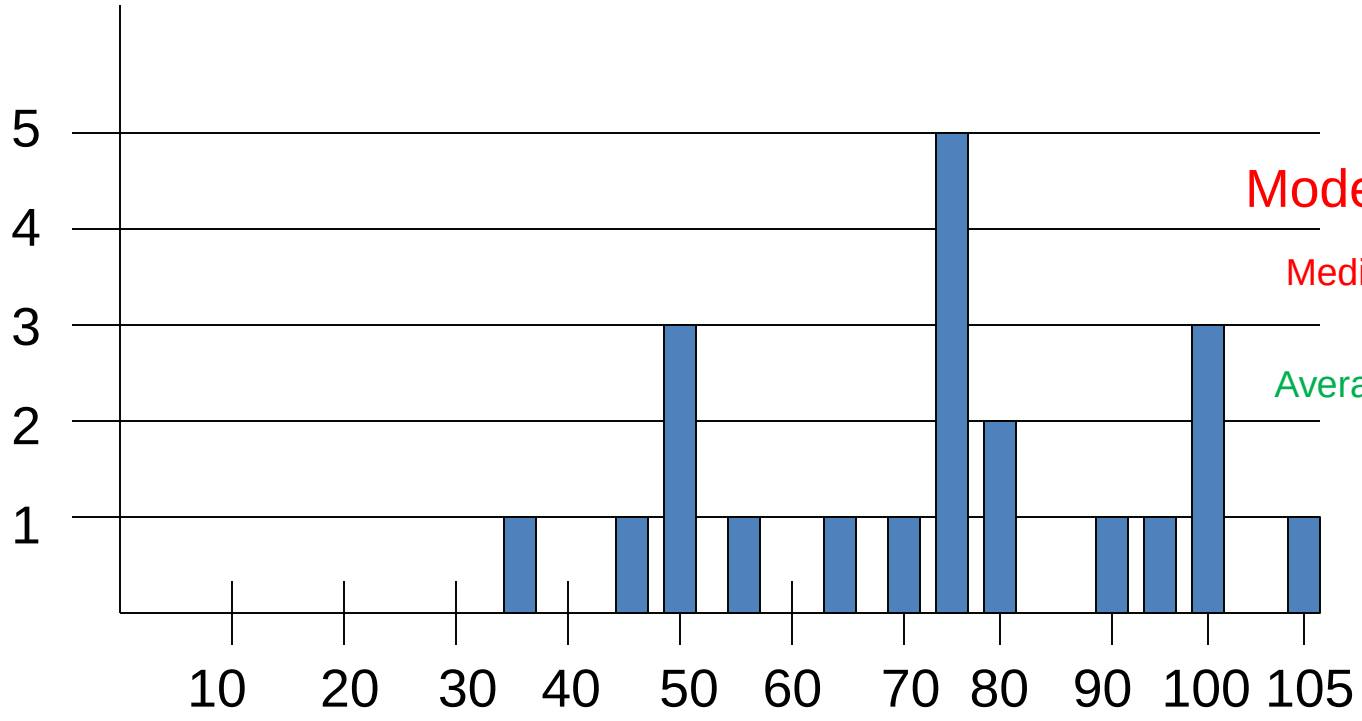


Box and Whisker Plot: Help us to compare data visually.

105	91	88	107
100	86	79	107
100	86	68	107
100	86	60	93
95	82	55	93
90	82	48	93
85	82	44	93
80	77	33	93
80	77	33	93
75	73	25	93
75	68	25	93
75	68	25	86
75	64	21	86
75	64	13	79
70	50	13	64
65	45	13	57
55		13	57
50		11	43
50			36
50			14
45			
35			



Frequency Distribution graph:



Mode = 75%

Median 75 %

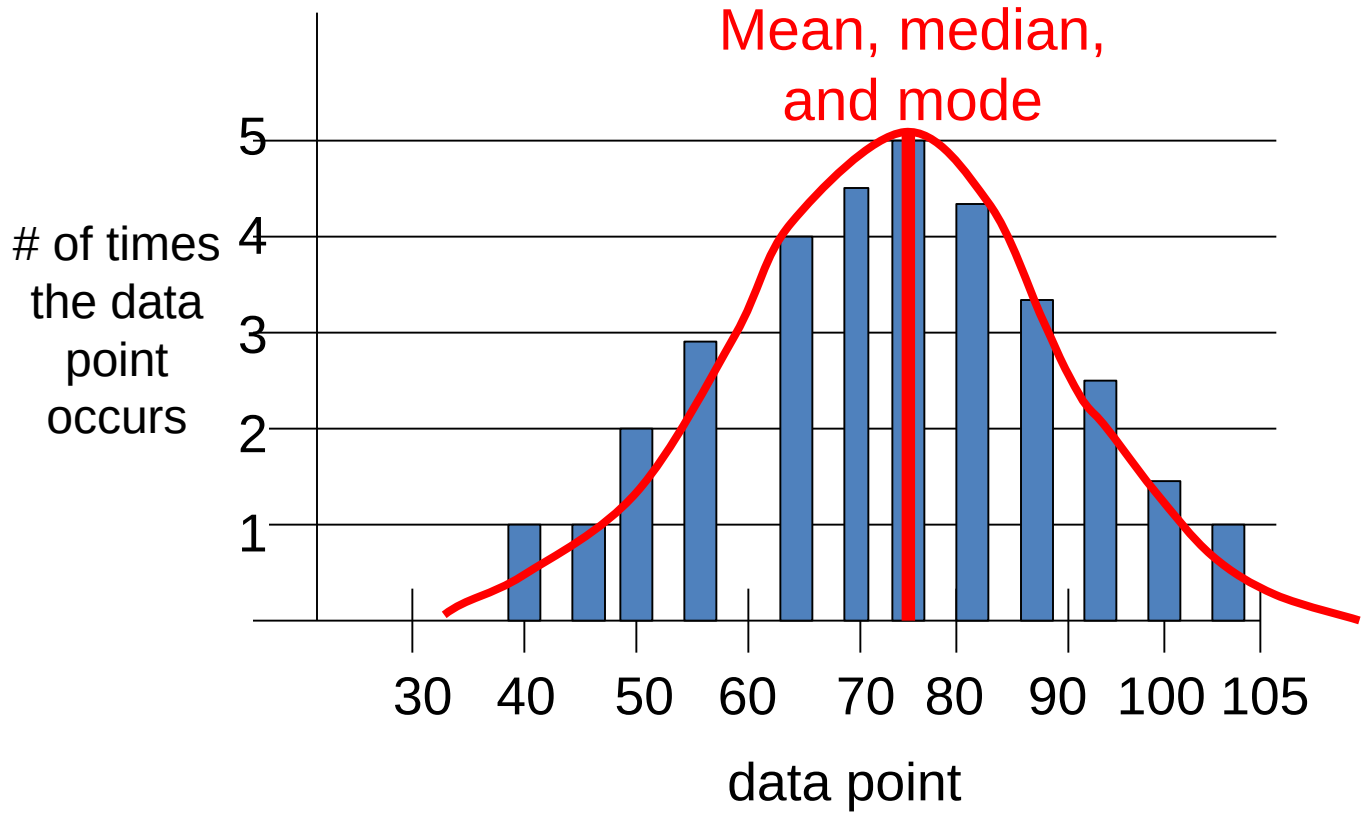
Average 74 %

105
100
100
100
95
90
85
80
80
75
75
75
75
75
70
65
55
50
50
50
45
35

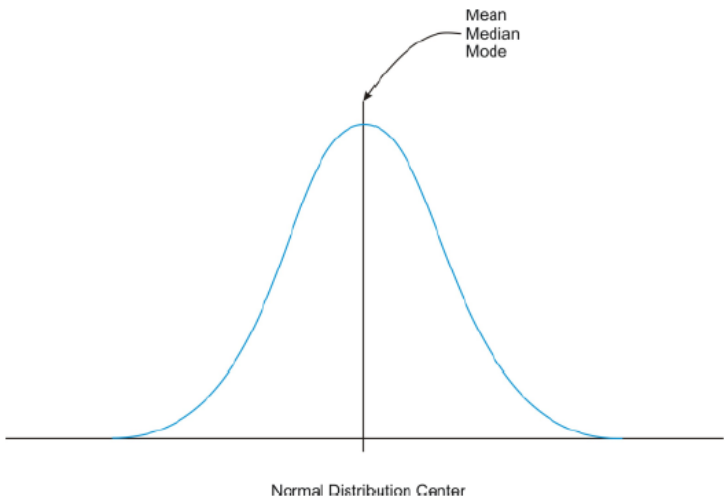
# of occurrences	1	1	3	1	1	1	5	2	1	1	1	3	1
Grade	35	45	50	55	65	70	75	80	85	90	95	100	105

Data Distribution

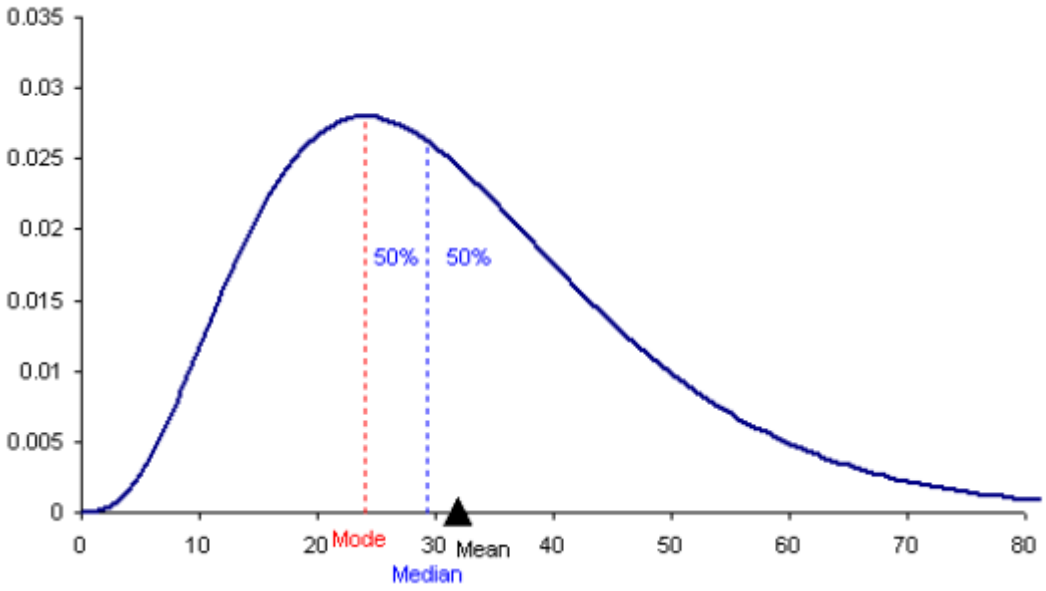
Bell curve: general shape of a frequency distribution curve that is "normally distributed" (when you have a lot of data).



Normal



Not Normal

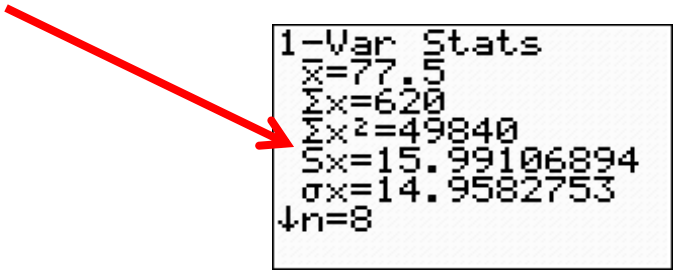


Standard Deviation

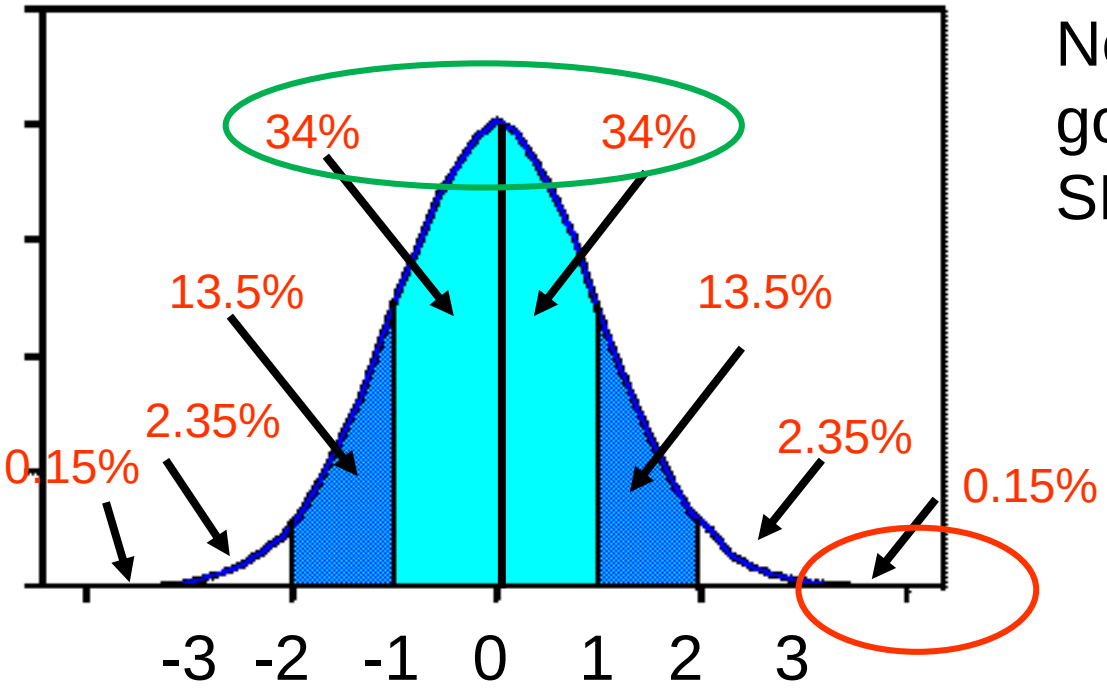
Standard deviation: a measurement of spread of the data from the mean. The calculator does this for you.

$$S = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}}$$

This gives the sdev of the data "sample".



To build the Normal Distribution Graph, we start off with the standard scale. The x-axis scale is labeled with #'s of standard deviations from the mean.



Notice: the scale only goes from -3 to +3 SDEV from the mean.

The portion of the data that falls within each region is labeled.

Only 0.15% of the data is greater than 3 sdev above the mean.

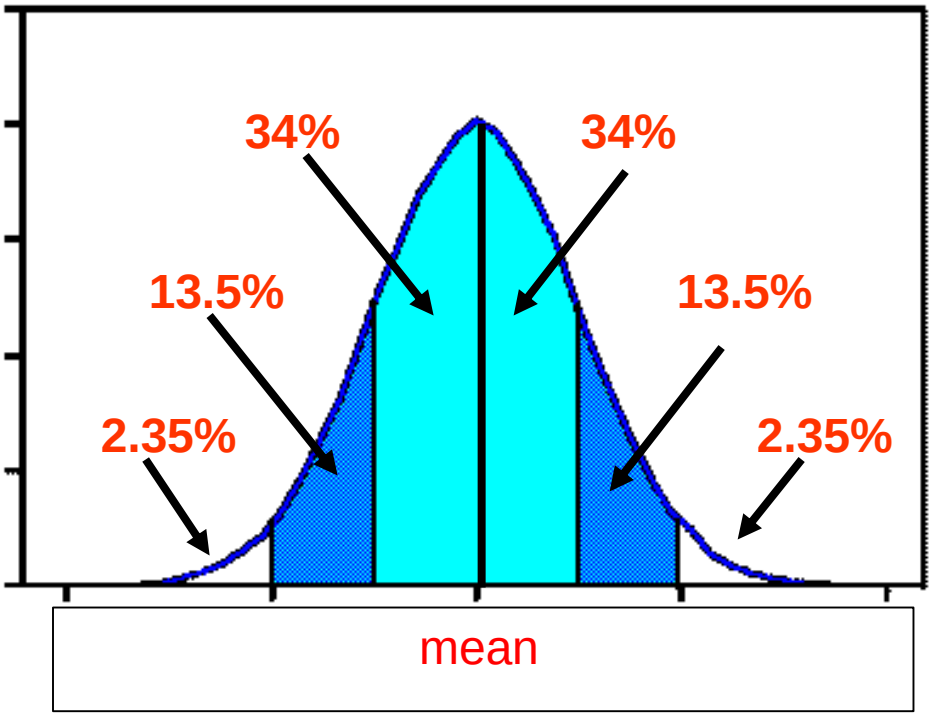
68% of the data falls between -1 sdev and +1 sdev of the mean.

Standard deviation a number that describes the spread of the data.

Standard deviation 68% of the data will be within one standard deviation of the mean.

probability of a data point being within two standard deviations of the mean.

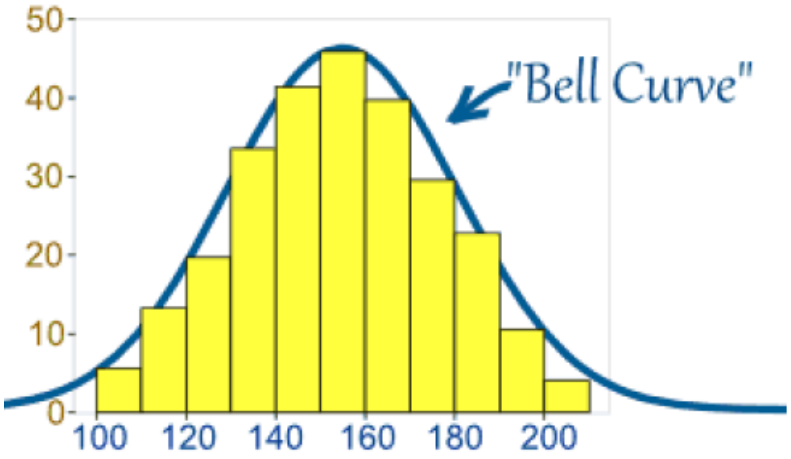
= 13.5 + 13.5 + 34 + 34 = 95%



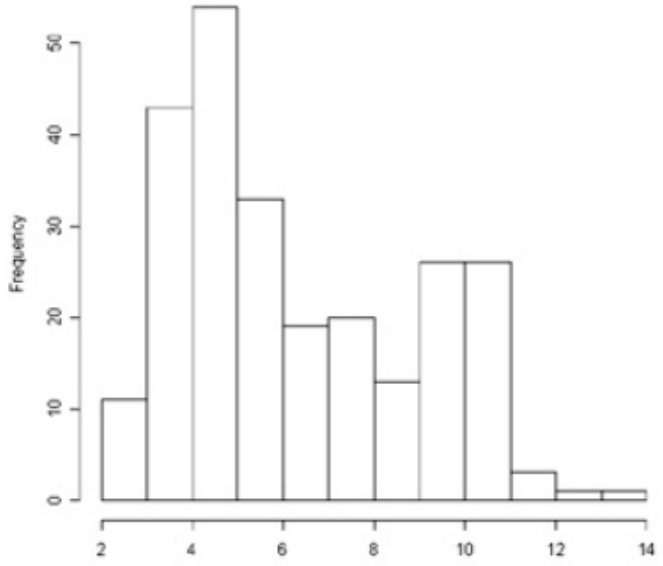
probability of a data point being within three standard deviations of the mean.

= 68 + 27 + 4.7 = 99.7 %

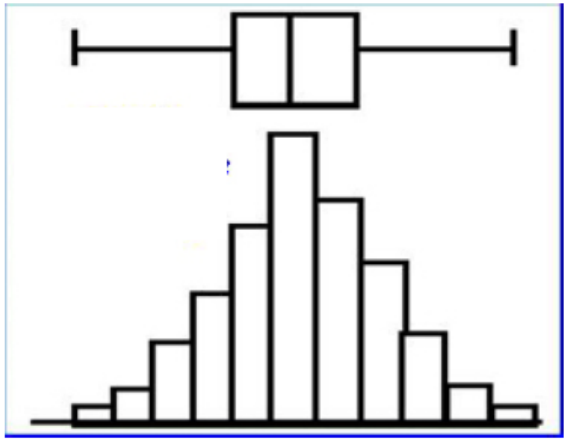
Normal



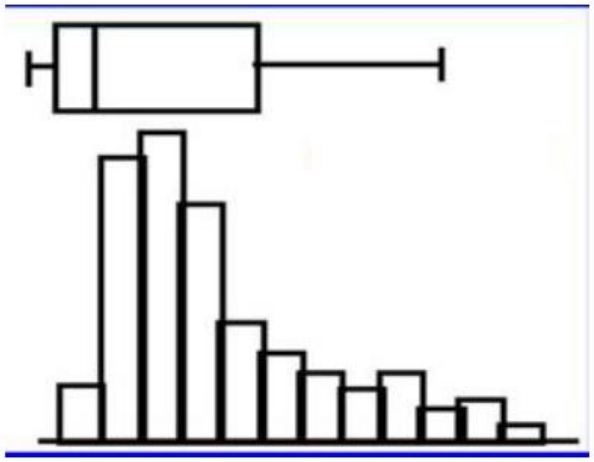
Not Normal



Normal

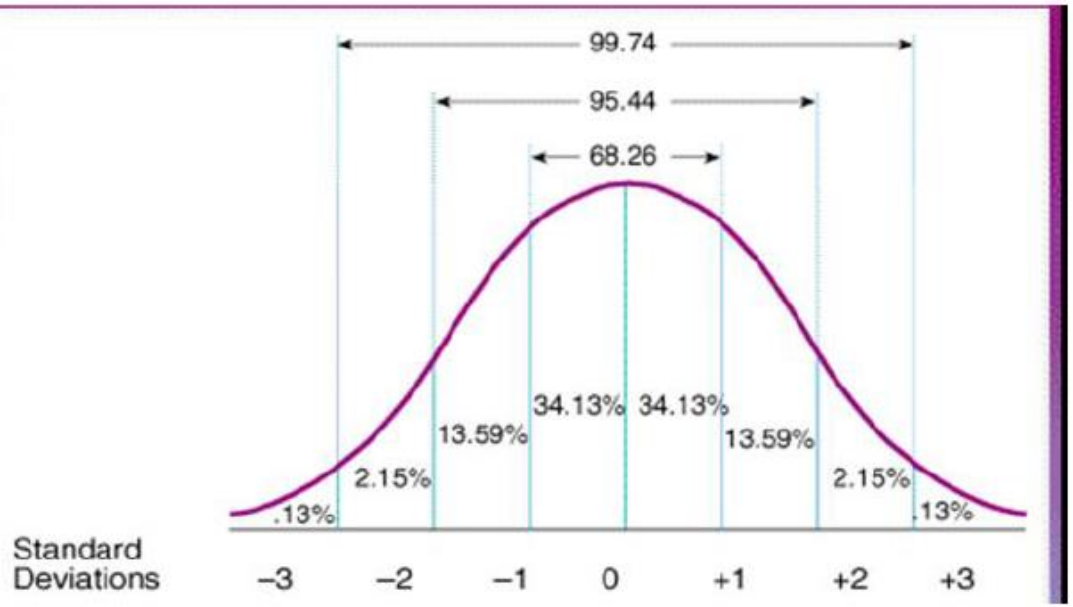


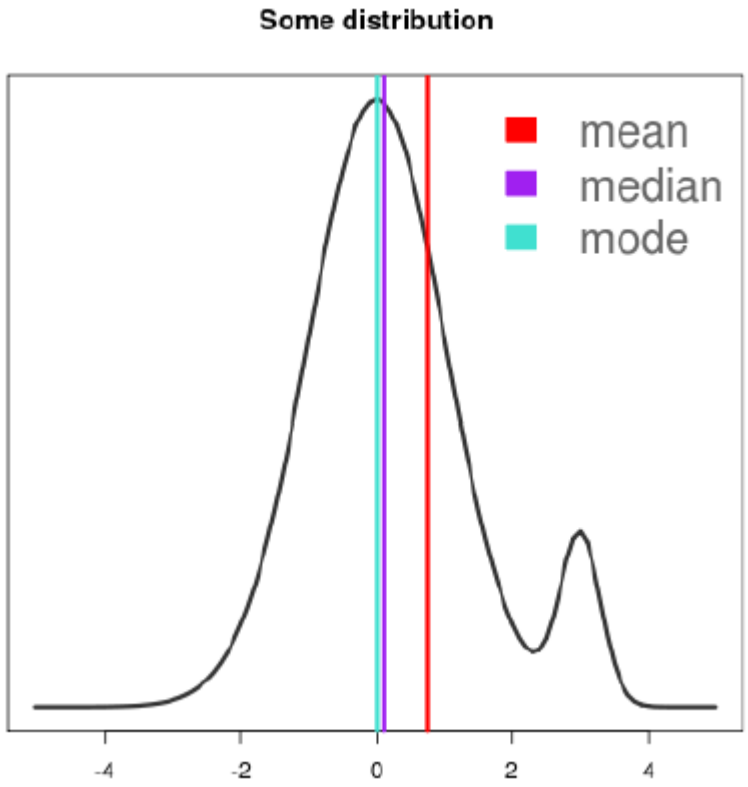
Not Normal



Normal

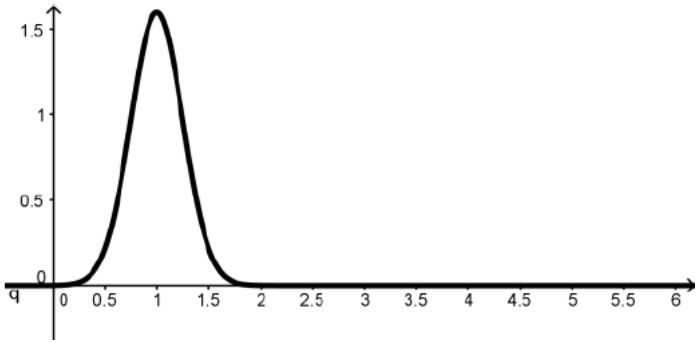
Not Normal



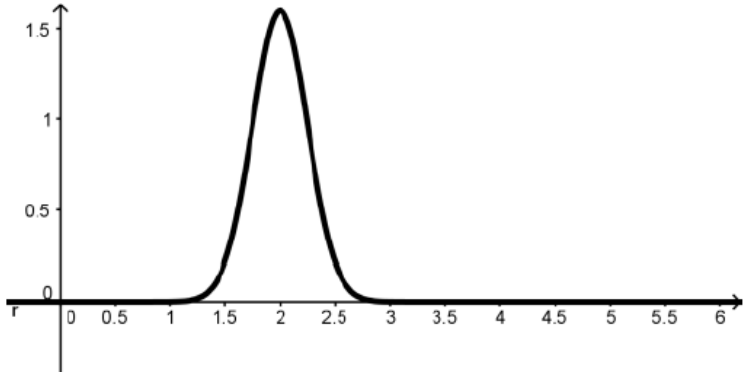


Same Std. Dev., different means

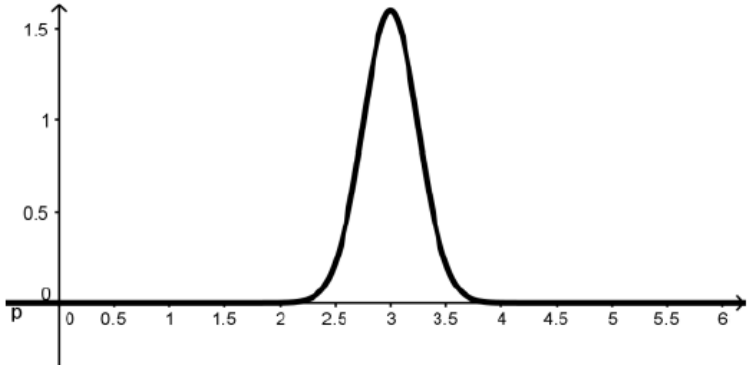
Mean = 1, Standard Deviation = 0.25



Mean = 2, Standard Deviation = 0.25

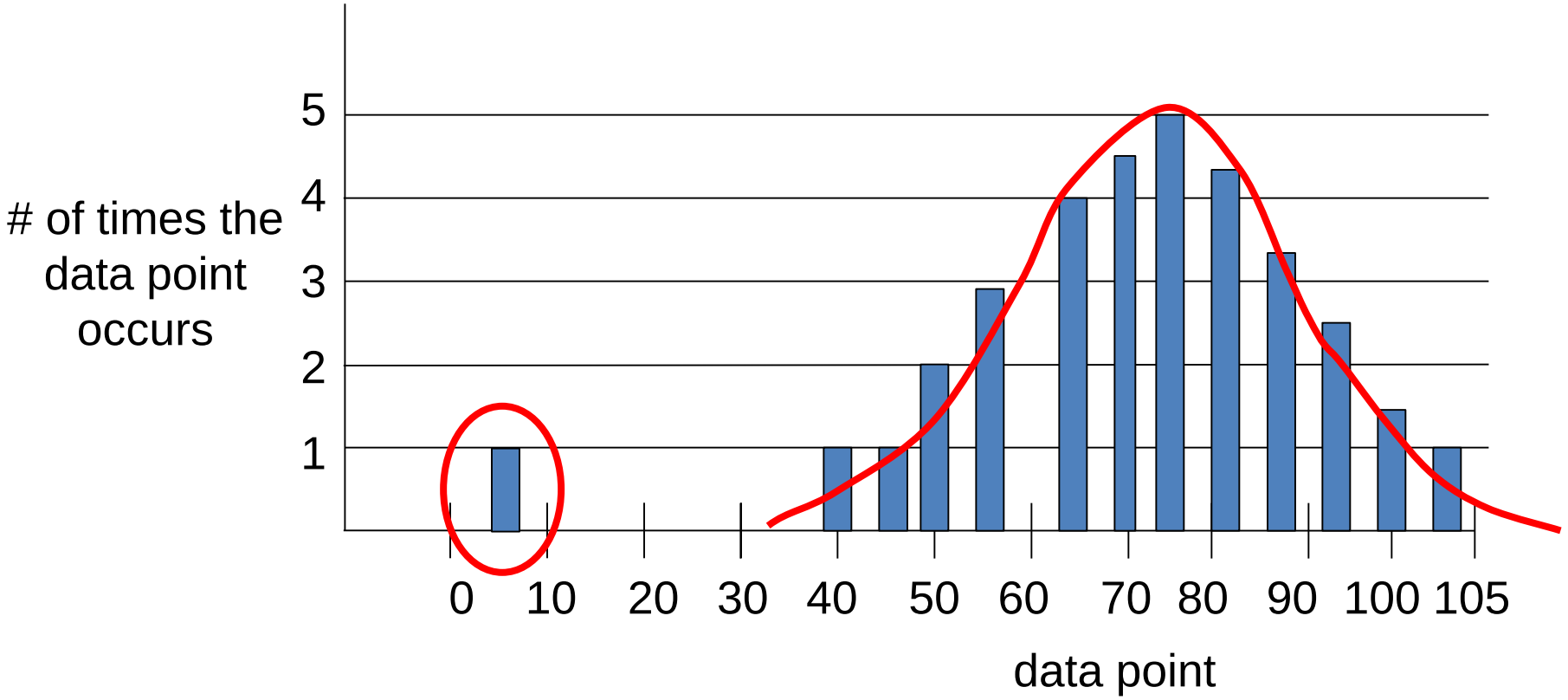


Mean = 3, Standard Deviation = 0.25

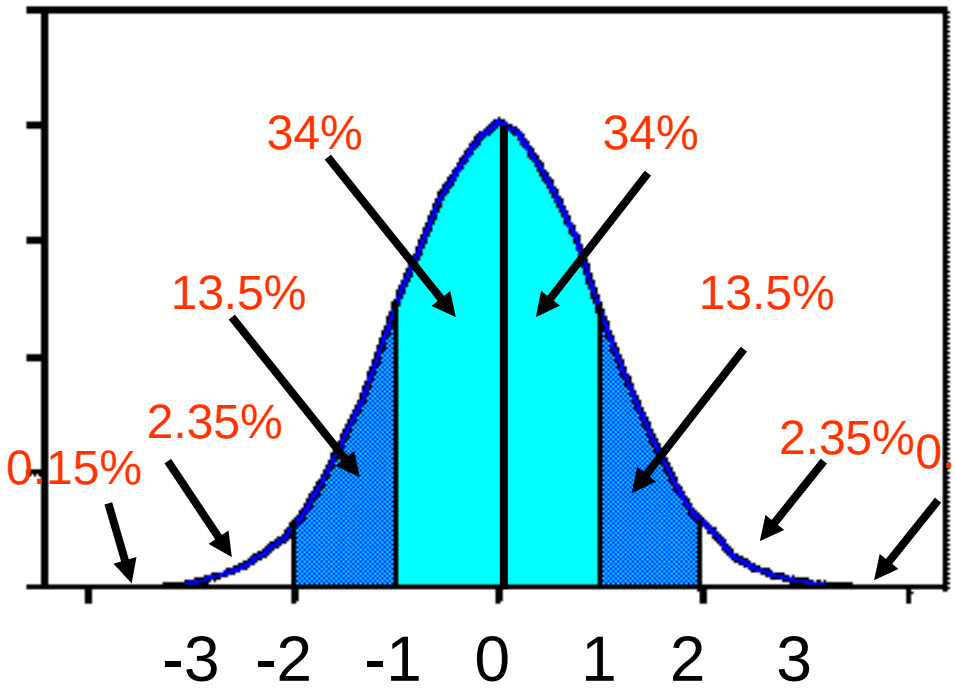


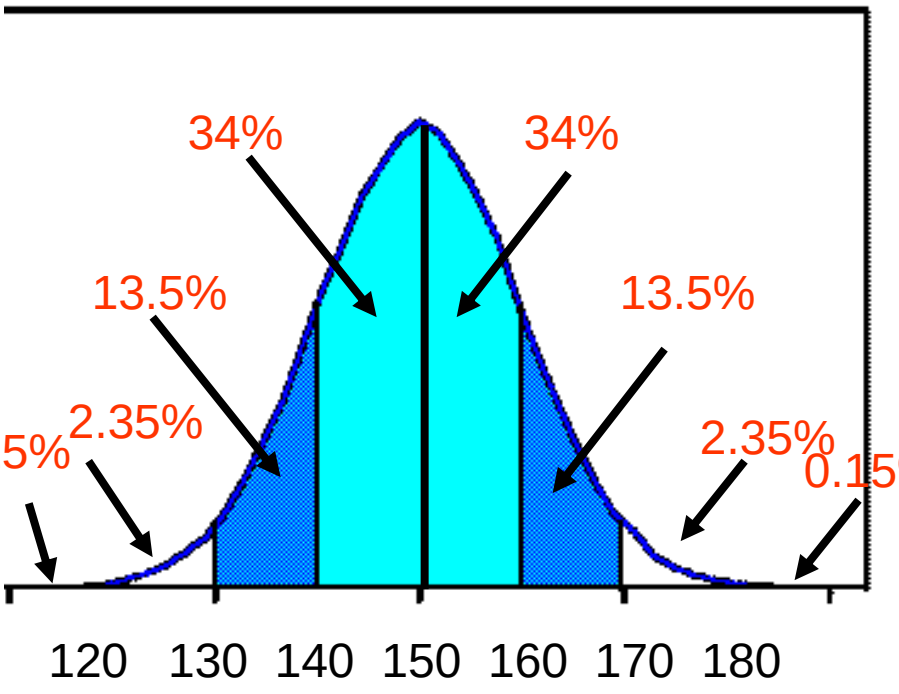
Same spread, different center point.

Outlier: a data point that is much higher or lower than the other data points.



To convert the standard scale of the Normal Distribution Graph to the data scale, we need (1) mean and (2) std. deviation. For example: $\bar{x} = 150 \quad S = 10$





68% of the data falls between -1 sdev and +1 sdev of the mean.

68% of the data falls between data values 140 and 160..

The standard deviation for some data is 7. The mean for this data is 42. Draw a bell curve and label the x-axis up to 3 standard deviations above and below the mean.

What is the probability that a data point will be in the range between 28 and 42?

What is the probability that a data point will be in the range between 21 and 28?

Comparing “apples to apples”

In math, Jordan scored a 53. The class average was 57. The standard deviation was 2. How many standard deviations below the mean did Jordan score? 2 std deviations

In science, Jordan scored a 114. The class average was 126. The standard deviation was 6. How many standard deviations below the mean did Jordan score? 2 std deviations

On which test did Jordan perform better on?

Even though his science test score was the higher of the two tests, based upon his score being the same number of standard deviations below the mean, he actually performed the same on both tests compared to the rest of the class (same portion of the class scored below him on both tests).