Learning Outcomes

Outcome 1. Compute the mean, median, mode, and weighted mean for a set of data and understand what these values represent.

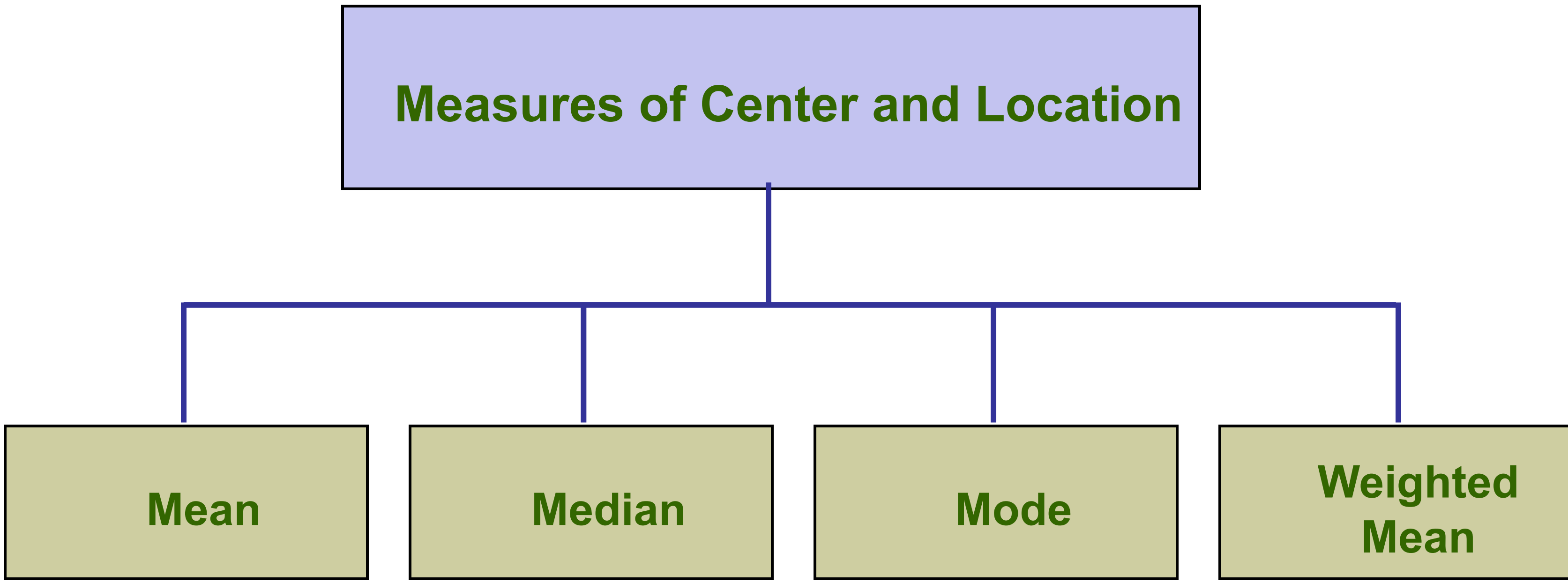
Outcome 2. Construct a box and whisker graph and interpret it.

Outcome 3. Compute the range, interquartile range, variance, and standard deviation and know what these values mean.

Outcome 4. Compute a z score and the coefficient of variation and understand how they are applied in decision-making situations.

Outcome 5. Understand the Empirical Rule and Tchebysheff's Theorem.

3.1 Measures of Center and Location



Parameter and Statistic

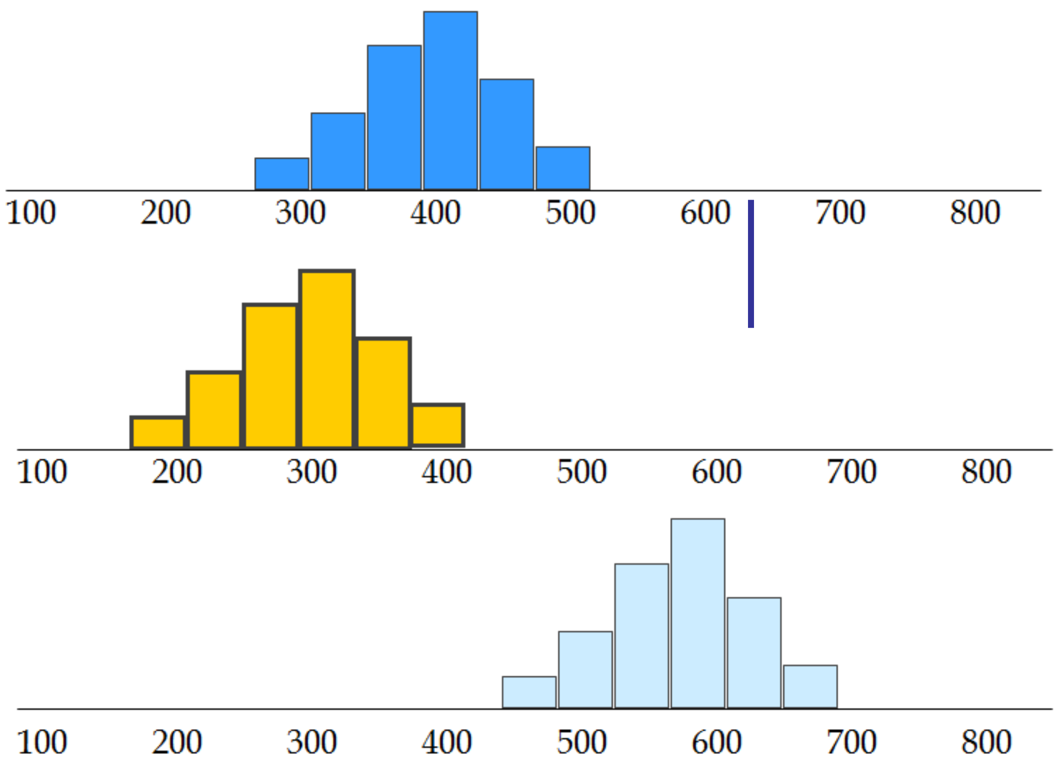
- Parameter
  - A measure computed from the entire population
  - As long as the population does not change, the value of the parameter will not change.
- Statistic
  - A measure computed from a sample that has been selected from a population
  - The value of the statistic will depend on which sample is selected.

Describing Data – What To Look For

- Measuring the Center
  - Data for a variable of interest forms a distribution
  - We need to be able to describe the center using a numerical measure.
- Measuring the Spread
  - Values for a variable of interest will take on different values –the data are spread out around the center.
  - We need to be able to measure the spread.

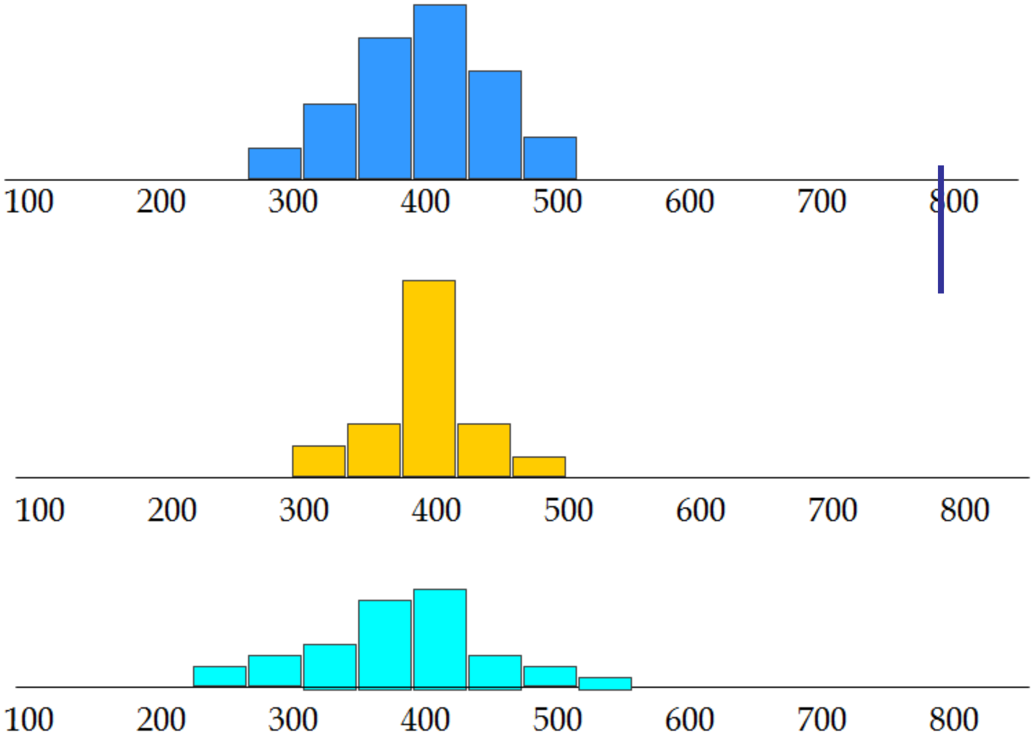
Data Distributions With Different Centers

Differences in Data



Data Distributions With Different Spreads

Data Differences



Descriptive Measures of the Center

- Mean
  - The arithmetic average of the data
- Median
  - The midpoint of the data
- Mode
  - The data value occurring most frequently

Population Mean (Parameter)

μ - Population mean (pronounced mu)
N - Population size
xi - ith individual value of variable x

- The average for <u>all values in the population</u> computed by dividing the sum of all values by the population size

Population Mean - Example

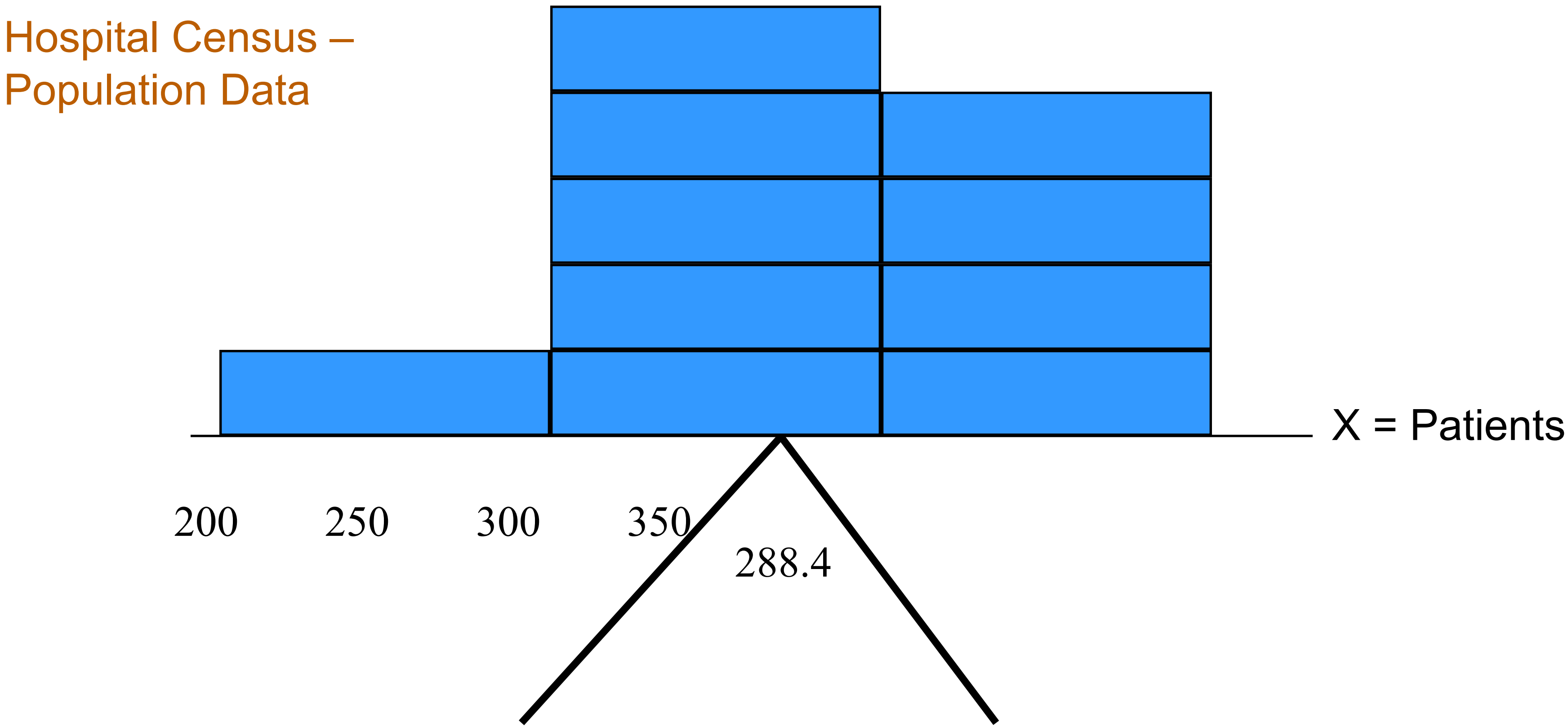
Each day, a local hospital counts the number of patients that are in the hospital. This is called the hospital census. The hospital is interested in only the census data for the past 10 days. This is the population of interest. The data are shown below:

216	255	330	254	348
317	292	267	310	295

$N = 10$

$$= \frac{216 + 255 + 330 + \ldots\ldots \quad 310 + 295}{10} = \frac{2{,}884}{10} = 288.4$$

The Population Mean - Graphic Representation



Population Mean - Example

N = 8 Weeks

TABLE 3.1 San Carlo Hotel Data

Week	Rooms Rented	Revenue	Complaints
1	22	$1,870	0
2	13	$1,590	2
3	10	$1,760	1
4	16	$2,345	0
5	23	$4,563	2
6	13	$1,630	1
7	11	$2,156	0
8	13	$1,756	0

x_1 = Total number of rooms rented

x_2 = Total dollar revenue from the room rentals

x_3 = Number of customer complaints that came from guests each Sunday

Population Mean - Example

Week	Rooms Rented
1	22
2	13
3	10
4	16
5	23
6	13
7	11
8	13

Calculate the Population Mean for the Variable, Rooms Rented

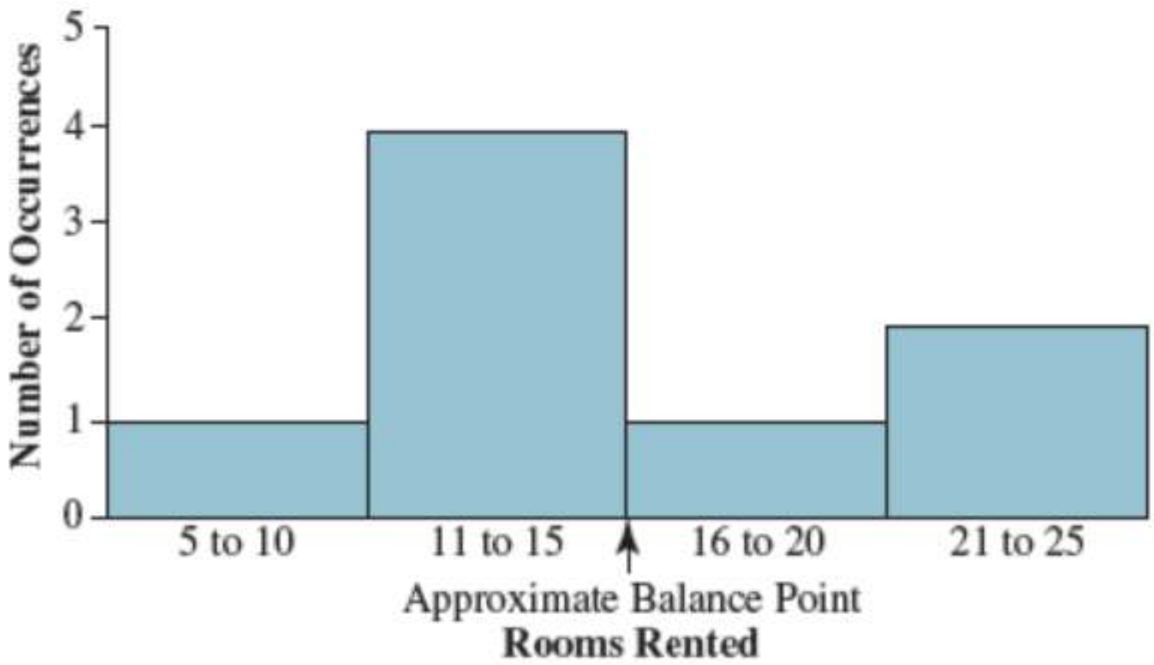
The Excel 2016 function for the mean is
=**Average(22,13,10,16,23, 13,11,13)**

$$\mu = \frac{\sum x}{N} = \frac{22 + 13 + 10 + 16 + 23 + 13 + 11 + 13}{8}$$

$$= \frac{121}{8}$$

$$\mu = 15.125$$

Mean – The Balance Point



Property of the Mean

TABLE 3.2 Deviations Around the Mean Using Hotel Data

x	$(x-\mu)$ = Deviation
22	$22 - 15.125 = 6.875$
13	$13 - 15.125 = -2.125$
10	$10 - 15.125 = -5.125$
16	$16 - 15.125 = 0.875$
23	$23 - 15.125 = 7.875$
13	$13 - 15.125 = -2.125$
11	$11 - 15.125 = -4.125$
13	$13 - 15.125 = \underline{-2.125}$
	$\Sigma(x-\mu) = 0.000 \leftarrow$ Sum of deviations equals zero.

Sum of Deviations from the mean always equals zero

Sample Mean (Statistic)

- Sample mean (x-Bar)

n - Sample size

- The average for all values <u>in the sample</u> computed by dividing the sum of all sample values by the sample size

Sample Mean - Example

Number of Sales Made By a Sales Representative

n = 10   sample of ten days during the past three years

20	**5**	**10**	**40**	**50**
15	**30**	**20**	**20**	**90**

Raw Data

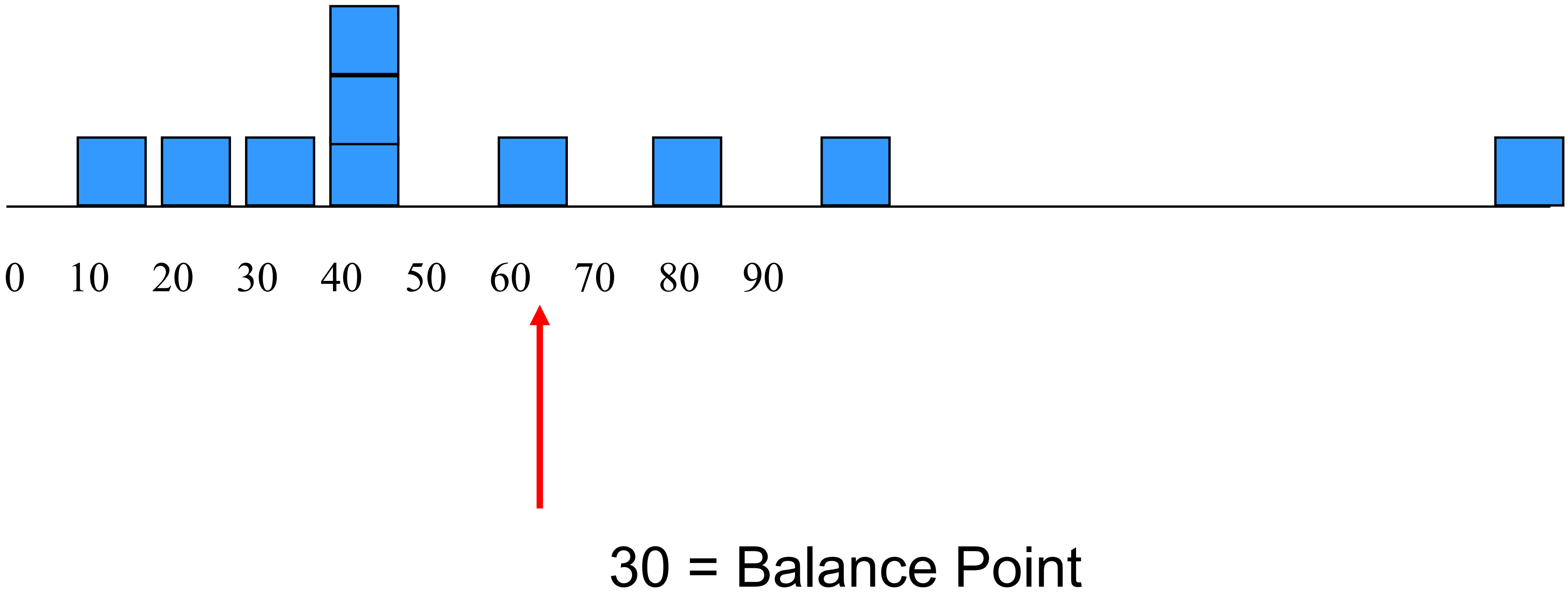
$$\bar{X} = \frac{20 + 5 + 10 + 40 + 50 + \ldots\ 90}{10}$$

$$= 300/10 = 30.0$$

X-Bar

Sample Mean – The Balance Point

Sales Example



Central Property of the Mean

Sales Example

x	
5	-25
10	-20
15	-15
20	-10
20	-10
20	-10
30	0
40	10
50	20
90	60

Property: The sum of the deviations from the mean will always equal 0.0.

When Extreme Values are Present Impact on the Mean -Sales Example

20	5	10	40	50
15	30	20	20	90

Original Data

= 30

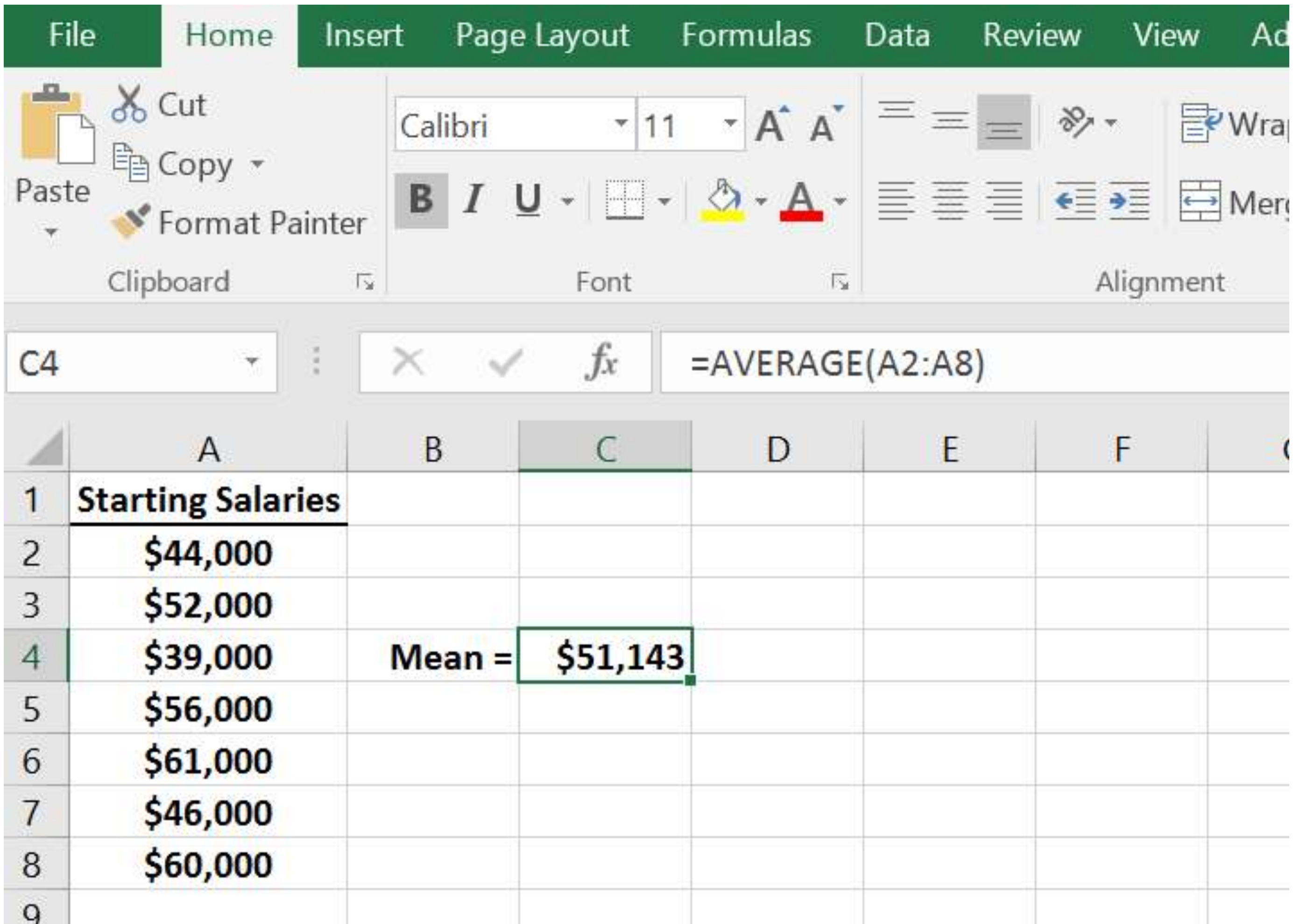
New Data (one value changed)

20	5	10	40	50
15	30	20	20	390

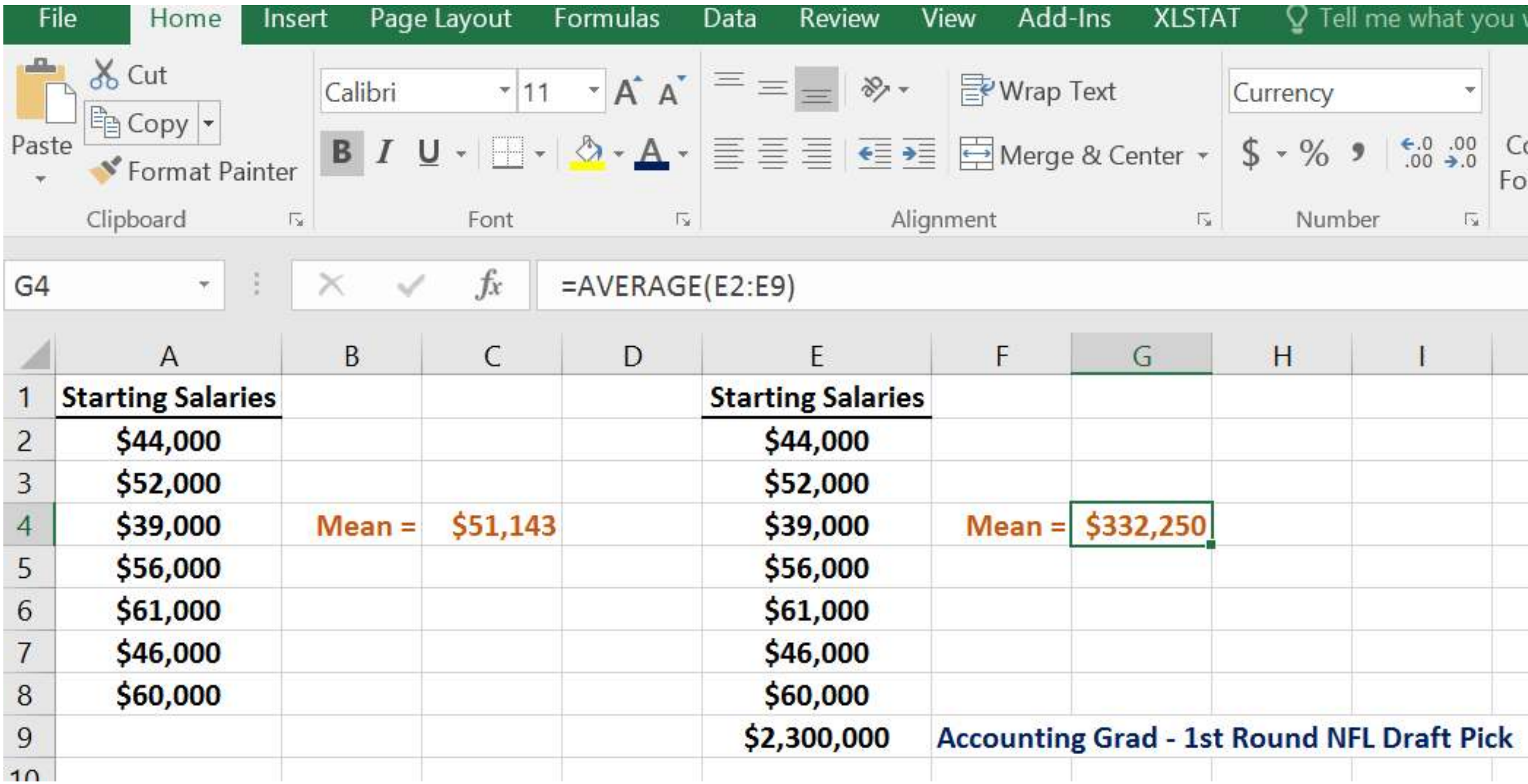
= 60

Conclusion: The Mean is Sensitive to Extreme Values.

Impact of Extreme Values on the Mean – Accounting Graduates Starting Salaries - Example



Impact of Extreme Values on the Mean –Accounting Graduates Starting Salaries Example Using Excel 2016



	A	B	C	D	E	F	G	H	I
1	Starting Salaries				Starting Salaries				
2	$44,000				$44,000				
3	$52,000				$52,000				
4	$39,000	Mean =	$51,143		$39,000	Mean =	$332,250		
5	$56,000				$56,000				
6	$61,000				$61,000				
7	$46,000				$46,000				
8	$60,000				$60,000				
9					$2,300,000	Accounting Grad - 1st Round NFL Draft Pick			

Characteristics of the Mean

- Uses all the Data
- Sensitive to Extreme Values
- The Balance Point
  - Sum of Deviations from the Mean Equals Zero

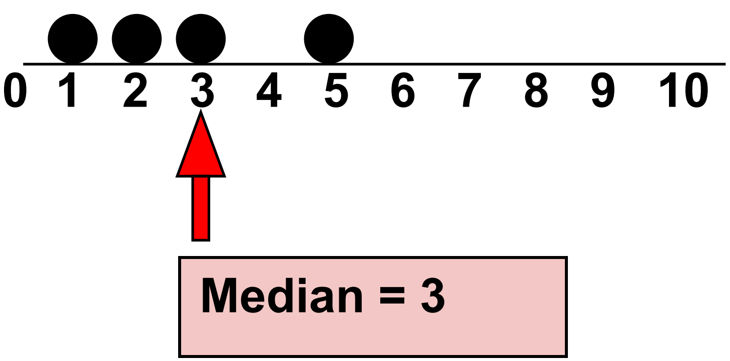
Median

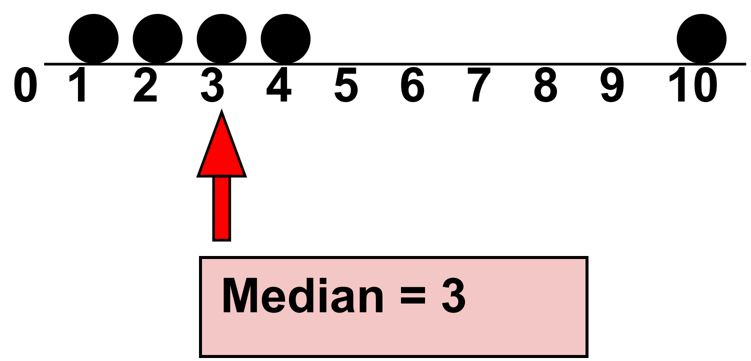
- The median is a center value that divides a data array into two halves (*Md*).
- Data Array
  - Data that have been arranged in numerical order
- Median Index

  - *i* = The index of the point in the data set corresponding to the median value
  - *n* = Sample size

Median

- In an ordered array (lowest to highest), the median is the "middle" number, i.e., the number that splits the distribution in half numerically.
  - 50% of the data is above the median, 50% is below
  - Represented as *Md*
- The median is <u>not</u> affected by extreme values.





Computing the Median

- Step 1: Collect the sample data.

Starting Salaries
$44,000
$52,000
$39,000
$56,000
$61,000
$46,000
$60,000

n = 7

Computing the Median

- Step 2: Sort data from smallest to largest.

Starting Salaries
$39,000
$44,000
$46,000
$52,000
$56,000
$60,000
$61,000

n = 7

Computing the Median

- Step 3: Calculate the median index.
  - If *i* is not an integer, round up to next highest integer.
  - If *i* is an integer, the median is the average of the values in position i and *i* + 1.

The Median is the 4th value from top or bottom of the data.

Computing the Median

- Step 4: Find the median.

Starting Salaries
$39,000
$44,000
$46,000
$52,000
$56,000
$60,000
$61,000

Median = Md = $52,000

Median Example - Even Number of Data

20	5	10	40	50
15	30	20	20	90

Sales Data
Number of Sales Made
$n = 10$

Sort the Data

5 10 15 20 20 20 30 40 50 90

Median = (20+20)/2 = 20

When there are an even number of data points, the median is the average of the middle two values.

Median - Impact of Extreme Values
Sales Data Example

5 10 15 20 20 20 30 40 50 90

Median = (20+20)/2 = 20

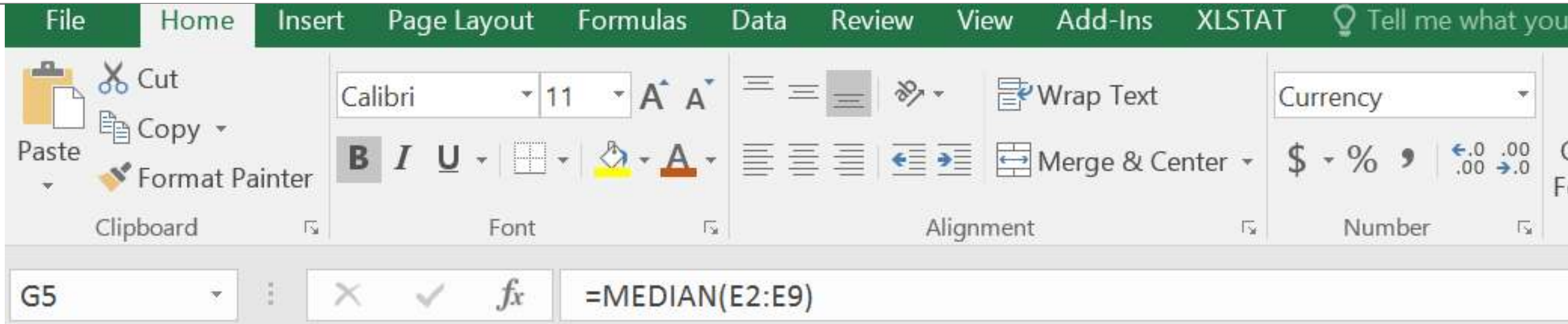
Extreme Value

5 10 15 20 20 20 30 40 50 **390**

Median = (20+20)/2 = 20

The Median is Unaffected by Extreme Values.

Median - Impact of Extreme Values Accounting Graduates Starting Salaries



	A	B	C	D	E	F	G	H	I
1	**Starting Salaries**				**Starting Salaries**				
2	**$39,000**				**$44,000**				
3	**$44,000**				**$39,000**				
4	**$46,000**	**Mean =**	**$51,143**		**$46,000**	**Mean =**	**$332,250**		
5	**$52,000**	**Median =**	**$52,000**		**$52,000**	**Median =**	**$54,000**		
6	**$56,000**				**$56,000**				
7	**$60,000**				**$60,000**				
8	**$61,000**				**$61,000**				
9					**$2,300,000**	**Accounting Grad - 1st Round NFL Draft Pick**			
10									

The only reason that the two medians are not equal is that the number of data values is different – Extreme salary did not affect the median.

Comparing the Mean and the Median

Starting Salaries				Starting Salaries					
$39,000				$44,000					
$44,000				$39,000					
$46,000	Mean =	$51,143		$46,000	Mean =	$332,250			
$52,000	Median =	$52,000		$52,000	Median =	$54,000			
$56,000				$56,000					
$60,000				$60,000					
$61,000				$61,000					
				$2,300,000	Accounting Grad - 1st Round NFL Draft Pick				

When extreme values are present, the mean gets pulled above the median.

Characteristics of the Median

- Not Sensitive to Extreme Values
- Uses only the Middle Value(s)

Mode

Mode: **The value in the data which appears most frequently**

Sales Data

20	5	10	40	50
15	30	20	20	90

Mode → 20

x	Freq(x)
5	**1**
10	**1**
15	**1**
20	**3**
30	**1**
40	**1**
50	**1**
90	**1**

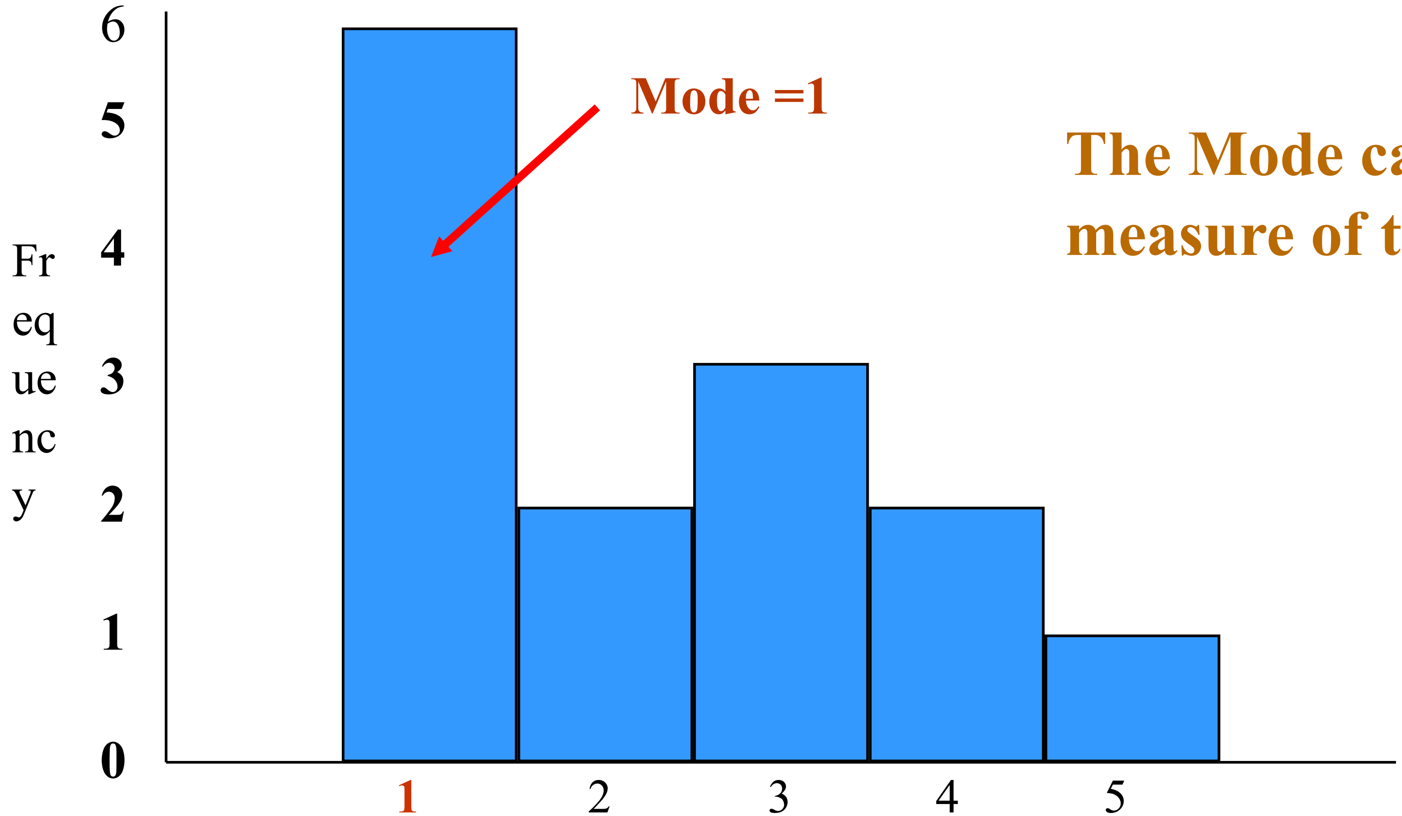
Mode = 20 (occurs 3 times)

The Mode

Number of Students Absent from Class Per Day

1	1	1	1	1	1	2	2	3	3	3	4	4	5

Sample Data



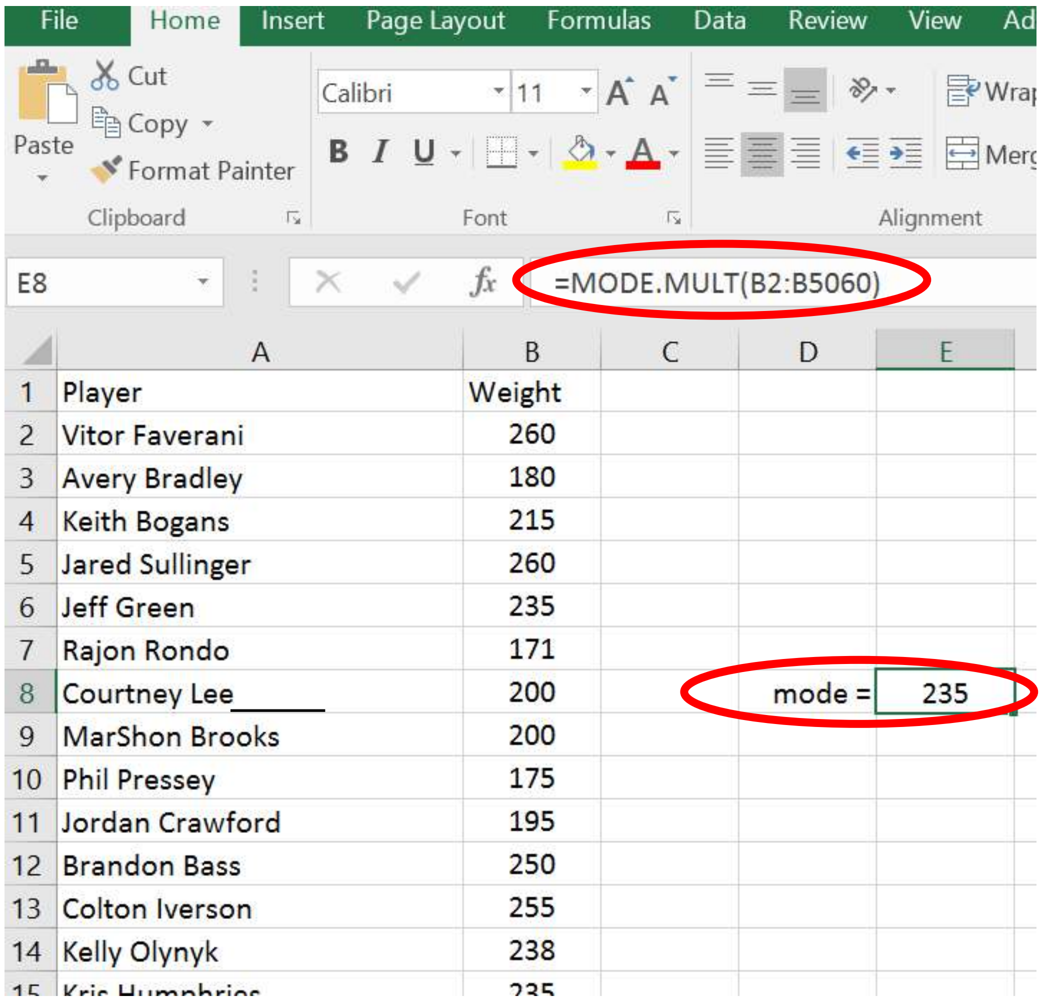
The Mode can be a weak measure of the center.

Mode: Example NBA Player Weights

Weight	Frequency	Weight	Frequency	Weight	Frequency	Weight	Frequency	Weight	Frequency
157	1	194	1	217	3	241	1	275	4
160	1	195	16	218	5	242	1	278	1
161	1	196	2	219	2	243	1	280	3
168	1	197	6	220	24	245	14	285	2
169	1	198	1	221	2	246	1	289	1
170	2	200	27	222	2	247	1	290	1
171	1	201	3	225	21	248	3		
172	2	202	1	226	3	250	22		
174	1	203	1	227	1	251	1		
175	10	204	1	228	8	252	2		
176	1	205	20	229	1	253	3		
178	1	206	2	230	22	254	1		
180	14	207	6	231	1	255	14		
181	2	208	3	232	5	257	1		
183	1	209	2	233	1	260	15		
185	20	210	19	235	33	261	1		
187	2	211	3	236	1	263	3		
188	1	212	2	237	3	265	6		
189	5	213	2	238	2	266	1		
190	16	214	2	239	1	267	1		
191	3	215	19	240	28	270	5		

Mode =235 lbs.

Mode: Example NBA Player Weights Using Excel 2016



	A	B	C	D	E
1	Player	Weight			
2	Vitor Faverani	260			
3	Avery Bradley	180			
4	Keith Bogans	215			
5	Jared Sullinger	260			
6	Jeff Green	235			
7	Rajon Rondo	171			
8	Courtney Lee	200		mode =	235
9	MarShon Brooks	200			
10	Phil Pressey	175			
11	Jordan Crawford	195			
12	Brandon Bass	250			
13	Colton Iverson	255			
14	Kelly Olynyk	238			
15	Kris Humphries	235			

Checking for Possible Multiple Modes

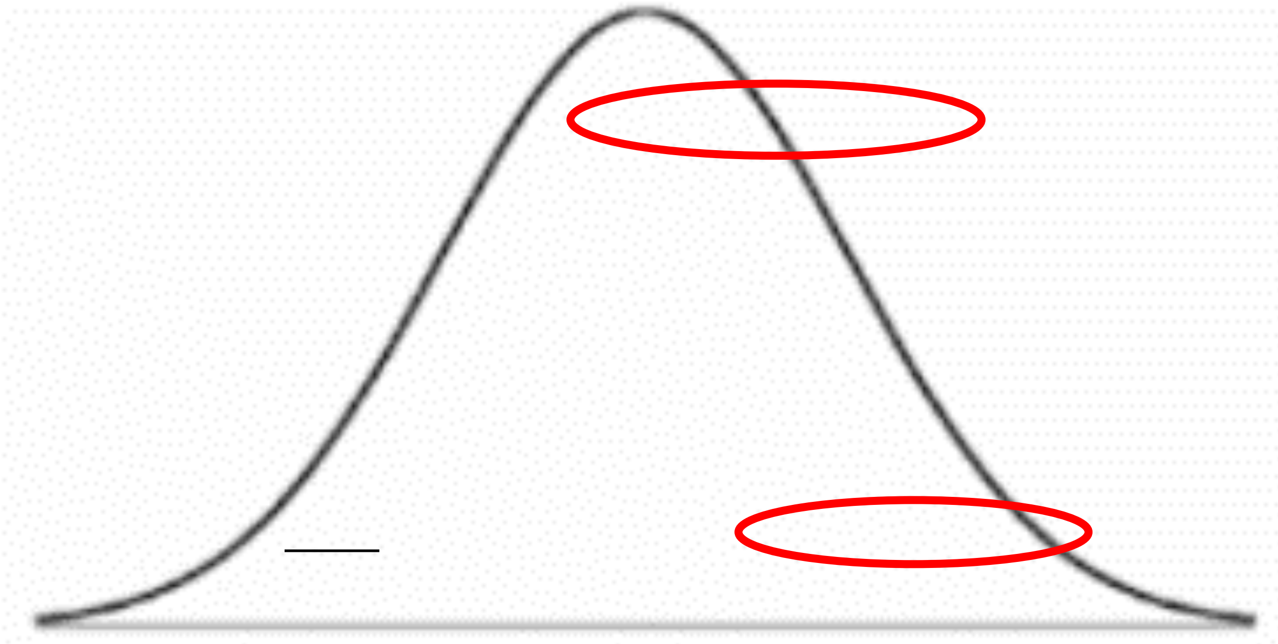
Mode

- The value in a data set that occurs most frequently
- Is not affected by extreme values.
- Can be used for both quantitative and qualitative data (ratio, interval, ordinal, nominal.)
- Can have more than one mode, or no mode.
- Distribution with two modes - bimodal

Distribution Shapes

Relationship to Mean, Median and Mode

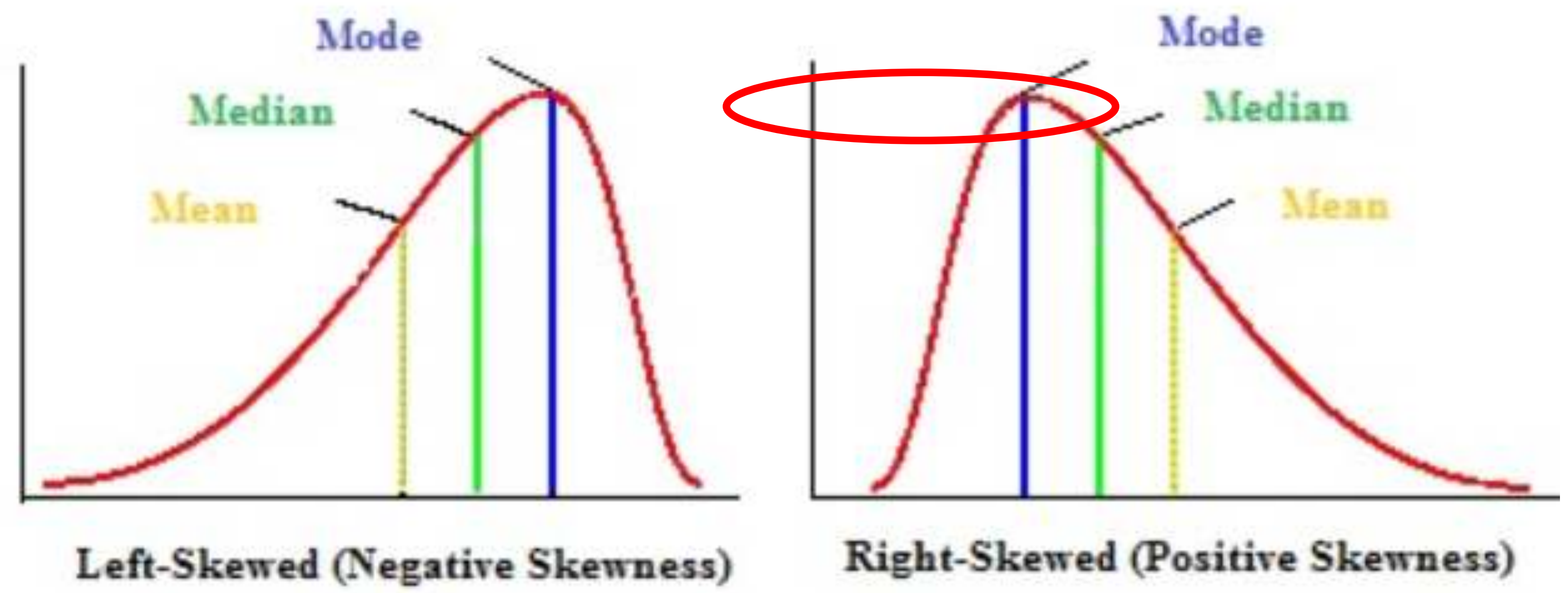
Symmetrical Distribution



Mean
Median
Mode

Distribution Shapes

Relationship to Mean, Median, and Mode



Weighted Mean

- The mean value of data values that have been weighted according to their relative importance

Weighted Mean for a Population

Weighted Mean for a Sample

w_i - The weight c

Weighted Mean Example: Financial Portfolio

Mutual Fund	Percentage Return	Shares
	x	w
Fidelity	7.2	2,000
Vanguard	8.3	5,000
Dimensional	5.4	12,000

Percentiles and Quartiles

Percentiles

The pth percentile
in a data array:

- p% are less than or equal to this value.
- $(100 - p)$% are greater than or equal to this value.

  (where $0 \le p \le 100$)
- 50th percentile is the median.

Quartiles

1st quartile = 25th percentile

2nd quartile = 50th percentile

Also the median

3rd quartile = 75th percentile

Calculating Percentiles

- Step 1: Sort the data in order from the lowest to highest value.
- Step 2: Determine the percentile location index:
- Step 3: If *i* is not ar then round to next highest integer. The *p*th percentile is located at the rounded index position. If *i* is an integer, the *p*th percentile is the average of the values at location index positions *i* and *i* + 1.

Calculating Percentiles Example

Moving Company Travel Distances - Miles

13.5	8.6	16.2	21.4	21.0	23.7	4.1	13.8	20.5	9.6
11.5	6.5	5.8	10.1	11.1	4.4	12.2	13.0	15.7	13.2
13.4	13.1	21.7	14.6	14.1	12.4	24.9	19.3	26.9	11.7

Determine the 80th Percentile.

Calculating Percentiles Example

STEP 1 **Sort the data from lowest to highest.**

4.1	4.4	5.8	6.5	8.6	9.6	10.1	11.1	11.5	11.7
12.2	12.4	13.0	13.1	13.2	13.4	13.5	13.8	14.1	14.6
15.7	16.2	19.3	20.5	21.0	21.4	21.7	23.7	24.9	26.9

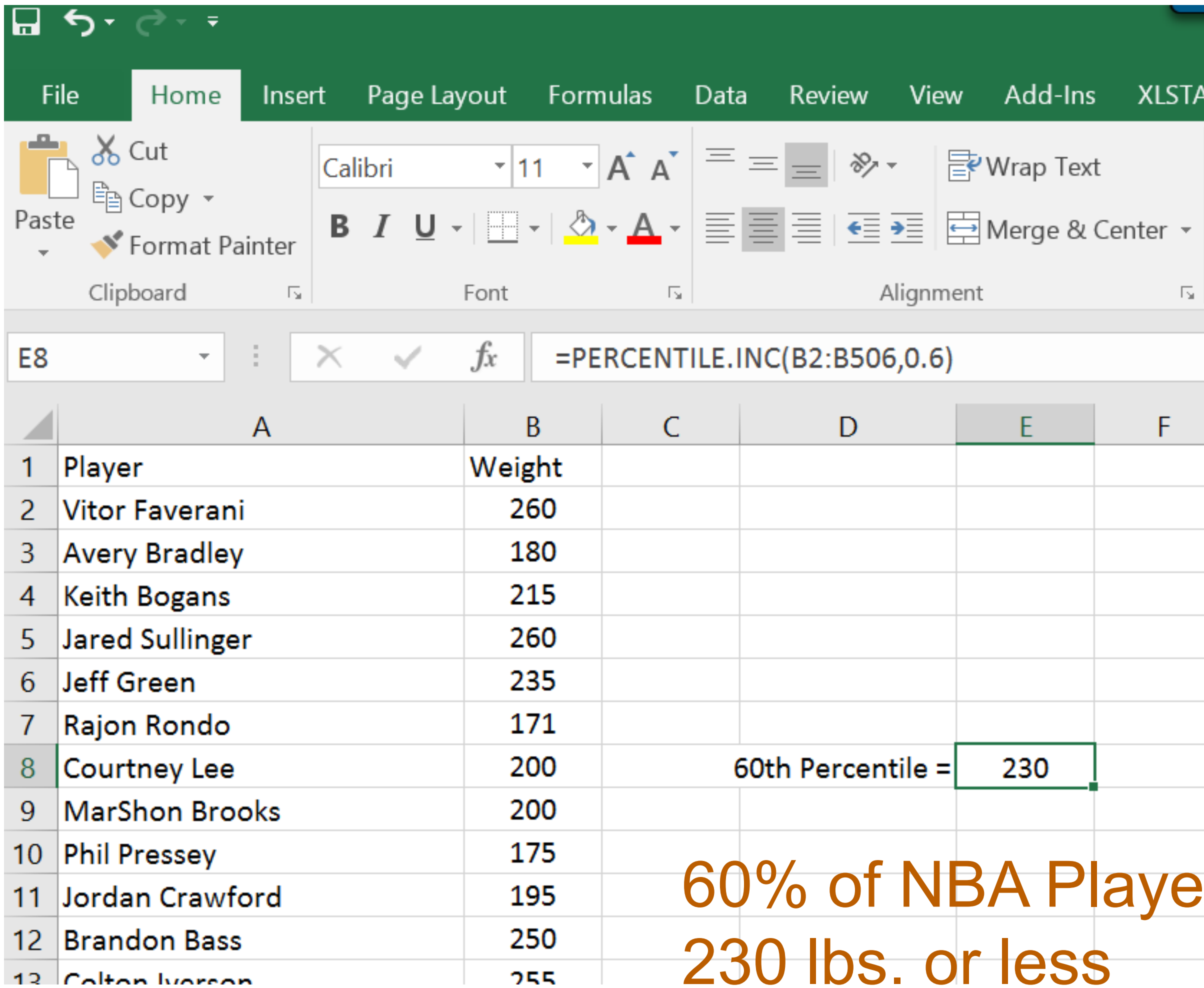
STEP 2 **Determine percentile location index, *i*, using Equation 3.6.**

The 80th percentile location index is

$$i = \frac{p}{100}(n) = \frac{80}{100}(30) = 24$$

80th percentile is the average of the 24th and 25th values.

Calculating Percentiles Using Excel 2016 - NBA Player Weights Example



	A	B	C	D	E	F
1	Player	Weight				
2	Vitor Faverani	260				
3	Avery Bradley	180				
4	Keith Bogans	215				
5	Jared Sullinger	260				
6	Jeff Green	235				
7	Rajon Rondo	171				
8	Courtney Lee	200		60th Percentile =	230	
9	MarShon Brooks	200				
10	Phil Pressey	175				
11	Jordan Crawford	195				
12	Brandon Bass	250				
13	Colton Iverson	255				

60% of NBA Players Weigh 230 lbs. or less

Calculating Quartiles

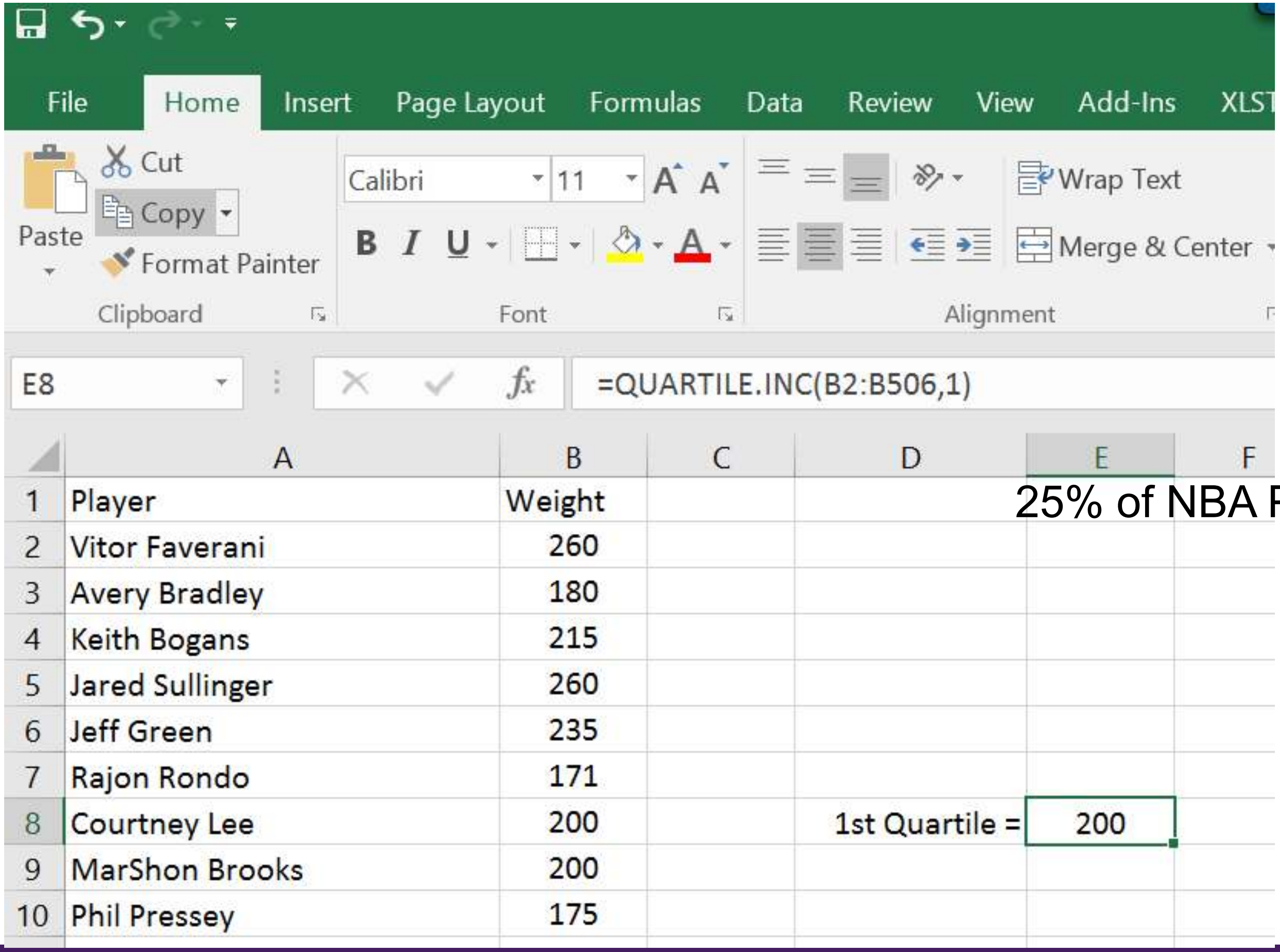
- Find the 1st quartile in an ordered array of 19 values.
- (36 40 42 46 51 56 62 65 71 74 78 82 84 87 88 90 92 95 97

$$i = \frac{q}{100}(n) = \frac{25}{100}(19) = 4.75$$

Use value at 5th position.

- 1st quartile Q1 equals 51.

Calculating Quartiles Using Excel 2016 - NBA Player Weights Example



25% of NBA Players

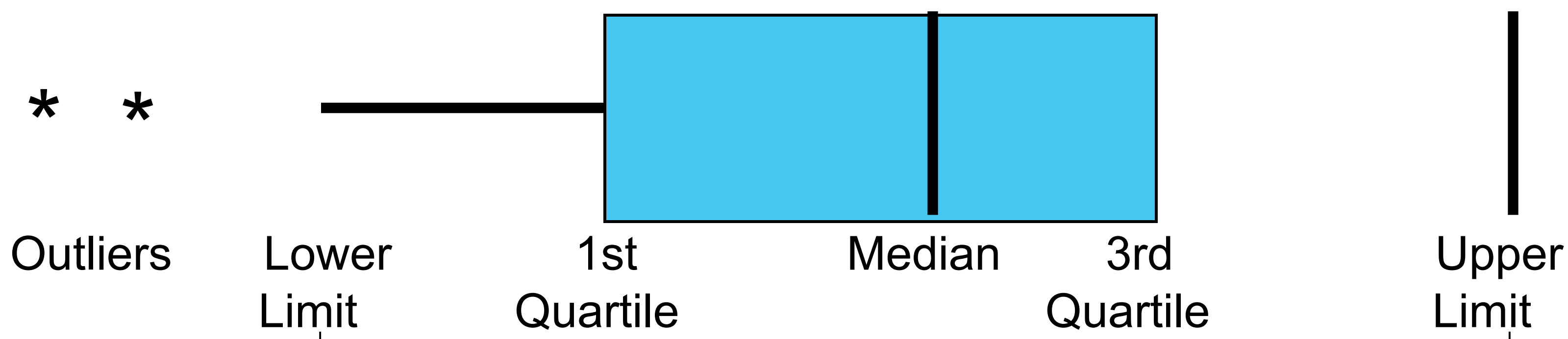
Box and Whisker Plot

- A graph that is composed of two parts: a box and the whiskers
- The box has a width that ranges from the first quartile ($Q1$) to the third quartile ($Q3$).
- A vertical line through the box is placed at the median.
- Limits are located at a value that is 1.5 multiplied by the difference between $Q1$ and $Q3$ below $Q1$ and above $Q3$.
- The whiskers extend to the left to the lowest value within the limits and to the right to the highest value within the limits.

Constructing a Box and Whisker Plot

- Step 1: Sort values from lowest to highest.
- Step 2: Find $Q1$, $Q2$, $Q3$.
- Step 3: Draw the box so that the ends are at $Q1$ and $Q3$.
- Step 4: Draw a vertical line through the median.
- Step 5: Calculate the interquartile range ($IQR = Q3 - Q1$).
- Step 6: Extend dashed lines from each end to the highest and lowest values within the limits.
- Step 7: Identify outliers with an asterisk (*).

Constructing a Box and Whisker Plot



The lower limit is $Q1 - 1.5\,(Q3 - Q1)$.

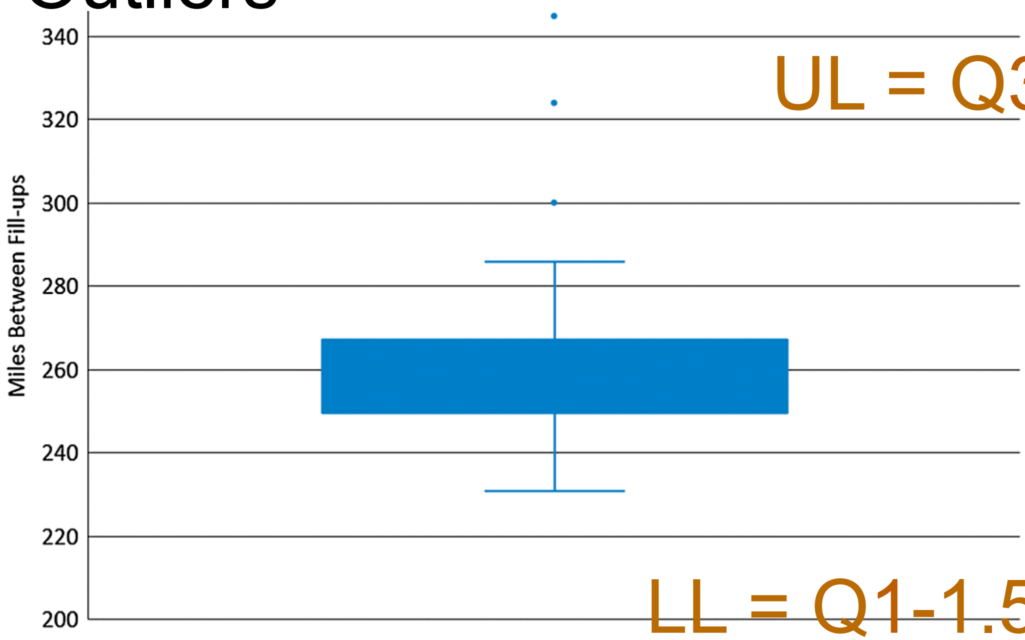
The upper limit is $Q3 + 1.5\,(Q3 - Q1)$.

- The center box extends from $Q1$ to $Q3$.
- The line within the box is the median.
- The whiskers extend to the smallest and largest values within the calculated limits.
- Outliers are plotted outside the calculated limits.

Box and Whisker Plot – Example Miles Between Fill-ups

231	236	241	242	242	243	243	243	248
248	249	250	251	251	252	252	254	255
255	256	256	257	259	260	260	260	260
262	262	264	265	265	265	266	268	268
270	276	277	277	280	286	300	324	345

Outliers



UL = Q3 +1.5(Q3-Q1) = 290

Q3 = 266 miles

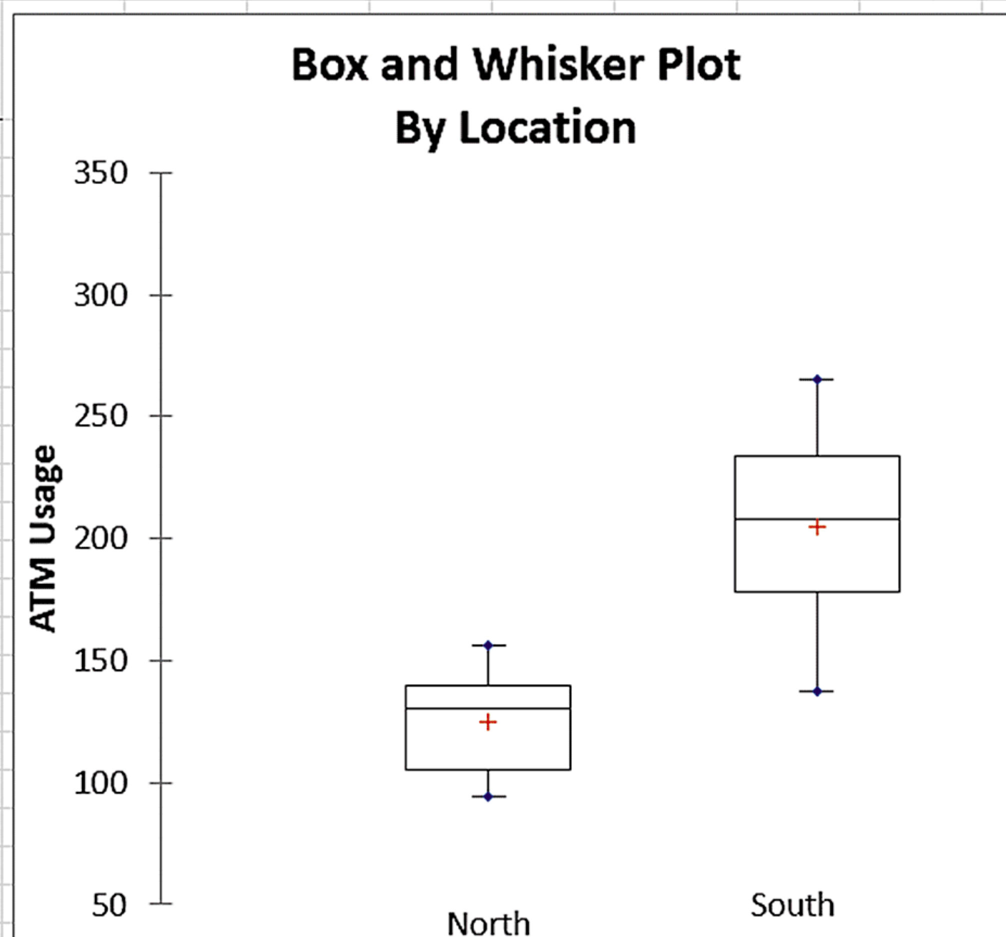
Md = 259 miles

Q1 = 250 miles

LL = Q1-1.5(Q3-Q1) = 226

Box and Whisker Plot – Example ATM Usage By Location

	Location	Daily ATM Usage
2	North	125
3	South	260
4	North	101
5	South	215
6	North	144
7	South	249
8	North	150
9	South	149
10	North	147
11	South	160
12	North	136
13	South	190
14	North	102
15	South	237
16	North	117
17	South	195
18	North	102
19	South	182
20	North	132
21	South	237
22	North	131
23	South	168



Data Level Issues

- Be aware of the level of data before computing a numerical measure.

The Mean - Data Level Requirements

Ratio or Interval Level Data

Variables Such As:

- Age
- Income
- Interest Rate
- Stock Prices
- Sales
- Number of Defects
- Temperature
- etc.

The Mean – Caution With Ordinal Data

Computing a Mean for Ordinal Data – Not Recommended

Example: Ordinal Data - Education Level

1 = Grade School
2 = High School
3 = Some College
4 = College Degree
5 = Graduate Degree

Sample of n = 5 people – Education Codes

1	4	3	4	2

$= 14/5 = 2.8$????

Does this imply that the mean education level in the sample is somewhere between High School and Some College?

The Mean – Caution With Ordinal Data

Example: 5 Point Scale Question

I believe that I spend enough time studying for my statistics class.

SA	A	Neutral	D	SD
5	4	3	2	1

$$= 27/9 = 3 \quad ????$$

Responses

2
4
3
5
2
1
3
3
4

n=9

Concerns:

Is there an equal distance between the 5 categories – is the distance in meaning between SA and A the same as between A and Neutral?

Is a response of A twice as high as a response of D?

The mean computation assumes the answer to both questions is "Yes".

The Mean – Definitely Not With Nominal Data

What is your Marital Status?

1 = Single
2 = Married
3 = Divorced
4 = Windowed

Mean = 1.93   Does this mean that on average, a person in the sample is almost married??

Don't compute the mean for ordinal data!

Marital Status		
2		
4		
2		
3	n =	
1	Mean =	1.93
2		
2		
1		
1		
3		
1		
1		
2		
2		

The Median - Data Level Requirements

Ratio or Interval Level Data

Variables Such As:
- **Age**
- **Income**
- **Interest Rate**
- **Stock Prices**
- **etc.**

Okay for Ordinal Data

Not for Nominal Data

Descriptive Measures - Summary

Descriptive Measure	Computation ~~Method~~	Data Level	Advantages/ Disadvantages
Mean	Sum of values divided by the number of values	Ratio Interval	• Numerical center of the data • Sum of deviations from the mean is zero • Sensitive to extreme values
Median	Middle value for data that have been sorted	Ratio Interval Ordinal	• Not sensitive to extreme values • Computed only from the center values • Does not use information from all the data
Mode	Value(s) that occur most frequently in the data	Ratio Interval Ordinal Nominal	• May not reflect the center • May not exist • Might have multiple modes

Comparing Distributions
Production Volumes Example

Output

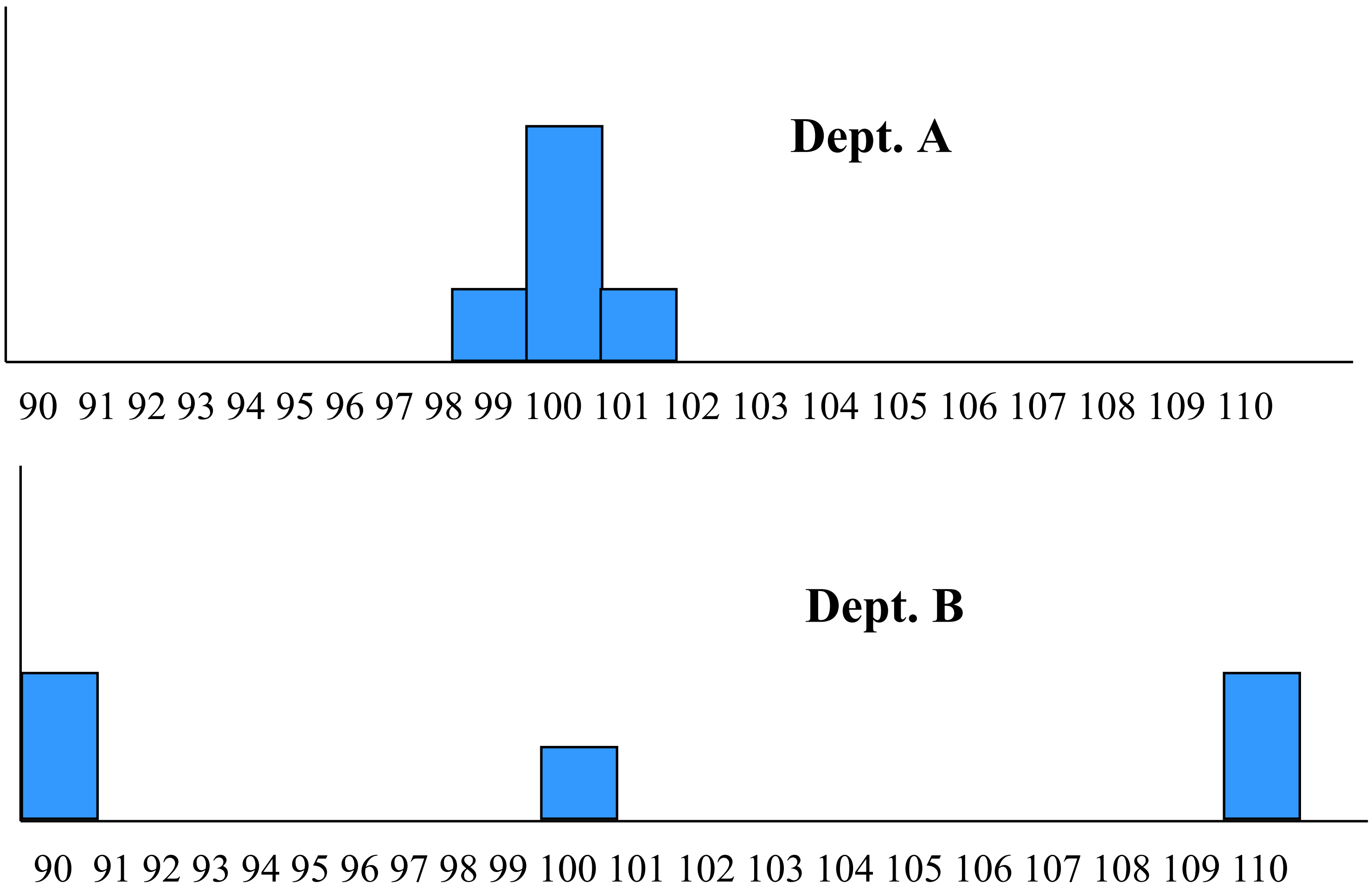
Dept. A	Dept. B
99	90
100	90
100	100
100	110
101	110

Mean = 100 **Mean = 100**
Median = 100 **Median =100**

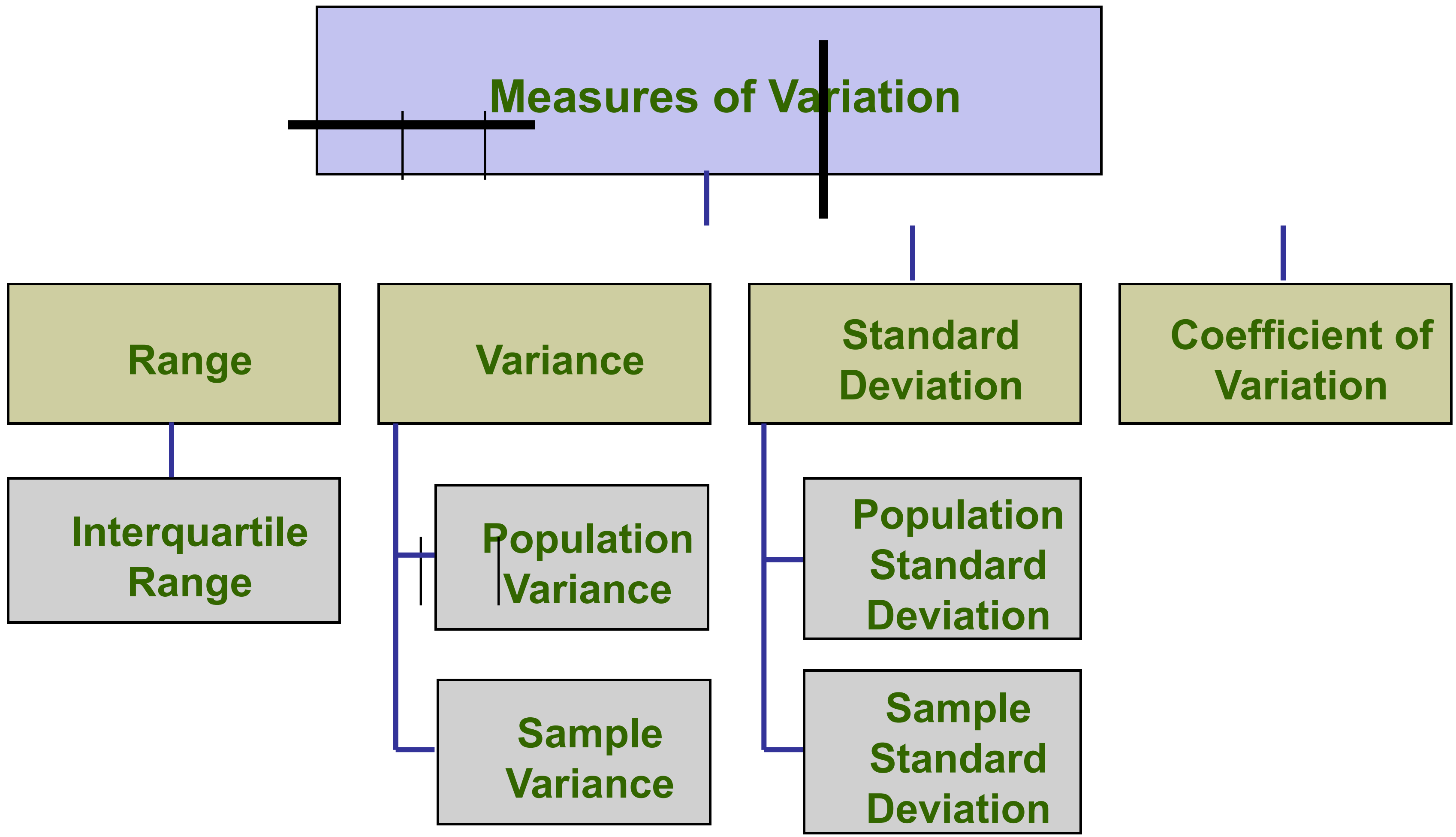
No Differences!

Production Distributions

What is the Difference?



3.2 Measures of Dispersion/Variation



Descriptive Statistics Measures of Dispersion

- Range
  - Simple range
  - Interquartile Range
- Variance
- Standard Deviation
- Coefficient of Variation

Range

- A measure of variation that is computed by finding the difference between the maximum and minimum values in a data set
- Simpl

$$R = \text{Maximum Value} - \text{Minimum Value}$$

- Is very sensitive to extreme values
- Ignores the data distribution

Measures of Dispersion - Staff Years of Experience

Raw Data

Values are the number of years experience for a sample of staff members.

13	16	16	13	13	9
12	15	8	12	13	14
11	6	15	11	15	13
12	12	13	13	13	11
17	14	11	14	13	15
11	11	13	15	9	10
11	12	11	16	11	10
13	13	11	13	10	13
11	10	8	11	11	17
10	12	13	10	10	15
14	11	9	15	13	12
11	14	12	10	12	14
14	12	11	11	12	13
11	11	11	13	12	11
12	10	9	12	10	13
12	14	10	10	16	12
11	11	14	14		

n = 100

Finding the Range

Years Experience

x		Frequency
6	Low	1
7		0
8		2
9		4
10		12
11		23
12		17
13		19
14		10
15		7
16		3
17	High	2

$\boxed{\text{Range = High - Low}}$

Range = 17 - 6 = 11

Characteristics of the Range

- Easy to Compute
- Uses Only Two Values (high and low)
- Sensitive to Extremes

Interquartile Range

- A measure of variation that is determined by computing the difference between the third and first quartiles
- Elimina

$$\textit{Interquartile Range} = Q3 - Q1$$

- Eliminates some high- and low-valued observations

Interquartile Range Example

Median

*X*minimum Q1 Q2 Q3

*X*maximum

25% 25% 25%
25%

12 30 45 57
70

Interquartile range =
57 – 30 = 27

Population Variance

- The average of the squared distances of the data values from the mean.

μ - population mean, N – population size

Population Variance - Example

Output

Dept. A	Dept. B
99	90
100	90
100	100
100	110
101	110

Mean = 100 **Mean = 100**

Measures of Dispersion
Production Volume Example

Dept. A		Dept. B	
x		x	
99		90	
100		90	
100	= 100	100	= 100
100		110	
101		110	

N = 5

N = 5

Departments the same in term of the center but differ in degree of dispersion

Computing the Variance
Department A

$$= \frac{2}{5} = .40$$

Dept. A

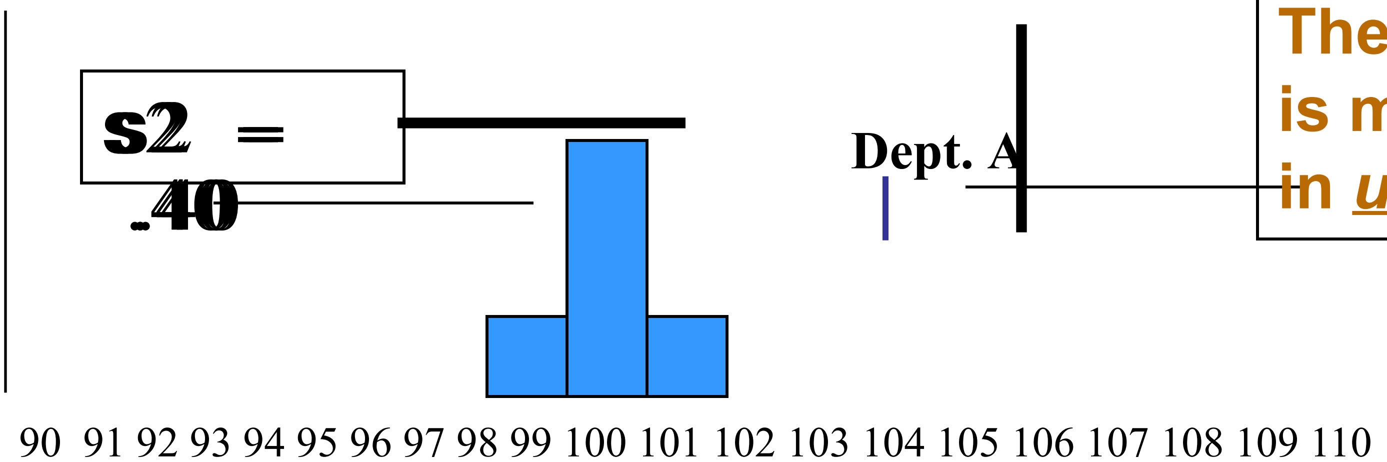
x	(x - m)	(x - m)2
99	99-100 = -1	1
100	100-100 = 0	0
100	100-100 = 0	0
100	100-100 = 0	0
101	101-100 = 1	1
	0	**2**

Computing the Variance Department B

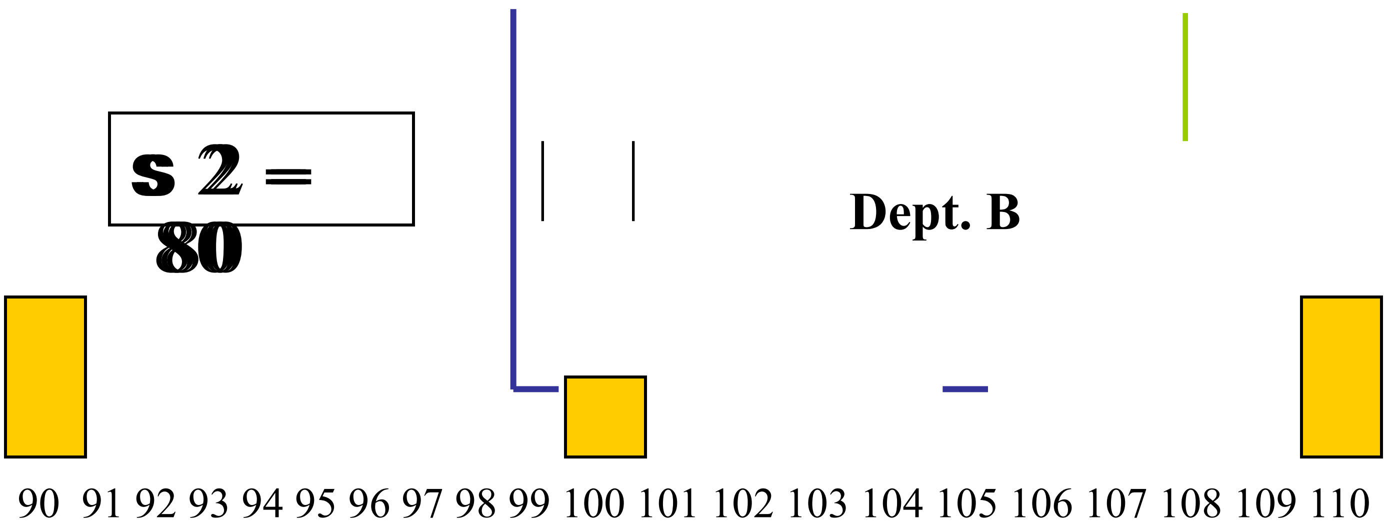
$$= \frac{400}{5} = 80$$

	x	(x - m)	x - m)2
Dept. B	90	90-100 = -10	100
	100	90-100 = -10	100
	100	100-100 = 0	0
	100	110-100 = 10	100
	101	110-100 = 10	100
		0	**400**

Comparing the Dispersion
Dept. A vs Dept. B



The variance is measured in *units squared*.



Population Standard Deviation

- The most commonly used measure of variation
- The positive square root of the variance
- Has the same units as the original data

Standard Deviation
(The Square Root of the Variance)

Population Standard Deviation

Dept. A

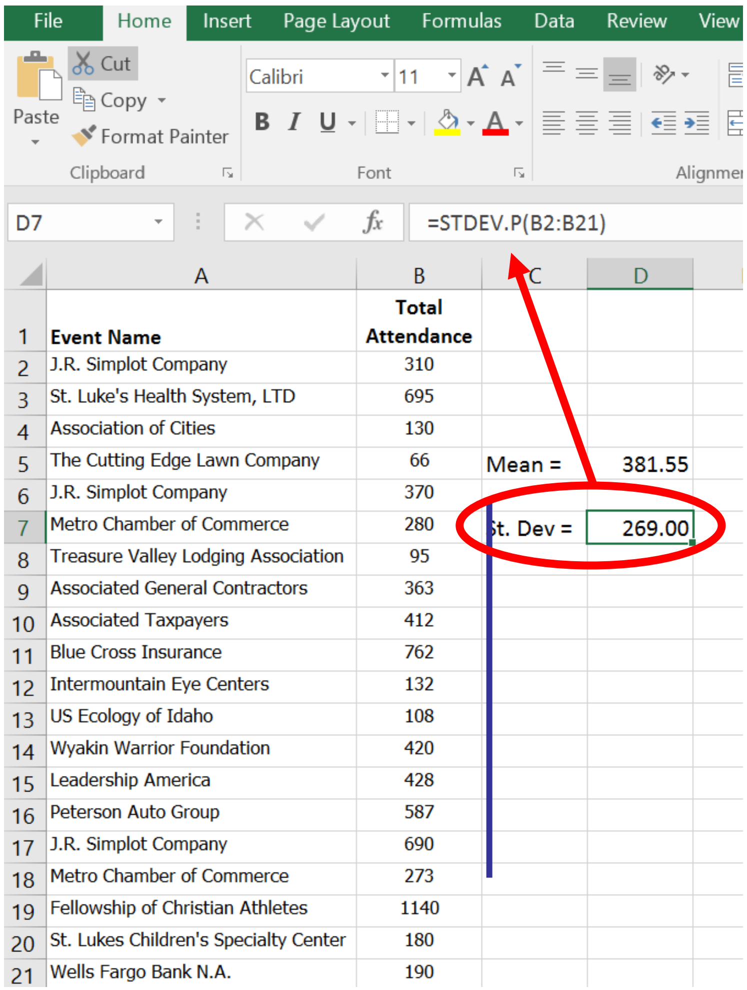
s2 = .40

Dept. B

s 2 = 80

Standard deviation is the measure of average deviation from the mean.

Population Standard Deviation - Example Using Excel 2016



D7 =STDEV.P(B2:B21)

	A	B	C	D
1	**Event Name**	**Total Attendance**		
2	J.R. Simplot Company	310		
3	St. Luke's Health System, LTD	695		
4	Association of Cities	130		
5	The Cutting Edge Lawn Company	66	Mean =	381.55
6	J.R. Simplot Company	370		
7	Metro Chamber of Commerce	280	St. Dev =	269.00
8	Treasure Valley Lodging Association	95		
9	Associated General Contractors	363		
10	Associated Taxpayers	412		
11	Blue Cross Insurance	762		
12	Intermountain Eye Centers	132		
13	US Ecology of Idaho	108		
14	Wyakin Warrior Foundation	420		
15	Leadership America	428		
16	Peterson Auto Group	587		
17	J.R. Simplot Company	690		
18	Metro Chamber of Commerce	273		
19	Fellowship of Christian Athletes	1140		
20	St. Lukes Children's Specialty Center	180		
21	Wells Fargo Bank N.A.	190		

The Convention Center hosted 20 events last year. This is the population.

The mean attendance is 381.55

The standard deviation is 269.00

Sample Standard Deviation

sample variance

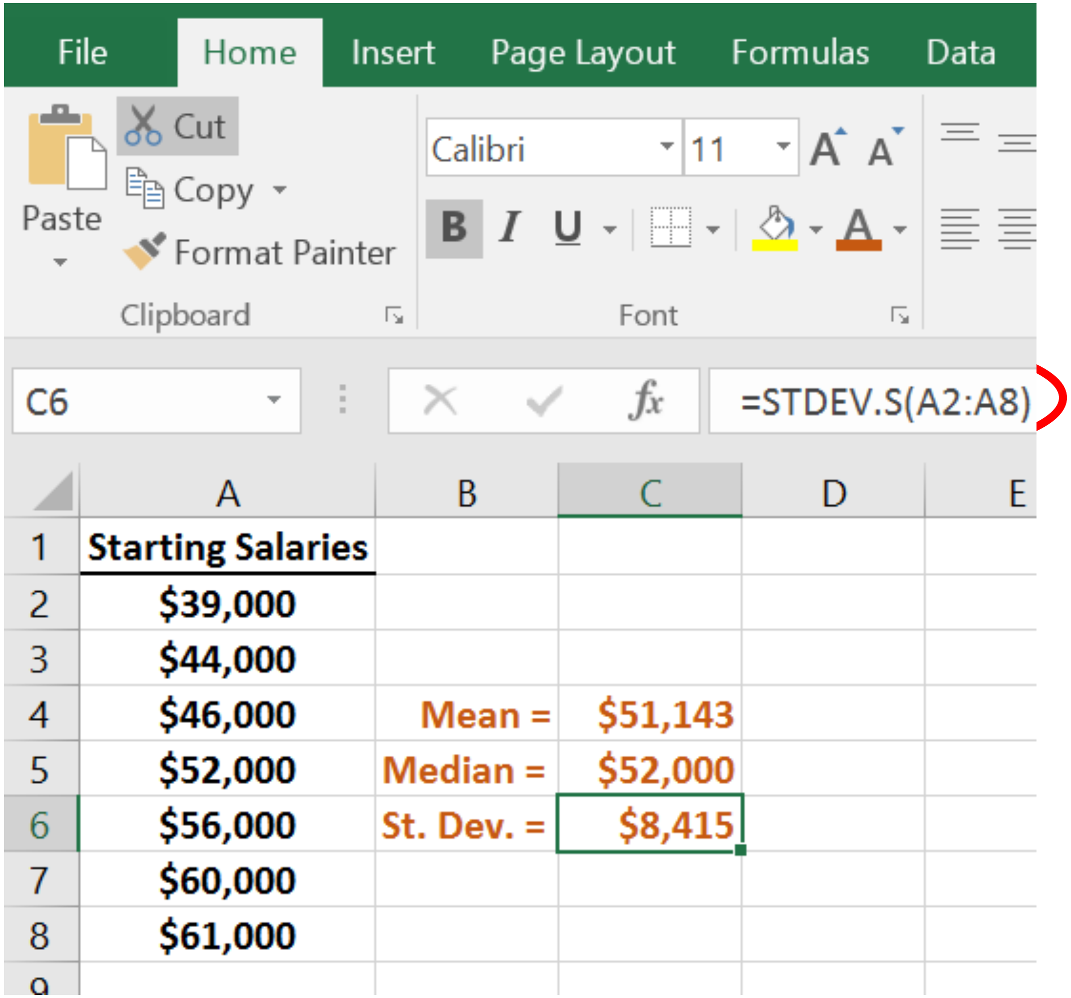
sample standard deviation

Sample Standard Deviation Restaurant Sales Example

Restaurant Sales Values (n = 7)

\$30	\$19.50	\$22.40	\$27	\$17.50	\$25.40	\$23.50

Sample Standard Deviation - Example Using Excel 2016



Sample of starting salaries for college graduates

n = 7

Sample Standard Deviation - Example Using Excel 2016

Fitness Club Membership Ages

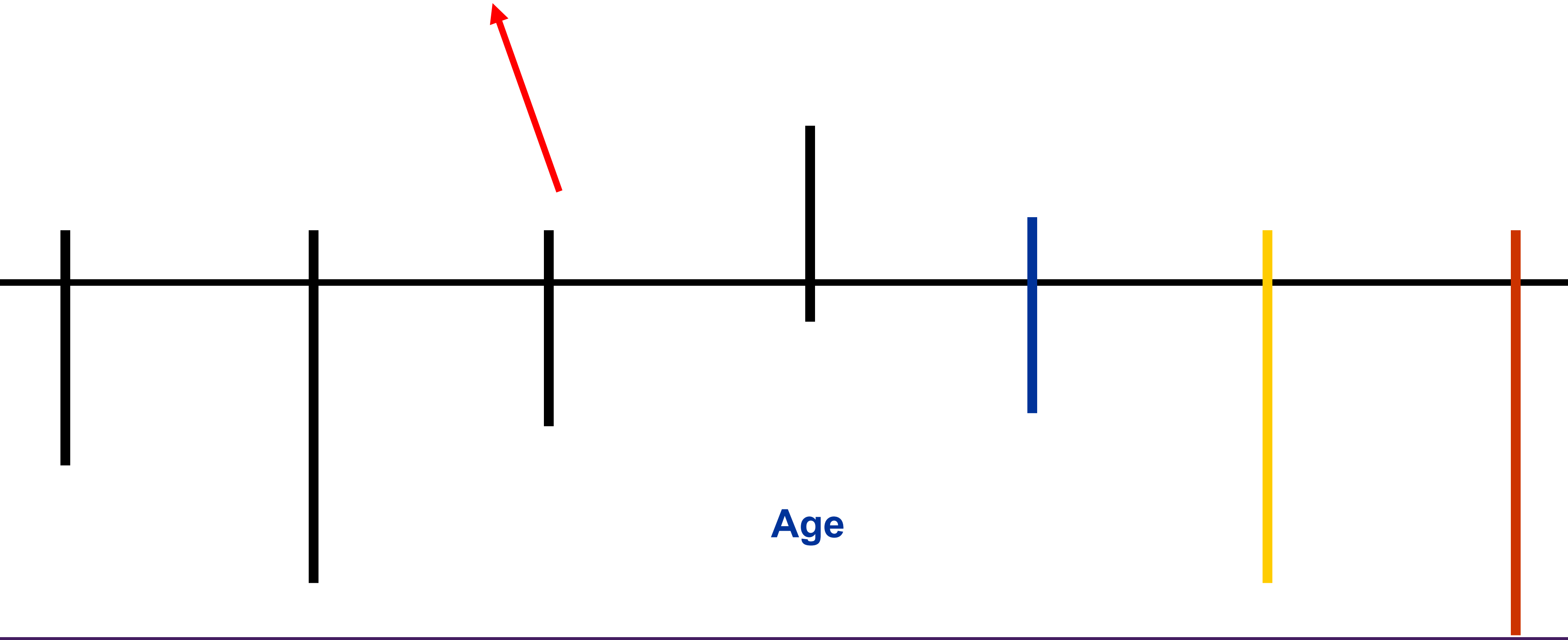
K6 =STDEV.S(H2:H1215)

	A	B	C	D	E	F	G	H
1	Weights & Exercise Equipment Satisfaction	Club Staff Satisfaction	Exercise Programs Satisfaction	Overall Service Satisfaction	Years With the Club	Gender	Typical Visits Per Week	Age
2	4	4	3	4	4	1	4	26
3	1	4	5	3	1	1	1	24
4	4	5	1	3	3	1	0	33
5	3	4	1	3	2	1	1	45
6	4	4	2	3	3	2	3	38
7	3	4	3	3	3.5	2	5	36
8	5	4	4	4	5	1	3	23
9	3	2	3	3	1	2	4	25
10	5	5	3	4	2	2	1	28
11	2	3	5	3	1.5	1	2	24
12	3	4	3	3	3.5	2	2	39
13	3	5	2	3	6	2	1	18

	J	K
4	n =	1,214
5	Mean =	34.84
6	St.Dev. =	11.18

Interpreting the Standard Deviation
Health Club Example

Standard deviation is a measure of distance from the mean.



Age

3.3 Using the Mean and Standard Deviation Together

- Coefficient of Variation (*CV*)
  - The ratio of the standard deviation to the mean expressed as a percentage. The coefficient of variation is used to measure variation relative to the mean.
  - Measures relative variation
  - Always expressed in percentage (%)
  - Shows variation relative to mean

Coefficient of Variation

- Is used to compare two or more sets of data measured in different units
  - Population *CV*
  - Sample *CV*

Comparing Coefficients of Variation

Stock A:

Average Price = $50 =

Standard Deviation = $5 = s

Stock B:

Average Price = $100 =

Standard Deviation = $5 = s

Both stocks have the same standard deviation, but stock B is less variable relative to its price.

The Empirical Rule

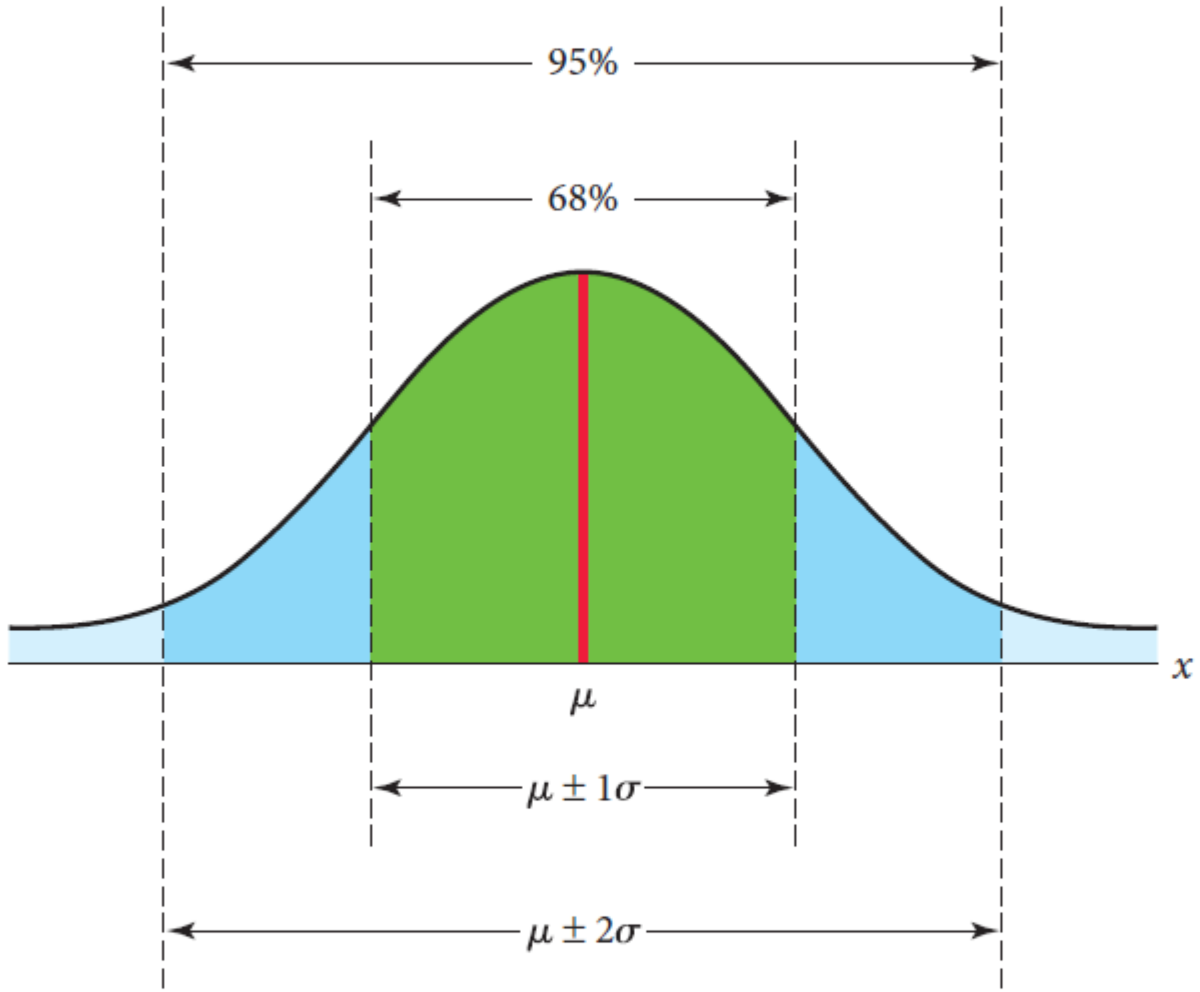
If the data distribution is bell shaped, then the interval

$\mu \pm 1\sigma$ contains approximately 68% of the values

$\mu \pm 2\sigma$ contains approximately 95% of the values

$\mu \pm 3\sigma$ contains virtually all of the data values.

The Empirical Rule



The Empirical Rule - Example

The travel distances for the Uber service in a major city is a bell shaped distributed with a mean of 15.1 miles and standard deviation of 3.1 miles per trip. Use the Empirical Rule to describe the distribution.

Assuming a bell shaped distribution, 68% percent of the trips should be between:

Assuming a bell shaped distribution, 95% percent of the trips should be between:

Standardized Data Values

- The number of standard deviations a value is from the mean
- Standardized data values are also referred to as *z* scores.
  - Population z score
  - Sample z score

x – data value
μ - population mean
σ – population standard deviation
$\bar{x}$ – sample mean
s – sample standard deviation

Converting Data to Standardized Values

- Step 1: Collect the population or sample values for the quantitative variable of interest.
- Step 2: Compute the population mean and standard deviation or the sample mean and standard deviation.
- Step 3: Convert the values to standardized *z*-values.

Standardized Value Calculation - Example

IQ Scores have a distribution thought to be bell shaped with a mean equal to 100 and standard deviation equal to 15. Suppose a person has an IQ of 121, calculate the standardized IQ.

A person with an IQ of 121 has an IQ that is 1.40 standard deviations higher than the population mean IQ.

Data Standardization - Example University Entrance Test

University A (100 point scale)	University B (1000 point scale)
89	890
60	450
45	356
88	798
56	560
92	880
.	.
.	.
.	.
.	.

Your Scores

How did you do on a relative basis at each university?

$m = 70$
$s = 10$

$m = 750$
$s = 70$

Entrance Test Scores

$$z = \frac{x - \mu}{\sigma}$$

University A

x	x - m	(x - m)/s
89	19	1.9 = Z
60	-10	-1.0
45	-25	-2.5
88	18	1.8
56	-14	-1.4
92	22	2.2
.	.	
.	.	

m = 70

s = 10

You

University B

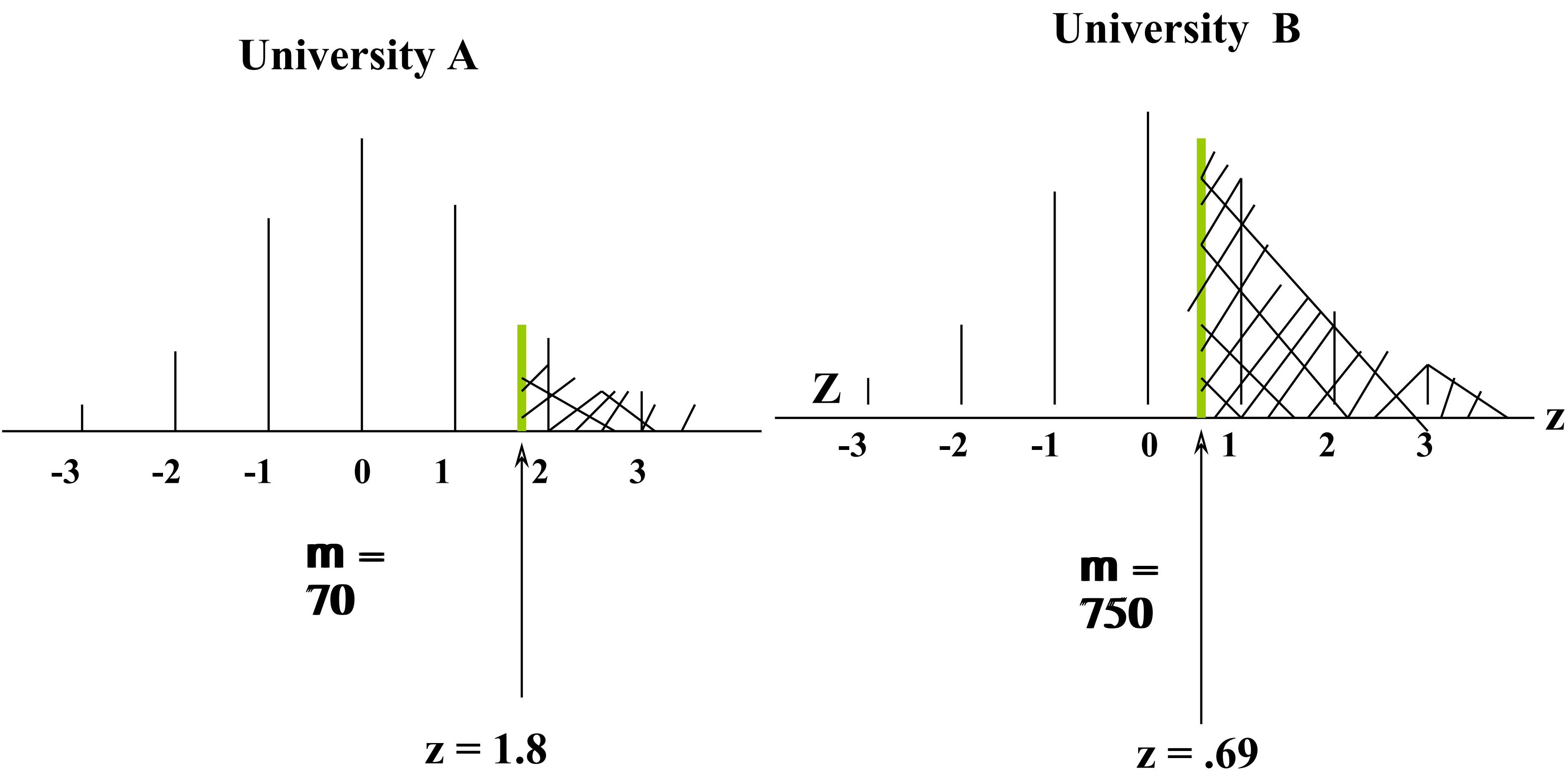
x	x - m	(x - m)/s
890	140	2.0
450	-300	-4.29
356	-394	-5.63
798	48	.69
560	-190	2.72
880	130	1.86
.		.
.		.

m = 750

s = 70

Entrance Score Was Relatively Better at University A than at University B

University Entrance Score Analysis



Coefficient of Variation (CV) – Example University Entrance Scores

The Coefficient of Variation measures the relative dispersion in a set of data. The CV is used to compare two or more sets of data with respect to variation.

Population

	University A	University B
	$m = 70$	$m = 750$
	$s = 10$	$s = 70$
	CV = 10/70 (100)	CV = 70/750 (100)
	= 14.2%	= 9.3%

University A **had more variable entrance scores.**

Calculator:

4: Sx → Sample Size

1: n → number of variables

2: $\bar{x}$ → Sample mean & $\overset{\text{sample}}{\text{Population}}$

3: σx → Population

	Population	Sample
Mean	$\mu = \frac{\Sigma X}{N}$	$\bar{X} = \frac{\Sigma X}{n}$
Var (varation)	$\sigma^2 = \frac{\Sigma (X - \mu)^2}{N}$	$S^2 = \frac{\Sigma (X - \bar{X})^2}{n-1}$
Standard deviation	$\sigma = \sqrt{\sigma^2}$	$S = \sqrt{S^2}$