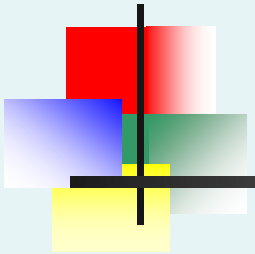
Statistics for Managers Using Microsoft Excel

7th Edition



Chapter 3

Numerical Descriptive Measures

Learning Objectives

In this chapter, you learn:

- To describe the properties of central tendency, variation, and shape in numerical data
- To compute descriptive summary measures for a population
- To construct and interpret a boxplot
- To calculate the covariance and the coefficient of correlation

Summary Definitions

DCOVA

- The **central tendency** is the extent to which all the data values group around a typical or central value.
- The **variation** is the amount of dispersion or scattering of values
- The **shape** is the pattern of the distribution of values from the lowest value to the highest value.

Measures of Central Tendency: The Mean

DCOVA

- The arithmetic mean (often just called the "mean") is the most common measure of central tendency

  - For a sample of size n:

Pronounced x-bar

The i^{th} value

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Sample size

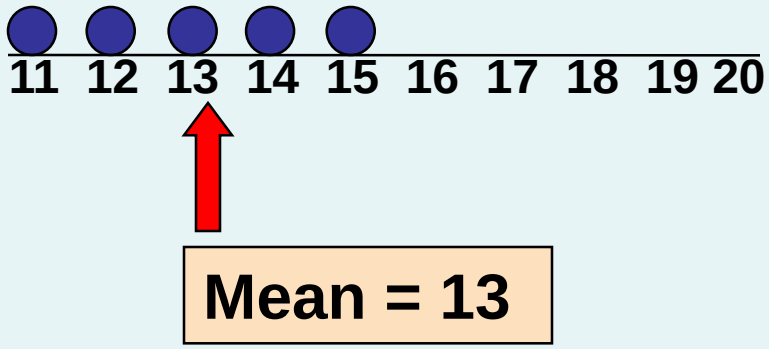
Observed values

Measures of Central Tendency: The Mean

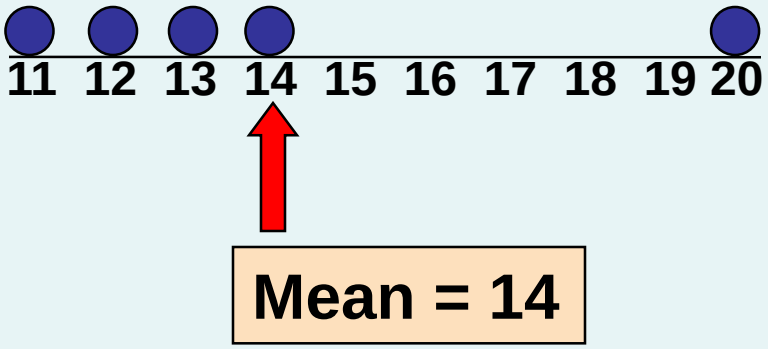
DCOVA

(continued)

- The most common measure of central tendency
- Mean = sum of values divided by the number of values
- Affected by extreme values (outliers)



$$\frac{11+12+13+14+15}{5}=\frac{65}{5}=13$$

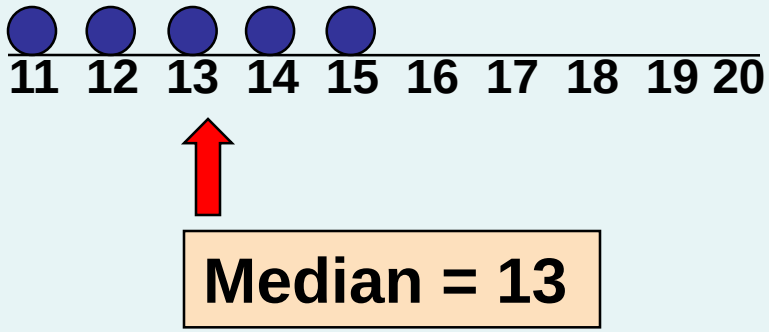


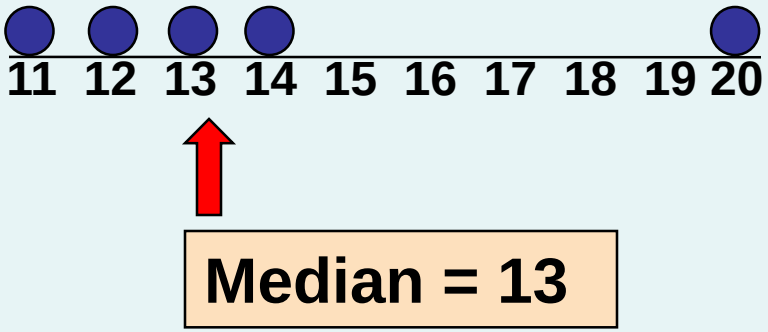
$$\frac{11+12+13+14+20}{5}=\frac{70}{5}=14$$

Measures of Central Tendency: The Median

DCOVA

- In an ordered array, the median is the “middle” number (50% above, 50% below)





- Not affected by extreme values

Measures of Central Tendency: Locating the Median

DCOVA

- The location of the median when the values are in numerical order (smallest to largest):

$$\text{Median position} = \frac{n+1}{2} \text{ position in the ordered data}$$

- If the number of values is odd, the median is the middle number

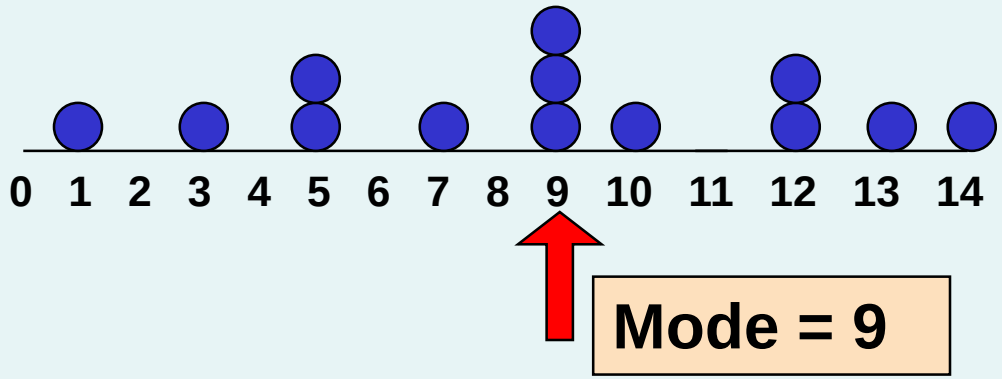
- If the number of values is even, the median is the average of the two middle numbers

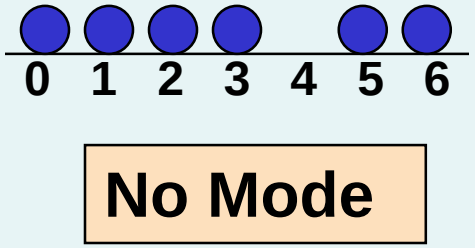
Note that $\frac{n+1}{2}$ is not the *value* of the median, only the *position* of the median in the ranked data

Measures of Central Tendency: The Mode

DCOVA

- Value that occurs most often
- Not affected by extreme values
- Used for either numerical or categorical (nominal) data
- There may may be no mode
- There may be several modes





Measures of Central Tendency: Review Example

DCOVA

House Prices:

$2,000,000
$ 500,000
$ 300,000
$ 100,000
$ 100,000

Sum $ **3,000,000**

- **Mean:** ($3,000,000/5)
  = **$600,000**
- **Median:** middle value of ranked data
  = **$300,000**
- **Mode:** most frequent value
  = **$100,000**

Measures of Central Tendency: Which Measure to Choose?

DCOVA

- The **mean** is generally used, unless extreme values (outliers) exist.
- The **median** is often used, since the median is not sensitive to extreme values. For example, median home prices may be reported for a region; it is less sensitive to outliers.
- In some situations it makes sense to report both the **mean** and the **median**.

Measure of Central Tendency For The Rate Of Change Of A Variable Over Time: The Geometric Mean & The Geometric Rate of Return

DCOVA

- Geometric mean
  - Used to measure the rate of change of a variable over time

$$\overline{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n}$$

- Geometric mean rate of return
  - Measures the status of an investment over time

$$\overline{R}_G = [(1+R_1) \times (1+R_2) \times \cdots \times (1+R_n)]^{1/n} - 1$$

  - Where R_i is the rate of return in time period i

The Geometric Mean Rate of Return: Example

DCOVA

An investment of $100,000 declined to $50,000 at the end of year one and rebounded to $100,000 at end of year two:

$$X_1 = \$100{,}000 \quad X_2 = \$50{,}000 \quad X_3 = \$100{,}000$$

50% decrease

100% increase

The overall two-year return is zero, since it started and ended at the same level.

The Geometric Mean Rate of Return: Example

(continued)

DCOVA

Use the 1-year returns to compute the arithmetic mean and the geometric mean:

Arithmetic mean rate of return:

$$\overline{X} = \frac{(-.5)+(1)}{2} = .25 = 25\%$$

Misleading result

Geometric mean rate of return:

$$\overline{R}_G = [(1+R_1)\times(1+R_2)\times\cdots\times(1+R_n)]^{1/n} - 1$$

$$= [(1+(-.5))\times(1+(1))]^{1/2} - 1$$

$$= [(.50)\times(2)]^{1/2} - 1 = 1^{1/2} - 1 = 0\%$$

More representative result

Measures of Central Tendency: Summary

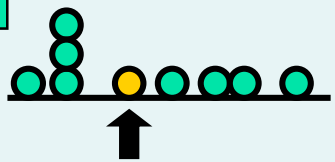
DCOVA

Central Tendency

Arithmetic Mean

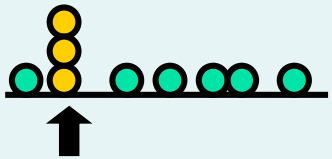
$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Median



Middle value in the ordered array

Mode



Most frequently observed value

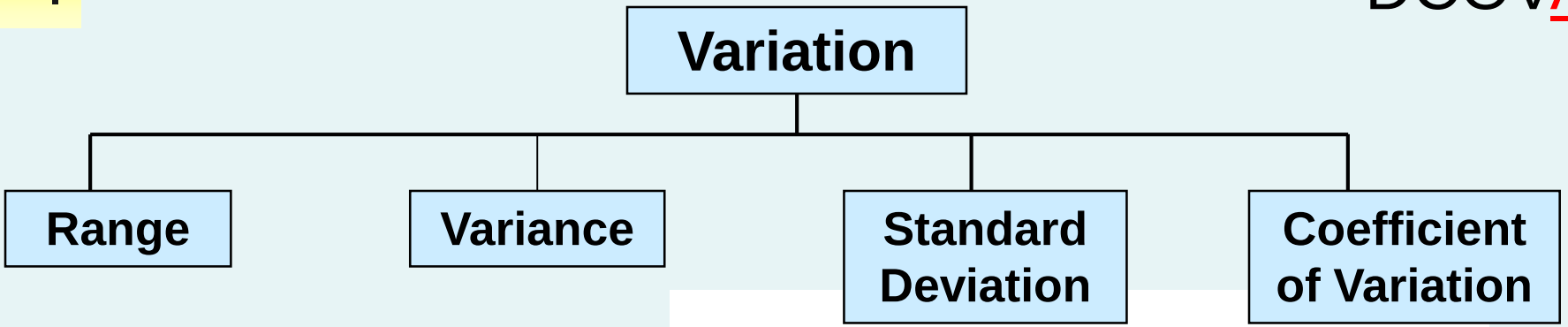
Geometric Mean

$$\overline{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n}$$

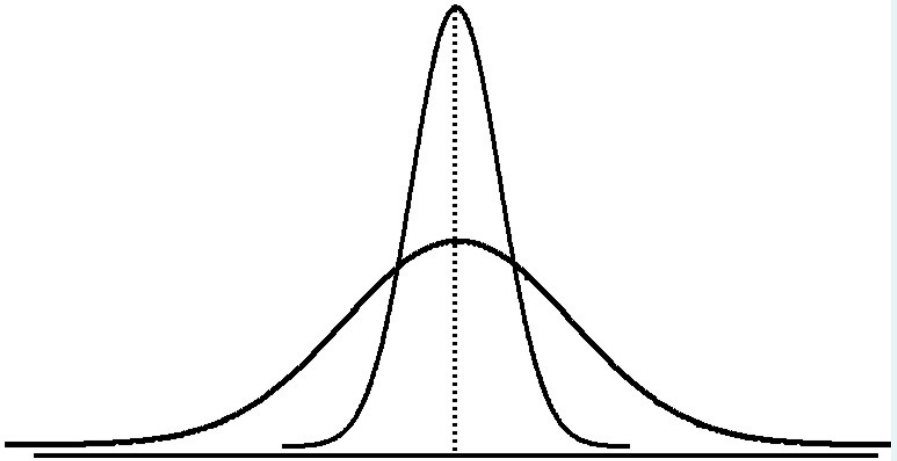
Rate of change of a variable over time

Measures of Variation

DCOVA



- Measures of variation give information on the **spread** or **variability** or **dispersion** of the data values.



Same center, different variation

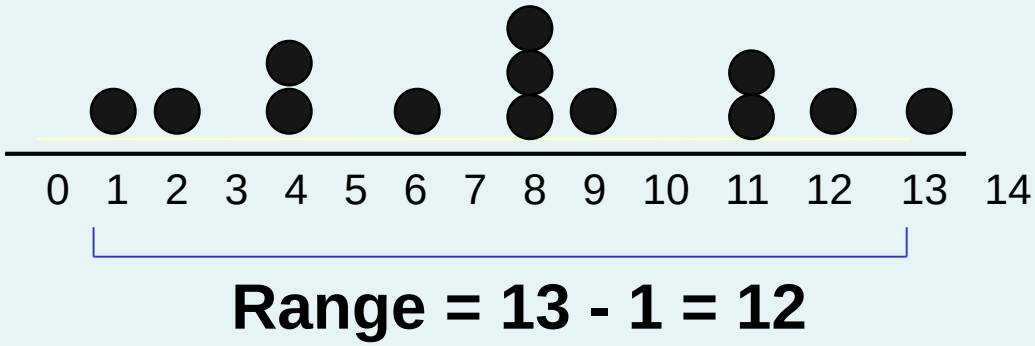
Measures of Variation: The Range

DCOVA

- Simplest measure of variation
- Difference between the largest and the smallest values:

$$\text{Range} = X_{\text{largest}} - X_{\text{smallest}}$$

Example:

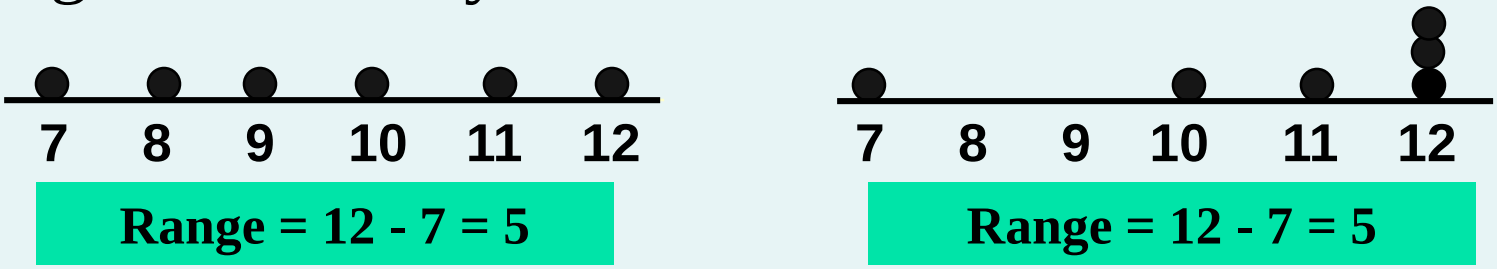


Range = 13 - 1 = 12

Measures of Variation: Why The Range Can Be Misleading

DCOVA

- Ignores the way in which data are distributed



- Sensitive to outliers

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,**5**

Range = 5 - 1 = 4

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,**120**

Range = 120 - 1 = 119

Measures of Variation: The Sample Variance

DCOVA

- Average (approximately) of squared deviations of values from the mean

  - Sample variance:

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

Where $\overline{X}$ = arithmetic mean

n = sample size

X_i = i^{th} value of the variable X

Measures of Variation: The Sample Standard Deviation

DCOVA

- Most commonly used measure of variation
- Shows variation about the mean
- Is the square root of the variance
- Has the same units as the original data

  - Sample standard deviation:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Measures of Variation: The Standard Deviation

DCOVA

Steps for Computing Standard Deviation

1. Compute the difference between each value and the mean.
2. Square each difference.
3. Add the squared differences.
4. Divide this total by n-1 to get the sample variance.
5. Take the square root of the sample variance to get the sample standard deviation.

Measures of Variation: Sample Standard Deviation: Calculation Example

DCOVA

Sample Data (X_i) :

10	12	14	15	17	18	18	24

n = 8 Mean = $\overline{X}$ = 16

$$S = \sqrt{\frac{(10-\overline{X})^2+(12-\overline{X})^2+(14-\overline{X})^2+\cdots+(24-\overline{X})^2}{n-1}}$$

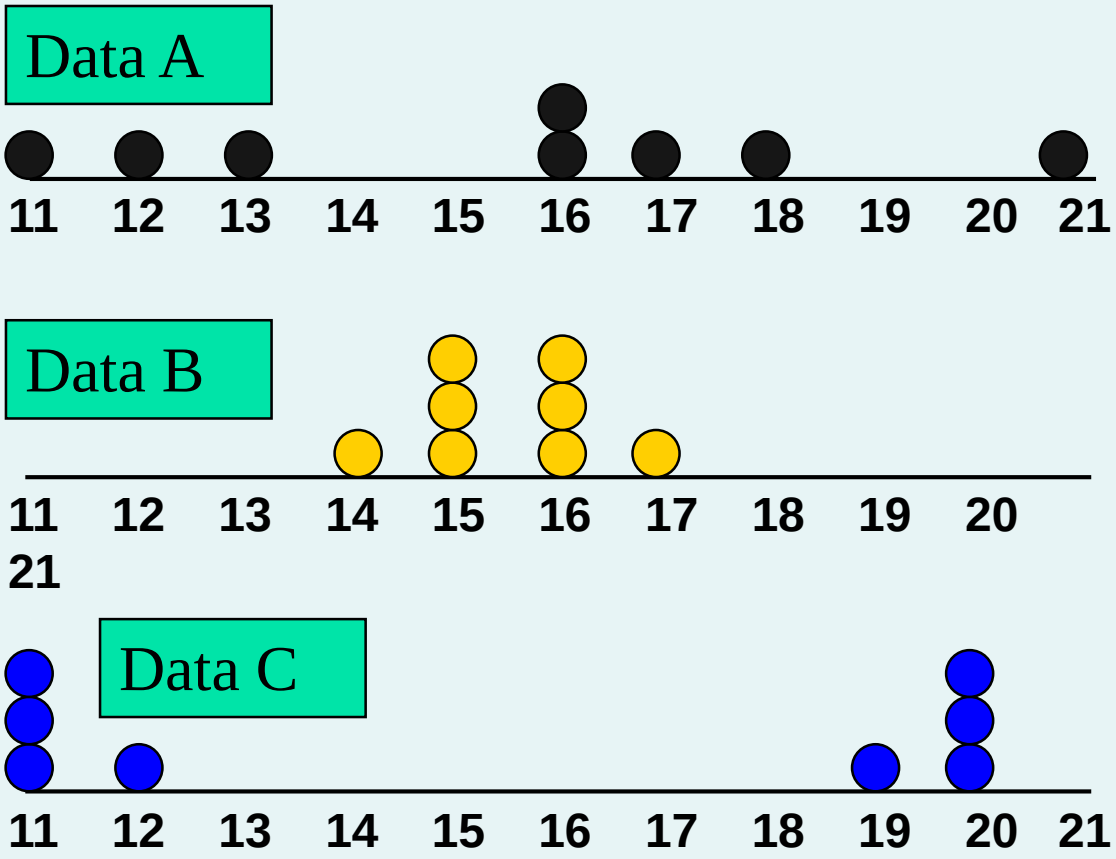
$$= \sqrt{\frac{(10-16)^2+(12-16)^2+(14-16)^2+\cdots+(24-16)^2}{8-1}}$$

$$= \sqrt{\frac{130}{7}} = 4.3095$$

A measure of the "average" scatter around the mean

Measures of Variation: Comparing Standard Deviations

DCOVA



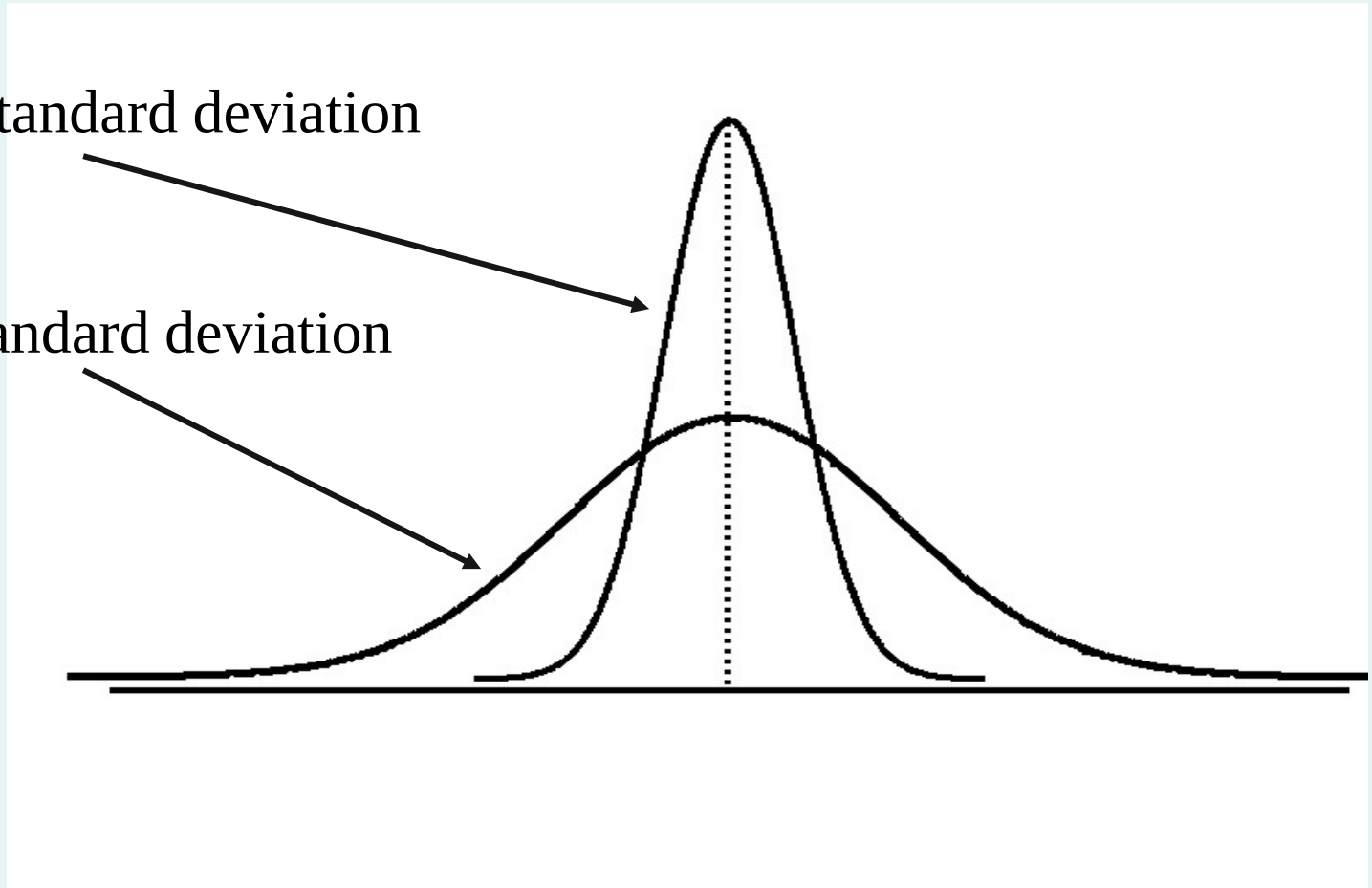
Mean = 15.5
S = 3.338

Mean = 15.5
S = 0.926

Mean = 15.5
S = 4.567

Measures of Variation: Comparing Standard Deviations

DCOVA



Measures of Variation: Summary Characteristics

DCOVA

- The more the data are spread out, the greater the range, variance, and standard deviation.
- The more the data are concentrated, the smaller the range, variance, and standard deviation.
- If the values are all the same (no variation), all these measures will be zero.
- None of these measures are ever negative.

Measures of Variation: The Coefficient of Variation

DCOVA

- Measures relative variation
- Always in percentage (%)
- Shows variation relative to mean
- Can be used to compare the variability of two or more sets of data measured in different units

$$CV = \left(\frac{S}{\bar{X}} \right) \cdot 100\%$$

Measures of Variation: Comparing Coefficients of Variation

DCOVA

- Stock A:
  - Average price last year = \$50
  - Standard deviation = \$5

$$CV_A = \left(\frac{S}{\bar{X}}\right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

- Stock B:
  - Average price last year = \$100
  - Standard deviation = \$5

$$CV_B = \left(\frac{S}{\bar{X}}\right) \cdot 100\% = \frac{\$5}{\$100} \cdot 100\% = 5\%$$

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

Measures of Variation: Comparing Coefficients of Variation

(continued)

DCOVA

- Stock A:
  - Average price last year = \$50
  - Standard deviation = \$5

$$CV_A = \left(\frac{S}{\overline{X}} \right) \cdot 100\% = \frac{\$5}{\$50} \cdot 100\% = 10\%$$

- Stock C:
  - Average price last year = \$8
  - Standard deviation = \$2

$$CV_C = \left(\frac{S}{\overline{X}} \right) \cdot 100\% = \frac{\$2}{\$8} \cdot 100\% = 25\%$$

Stock C has a much smaller standard deviation but a much higher coefficient of variation

Locating Extreme Outliers: Z-Score

DCOVA

- To compute the Z-**score** of a data value, subtract the mean and divide by the standard deviation.
- The Z-score is the number of standard deviations a data value is from the mean.
- A data value is considered an extreme outlier if its Z-score is less than -3.0 or greater than +3.0.
- The larger the absolute value of the Z-score, the farther the data value is from the mean.

Locating Extreme Outliers: Z-Score

DCOVA

$$Z = \frac{X - \overline{X}}{S}$$

where X represents the data value

$\overline{X}$ is the sample mean

S is the sample standard deviation

Locating Extreme Outliers: Z-Score

- Suppose the mean math SAT score is 490, with a standard deviation of 100.
- Compute the Z-score for a test score of 620.

$$Z = \frac{X - \overline{X}}{S} = \frac{620 - 490}{100} = \frac{130}{100} = 1.3$$

A score of 620 is 1.3 standard deviations above the mean and would not be considered an outlier.

Shape of a Distribution

DCOVA

- Describes how data are distributed
- Two useful shape related statistics are:
  - Skewness
    - Measures the extent to which data values are not symmetrical
  - Kurtosis
    - Kurtosis affects the peakedness of the curve of the distribution—that is, how sharply the curve rises approaching the center of the distribution

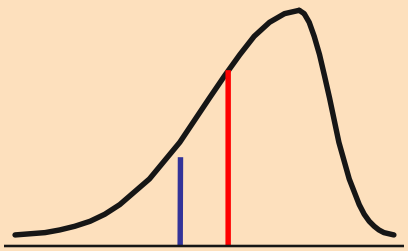
Shape of a Distribution (Skewness)

DCOVA

- Measures the extent to which data is not symmetrical

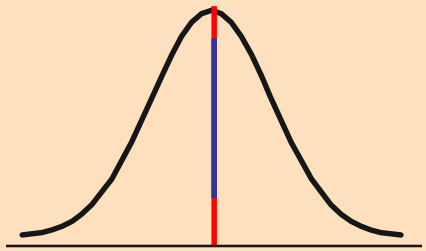
Left-Skewed

Mean < Median



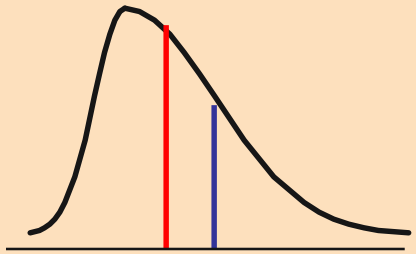
Symmetric

Mean = Median



Right-Skewed

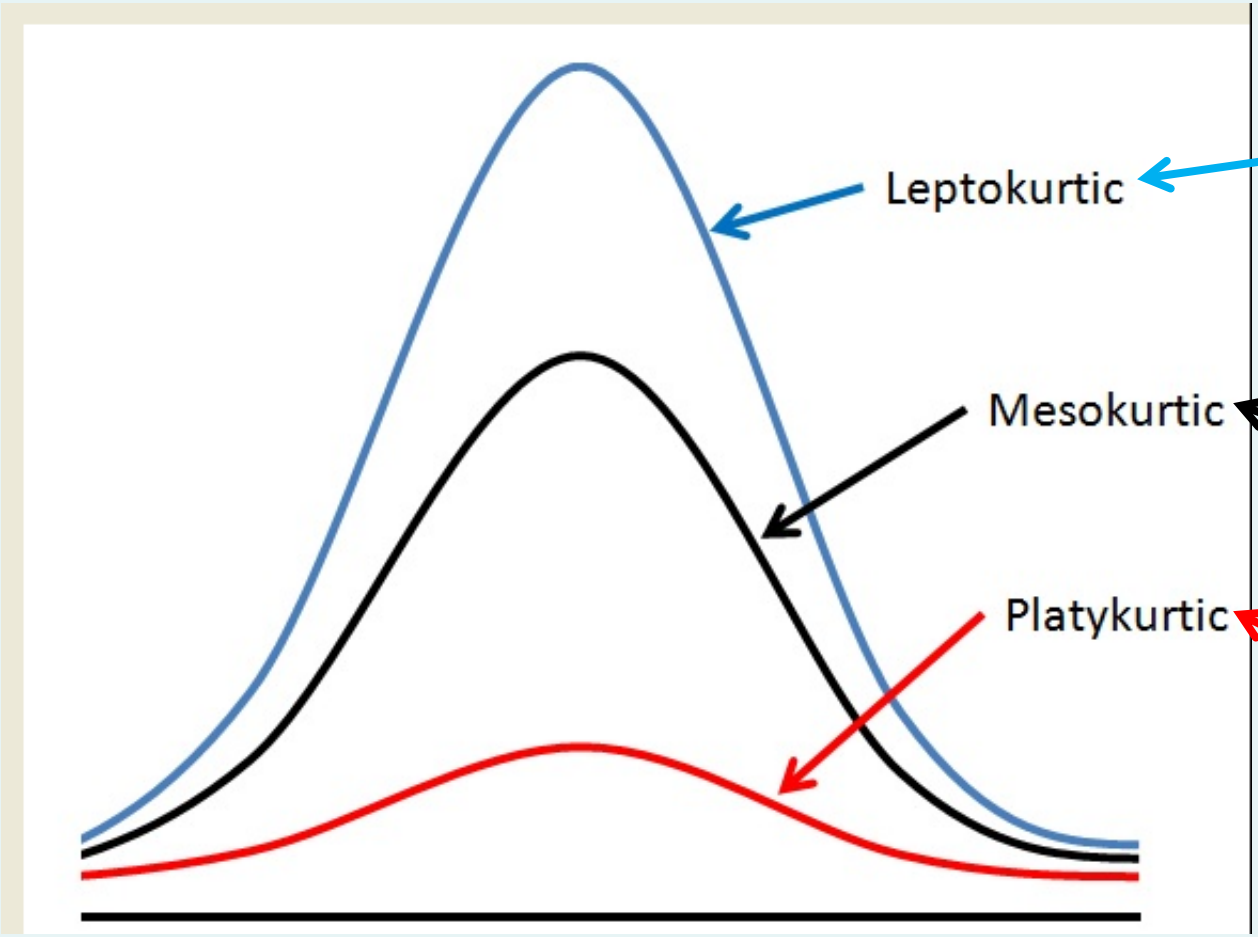
Median < Mean



Skewness Statistic	< 0	0	>0

Shape of a Distribution -- Kurtosis measures how sharply the curve rises approaching the center of the distribution)

DCOVA



Sharper Peak Than Bell-Shaped (Kurtosis > 0)

Bell-Shaped (Kurtosis = 0)

Flatter Than Bell-Shaped (Kurtosis < 0)

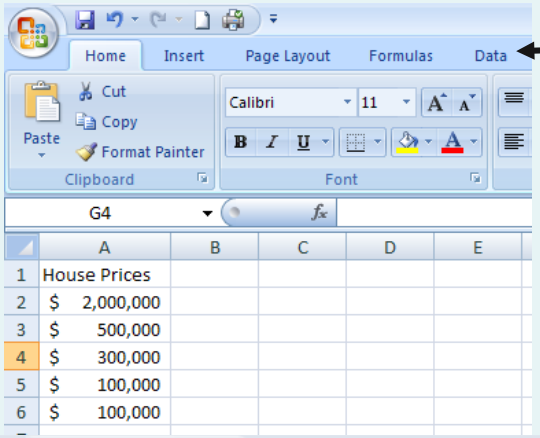
General Descriptive Stats Using Microsoft Excel Functions

DCOVA

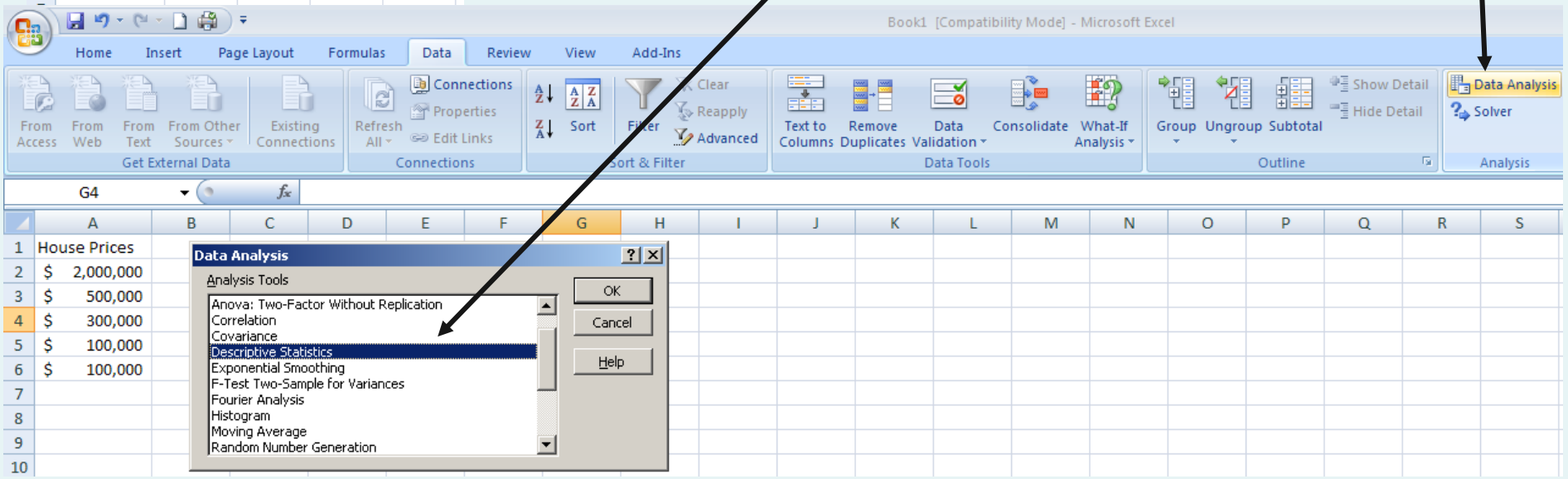
House Prices		Descriptive Statistics		
$ 2,000,000		Mean	$ 600,000	=AVERAGE(A2:A6)
$ 500,000		Standard Error	$ 357,770.88	=D6/SQRT(D14)
$ 300,000		Median	$ 300,000	=MEDIAN(A2:A6)
$ 100,000		Mode	$ 100,000.00	=MODE(A2:A6)
$ 100,000		Standard Deviation	$ 800,000	=STDEV(A2:A6)
		Sample Variance	640,000,000,000	=VAR(A2:A6)
		Kurtosis	4.1301	=KURT(A2:A6)
		Skewness	2.0068	=SKEW(A2:A6)
		Range	$ 1,900,000	=D12 - D11
		Minimum	$ 100,000	=MIN(A2:A6)
		Maximum	$ 2,000,000	=MAX(A2:A6)
		Sum	$ 3,000,000	=SUM(A2:A6)
		Count	5	=COUNT(A2:A6)

General Descriptive Stats Using Microsoft Excel Data Analysis Tool

DCOVA



1. Select Data.
2. Select Data Analysis.
3. Select Descriptive Statistics and click OK.



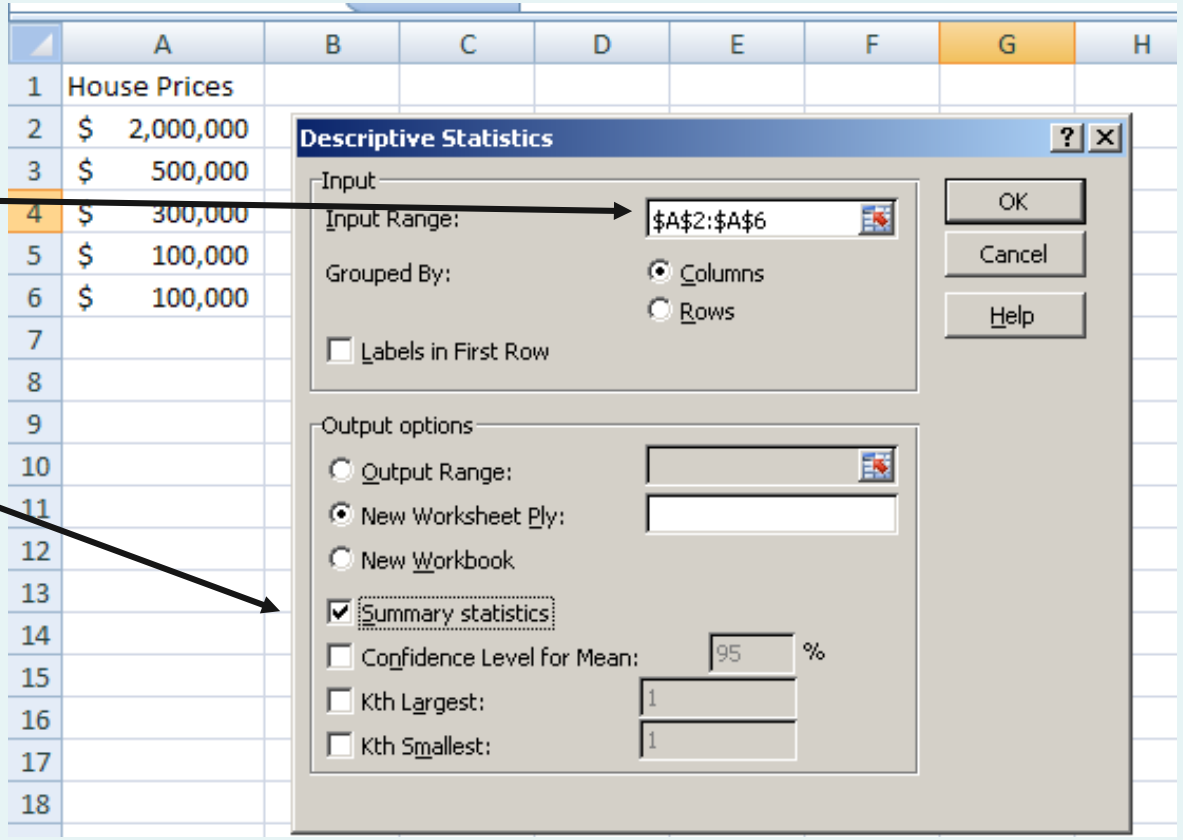
General Descriptive Stats Using Microsoft Excel

DCOVA

4. Enter the cell range.

5. Check the Summary Statistics box.

6. Click OK



Excel output

DCOVA

Microsoft Excel descriptive statistics output, using the house price data:

House Prices:

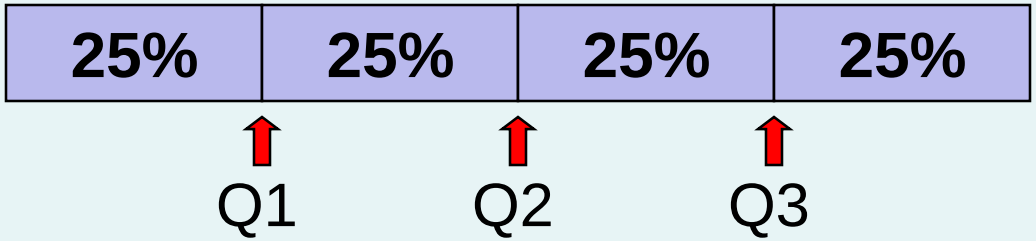
$2,000,000
500,000
300,000
100,000
100,000

House Prices	
Mean	600000
Standard Error	357770.8764
Median	300000
Mode	100000
Standard Deviation	800000
Sample Variance	640,000,000,000
Kurtosis	4.1301
Skewness	2.0068
Range	1900000
Minimum	100000
Maximum	2000000
Sum	3000000
Count	5

Quartile Measures

DCOVA

- Quartiles split the ranked data into 4 segments with an equal number of values per segment



- The first quartile, Q_1, is the value for which 25% of the observations are smaller and 75% are larger
- Q_2 is the same as the median (50% of the observations are smaller and 50% are larger)
- Only 25% of the observations are greater than the third quartile

Quartile Measures: Locating Quartiles

DCOVA

Find a quartile by determining the value in the appropriate position in the ranked data, where

First quartile position: $Q_1 = (n+1)/4$ ranked value

Second quartile position: $Q_2 = (n+1)/2$ ranked value

Third quartile position: $Q_3 = 3(n+1)/4$ ranked value

where **n** is the number of observed values

Quartile Measures: Calculation Rules

- When calculating the ranked position use the following rules
  - If the result is a whole number then it is the ranked position to use
  - If the result is a fractional half (e.g. 2.5, 7.5, 8.5, etc.) then average the two corresponding data values.
  - If the result is not a whole number or a fractional half then round the result to the nearest integer to find the ranked position.

Quartile Measures: Locating Quartiles

DCOVA

Sample Data in Ordered Array: **11 12 13 16 16 17 18 21 22**

(n = 9)

Q_1 is in the (9+1)/4 = 2.5 position of the ranked data

so use the value half way between the 2nd and 3rd values,

so **$Q_1 = 12.5$**

Q_1 and Q_3 are measures of non-central location
Q_2 = median, is a measure of central tendency

Quartile Measures
Calculating The Quartiles: Example

DCOVA

Sample Data in Ordered Array: 11 12 13 16 16 17 18 21 22

(n = 9)

Q_1 is in the (9+1)/4 = 2.5 position of the ranked data,

so **Q_1 = (12+13)/2 = 12.5**

Q_2 is in the (9+1)/2 = 5th position of the ranked data,

so **Q_2 = median = 16**

Q_3 is in the 3(9+1)/4 = 7.5 position of the ranked data,

so **Q_3 = (18+21)/2 = 19.5**

Q_1 and Q_3 are measures of non-central location
Q_2 = median, is a measure of central tendency

Quartile Measures: The Interquartile Range (IQR)

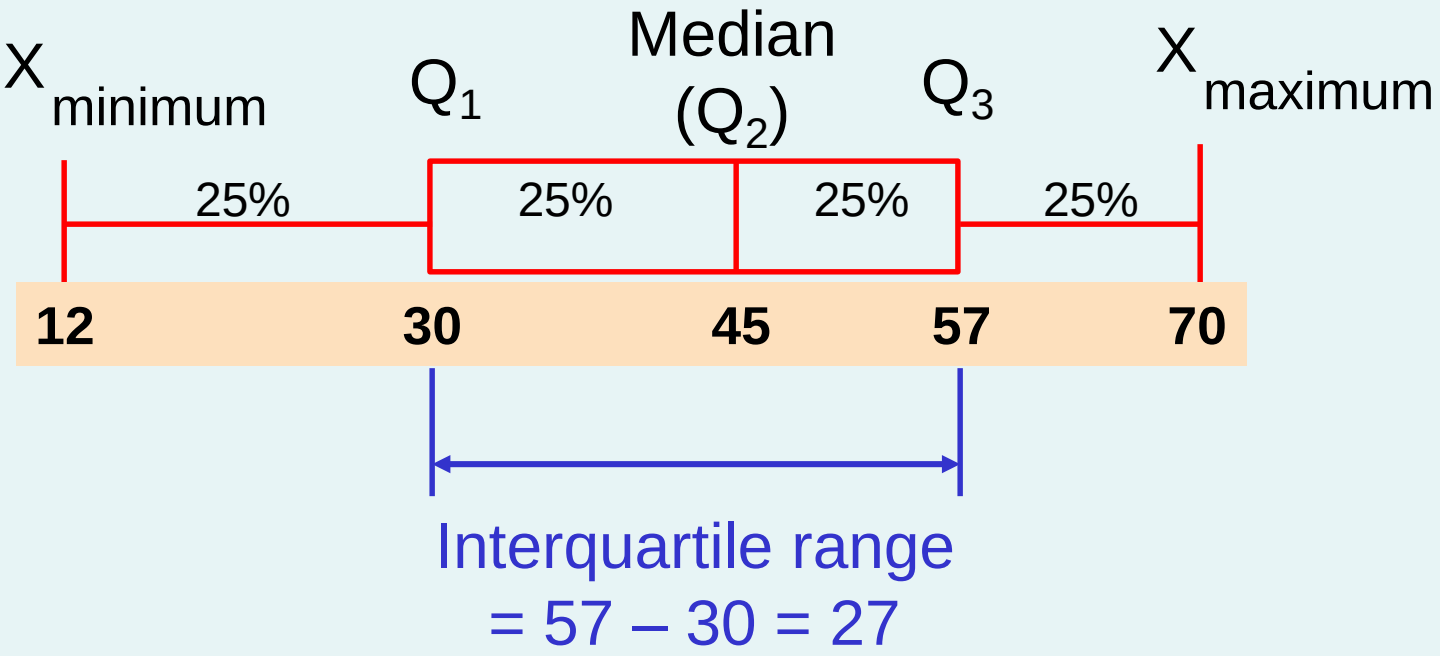
DCOVA

- The IQR is $Q_3 - Q_1$ and measures the spread in the middle 50% of the data
- The IQR is also called the midspread because it covers the middle 50% of the data
- The IQR is a measure of variability that is not influenced by outliers or extreme values
- Measures like Q_1, Q_3, and IQR that are not influenced by outliers are called resistant measures

Calculating The Interquartile Range

DCOVA

Example:



The Five Number Summary

DCOVA

The five numbers that help describe the center, spread and shape of data are:

- $X_{smallest}$
- First Quartile (Q_1)
- Median (Q_2)
- Third Quartile (Q_3)
- $X_{largest}$

Relationships among the five-number summary and distribution shape

DCOVA

Left-Skewed	Symmetric	Right-Skewed
$\text{Median} - X_{\text{smallest}}$	$\text{Median} - X_{\text{smallest}}$	$\text{Median} - X_{\text{smallest}}$
$>$	$\approx$	$<$
$X_{\text{largest}} - \text{Median}$	$X_{\text{largest}} - \text{Median}$	$X_{\text{largest}} - \text{Median}$
$Q_1 - X_{\text{smallest}}$	$Q_1 - X_{\text{smallest}}$	$Q_1 - X_{\text{smallest}}$
$>$	$\approx$	$<$
$X_{\text{largest}} - Q_3$	$X_{\text{largest}} - Q_3$	$X_{\text{largest}} - Q_3$
$\text{Median} - Q_1$	$\text{Median} - Q_1$	$\text{Median} - Q_1$
$>$	$\approx$	$<$
$Q_3 - \text{Median}$	$Q_3 - \text{Median}$	$Q_3 - \text{Median}$

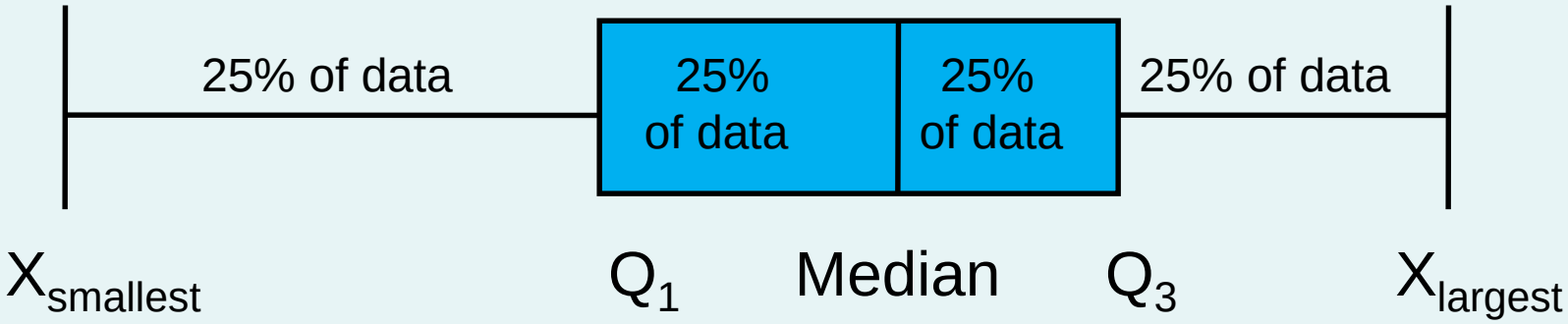
Five Number Summary and The Boxplot

DCOVA

- The Boxplot: A Graphical display of the data based on the five-number summary:

$X_{smallest}$ -- Q_1 -- Median -- Q_3 -- $X_{largest}$

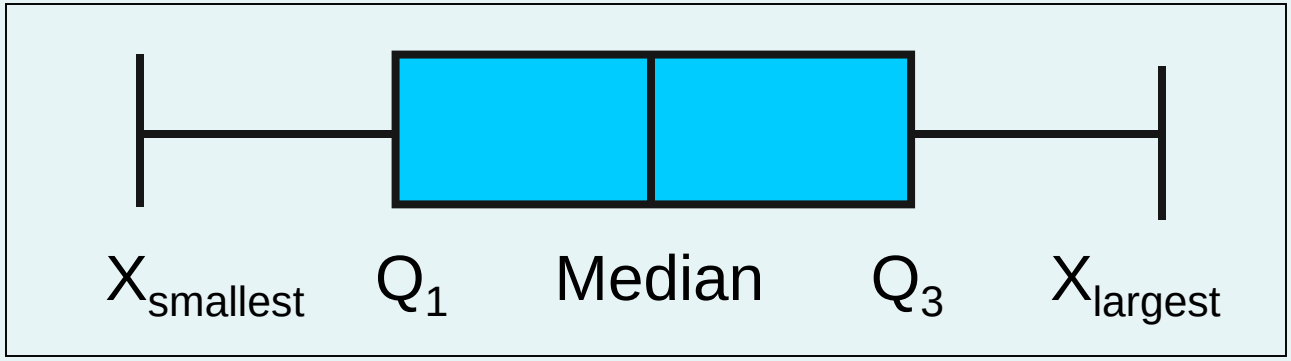
Example:



Five Number Summary: Shape of Boxplots

DCOVA

- If data are symmetric around the median then the box and central line are centered between the endpoints

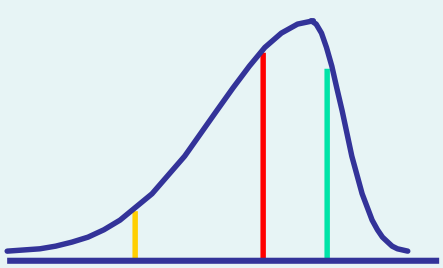


- A Boxplot can be shown in either a vertical or horizontal orientation

Distribution Shape and The Boxplot

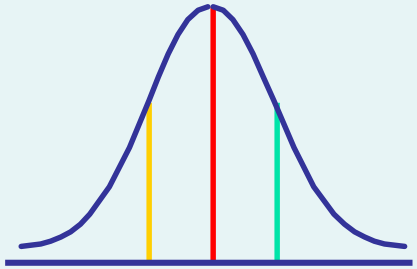
DCOVA

Left-Skewed



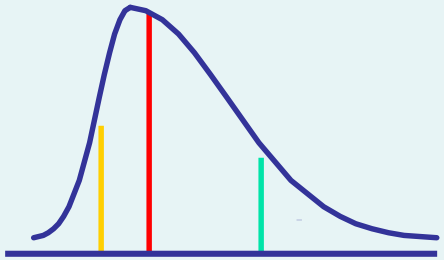
Q_1 Q_2 Q_3

Symmetric

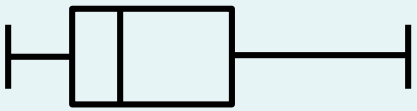


Q_1 Q_2 Q_3

Right-Skewed



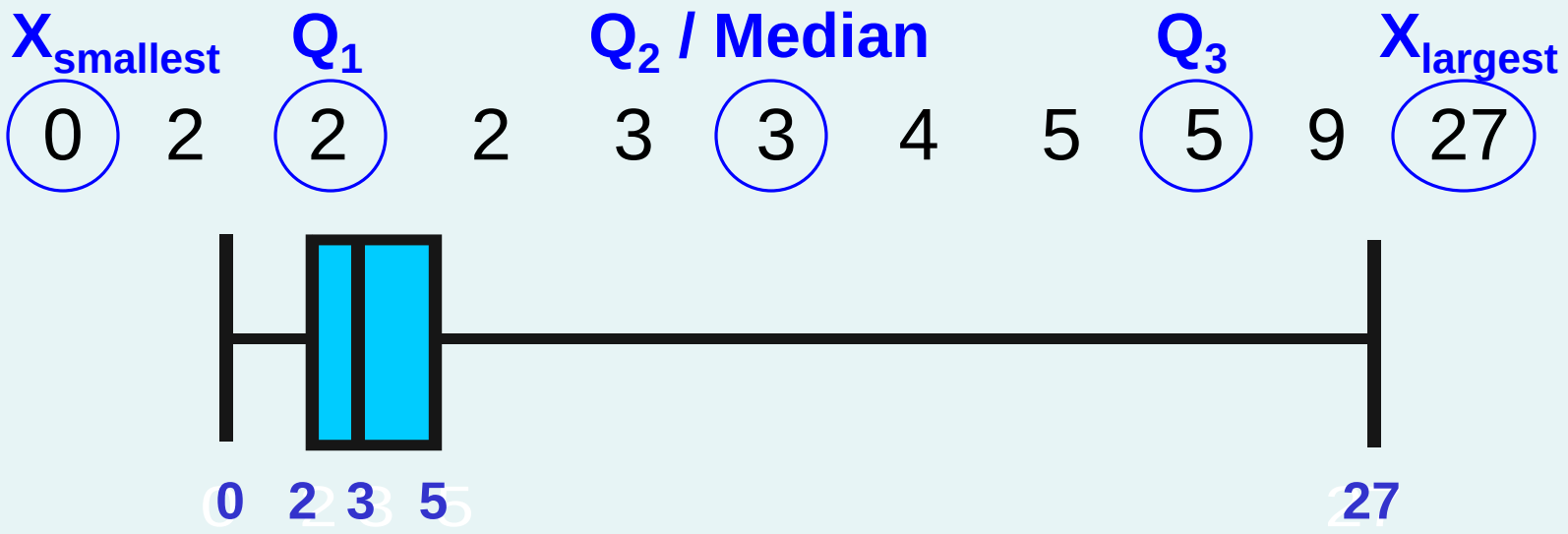
Q_1 Q_2 Q_3



Boxplot Example

DCOVA

- Below is a Boxplot for the following data:



- The data are right skewed, as the plot depicts

Numerical Descriptive Measures for a Population

DCOVA

- Descriptive statistics discussed previously described a *sample*, not the *population*.
- Summary measures describing a population, called **parameters**, are denoted with Greek letters.
- Important population parameters are the population mean, variance, and standard deviation.

Numerical Descriptive Measures for a Population: The mean μ

- The population mean is the sum of the values in the population divided by the population size, N

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

Where μ = population mean

N = population size

X_i = i^{th} value of the variable X

Numerical Descriptive Measures For A Population: The Variance σ^2

DCOVA

- Average of squared deviations of values from the mean

  - Population variance:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

Where μ = population mean

N = population size

X_i = i^{th} value of the variable X

Numerical Descriptive Measures For A Population: The Standard Deviation σ

DCOVA

- Most commonly used measure of variation
- Shows variation about the mean
- Is the square root of the population variance
- Has the same units as the original data

  - Population standard deviation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$$

Sample statistics versus population parameters

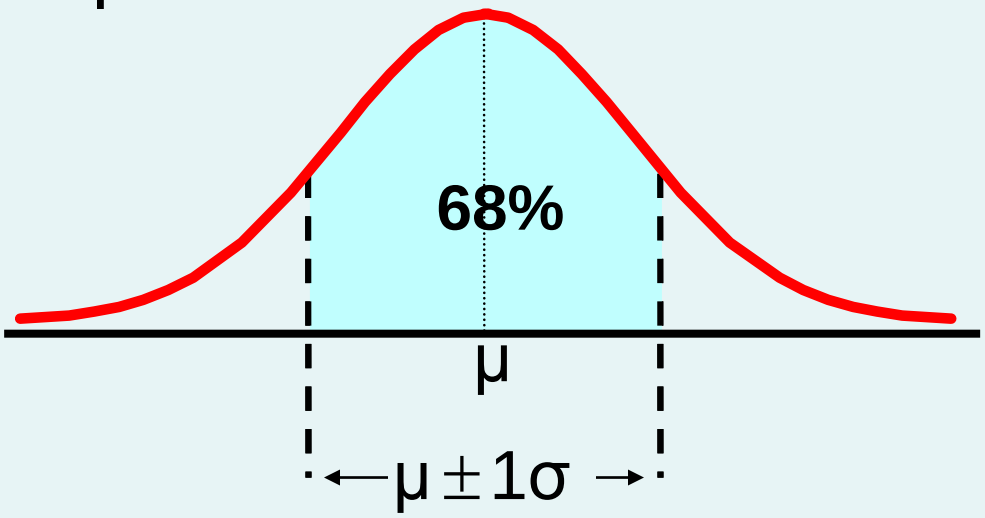
DCOVA

Measure	Population Parameter	Sample Statistic
Mean	μ	$\overline{X}$
Variance	σ^2	S^2
Standard Deviation	σ	S

The Empirical Rule

DCOVA

- The empirical rule approximates the variation of data in a bell-shaped distribution
- Approximately 68% of the data in a bell shaped distribution is within 1 standard deviation of the mean or $\mu \pm 1\sigma$

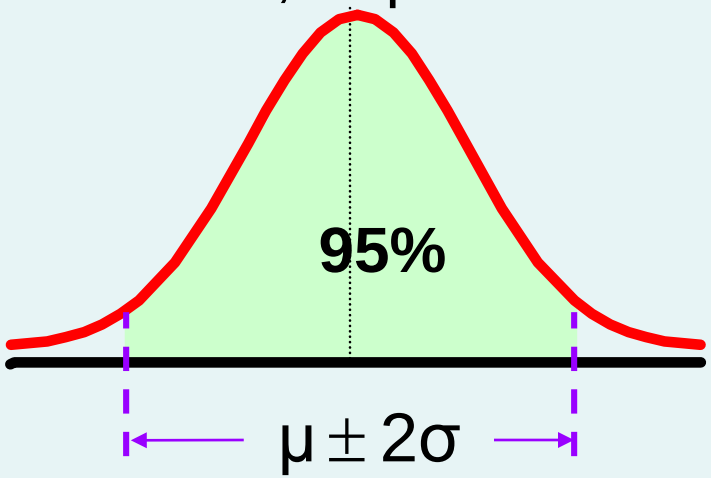


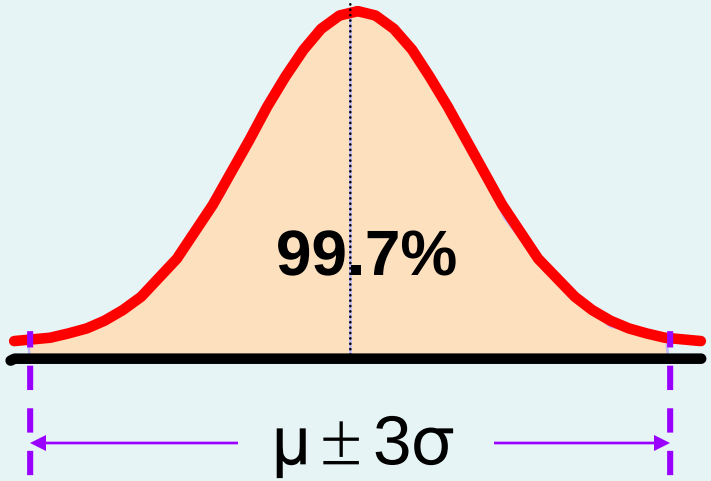
The Empirical Rule

DCOVA

- Approximately 95% of the data in a bell-shaped distribution lies within two standard deviations of the mean, or $\mu \pm 2\sigma$

- Approximately 99.7% of the data in a bell-shaped distribution lies within three standard deviations of the mean, or $\mu \pm 3\sigma$





Using the Empirical Rule

DCOVA

- Suppose that the variable Math SAT scores is bell-shaped with a mean of 500 and a standard deviation of 90. Then,
  - 68% of all test takers scored between 410 and 590 (500 ± 90).
  - 95% of all test takers scored between 320 and 680 (500 ± 180).
  - 99.7% of all test takers scored between 230 and 770 (500 ± 270).

Chebyshev Rule

DCOVA

- Regardless of how the data are distributed, at least $(1 - 1/k^2) \times 100\%$ of the values will fall within k standard deviations of the mean (for $k > 1$)
  - Examples:

	At least	Within
$(1 - 1/2^2) \times 100\% =$	75%	$k=2$ $(\mu \pm 2\sigma)$
$(1 - 1/3^2) \times 100\% =$	88.89%	$k=3$ $(\mu \pm 3\sigma)$

We Discuss Two Measures Of The Relationship Between Two Numerical Variables

- Scatter plots allow you to visually examine the relationship between two numerical variables and now we will discuss two quantitative measures of such relationships.

- The Covariance
- The Coefficient of Correlation

The Covariance

DCOVA

- The covariance measures the strength of the linear relationship between **two numerical variables** (X & Y)

- The sample covariance:

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

- Only concerned with the strength of the relationship
- No causal effect is implied

Interpreting Covariance

DCOVA

- **Covariance** between two variables:

cov(X,Y) > 0 → X and Y tend to move in the same direction

cov(X,Y) < 0 → X and Y tend to move in opposite directions

cov(X,Y) = 0 → X and Y are independent

- The covariance has a major flaw:
  - It is not possible to determine the relative strength of the relationship from the size of the covariance

Coefficient of Correlation

DCOVA

- Measures the relative strength of the linear relationship between two numerical variables
- Sample coefficient of correlation:

$$r = \frac{\text{cov}(X,Y)}{S_X S_Y}$$

where

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}}$$

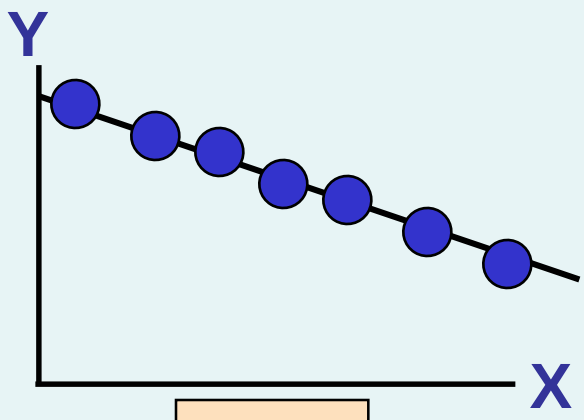
Features of the Coefficient of Correlation

DCOVA

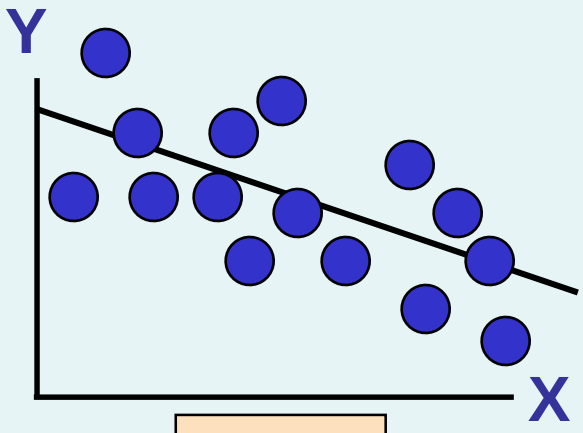
- The population coefficient of correlation is referred as ρ.
- The sample coefficient of correlation is referred to as r.
- Either ρ or r have the following features:
  - Unit free
  - Ranges between –1 and 1
  - The closer to –1, the stronger the negative linear relationship
  - The closer to 1, the stronger the positive linear relationship
  - The closer to 0, the weaker the linear relationship

Scatter Plots of Sample Data with Various Coefficients of Correlation

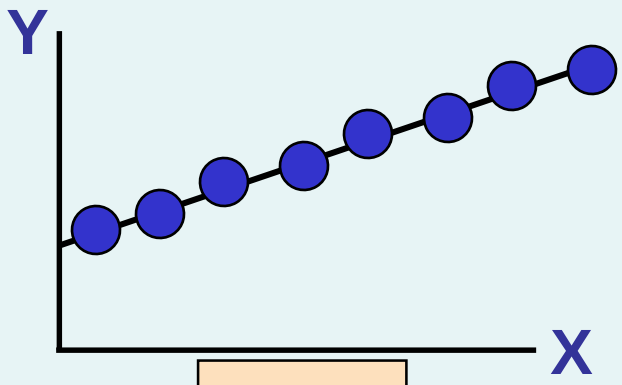
DCOVA



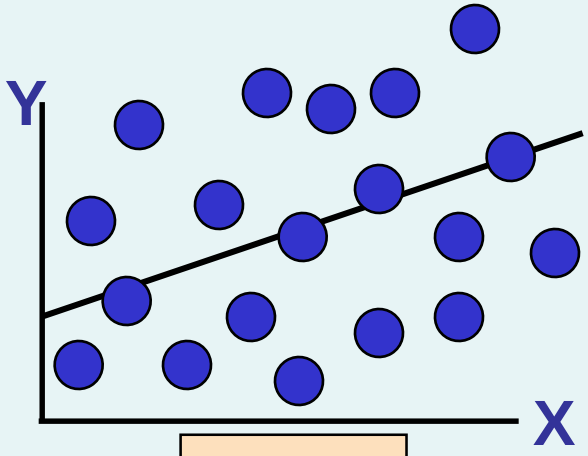
r = -1



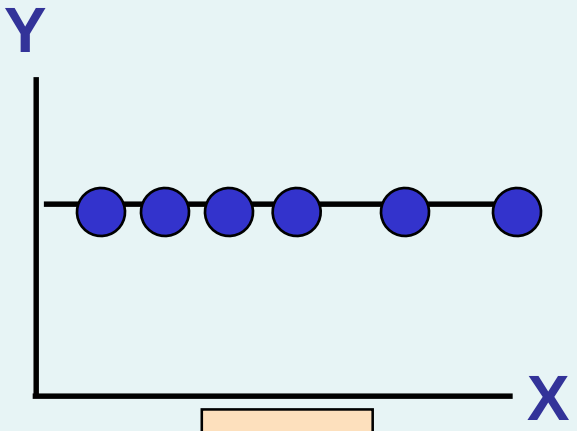
r = -.6



r = +1



r = +.3



r = 0

The Coefficient of Correlation Using Microsoft Excel Function

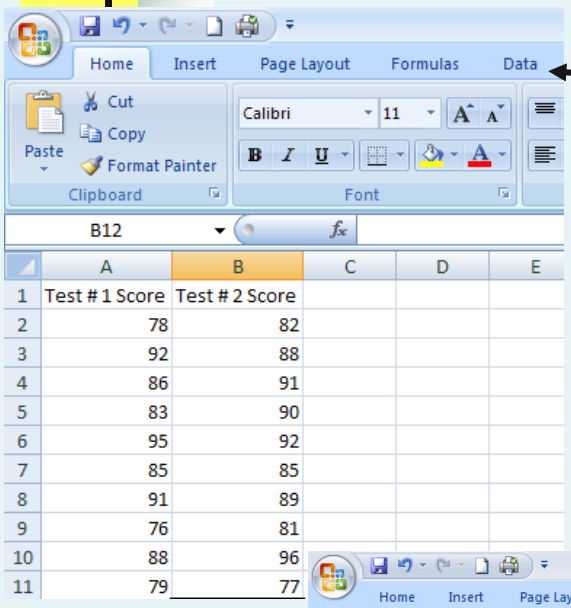
DCOVA

Test #1 Score	Test #2 Score		**Correlation Coefficient**	
78	82		0.7332	=CORREL(A2:A11,B2:B11)
92	88			
86	91			
83	90			
95	92			
85	85			
91	89			
76	81			
88	96			
79	77			

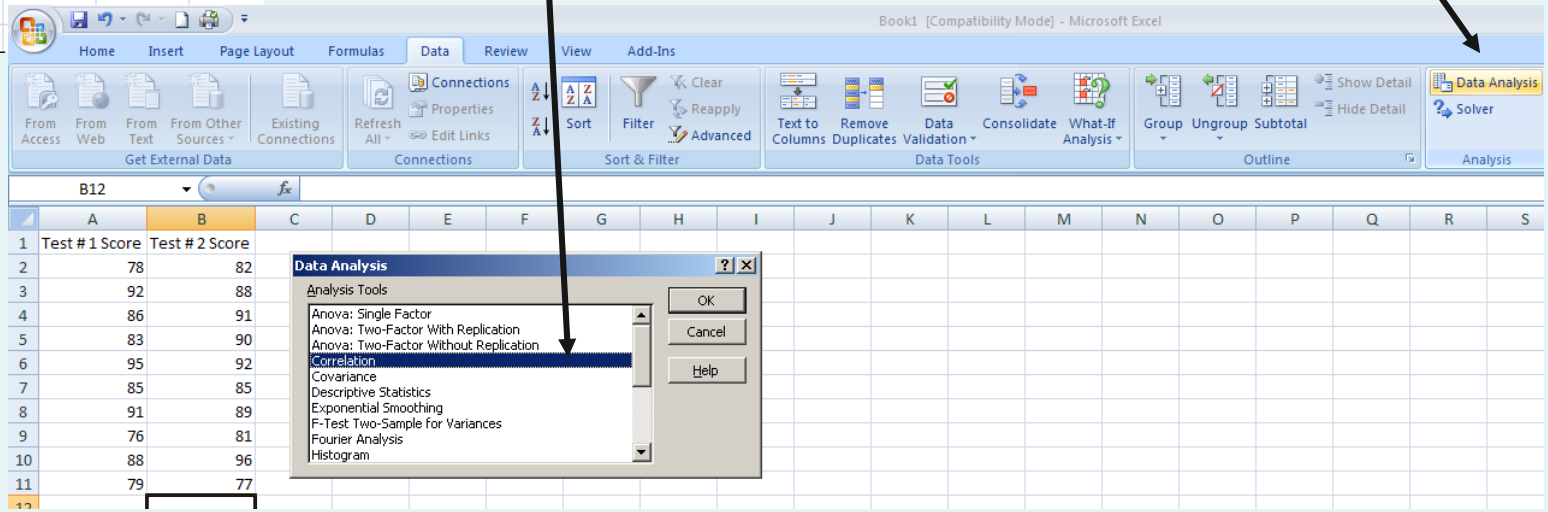
The Coefficient of Correlation Using Microsoft Excel Data Analysis Tool

DCOVA

1. Select Data
2. Choose Data Analysis
3. Choose Correlation & Click OK

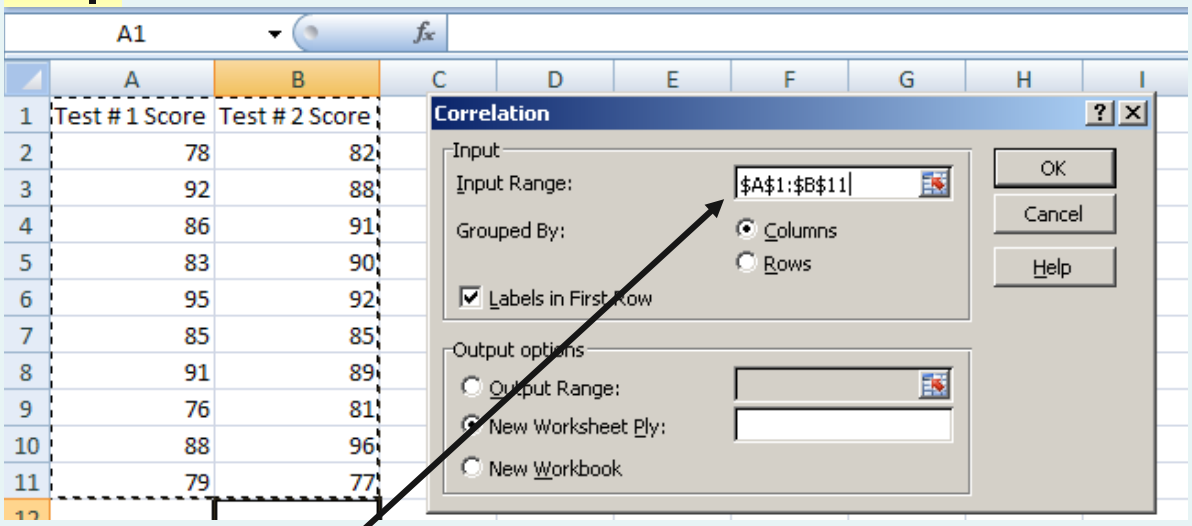


	A	B	C	D	E
1	Test # 1 Score	Test # 2 Score			
2	78	82			
3	92	88			
4	86	91			
5	83	90			
6	95	92			
7	85	85			
8	91	89			
9	76	81			
10	88	96			
11	79	77			



The Coefficient of Correlation Using Microsoft Excel

DCOVA



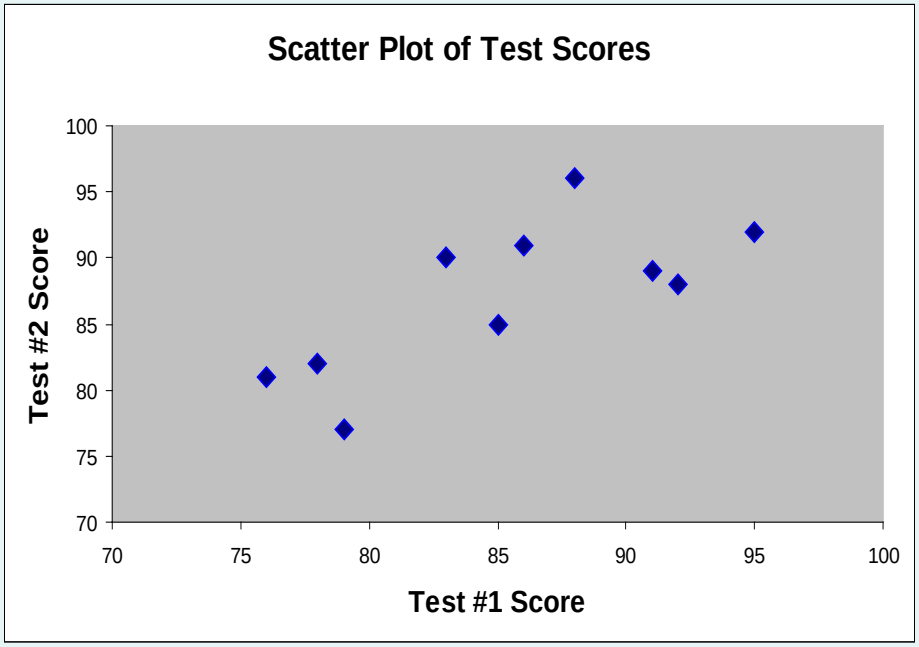
4. Input data range and select appropriate options
5. Click OK to get output

	A	B	C
1		*Test # 1 Score*	*Test # 2 Score*
2	Test # 1 Score	1	
3	Test # 2 Score	0.733243705	1
4			

Interpreting the Coefficient of Correlation Using Microsoft Excel

DCOVA

- r = .733
- There is a relatively strong positive linear relationship between test score #1 and test score #2.
- Students who scored high on the first test tended to score high on second test.



Pitfalls in Numerical Descriptive Measures

DCOVA

- Data analysis is objective
  - Should report the summary measures that best describe and communicate the important aspects of the data set
- Data interpretation is subjective
  - Should be done in fair, neutral and clear manner

Ethical Considerations

DCOVA

Numerical descriptive measures:

- Should document both good and bad results
- Should be presented in a fair, objective and neutral manner
- Should not use inappropriate summary measures to distort facts

Chapter Summary

In this chapter we discussed

- Measures of central tendency
  - Mean, median, mode, geometric mean
- Measures of variation
  - Range, interquartile range, variance and standard deviation, coefficient of variation, Z-scores
- The shape of distributions
  - Skewness & Kurtosis
- Describing data using the 5-number summary
  - Boxplots

Chapter Summary

(continued)

- Covariance and correlation coefficient
- Pitfalls in numerical descriptive measures and ethical considerations

This work is protected by United States copyright laws and is provided solely for the use of instructors in teaching their courses and assessing student learning. Dissemination or sale of any part of this work (including on the World Wide Web) will destroy the integrity of the work and is not permitted. The work and materials from it should never be made available to students except by instructors using the accompanying text in their classes. All recipients of this work are expected to abide by these restrictions and to honor the intended pedagogical purposes and the needs of other instructors who rely on these materials.