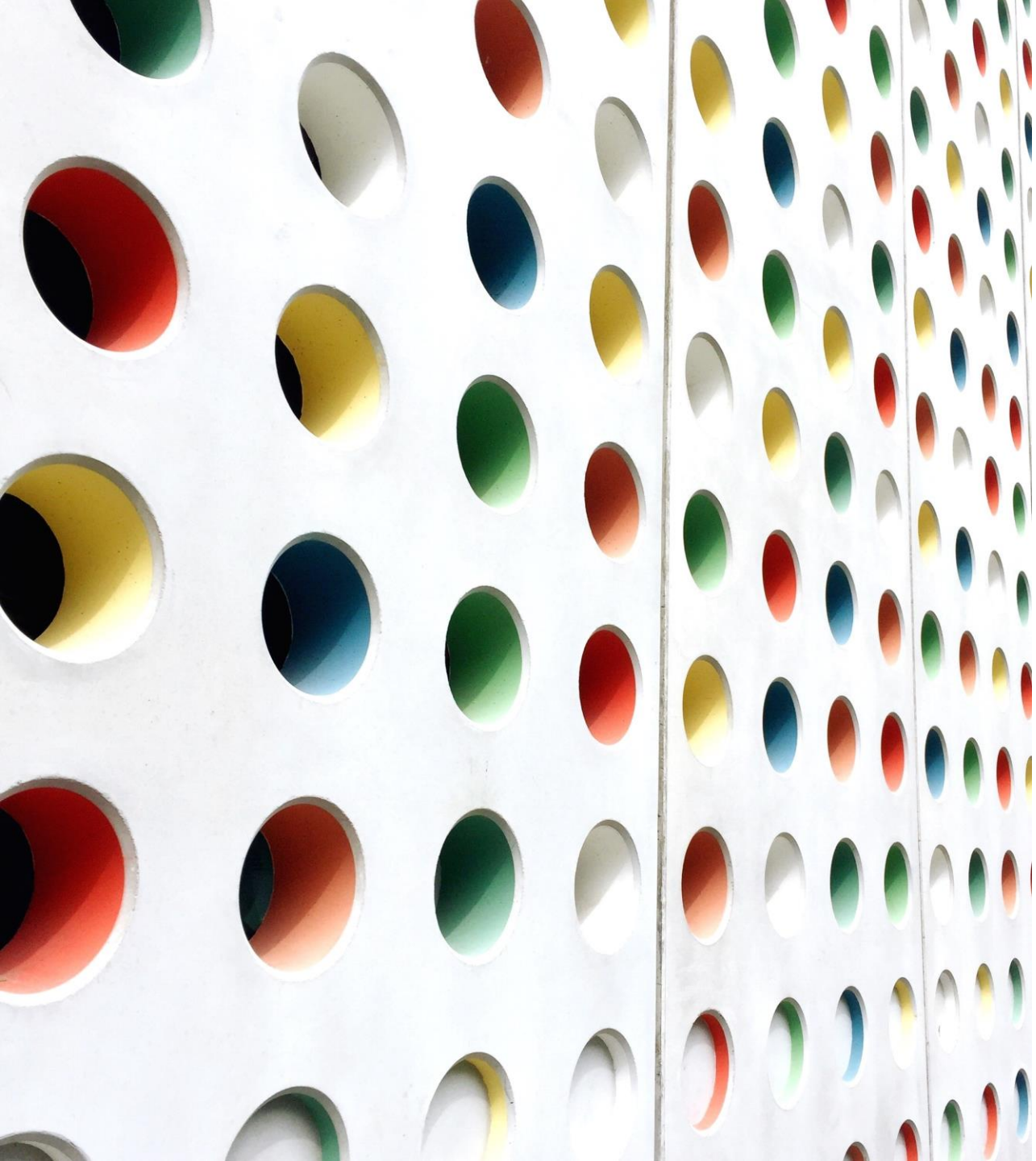


Numerical Descriptive Measures

BM 4419

BUSINESS ANALYTICS

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.

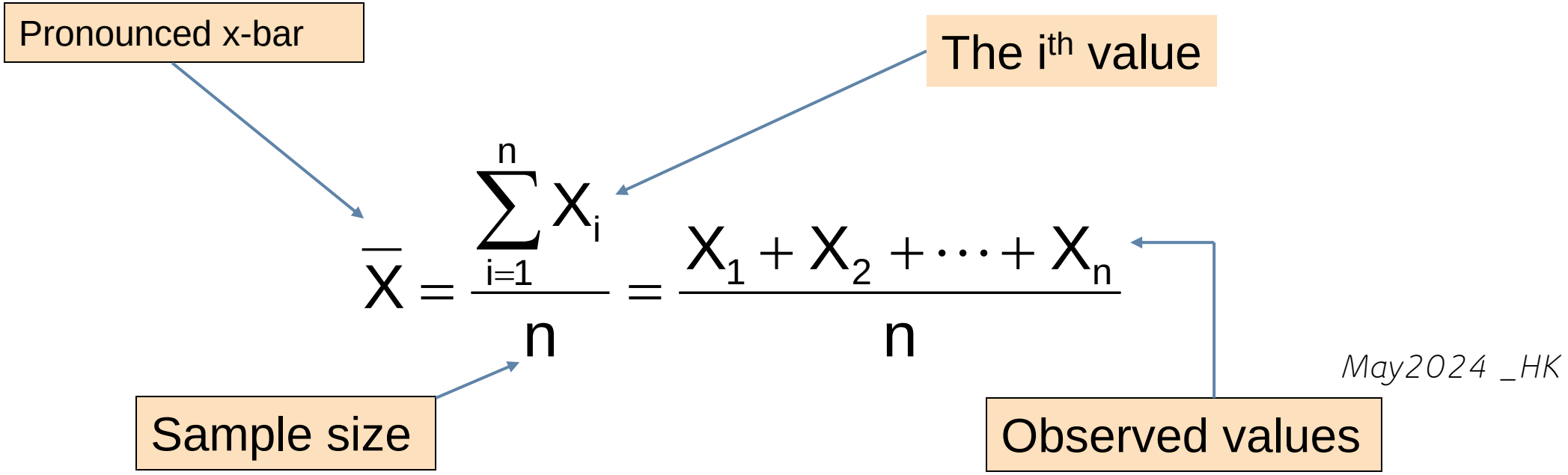
May2024 _HK

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:



May2024 _HK

Measures of Central Tendency: The Mean

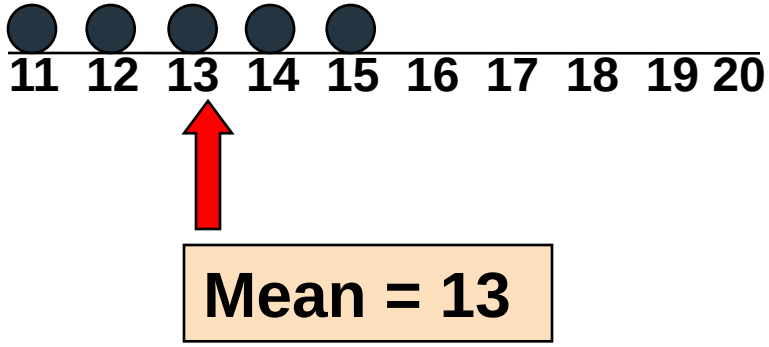
DCOVA

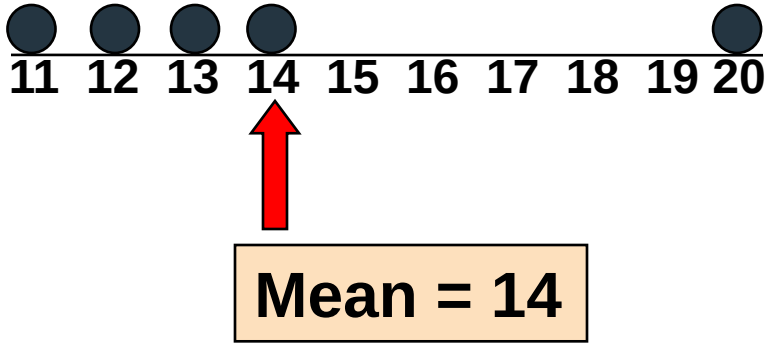
(continued)

The most common measure of central tendency

Mean = sum of values divided by the number of values

Affected by extreme values (outliers)





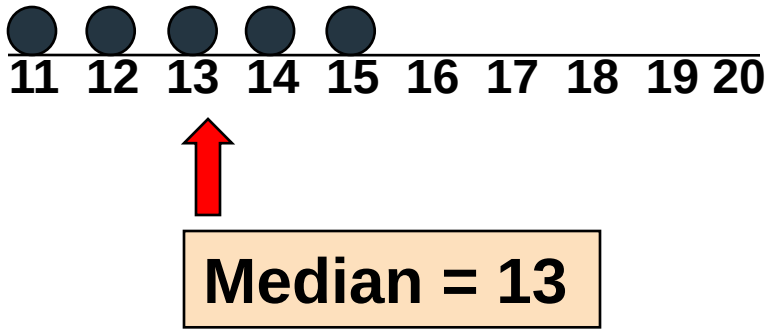
$$\frac{11+12+13+14+15}{5}=\frac{65}{5}=13$$

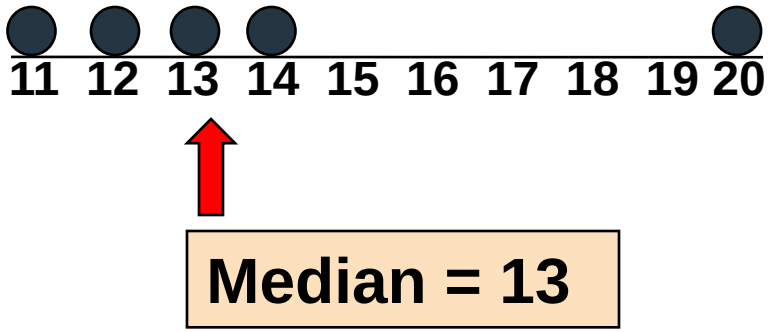
$$\frac{11+12+13+14+20}{5}=\frac{70}{5}=14$$

Measures of Central Tendency: The Median

DCOVA

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





May2024 _HK

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

May2024 _HK

Measures of Central Tendency: The Mode

DCOVA

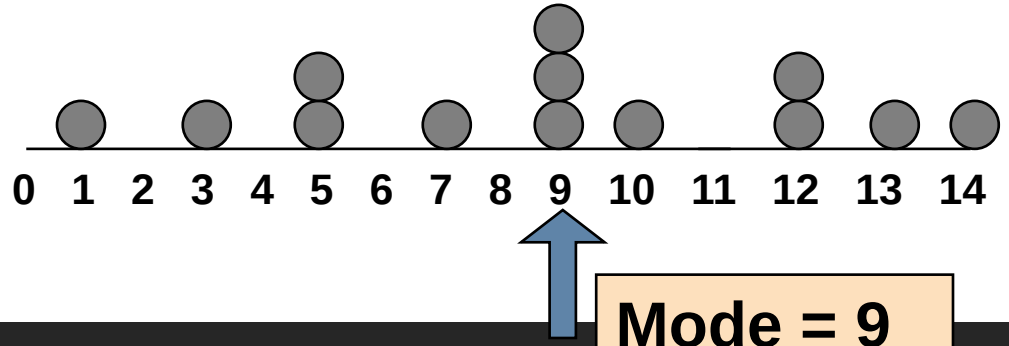
Value that occurs most often

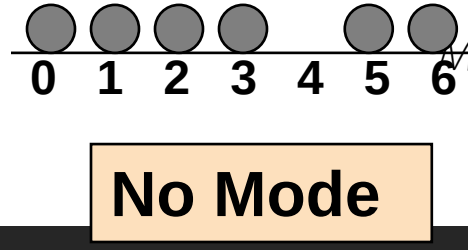
Not affected by extreme values

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

May2024 _HK

Measures of Central Tendency: Which Measure to Choose?

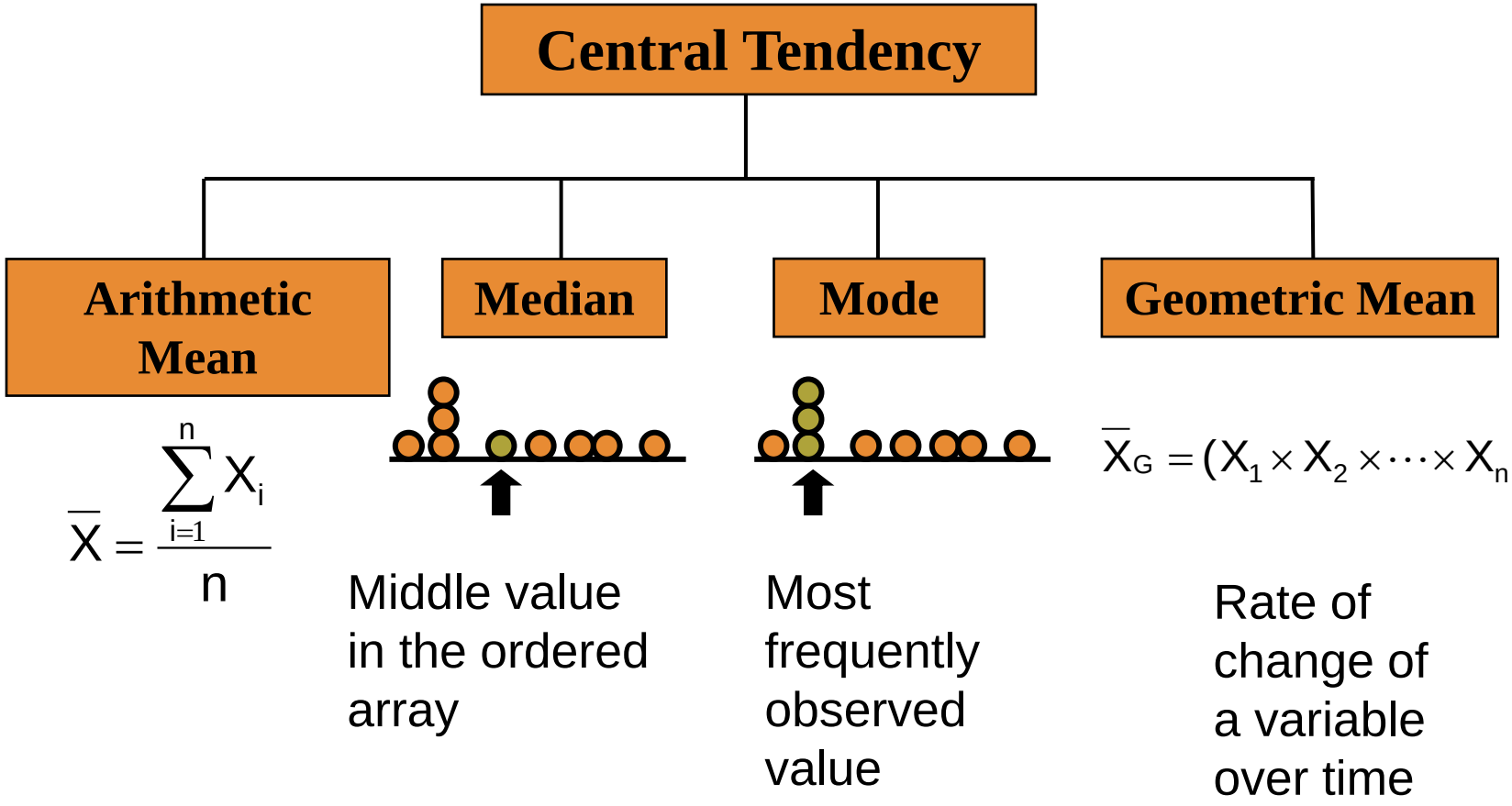
DCOVA

- 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**.

May2024 _HK

Measures of Central Tendency: Summary

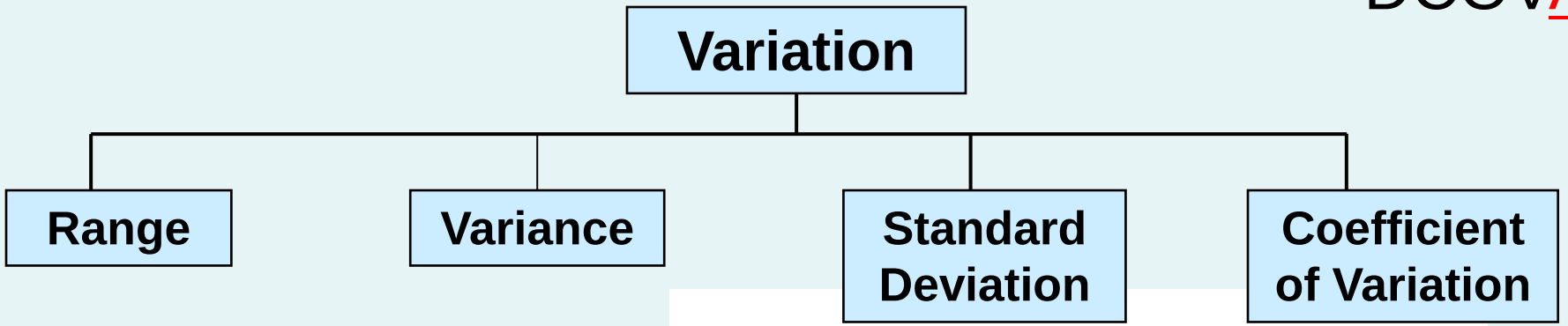
DCOVA



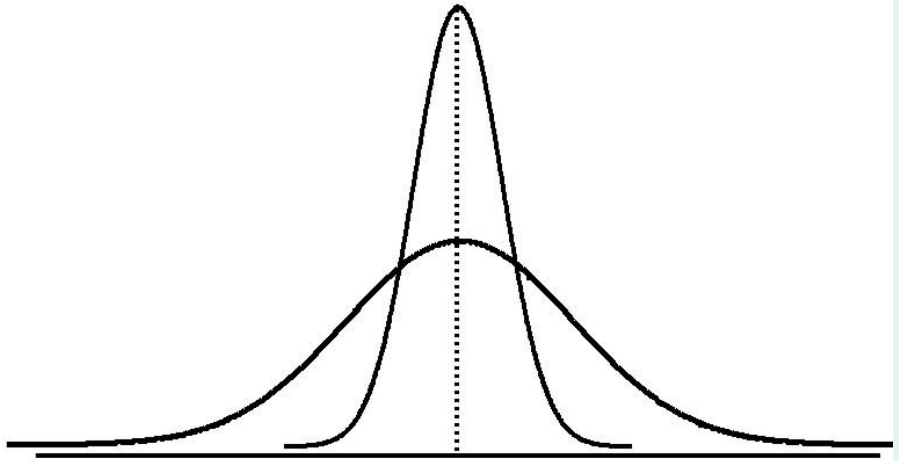
May2024 _HK

Measures of Variation

DCOVA



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



Same center, different variation

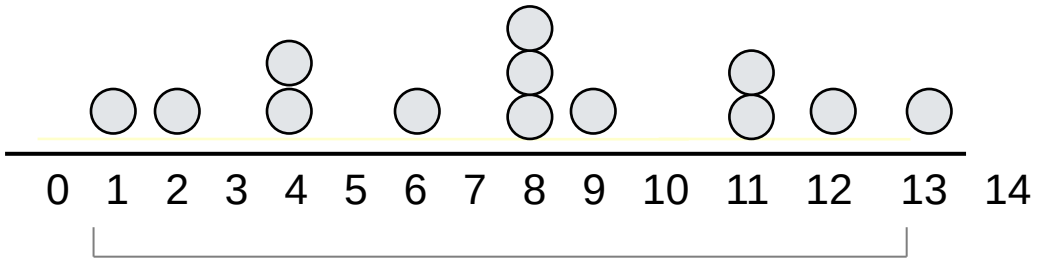
Measures of Variation: The Range

DCOVA

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

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

Example:



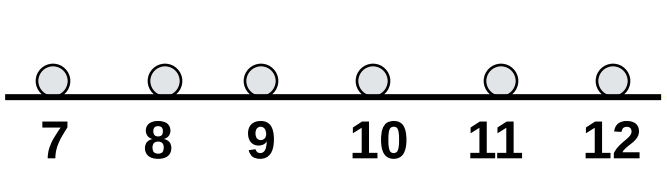
Range = 13 - 1 = 12

May2024 _HK

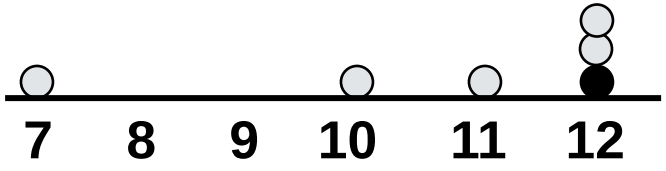
Measures of Variation: Why The Range Can Be Misleading

DCOVA

- Ignores the way in which data are distributed



Range = 12 - 7 = 5



Range = 12 - 7 = 5

- 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

May2024 _HK

Measures of Variation: The Sample Variance

DCOVA

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

- Sample variance:

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

Where $\bar{X}$ = arithmetic mean

n = sample size

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

May2024 _HK

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:

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

May2024 _HK

Measures of Variation: Sample Standard Deviation: Calculation Example

DCOVA

Sample Data (X_i) :

10	12	14	15	17	18	18	24

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

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

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

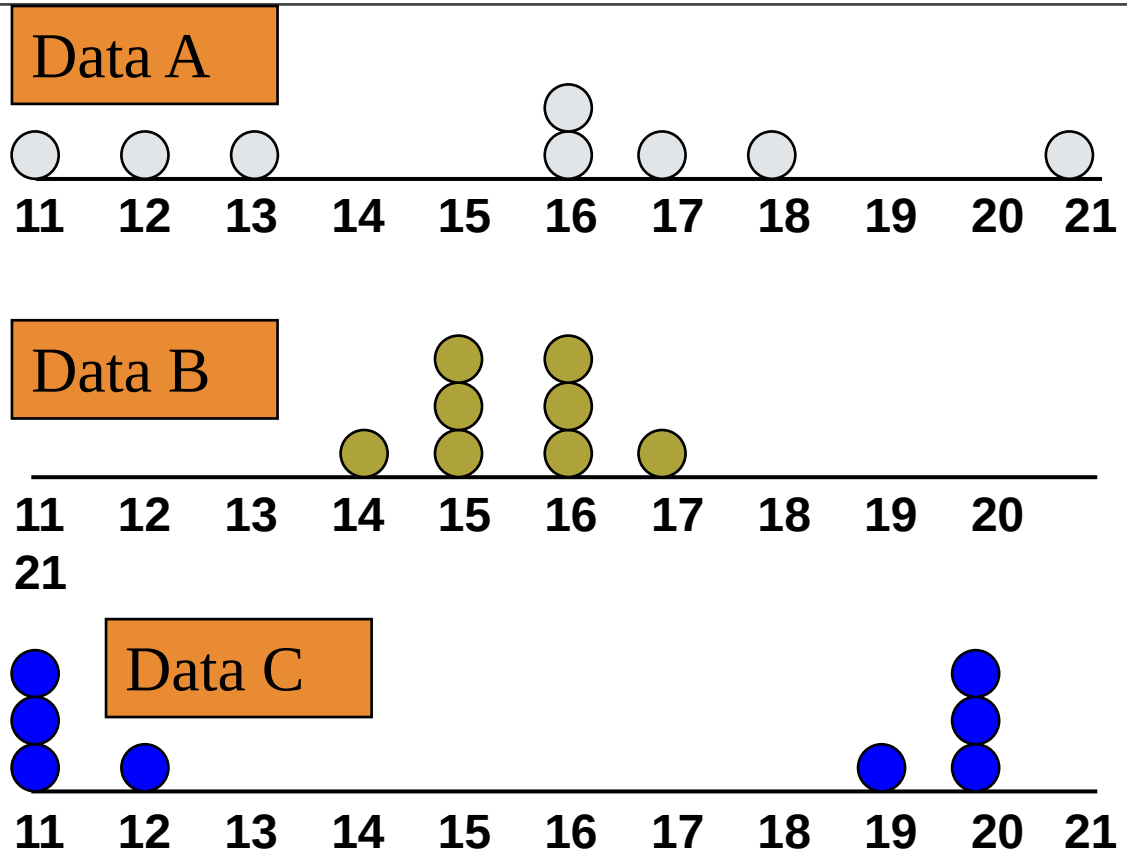
$$= \sqrt{\frac{130}{7}} = 4.3095$$

A measure of the "average" scatter around the mean

May2024 _HK

Measures of Variation: Comparing Standard Deviations

DCOVA

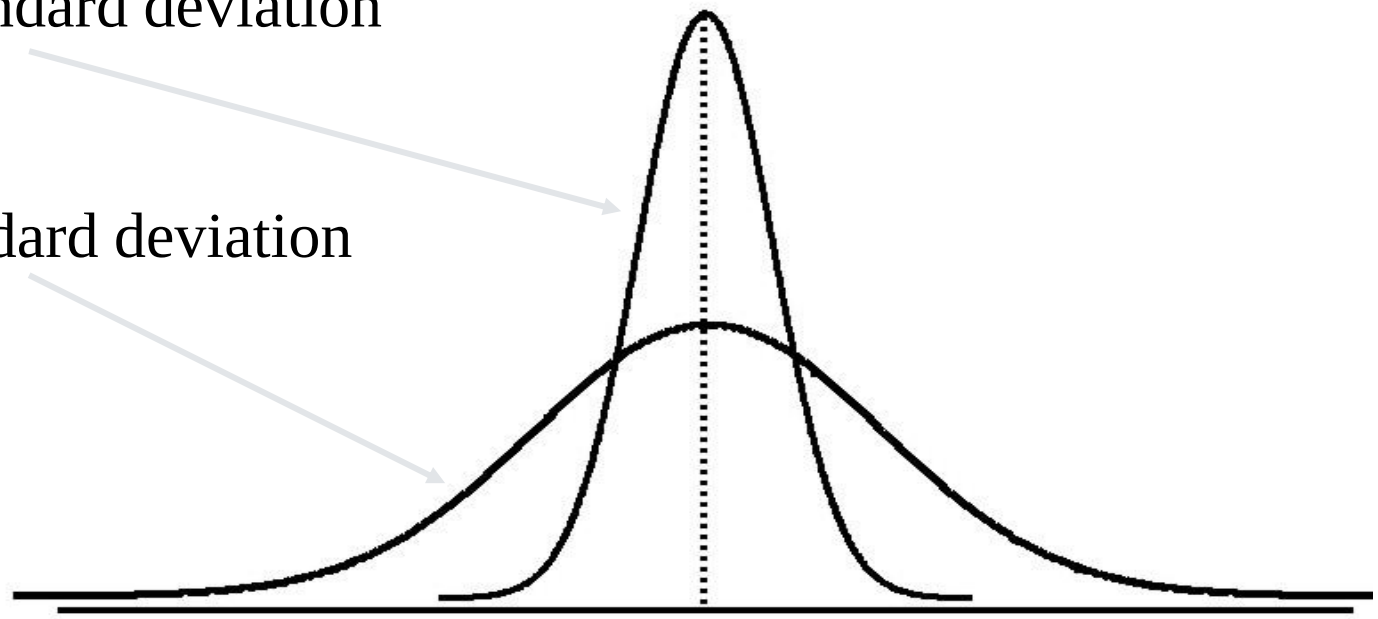


Measures of Variation: Comparing Standard Deviations

DCOVA

Smaller standard deviation

Larger standard deviation



2024 _HK

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.

May2024 _HK

- None of these measures are ever negative.

Shape of a Distribution

DCOVA

Describes how data are distributed

Two useful shape-related statistics are:

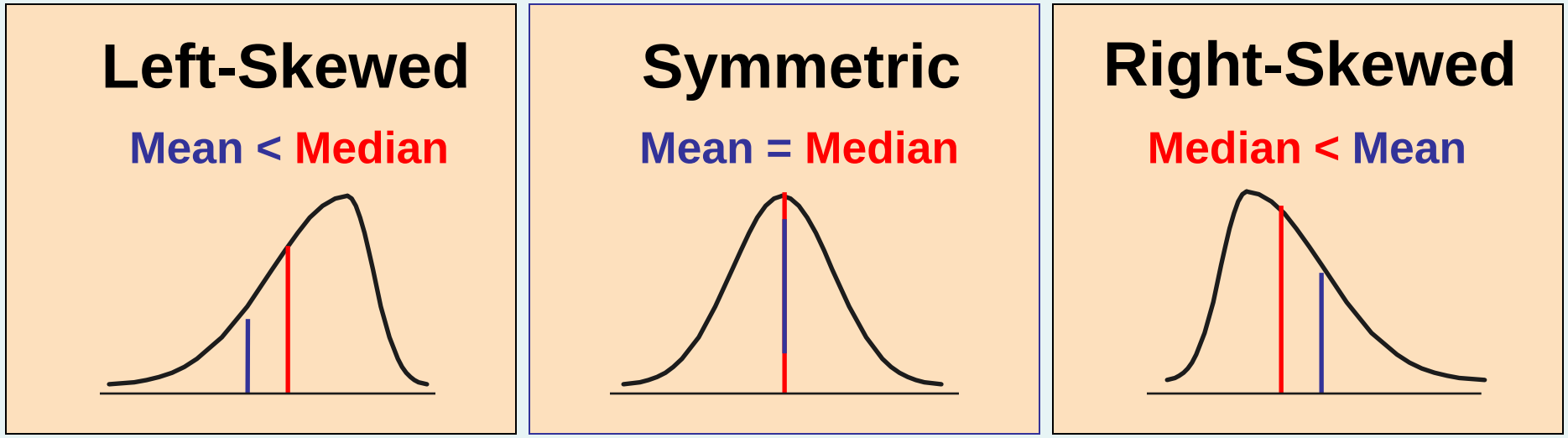
- Skewness
  - Measures the extent to which data values are not symmetrical
- Kurtosis
  - Kurtosis affects the peakedness of the curve of the distribution–that is, how sharply the curve rises approaching the center of the distribution

May2024 _HK

Shape of a Distribution (Skewness)

DCOVA

- Measures the extent to which data is not symmetrical



Skewness Statistic	< 0	0	>0

Quick Test

State the type of skewness for the following sets of distribution:

a) mean=35, median=35, mode= 35

b) mean = 32, median = 36, mode = 40

c) mean= 70, median 68, mode= 65

2024 _HK

General Descriptive Stats Using Microsoft Excel Functions

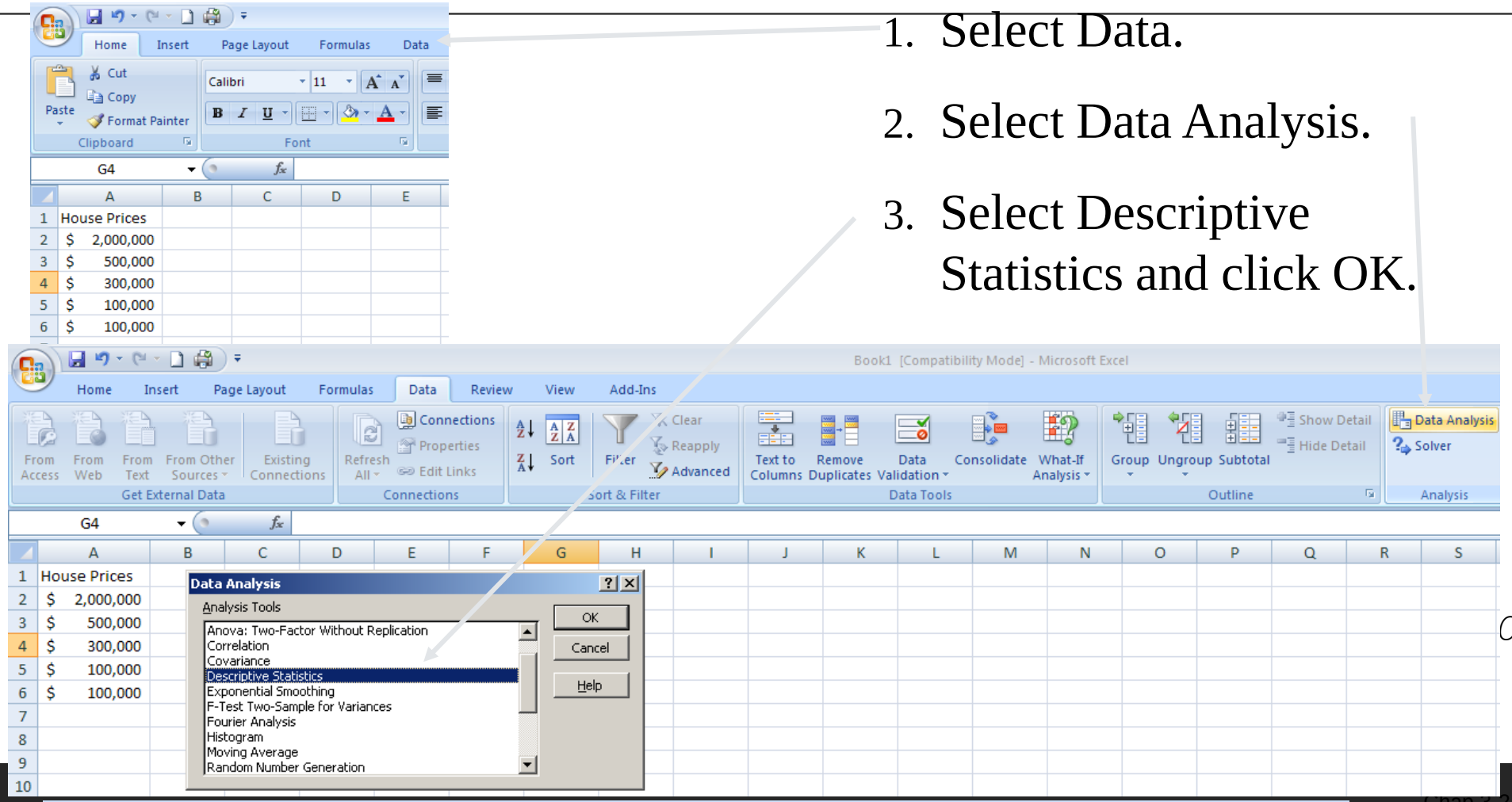
DCOVA

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)

ay2024 _HK

General Descriptive Stats Using Microsoft Excel Data Analysis Tool

DCOVA



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

024 _HK

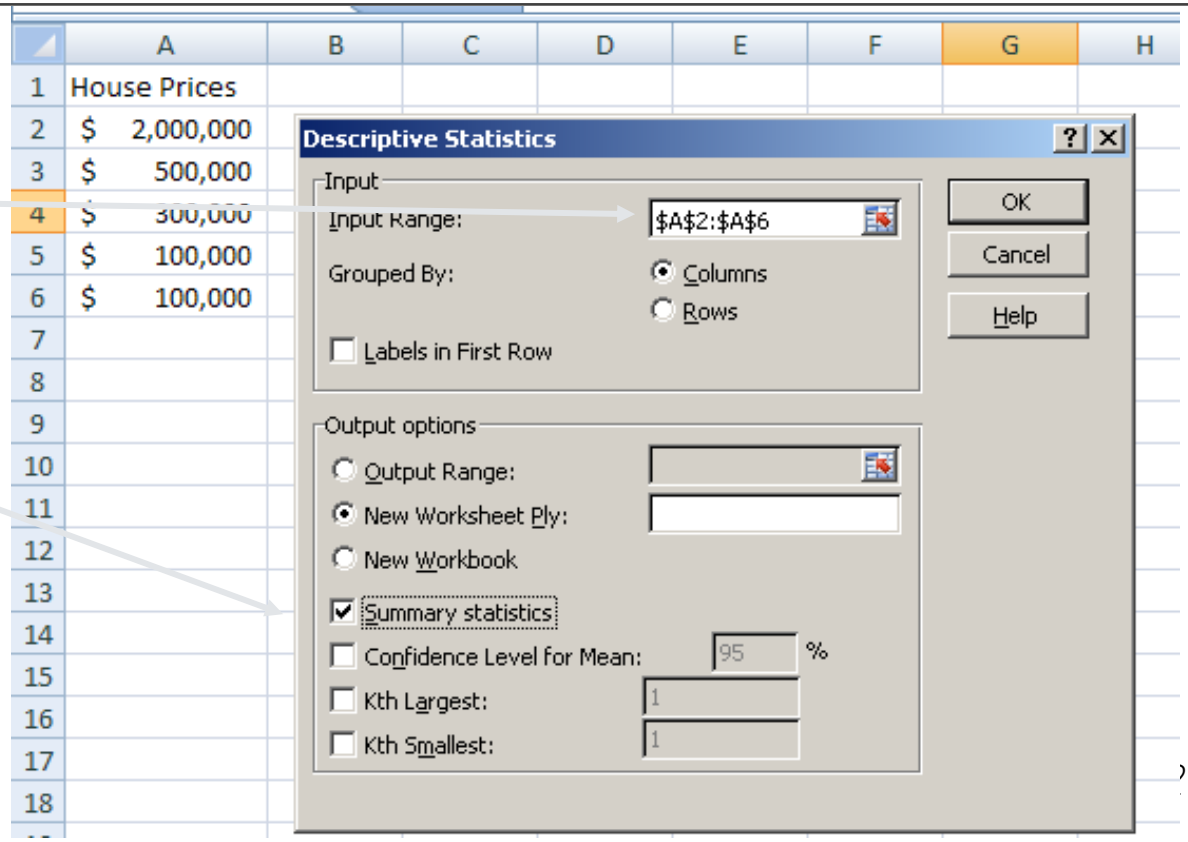
General Descriptive Stats Using Microsoft Excel

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

May2024 _HK

SELF-TEST

The following data (see the drink.xls file) represent the amount of soft-drink filled in a sample of 50 consecutive 2-liter bottles. The results, listed horizontally in the order of being filled, were:

2.109	2.086	2.066	2.075	2.065	2.057	2.052	2.044	2.036	2.038
2.031	2.029	2.025	2.029	2.023	2.02	2.015	2.014	2.013	2.014
2.012	2.012	2.012	2.01	2.005	2.003	1.999	1.996	1.997	1.992
1.994	1.986	1.984	1.981	1.973	1.975	1.971	1.969	1.966	1.967
1.963	1.957	1.951	1.951	1.947	1.941	1.941	1.938	1.908	1.894
2.109	2.086	2.066	2.075	2.065	2.057	2.052	2.044	2.036	2.038

Construct the Descriptive Statistics for the above data Using Microsoft Excel Functions and draw two important conclusions based on the measurement outcome.