## Able to understand the Business Analytics.

## Activity 4

**Aim:** Calculate the arithmetic mean, geometric mean and Harmonic mean (5Hrs).

# **Learning outcome**: Able to understand the Business Analytics.

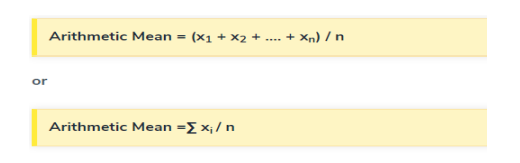
###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

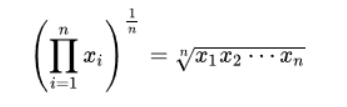
# Arithmetic mean

Arithmetic Mean, commonly used term in statistics, is the average of the numerical values set and is calculated by firstly calculating the sum of number in the set and then dividing resultant by count of those numbers.

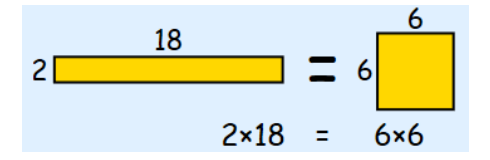


# Geometric mean

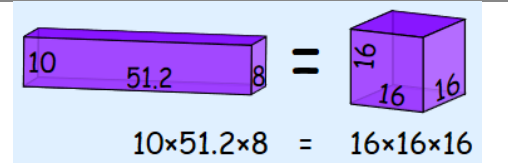
In mathematics, the geometric mean is a mean or average, which indicates the central tendency or typical value of a set of numbers by using the product of their values (as opposed to the arithmetic mean which uses their sum). The geometric mean is defined as the nth root of the product of n numbers, i.e., for a set of numbers x1, x2, ..., an, the geometric mean is defined as



In two dimensions, it is the equivalent of finding the equivalent square with the same area as the rectangle given by the two dimensions cited:



In three dimensions, it is the equivalent of finding the equivalent cube with the same volume as the given hexahedron with the three dimensions cited:



The idea continues in n dimensions.

The Excel function GEOMEAN returns the geometric mean of an array or range of positive data. For example, you can use GEOMEAN to calculate average growth rate given compound interest with variable rates. It has the following syntax:

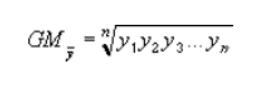
GEOMEAN (number1, [number2], ...)

The GEOMEAN function has the following arguments:

* number1, number2...where number1 is required, and subsequent numbers are optional. There can be between one (1) and 255 numbers. You can also use a single array or a reference to an array instead of arguments separated by commas.

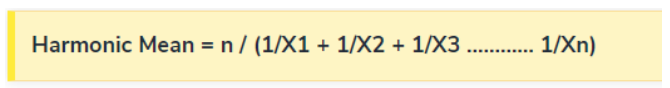
It should be further noted that:

* arguments can either be numbers or names, arrays, or references that contain numbers
* logical values and text representations of numbers that you type directly into the list of arguments are counted
* of an array or reference argument contains text, logical values or empty cells, those values are ignored; however, cells with the value zero are included
* arguments that are error values or text that cannot be translated into numbers cause errors
* if any data point ≤ 0, GEOMEAN returns the #NUM! error value
* the equation for the geometric mean is:



# Harmonic mean

The Excel HARMEAN function returns the harmonic mean for a set of numeric values. The harmonic mean is the reciprocal of the arithmetic mean of reciprocals. Harmonic mean can be used to calculate a mean that reduces the impact of outliers.



=HARMEAN (number1, [number2], ...)

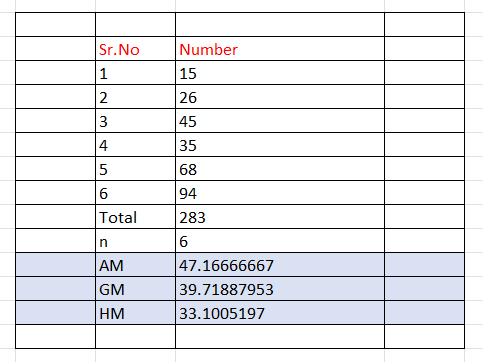
● number1 - First value or reference.

● number2 - [optional] Second value or reference. Where:

X1, X2…Xn – Data Points

n – Total number of data points

# Arithmetic Mean (AM), Geometric Mean (GM) and Harmonic Mean (HM):



## Activity 5

**Aim:** Calculate median from raw & grouped data.

# **Learning outcome**: Able to understand the Business Analytics.

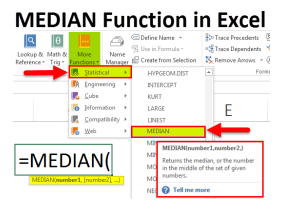
###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

Code/Program/Procedure (with comments):

Median Function in Excel



1. Number 1 (required argument) – The number arguments are a set of single or more numeric values (or arrays of

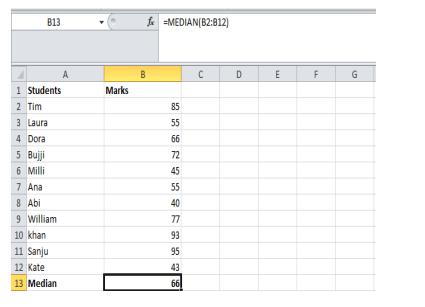
numeric values), for which you want to calculate the median.

1. Number 2 (optional argument)

Median is a function which is used to find the middle number in each range of numbers. When you are finding median manually, you need to sort the data in ascending order but in Excel, you can simply use the Median function and select the range and you will find your median. We take the same example as above to find the median of marks obtained by students. So, we use = MEDIAN (B2: B12).

##### Output/Result sippet:

Median



## Activity 6

**Aim:** Calculate mode for row & grouped data.

# **Learning outcome**: Able to understand the Business Analytics.

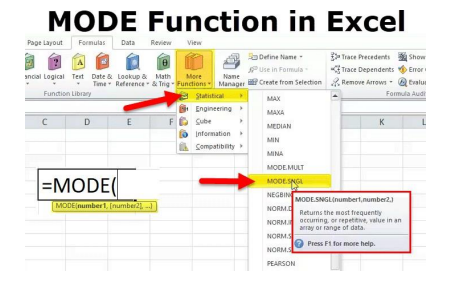
###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

Code/Program/Procedure (with comments):0

MODE in Excel

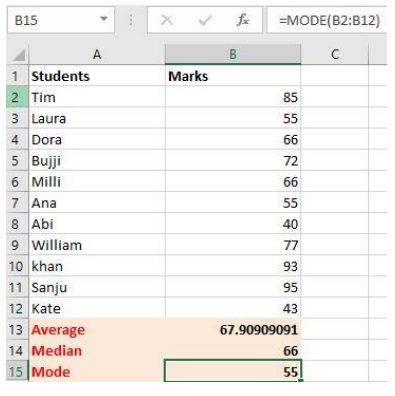


1. number1 (compulsory OR required argument) – Arrays of cell reference or numericvalues (set of one or more numeric values) for which we have to calculate the mode.
2. number2 (Optional OR not required) – Arrays of cell reference or numeric values (set ofone or more numeric values) for which we have to calculate the mode.

Mode helps you to find out the value that occurs the most number of times. When you are workingon a large amount of data, this function can be a lot of help. To find the most occurring value in Excel, use the MODE function and select the range you want to find the mode of. In our example below, we use =MODE (B2: B12) and since 2 students have scored 55, we get the answer as 55.

Output/Results snippet:

Mode



# Able to understand business analytics and develop business intelligence.

## Activity 1

**Aim:** Calculate standard deviation for set of data (2.5Hrs).

# **Learning outcome**: Able to business analytics and develop business intelligence.

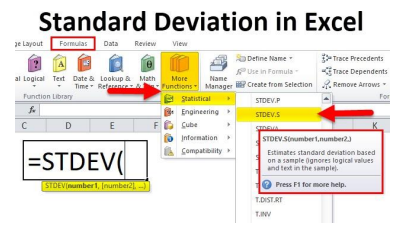
###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

Code/Program/Procedure (with comments):

Standard Deviation in Excel

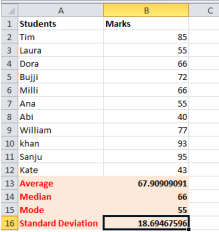


Standard deviation formula in Excel

1. number1: (Compulsory or mandatory argument) It is the first element of the sample of a.
2. number2: (Optional argument) It is a number of arguments from 2 to 254 corresponding to a sample of a population

The standard deviation in Excel helps you to understand, how much your values deviate from the Average or Mean that is it tells you whether your data is somewhere close to the average or fluctuates a lot. If the value received is on the higher side then that means that your data has a lot 37 of fluctuations and vice versa. To calculate standard deviation in excel we use the STDEV function. In the same example, we shall use the STDEV function so our formula will be = STDEV (B2: B12). Our answer is around 20 which indicates that the marks of the students fluctuate a lot.

Output/Results snippet: Standard Deviation



## Activity 2

**Aim:** Calculate standard variance for a set of data (2.5Hrs).

# **Learning outcome**: Able to business analytics and develop business intelligence.

###### Duration: 5 hours

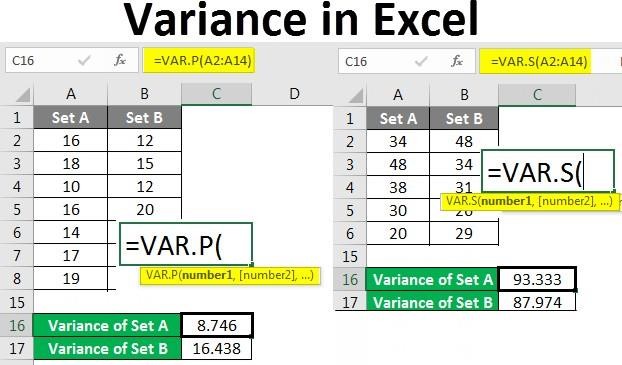
# List of Hardware/Software requirements:

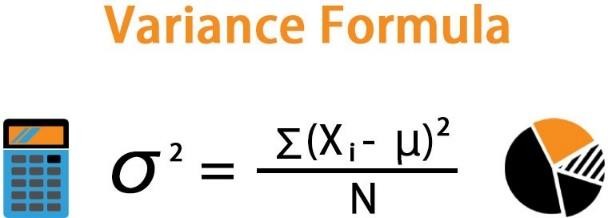
* Laptop/Computer with Windows
* MS Office with Excel

Code/Program/Procedure (with comments):

**Variance in Excel**

The term “variance” refers to the extent of dispersion of the data points of a data set from its mean, which is computed as the average of the squared deviation of each data point from the population mean.





## Variance formula in Excel

The formula for a variance can be derived by summing up the squared deviation of each data point and then dividing the result by the total number of data points in the data set.

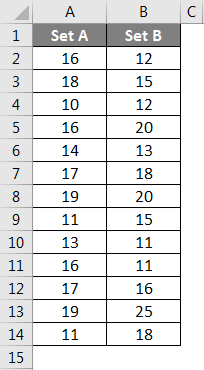
Mathematically, it is represented as,

## σ2 = ∑ (Xi – μ)^2 / N

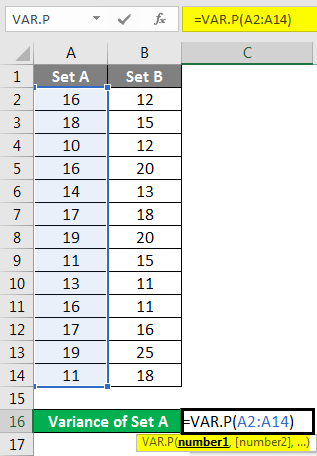
where,

Xi = ith data point in the data set μ = Population mean N = Number of data points in the population

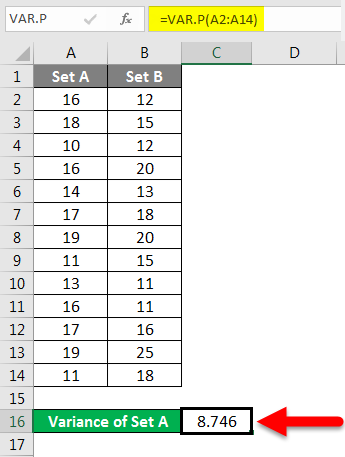
* Step 1 – Enter the data set in the columns.



* + Step 2 – Insert the VAR.P function and choose the range of the data set. Here one thing should be noted that if any cell has an error, then that cell will be ignored.



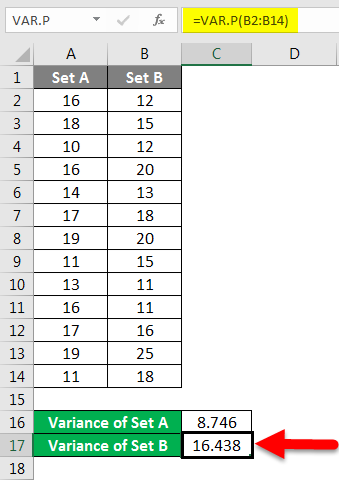
* + Step 3 – After pressing the Enter key we will get the variance.



We have calculated the variance of Set B by following the same steps given above.

## Output/Results snippet:

**Var.p**



## Activity 3

**Aim:** Calculate standard deviation for set of data (5Hrs).

# **Learning outcome**: Able to business analytics and develop business intelligence.

###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

Code/Program/Procedure (with comments):

**VLOOKUP function**

The VLOOKUP function in Excel is a tool for looking up a piece of information in a table or data set and extracting some corresponding data/information. In simple terms, the VLOOKUP function says the following to Excel: “Look for this piece of information (e.g., bananas), in this data set (a table), and tell me some corresponding information about it (e.g., the price of bananas)”.

VLOOKUP Formula

## =VLOOKUP(lookup\_value, table\_array, col\_index\_num, [range\_lookup])

To translate this to simple English, the formula is saying, “Look for this piece of information, in the following area, and give me some corresponding data from another column”.

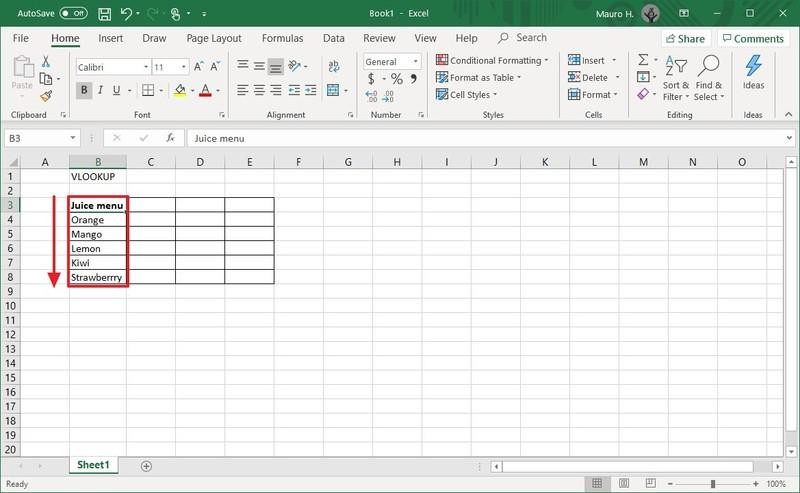
The VLOOKUP function uses the following arguments:

* Lookup\_value (required argument) – Lookup\_value specifies the value that we want to look up in the first column of a table.
* Table\_array (required argument) – The table array is the data array that is to be searched. The VLOOKUP function searches in the left-most column of this array.
* Col\_index\_num (required argument) – This is an integer, specifying the column number of the supplied table\_array, that you want to return a value from.
* Range\_lookup (optional argument) – This defines what this function should return in the event that it does not find an exact match to the lookup\_value. The argument can be set to TRUE or FALSE, which means:
* TRUE – Approximate match, that is, if an exact match is not found, use the closest match below the lookup\_value.
* FALSE – Exact match, that is, if an exact match not found, then it will return an error.

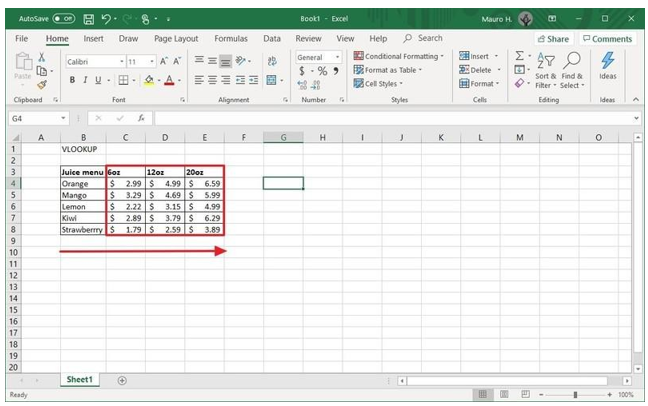
## Write VLOOKUP function in Excel

To write a VLOOKUP function manually in Excel, use these steps:

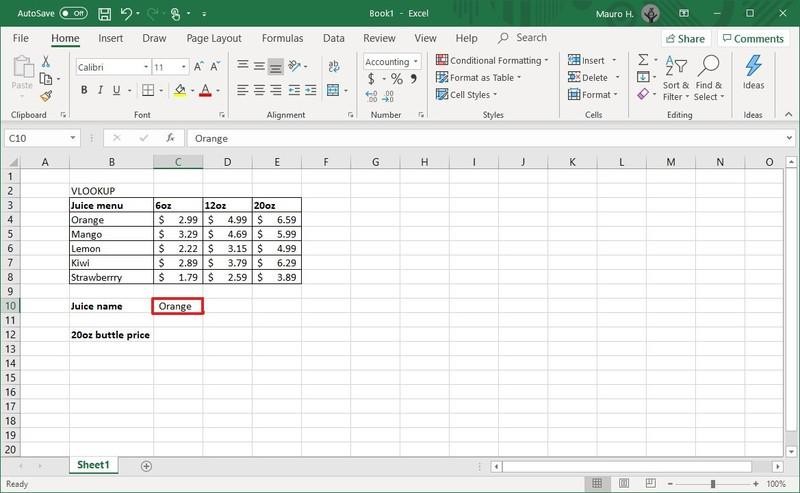
1. Open Excel.
2. Create the first column with items that will work as unique identifiers (required).



1. Create one or more additional columns (on the right side) with the different values for each item from the first column (on the left side).

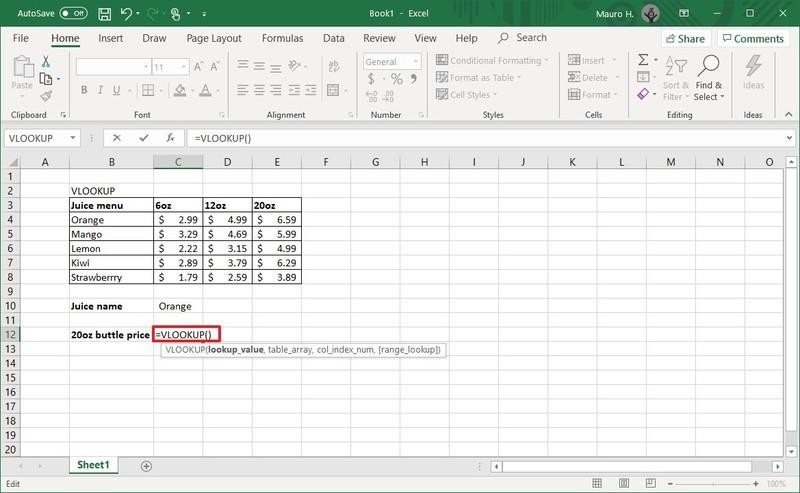


1. Select an empty cell in the spreadsheet and specify the name of the item you want to find an answer to—for example, Orange.



1. Select an empty cell to store the formula and returned value.
2. In the empty cell, type the following syntax to create a VLOOKUP formula and press Enter:

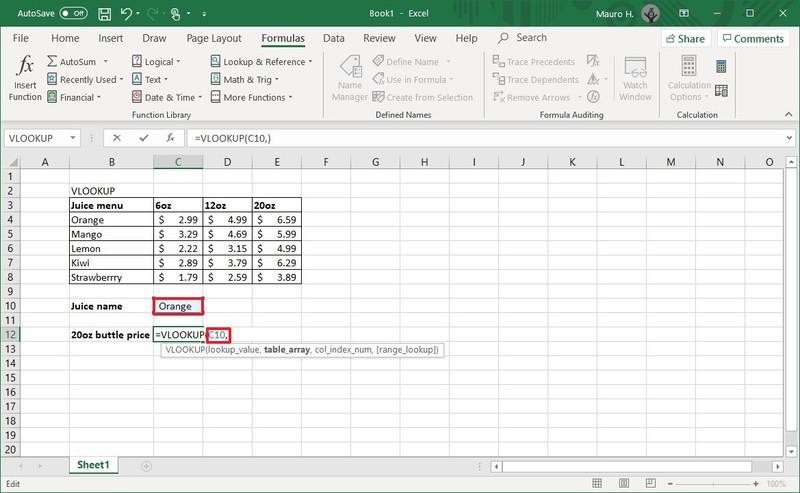
=VLOOKUP()



1. Type the following arguments inside the parenthesis "()" to write the function and press Enter:

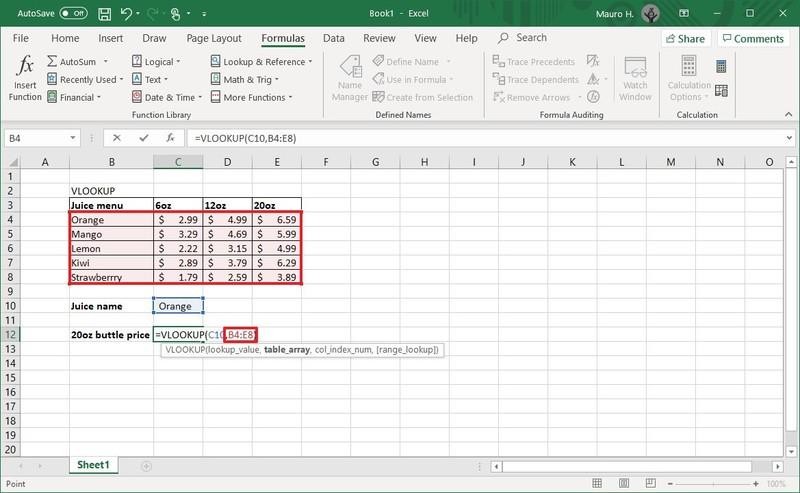
=VLOOKUP(lookup\_value,table\_array,col\_index\_num,range\_lookkup)

* + lookup\_value: defines the cell that includes the product identifier from the first column on the left.

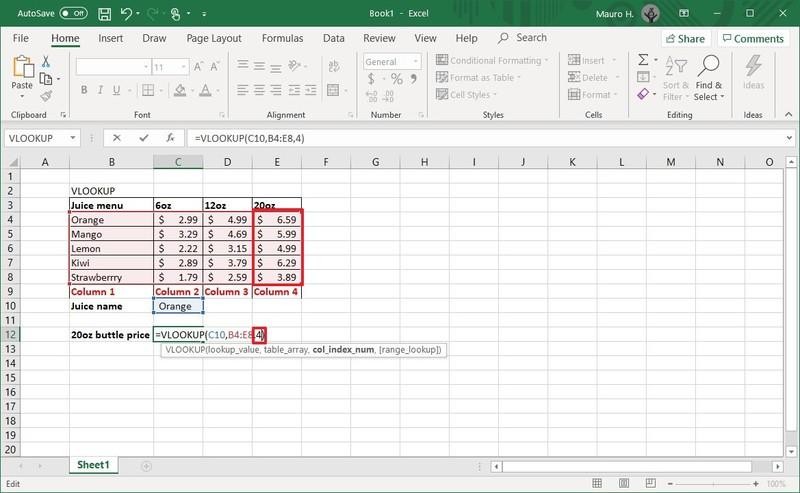


* + table\_array: defines the range of data where you want to perform a search. Typically, you would

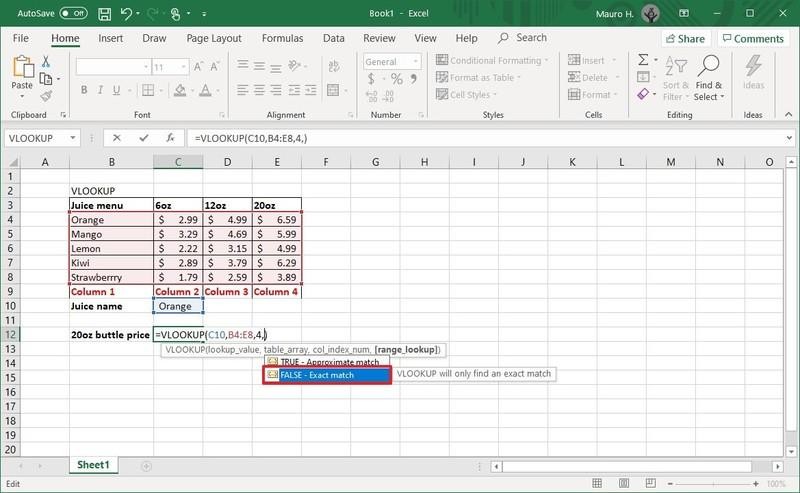
select the entire Excel table.



* + col\_index\_num: defines the column number that the function will look to find a value. When specifying multiple columns, you should do from left to right.



* + range\_lookkup: includes two options: "false" for exact match or "true" for an approximate match. Usually, you want to use the false option.

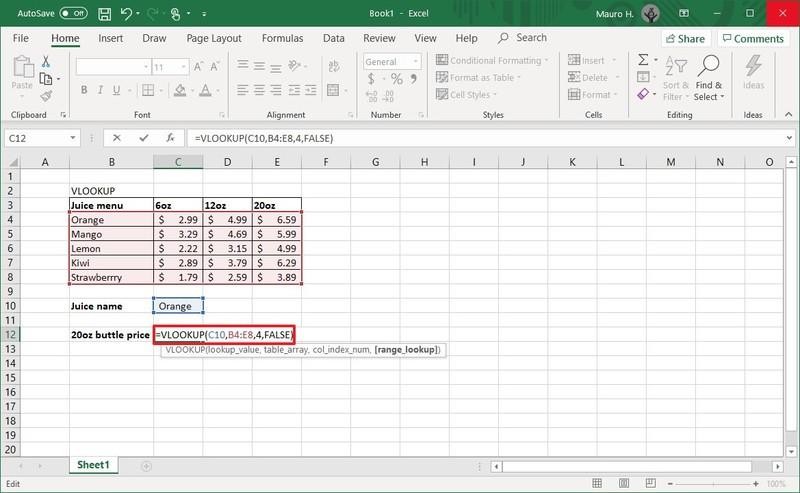


* + Quick note: If you don't specify a value, then the "true" option will be applied by default. Sometimes, when using the "true" option, the first column needs to be shorted, which may cause an unexpected result. If you're not getting the correct value, you should use the "false" option or sort the first column alphabetically or numerically.

In the command, make sure to update the variables inside the parenthesis with the information you want to query. Also, remember to use a comma to separate each value in the function. You do not need a space between each comma.

Here's an example that returns the price for the 20oz bottle of orange juice:

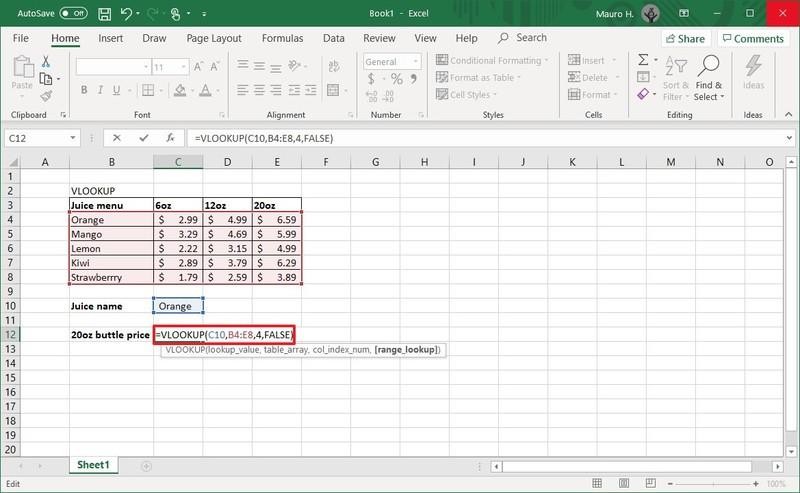
=VLOOKUP(C10,B4:E8,4,FALSE)



Once you complete the steps, the feature will return the value for the item you specified on step No. 4. If you receive the "#NAME?" error value, then it means that the formula is missing one or multiple quotes.

If you are trying to find data for another item, update the name of the cell on step No. 4. For example, if you want to see the price for the "20oz" bottle of Kiwi juice, then replace "Orange" with "Kiwi" in the "lookup\_value" cell and press Enter to update the result.

## Output/Results snippet:



## Activity 4

**Aim:** Plot basic charts in excel over numeric data series (5Hrs).

# **Learning outcome**: Able to business analytics and develop business intelligence.

###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

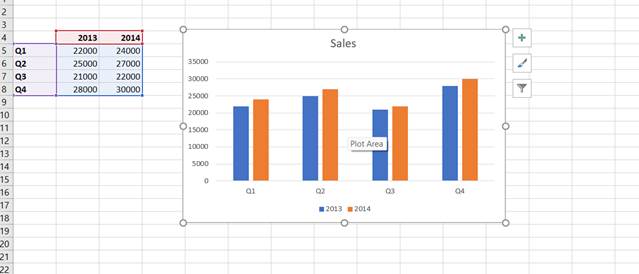
## Code/Program/Procedure (with comments):

1.     Open your Excel worksheet and create numeric data series.

2.     Select data series click on insert and select chart.

|  |
| --- |
|  |
|  |  |

3.     Choose required chart type.

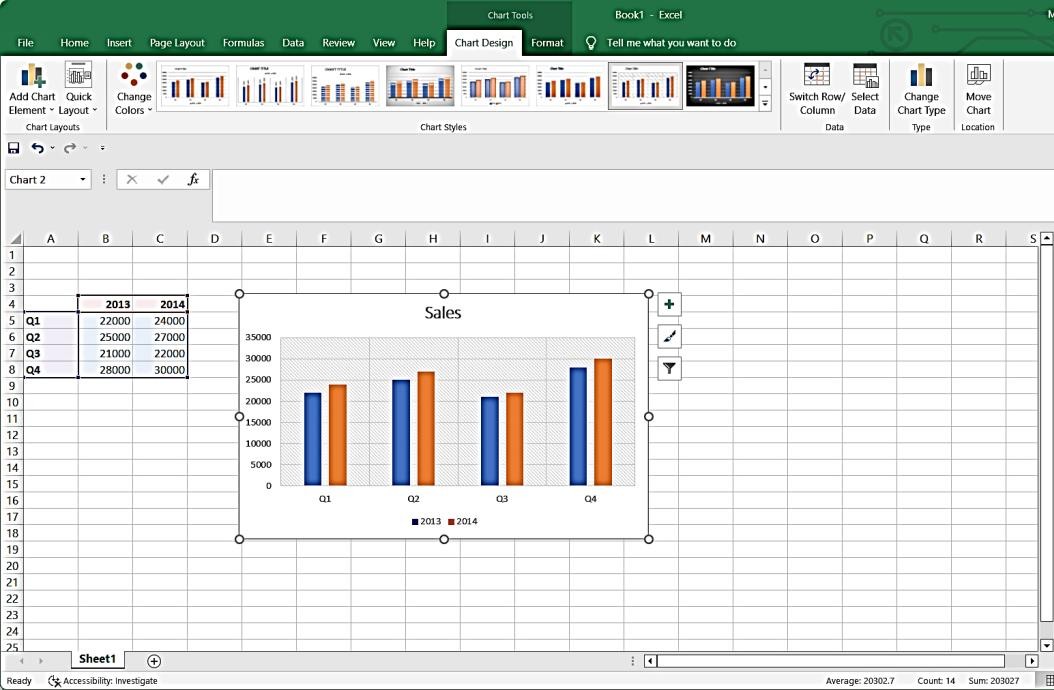


|  |
| --- |
|  |
|  |  |

4. Customize the chart by clicking the "Design," "Layout" and "Format" tabs of the Ribbon. Change the color with the Chart Styles options of the Design tab, add data labels, titles and shapes from the Layout tab and modify the colors, fill and effects from the Format tab. Save your Excel spreadsheet when complete.

|  |
| --- |
|  |
|  |  |

## Output/Results snippet:



## Activity 5

**Aim:** Plot uniform and binomial distributions in excel.(5hour)

# **Learning outcome**: Able to business analytics and develop business intelligence.

###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

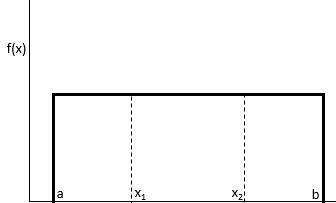
## Code/Program/Procedure (with comments):

**Uniform distribution**

A uniform distribution is a probability distribution in which every value between an interval from a to b is equally likely to be chosen.

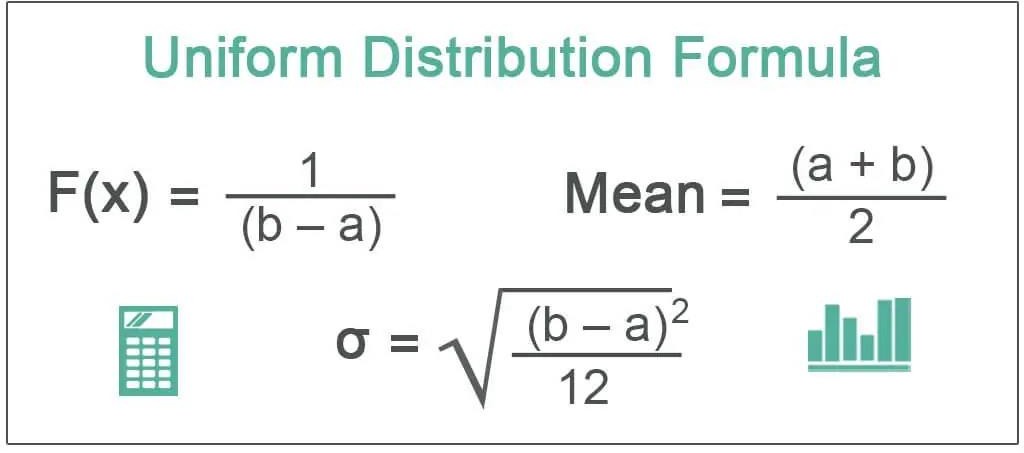
The probability that we will obtain a value between x1 and x2 on an interval from a to b can be found using the formula

P(obtain value between x1 and x2) = (x2 – x1) / (b – a)



The uniform distribution has the following properties:

* The mean of the distribution is μ = (a + b) / 2
* The variance of the distribution is σ2 = (b – a)2 / 12
* The standard deviation of the distribution is σ = √σ2



The following examples show how to calculate probabilities for uniform distributions in Excel.

Note: You can double check the solution to each example below using the Uniform Distribution Calculator.

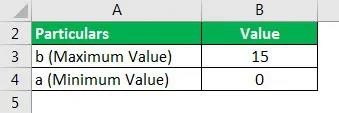
## Uniform Distribution in Excel

Let us take the example of an employee of company ABC. He normally takes up the services of the cab or taxi for the purpose of traveling from home and office. The duration of the wait time of the cab from the nearest pickup point ranges from zero and fifteen minutes.

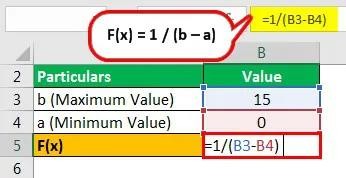
Help the employee determine the probability that he would have to wait for approximately less than 8 minutes. Additionally, determine the mean and standard deviation with respect to the wait time. Determine the probability density function as displayed below wherein for a variable X; the following steps should be performed:

## Solution

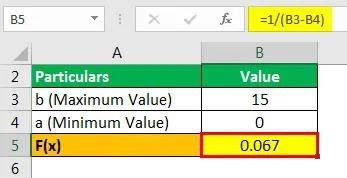
Use the given data for the calculation of uniform distribution.

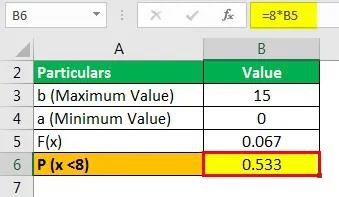


Calculation of the probability of the employee waiting for less than 8 minutes.



Uniform Distribution Formula Example 1.1

= 1 / (15 – 0)

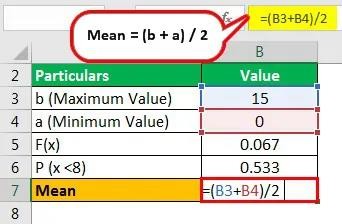
Uniform Distribution Formula Example 1.2 F(x) = 0.067

P (x < k) = base x height P (x <8) = (8) x 0.067

P (x <8) = 0.533

Therefore, for a probability density function of 0.067, the probability that the waiting time for the individual would be less than 8 minutes is 0.533.

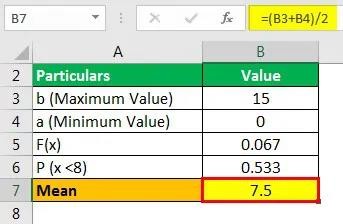
Calculation of mean of the distribution –



Uniform Distribution Formula Example

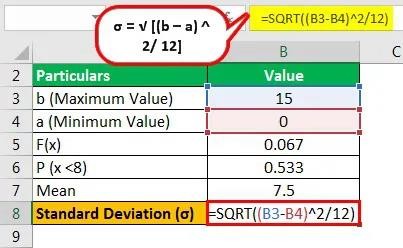
= (15 + 0) /2

Mean will be –



Mean = 7.5 minutes.

Calculation of standard deviation of the distribution –



σ = √ [(b – a) ^ 2/ 12]

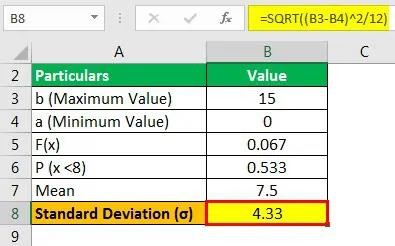
= √ [(15 – 0) ^ 2/ 12]

= √ [(15) ^ 2/ 12]

= √ [225 / 12]

= √ 18.75

Standard Deviation will be –



σ = 4.33

Therefore, the distribution shows a mean of 7.5 minutes with a standard deviation of 4.3 minutes.

## Binomial distributions

The BINOM.DIST function is categorized under Excel Statistical functions. It calculates the binomial distribution probability for the number of successes from a specified number of trials. This binomial distribution Excel guide will show you how to use the function, step by step.

The binomial distribution is a statistical measure that is frequently used to indicate the probability of a specific number of successes occurring from a specific number of independent trials. The two forms used are:

The Probability Mass Function – Calculates the probability of there being exactly x successes from n independent trials

The Cumulative Distribution Function – Calculates the probability of there being at most x successes from n independent trials

In financial analysis, the BINOM.DIST function can be useful in finding out, for example, the probability of publishing a best-selling book from a range of books to be published by a company.

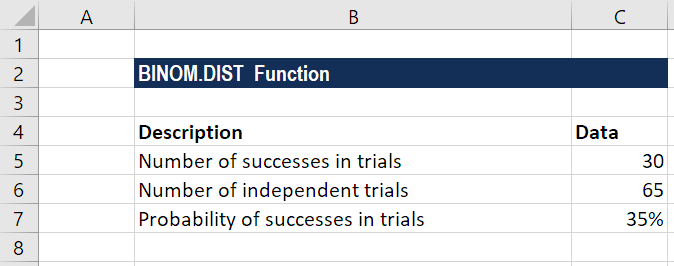
BINOM.DIST function is an updated version of the BINOMDIST function.

## Formula for Binomial Distribution

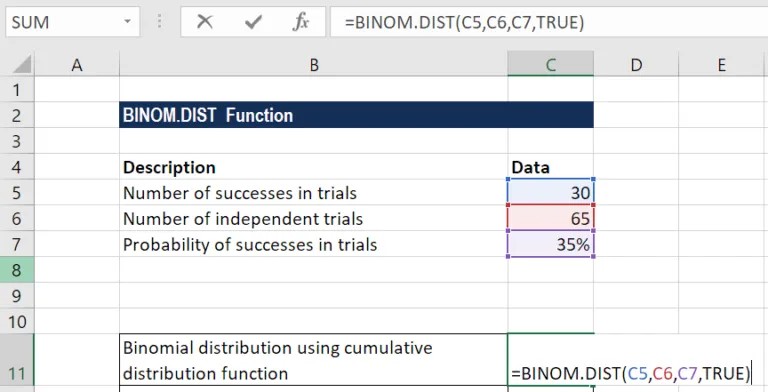
=BINOM.DIST(number\_s,trials,probability\_s,cumulative) The BINOM.DIST uses the following arguments:

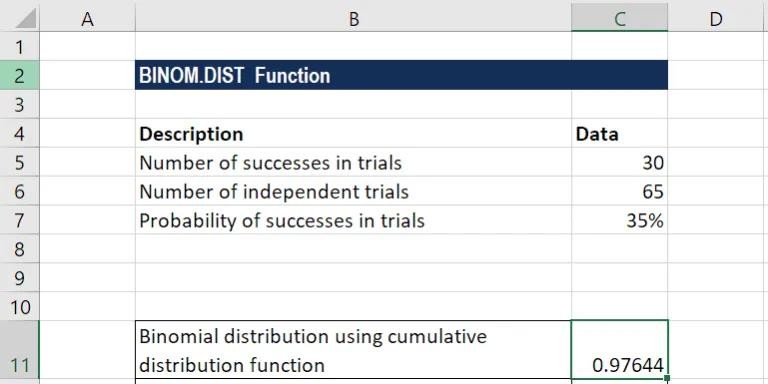
* + Number\_s (required argument) – This is the number of successes in trials.
  + Trials (required argument) – This is the number of independent trials. It must be greater than or equal to 0.
  + Probability\_s (required argument) – This is the probability of success in each trial.
  + Cumulative (required argument) – This is a logical value that determines the form of the function. It can either be:
  + TRUE – Uses the cumulative distribution function.
  + FALSE – Uses the probability mass function.

1. Suppose we are given the following data:

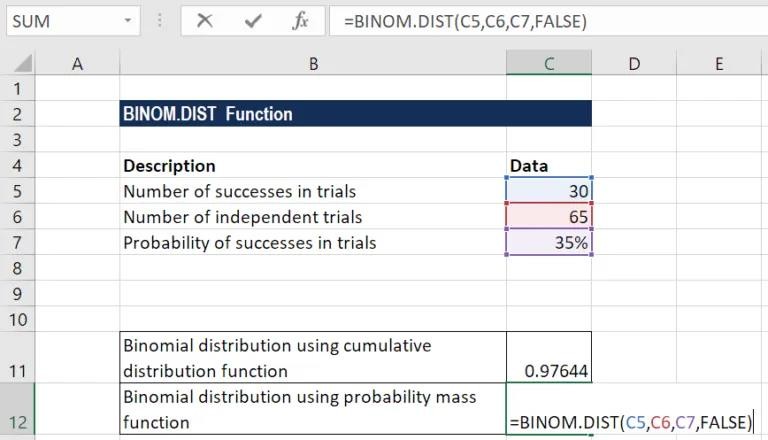


1. Apply the formula for calculating binomial distribution using the cumulative distribution function as shown below:



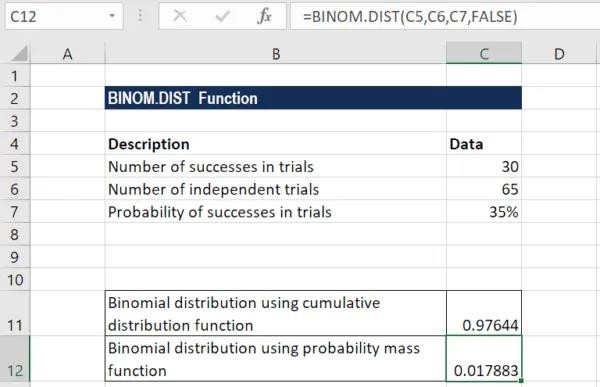
1. After applying the formula, we get the result below:
2. The

formula for calculating binomial distribution using the probability mass function is shown below:



We get the result below:

## Output:



## Activity 5

**Aim:** Implement Central limit theorem in excel. (5 hour)

# **Learning outcome**: Able to business analytics and develop business intelligence.

###### Duration: 5 hours

# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

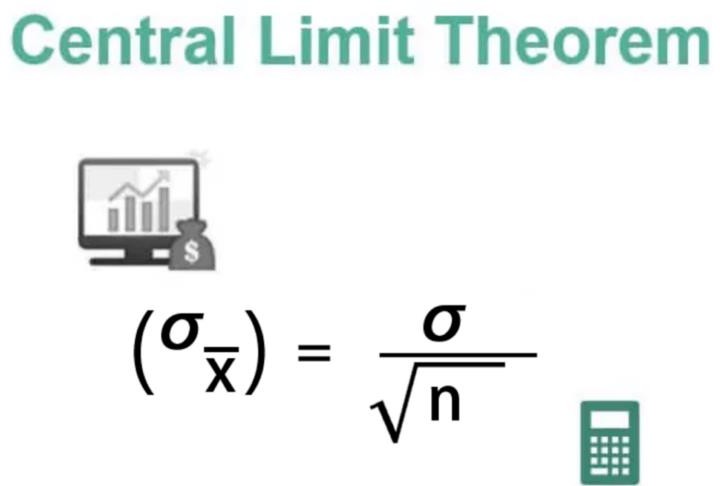
1. Windows 7/Windows 10
2. MS Office 2010 with Excel or Latest Version

## Code/Program/Procedure (with comments):

**Central Limit Theorem**

The central limit theorem is a sampling distribution theory. It states that normal distribution can be attained by increasing sample size. Thus, the population mean is represented by the average of random sample means.

## Central Limit Theorem Formula



The central limit theorem sets forth that the average of the sample means gives the population mean.



The central limit theorem is calculated using the following formula.

The sample’s standard deviation is computed by dividing the population’s standard deviation by the square root of sample size:

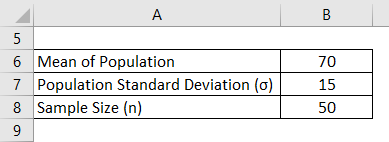


Here,

σ is the population standard deviation, σx is the sample standard deviation; and n is the sample size

Example:

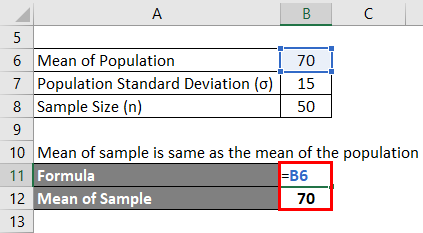
In a country located in the middle east region, the recorded weights of the male population are following a normal distribution. The mean and the standard deviations are 70 kg and 15 kg respectively. If a person is eager to find the record of 50 males in the population then what would mean and the standard deviation of the chosen sample?



## Solution:

* Calculation of Mean of Sample for example 1

Mean of Sample is the same as the mean of the population.

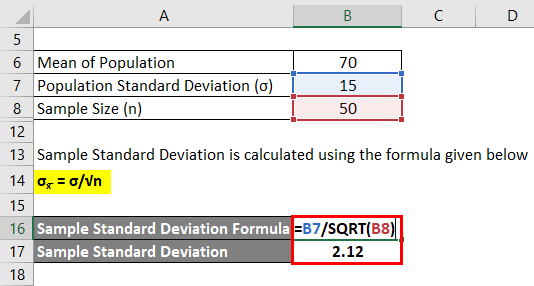


The mean of the population is 70 since the sample size > 30.

* Calculation of Sample Standard Deviation for example

Sample Standard Deviation is calculated using the formula given below

## σx= σ/√n



Sample Standard Deviation = 15 / √50 Sample Standard Deviation = 2.12

## Activity 7

**Aim:** Implement Central limit theorem in excel. (5hour)

# **Learning outcome**: Able to business analytics and develop business intelligence.

###### Duration: 5 hours

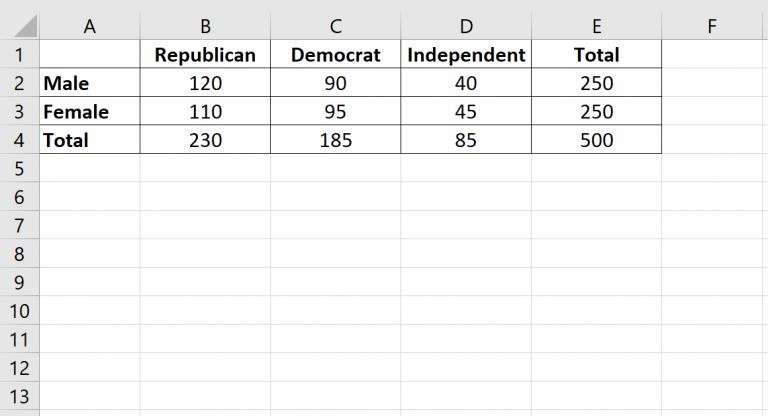
# List of Hardware/Software requirements:

* Laptop/Computer with Windows
* MS Office with Excel

## Code/Program/Procedure (with comments):

A Chi-Square Test of Independence is used to determine whether or not there is a significant association between two categorical variables.

## Chi-Square Test of Independence in Excel

Suppose we want to know whether or not gender is associated with political party preference. We take a simple random sample of 500 voters and survey them on their political party preference. The following table shows the results of the survey:

## Steps to perform a Chi-Square test of independence to determine if gender is associated with political party preference.

* Step 1: Define the hypotheses.

We will perform the Chi-Square test of independence using the following hypotheses:

H0: Gender and political party preference are independent.

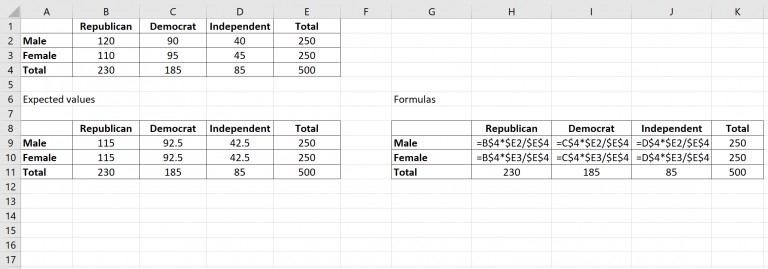
H1: Gender and political party preference are not independent.

* Step 2: Calculate the expected values.

Next, we will calculate the expected values for each cell in the contingency table using the following formula:

Expected value = (row sum \* column sum) / table sum.

For example, the expected value for Male Republicans is: (230\*250) / 500 = 115. We can repeat this formula to obtain the expected value for each cell in the table:



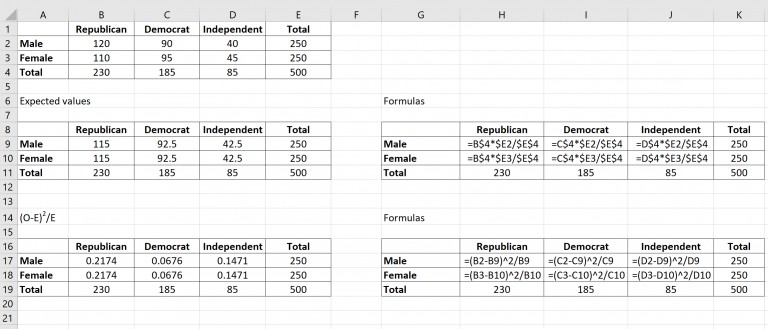
* Step 3: Calculate (O-E)2 / E for each cell in the table.

Next, we will calculate (O-E)2 / E for each cell in the table where: O: observed value

E: expected value

For example, Male Republicans would have a value of: (120-115)2 /115 = 0.2174.

We can repeat this formula for each cell in the table:



* Step 4: Calculate the test statistic X2 and the corresponding p-value. The test statistic X2 is simply the sum of the values in the last table.

The p-value that corresponds to the test statistic X2 can be found by using the formula:

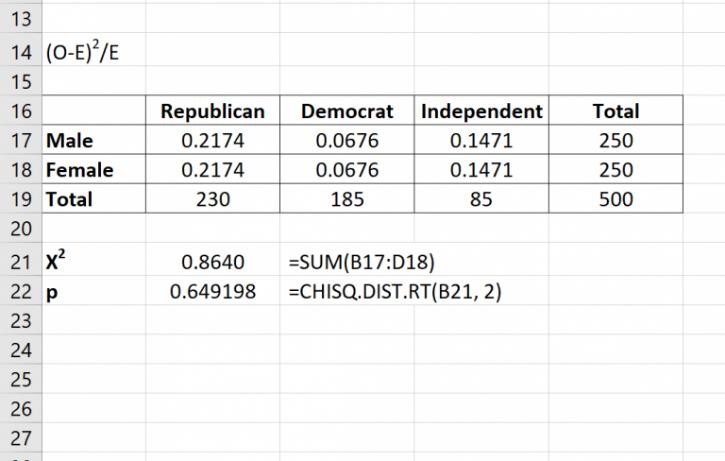
## =CHISQ.DIST.RT(x, deg\_freedom)

where:

x: test statistic X2

deg\_freedom: degrees of freedom, calculated as (#rows-1) \* (#columns-1)

The test statistic X2 turns out to be 0.8640 and the corresponding p-value is 0.649198.



* Step 5: Draw a conclusion.

Since this p-value is not less than 0.05, we fail to reject the null hypothesis. This means we do not have sufficient evidence to say that there is an association between gender and political party preference.