STATISTICAL TECHNIQUES

Interpreting Basic Statistical Values

Sample representative

How would one represent the "average" or "typical" piece of information from a given set of data? There are many different procedures that can be used to summarize the information depending on the type of question asked and the nature of data being summarized.

For example, the measures of a location provides' information about the location of the central tendency within a group of numbers. The measures of location presented in this unit of ungrouped (raw) data are the mean, median, and mode. So, what do these terms mean?

Mean

The arithmetic mean $\bar{x}$ is calculated by summing all of the observed values (x_i) or numbers and dividing it by the total number of observations (n). Note that the mean is sensitive to extremely large or small values as it can distort the mean and pull it towards the extreme value.

Median

The median is the middle value when the data is presented in an ordered manner i.e. usually in ascending order. If there are odd numbers of observations, the median is the middle number. However, if there is an even number of observations, the median is the average of the two middle numbers.

The median is used to summarize the distribution of an outcome. For example, if the distribution is skewed to the left or right, the median may be more appropriate to be used as compared to the mean.

In most developing countries where the household income disparity may be great between the "have" and "have not", the median income may be more appropriate as a measure compared to the mean. If the median is less than the mean, the data set is skewed to the right. If the median is greater than the mean, the data set is skewed to the left.

Mode

The mode refers to the most frequently occurring value in a data set. There are many reasons to use the mode compared to using the mean or median.

For example, if you look at the men's shirt sizes, the general Asian may use the medium size whereas their Caucasian counterparts may use the large size. So if the shirt manufacturer is catering for the export market, he will need to look into these matters. This is where the mode comes in place. To decide what sizes are purchased more, can use the mode for comparison purposes.

The next section onwards focuses on how you can perform and interpret some of the statistical calculations using Microsoft Excel's (Excel in short) built in functions.

Excel functions – An introduction

Excel consists of hundreds of built-in functions. When you install Microsoft Excel, these functions are installed too (with the exception of the **Add-In** functions).

So, what then is a function? A function in Excel is a predefined formula that returns a specific result. It mostly performs calculations that would otherwise be difficult in doing it manually.

Without the use of functions, the normal way of performing calculations is using the arithmetic operators such as plus sign (" **+** "), minus sign (" **-** "), multiplication or asterisk sign (" ***** "), division or forward slash (" **/** "), exponentiation or caret sign (" **^** "), and the percentage sign (" **%** "). Examples of these signs are illustrated below. Assume cells **A1** and **A2** have the values **6** and **3** respectively.

Arithmetic Operator	Example usage	Excel formula
+ (plus sign)	Addition (6+3)	**=A1+A2**
- (minus sign)	Subtraction (6-3) and Negation (-6)	**=A1-A2, -A1**
* (asterisk sign)	Multiplication (6*3)	**=A1*A2**
/ (forward slash)	Division (6/3)	**=A1/A2**
^ (caret sign)	Exponentiation (6^3)	**=A1^A2**
% (percentage sign)	Percentage (30%)	**=A1% or =A2%**

With the built-in functions, you do not have to manually construct formulas using the above operators. For example, instead of manually calculating the sum of two or more numbers, you could use the **SUM** function to calculate the result. Similarly, to find the average of two or more numbers, you could use the **AVERAGE** function. These are simple examples only.

More examples on the other functions available and how they will be used are explained later.

Authored by: Palani Murugappan

To understand how functions work, a simple exercise of calculating the summation of two numbers is explained. In this example, the **SUM** function is selected. Assume cell **A2** has the value **6** and cell **A3** the value **3**. Select cell **A4**. This is where the value of the **SUM** function will be pasted.

Before proceeding with the selection of the **SUM** function, one should know that functions in Excel fall into several categories. To view these categories, click the **Insert** Menu. Select **Function**.
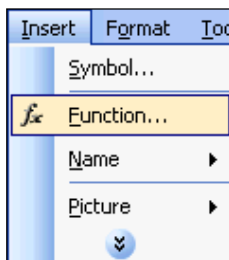
The **Insert Function** dialog box is opened as illustrated.

Authored by: Palani Murugappan
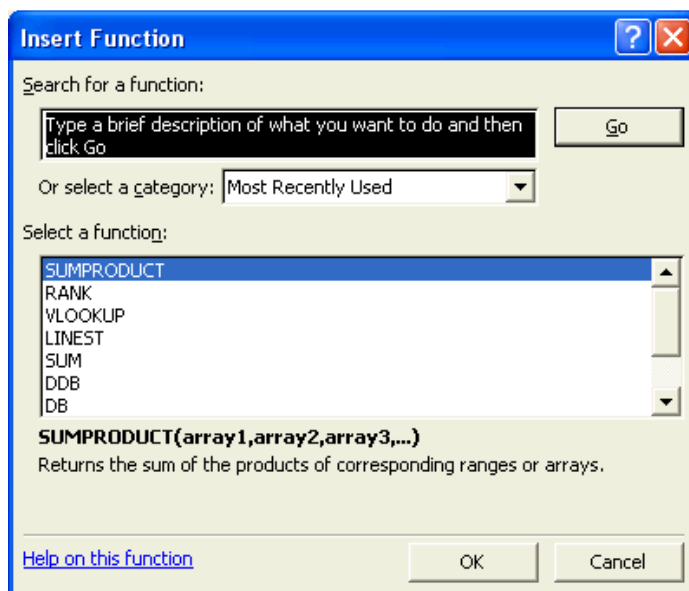
Note that in the above **Insert Function** dialog box, the last function used will be displayed on the top of the **Select a function** list. In this case, it is the **SUMPRODUCT** function. Your screen may show another function.

To view the other function categories, click the **Or select a category** drop down list. This will produce the following categories as illustrated.
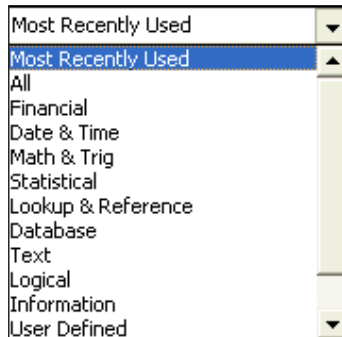

Notice that for every function selected, there is a description of that selected function. For example, if the **SUM** function is selected from the **Math & Trig** category, a brief description is outlined below the **Select a function** list as illustrated.
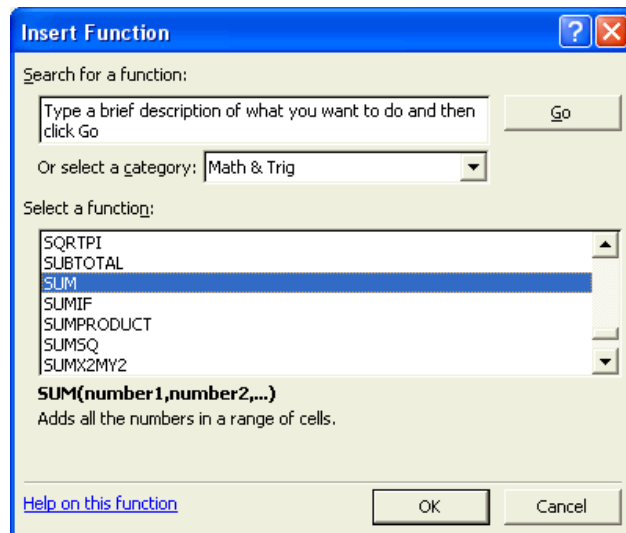
A point to note is that all functions consists of two elements i.e. the function name and an argument list. In the above **SUM** function, the function name is **SUM**. The argument list is **number1**, **number2**, etc. The argument list usually includes one or more operands enclosed in parentheses. Depending on the argument, it can be numeric or text constant, or a single reference or multiple cell references.
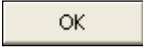
Authored by: Palani Murugappan

Coming back to the **Insert Function** dialog box, upon selecting the function, click the

OK button. The **Function Arguments** dialog box is displayed.
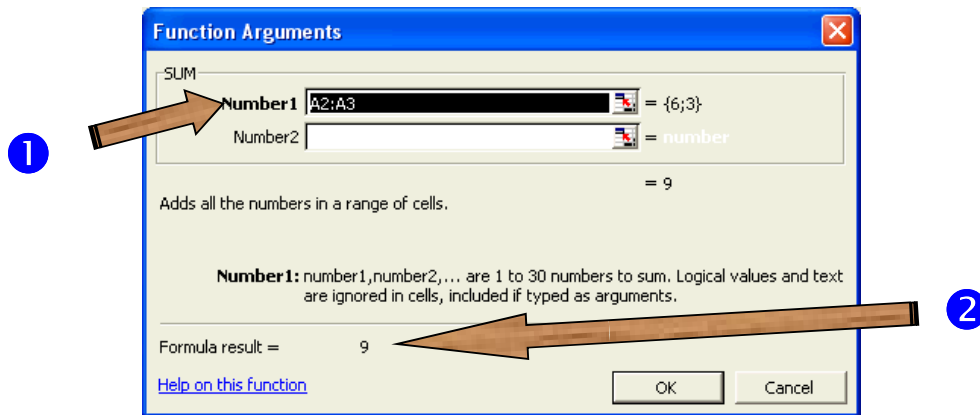
Observe that in the above **Function Arguments** dialog box, Excel has selected the range of the two cells and calculated the summation value of the cells **A2** and **A3** (❶). The argument list, in this case the values of **6** and **3** is selected as the first argument i.e. **Number1**. The final value of **9** is displayed as the **Formula result** (❷).

Also, observe the **Formula Bar**. The selected function, **SUM**, is displayed with the argument list i.e. **(A2:A3)**. This is illustrated below.

SUM function displayed here
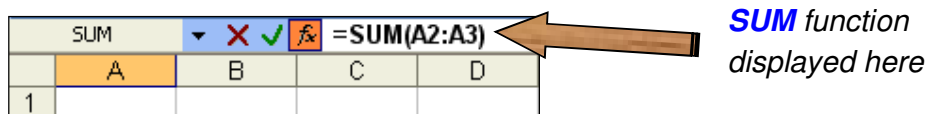
Click the OK button to close the **Function Arguments** dialog box. The final result of **9** is displayed in cell **A4** as illustrated.

The above example illustrates the basis of using Excel functions. Some of the other points that you should know of functions is that all functions must be preceded by an equal sign (" **=** "). Next, the function name and arguments can be typed in either uppercase or lowercase characters. Multiple arguments must be separated by commas. Note that certain functions do not require arguments. However, it is still necessary to include and empty argument list i.e. ( ).

Authored by: Palani Murugappan

Order of operators

This was explained in the earlier part. However, to refresh your memory, here are some examples to enhance your understanding.

Before performing any calculations, you should know the order of the operators. For example, assume you have the following data in cells **A1**, **A2**, and **A3** as illustrated.

	A	B
1	6	
2	3	
3	2	
4		

If you want to calculate the result of **6 + 3 x 2**, the normal mathematical way of interpreting it is reading the numbers from left to right and performing the calculations one at a time. The answer to this is **18**. However, in Excel, this method does not hold. Assume you had typed the formula as the way it is read in cell **C1** i.e. this would be interpreted as **=A1+A2*A3**.

C1		▼	f_x =A1+A2*A3	
	A	B	C	D
1	6		12	
2	3			
3	2			
4				

Observe that the value **12** is returned in cell **C1**, which is not quite the expected answer. To get the right answer, use parentheses or brackets. Thus, the above formula should read as **=(A1+A2)*A3**. Type this formula in cell **C2** and observe the result. It should be as illustrated below.

C2		▼	f_x =(A1+A2)*A3	
	A	B	C	D
1	6		12	
2	3		18	
3	2			
4				

The example above shows that if the brackets are not specified, Excel calculates the result of **A2*A3** and adds this value to cell **A1**, which gives a final value of **12**. Thus, it is important for you to know the order of the operators.

Another example using the above data is calculating the negated value of a cell and adding it to the exponentiation result. For example, how would you write the formula for calculating the negated value of cell **A1** and subtracting it from the exponentiation value of cells **A2** and **A3**, and placing the result in cell **C3**? Think about it first.

To answer the above question, first learn to use brackets whenever possible. The above question can be translated as the formula i.e. **=-A1-(A2^A3)**, as illustrated.

C3		f_x =-A1-(A2^A3)	
	A	B	C	D
1	6		12	
2	3		18	
3	2		-15	
4				

A result of **-15** is achieved in cell **C3**. Observe that in this example, if the brackets were removed, you would still get a final answer of **-15** as illustrated.

C3		f_x =-A1-A2^A3	
	A	B	C	D
1	6		12	
2	3		18	
3	2		-15	
4				

Why is this so? Excel gives higher precedence to the calculation of the exponentiation (**^**) operator. Thus, in the above formula, Excel calculates the result of **A2^A3** first, before subtracting this value from the negated value of cell **A1**.

Precedence of operators

Below is a summary of the order of operators that you will need to keep in mind when performing any calculations. The list is from the highest order (parentheses or brackets) to the lowest order (addition and subtraction). Note that multiplication (*) and division (/) both have the same order of precedence. The same applies to addition (+) and subtraction (-). However, multiplication and division have higher precedence to addition and subtraction.

Authored by: Palani Murugappan

	Operator	Meaning	Example
High	( )	Parentheses or brackets	=(A1+A2)*A3-(A3-A1)/A3
	-	Negation	=-A1
	%	Percentage	=A2%
	^	Exponentiation	=A2^A3
	*	Multiplication	=A1*A3
	/	Division	=A1/A3
	+	Addition	=A1+A2+A3
Low	-	Subtraction	=A1-A2

Naming cells

The assigning of descriptive names to cells and cell ranges is encouraged. Why should you name a cell? Naming a cell helps to identify which value or text is being referred to by cell addresses in the formulas. Excel has the capability of assigning names based on the user's choice to cells and ranges of cells using these names in formulas instead of cell addresses. To illustrate the use of cell names, the previous example will be used.

1. Assume you want to name the cell **C9** to **Total_Sales**. First select cell **C9** as illustrated.

	A	B	C	D	E
	C9	▼	fx =SUM(C4:C8)		
1	Sales by Region				
2					
3	Region	Quantity	Sales ($)	Percentage	
4	North	85	255	23.48%	
5	South	92	276	25.41%	
6	East	64	192	17.68%	
7	West	121	363	33.43%	
8					
9	Total:	362	1086		
10					

Authored by: Palani Murugappan

2. In the **Name Box**, type **Total_Sales** and press the **Enter** key.

Total_Sales	▼	*fx* =SUM(C4:C8)	
Name Box	B	C	D
1	Sales by Region		

3. Change the formula of cell **D4** so that it reads as **=C4/Total_Sales*100**. Note that if your cell displays some other number, right click on the cell; select **Format Cells**; ensure that the **Number** tab is selected; and select the **Number** category. Click and drag the **Fill Handle** to copy the formula to cells **D5**, **D6**, and **D7**. The following should be observed.

D4		▼	*fx* =C4/Total_Sales*100		
	A	B	C	D	E
1	Sales by Region				
2					
3	Region	Quantity	Sales ($)	Percentage	
4	North	85	255	23.48	
5	South	92	276	25.41	
6	East	64	192	17.68	
7	West	121	363	33.43	
8					
9	Total:		362	1086	
10					

4. Observe the formulas in cells **D5** to **D7**. Excel has copied the cell name across to the other cells too. By using cell names, there is no need for the user to make a column or row absolute. This is the advantage of naming a cell.

5. On similar grounds, you can name a range of cells too. For example, to name the range of cells from **B4** to **B7** as **Qty_by_region**, first select these cells. In the **Name Box**, type **Qty_by_region** and press the **Enter** key. This is illustrated below.

Qty_by_region	▼	*fx* 85		
Name Box	B	C	D	E
1	Sales by Region			
2				
3	Region	Quantity	Sales ($)	Percentage
4	North	85	255	23.48
5	South	92	276	25.41
6	East	64	192	17.68
7	West	121	363	33.43
8				
9	Total:	362	1086	
10				

Authored by: Palani Murugappan

6. Delete the formula in cell **B9**. Now click the **AutoSum** ([Σ ▾]) button on the **Standard** toolbar. By default, Excel selects the entire range from **B4** to **B8** to perform the summation. This is illustrated below.

	A	B	C	D	E
	SUM	▾ X ✓ ƒ×	=SUM(B4:B8)		
1	Sales by Region				
2					
3	Region	Quantity	Sales ($)	Percentage	
4	North	85	255	23.48	
5	South	92	276	25.41	
6	East	64	192	17.68	
7	West	121	363	33.43	
8					
9	Total:	=SUM(B4:B8)			
10		SUM(**number1**, [number2], ...)			
11					

7. However, we only want the summation of cells **B4** to **B7**. Press and hold the **Shift** key and press the **Up** (↑) arrow key once. This procedure will ensure that only cells **B4** to **B7** are selected as illustrated.

	A	B	C	D	E
	SUM	▾ X ✓ ƒ×	=SUM(Qty_by_region)		
1	Sales by Region				
2					
3	Region	Quantity	Sales ($)	Percentage	
4	North	85	255	23.48	
5	South	92	276	25.41	
6	East	64	192	17.68	
7	West	121	363	33.43	
8			4R x 1C		
9	Total:	=SUM(Qty_by_region)			
10		SUM(**number1**, [number2], ...)			
11					

8. Observe that by performing the above, the arguments of the **SUM** function is now replaced by a cell name assigned to the range selected. The formula now reads as **=SUM(Qty_by_region)**.

When selecting a name for a cell or range of cells, bear in mind of the following:
- ➢ Name should begin with a letter
- ➢ Name can contain up to 255 characters
- ➢ Name may consist of letters, numbers, and underscores (_)
- ➢ Name must not have spaces between characters
- ➢ No arithmetic operators or punctuation characters can be used
- ➢ Always select a name that makes sense, and a name that you will remember
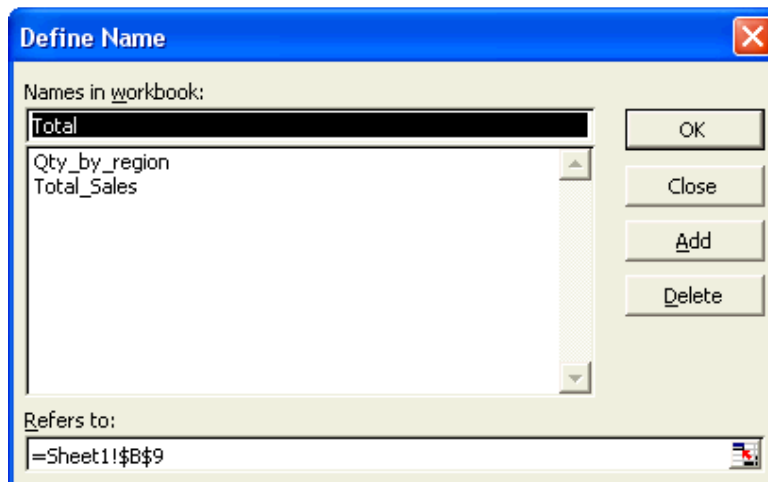
Authored by: Palani Murugappan

Deleting a cell name

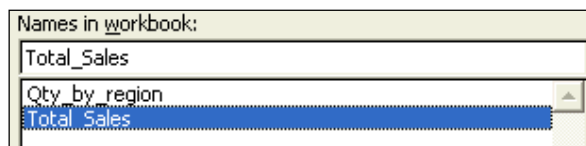How would you delete a cell name? Here is how to do so.

1. First, click the **Insert** Menu. Select **Name**, followed by **Define**. This step is illustrated below.
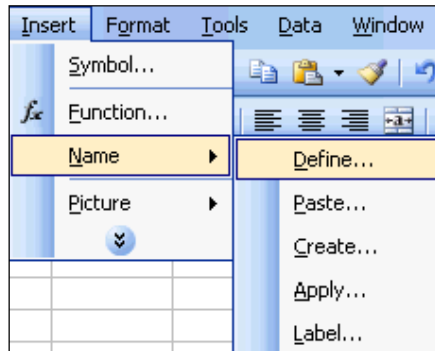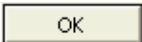
2. The **Define Name** dialog box is opened as illustrated (assuming the above cell **C9** is still selected).

3. To delete the **Total_Sales** cell name, click on its name in the **Names** in workbook list.

Authored by: Palani Murugappan

4. Click the [Delete] button. The selected cell name is deleted. Click the [OK] button to close the **Define Name** dialog box. Select cell **D4** and observe your data. There appears to be errors as illustrated.

D4		▼	*fx* =C4/Total_Sales*100		
	A	B	C	D	E
1	**Sales by Region**				
2					
3	**Region**	**Quantity**	**Sales ($)**	**Percentage**	
4	North	85	◈ 55	#NAME?	
5	South	92	276	#NAME?	
6	East	64	192	#NAME?	
7	West	121	363	#NAME?	
8					
9	Total:		362	1086	
10					

5. The cells **D4** to **D7** is still referring to the cell name **Total_Sales**. Edit these formulas as it was before applying the cell name. It should be as follows. You may use either the mixed or absolute cell referencing here. Use the **Fill Handle** to copy the edited formula in cell **D4** to the cells **D5**, **D6**, and **D7**.

D4		▼	*fx* =C4/C$9*100		
	A	B	C	D	E
1	**Sales by Region**				
2					
3	**Region**	**Quantity**	**Sales ($)**	**Percentage**	
4	North	85	255	23.48	
5	South	92	276	25.41	
6	East	64	192	17.68	
7	West	121	363	33.43	
8					
9	Total:		362	1086	
10					

Circular Reference

The other aspect of referencing is the notion called circular referencing. A formula that depends on its current cell value is known to have a circular reference. In other words, the cell contains a formula that refers to itself.
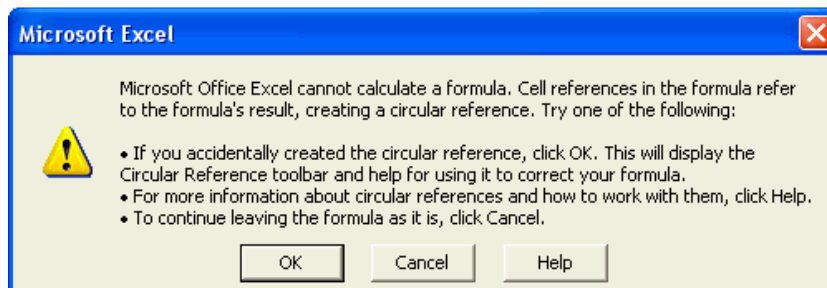
The following example will highlight this.

1. Suppose you wanted to calculate the total **Percentage** in cell **D9**. Somehow due to human error, you had typed the formula to include the cell **D9** too. The formula in cell **D9** thus reads as **=SUM(D4:D9)** as illustrated.
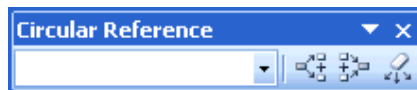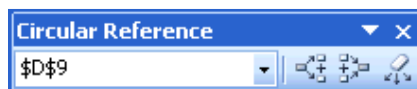
Authored by: Palani Murugappan

2. The moment you press the **Enter** key, you will be prompted with the following dialog box.

3. Click the [OK] button. The following **Circular Reference** toolbar is displayed.

4. Select cell **D9**. Note that upon selecting cell **D9**, the absolute value of **D9** i.e. **D9** is displayed in the **Circular Reference**.

5. Click the **Trace Precedents** ( ) button on the **Circular Reference** toolbar. This will highlight the cells used in performing the calculation as illustrated below.

Authored by: Palani Murugappan

Trace
Precedents
highlights
cells
involved

6. Click the **Remove All Arrows** ( ) button on the **Circular Reference** toolbar. This will remove the arrows based on the calculation of the formula.

7. Next, edit the formula in cell **D9** to read as **=SUM(D4:D7)**, and press the **Enter** key. The final total of cell **D9** should be **100.00** as illustrated.

The above examples have been shown to ensure that you have the basic necessary foundation to perform some of the many functions that will be covered in this book later. Also, some of the concepts covered above will help to refresh the knowledge of the many new users and those who have not used it for some time.

Common error messages

| Error value	Meaning
#DIV/0!	An attempt by the formula or function to divide by 0.
#N/A	A non-existing value i.e. a blank cell or cell that is not available is referred to.
#NAME?	Use of unrecognized Excel name.
#NULL!	Specifies an invalid intersection of two areas.
#NUM!	The incorrect use of a number in the argument of a function. Also displayed if the numbers are not understood by Excel (numbers too large or small).
#REF!	Refers to a non-existing cell.
#VALUE!	Incorrect use of argument.
#####	Result is too long to be fitted in a cell. To rectify this, increase the width of the cell.

Common Statistical functions

Most users of Excel are already familiar with the frequently used function **SUM**. The other commonly used functions are **AVERAGE**, **COUNT**, **MAX**, and **MIN**. These functions fall under the **Statistical** function category. To view the details of these functions, select an empty cell and click the **Insert** Menu. Select **Function**.

Authored by: Palani Murugappan
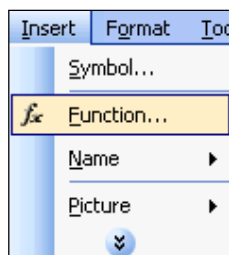
The **Insert Function** dialog box is opened. Select the **Statistical** category from the **Or select a category** drop down list.
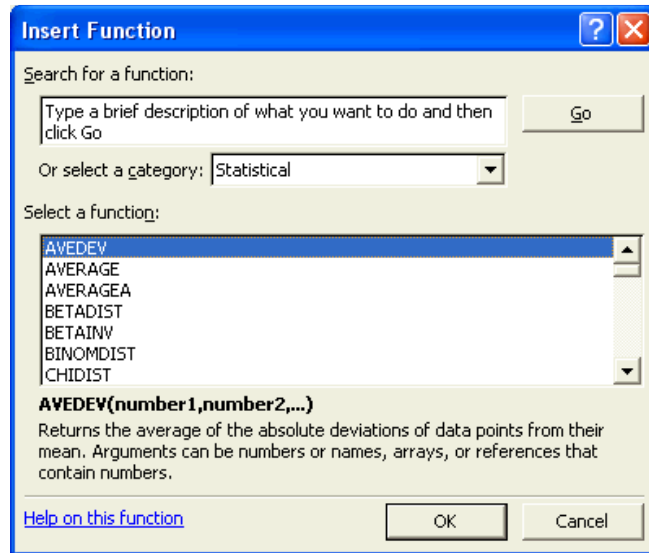
To understand how these commonly used functions work, a common set of data will be used to illustrate them.
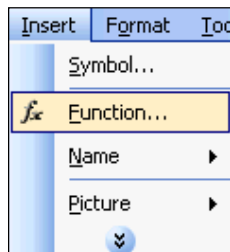
AVERAGE function

The **AVERAGE** function calculates the unweighted arithmetic mean for a selected range of data. The following example will illustrate the use of this function.

1. Assume you have the following set of sales data for a specific branch.

	A	B	C	D	E
1	Sales by Month				
2	Northern Region				
3					
4	Month	Sales ($)			
5	January	131,500		Total Sales:	
6	February	148,600		Average Sales:	
7	March	133,100		Total Count:	
8	April	125,500		Maximum Sales:	
9	May	100,500		Minimum Sales:	
10	June	102,600			
11	July	114,700			
12	August	122,000			
13	September	126,000			
14	October	132,800			
15	November	141,200			
16	December	134,600			
17					

Authored by: Palani Murugappan

2. To begin with, select cell **E5**. To calculate the sum of sales for the **12** months, click the **AutoSum** (Σ) drop down list button on the **Standard** toolbar. Select the **SUM** function.

3. By default, Excel selects the values in row **5** as illustrated.

4. We do not want this selected range. Select the range of cells from **B5** to **B16**.

Authored by: Palani Murugappan

5. Press the **Enter** key after selecting the range. The summation of the total sales for 12 months is displayed in cell **E5** with a value of **1,513,100**.

	A	B	C	D	E
	Sales by Month				
1	Sales by Month				
2	Northern Region				
3					
4	Month	Sales ($)			
5	January	131,500		Total Sales:	1,513,100
6	February	148,600		Average Sales:	
7	March	133,100		Total Count:	
8	April	125,500		Maximum Sales:	
9	May	100,500		Minimum Sales:	
10	June	102,600			
11	July	114,700			
12	August	122,000			
13	September	126,000			
14	October	132,800			
15	November	141,200			
16	December	134,600			
17					

6. Now, to calculate the **AVERAGE** value for the 12 months of sales, select cell **E6**. Click the **AutoSum** drop down list button on the **Standard** toolbar. Be careful not to click on the **AutoSum** button, as this will invoke the **SUM** function. Select **Average** from the function list.

```
Σ ▼  A↓ Z↓  📊 📊 100
      Sum
      Average
      Count
      Max
      Min
      More Functions...
```

7. Again, select the range of cells from **B5** to **B16** as illustrated.

8. Press the **Enter** key. The average value for the **12** months is calculated as **126,092**.

9. Note that the average value can also be calculated using the **SUM** function and dividing the value by 12 (or using the **COUNT** function – to be covered later). To do a comparison, select cell **F6** and type **=SUM(B5:B16)/12**. A value of **126,091.67** is obtained as illustrated.

Authored by: Palani Murugappan

10. Note that the value of cell **E6** is rounded when using the **AVERAGE** function as compared to the value of cell **F6**, which is displayed in 2 decimal places. Alternatively, you may use the combination of **SUM** and **COUNT** functions to perform the average calculation. To do this, first select and delete the value of cell **F6**. Now type the formula **=SUM(B5:B16)/COUNT(B5:B16)**, as illustrated.

	F6	▼	fx	=SUM(B5:B16)/COUNT(B5:B16)		
	A	B	C	D	E	F
1	Sales by Month					
2	Northern Region					
3						
4	Month	Sales ($)				
5	January	131,500		Total Sales:	1,513,100	
6	February	148,600		Average Sales:	126,092	126091.67
7	March	133,100		Total Count:		
8	April	125,500		Maximum Sales:		
9	May	100,500		Minimum Sales:		

COUNT function

The **COUNT** function counts the number of cells with values only for a specified range. It ignores empty cells within the range. The following example will illustrate this.

1. Select cell **E7** in the given set of data.

	E7	▼	fx		
	A	B	C	D	E
1	Sales by Month				
2	Northern Region				
3					
4	Month	Sales ($)			
5	January	131,500		Total Sales:	1,513,100
6	February	148,600		Average Sales:	126,092
7	March	133,100		Total Count:	
8	April	125,500		Maximum Sales:	
9	May	100,500		Minimum Sales:	

2. Click the **AutoSum** drop down list. Select **Count** from the function list.

Σ ▼ | A↓ Z↓ | 100

Sum

Average

Count

Max

Min

More Functions...

Authored by: Palani Murugappan

3. Select the range **B5** to **B16**, as illustrated.

4. Press the **Enter** key. The **COUNT** function has counted the number of cells in the selected range as **12**.

5. Assume for some reason, there were no sales for the months of **August** and **September** i.e. delete the values in cells **B12** and **B13**. Observe that the value of **COUNT** function in cell **E7** is reduced to **10**, as illustrated below. The **COUNT** function is clever enough to eliminate counting any blank cells in a given range.

Authored by: Palani Murugappan

	A	B	C	D	E
	B12		fx		
1	Sales by Month				
2	Northern Region				
3					
4	Month	Sales ($)			
5	January	131,500		Total Sales:	1,265,100
6	February	148,600		Average Sales:	126,510
7	March	133,100		Total Count:	10
8	April	125,500		Maximum Sales:	
9	May	100,500		Minimum Sales:	
10	June	102,600			
11	July	114,700			
12	August				
13	September				
14	October	132,800			
15	November	141,200			
16	December	134,600			
17					

MAX function

The **MAX** function locates the largest or maximum value in a specified range. It is useful especially if you have a large set of data and need to find the largest value. The following example will illustrate the use of it.

1. Assume you have the following data. Select cell **E8** as illustrated.

	A	B	C	D	E
	E8		fx		
1	Sales by Month				
2	Northern Region				
3					
4	Month	Sales ($)			
5	January	131,500		Total Sales:	1,513,100
6	February	148,600		Average Sales:	126,092
7	March	133,100		Total Count:	12
8	April	125,500		Maximum Sales:	
9	May	100,500		Minimum Sales:	
10	June	102,600			

2. Click the **AutoSum** drop down list on the **Standard** toolbar. Select **MAX** from the function list.

Σ ▾ | A↓ Z↓ | 🕮 | 🕮 | 10(
- Sum
- Average
- Count
- Max
- Min
- More Functions...

Authored by: Palani Murugappan

3. Select the range of cells from **B5** to **B16**.

4. Press the **Enter** key. The largest or maximum value for the range is located and its value i.e. **148,600** displayed in cell **E8** as illustrated.

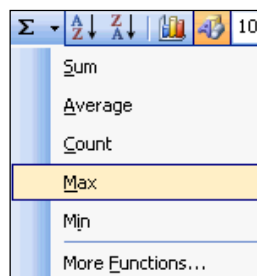# MIN function

The **MAX** function locates the smallest or minimum value in a specified range. It is useful if you have a large set of data and need to find the smallest value. These values are sometimes considered insignificant and may be omitted in other calculations. Thus, the use of **MIN** function. The following example will illustrate the use of it.

Authored by: Palani Murugappan

1. Assume you have the following data. Select cell **E9** as illustrated.

	E9	▼	fx			
	A	B	C	D	E
1	Sales by Month				
2	Northern Region				
3					
4	Month	Sales ($)			
5	January	131,500		Total Sales:	1,513,100
6	February	148,600		Average Sales:	126,092
7	March	133,100		Total Count:	12
8	April	125,500		Maximum Sales:	148,600
9	May	100,500		Minimum Sales:	
10	June	102,600			
11	July	114,700			
12	August	122,000			
13	September	126,000			
14	October	132,800			
15	November	141,200			
16	December	134,600			
17					

2. Click the **AutoSum** drop down list on the **Standard** toolbar. Select **MIN** from the function list.

Σ ▾	A↓ Z↓		100
	Sum	
	Average	
	Count	
	Max	
	Min	
	More Functions...	

3. Select the range of cells from **B5** to **B16** as illustrated.

	SUM	▼ X ✓ fx	=MIN(B5:B16)			
	A	B	C	D	E
1	Sales by Month				
2	Northern Region				
3					
4	Month	Sales ($)			
5	January	131,500		Total Sales:	1,513,100
6	February	148,600		Average Sales:	126,092
7	March	133,100		Total Count:	12
8	April	125,500		Maximum Sales:	148,600
9	May	100,500		Minimum Sales:	=MIN(B5:B16)
10	June	102,600		MIN(**number1**, [number2], ...)	
11	July	114,700			
12	August	122,000			
13	September	126,000			
14	October	132,800			
15	November	141,200			
16	December	134,600			
17					

Authored by: Palani Murugappan

4. Press the **Enter** key. The smallest or minimum value for the range is located and its value i.e. **100,500** displayed in cell **E9** as illustrated.

|	E9		▼		f_x	=MIN(B5:B16)	
	A	B	C	D		E	
1	Sales by Month						
2	Northern Region						
3							
4	Month	Sales ($)					
5	January	131,500		Total Sales:		1,513,100	
6	February	148,600		Average Sales:		126,092	
7	March	133,100		Total Count:		12	
8	April	125,500		Maximum Sales:		148,600	
9	May	100,500		Minimum Sales:		100,500	
10	June	102,600					
11	July	114,700					
12	August	122,000					
13	September	126,000					
14	October	132,800					
15	November	141,200					
16	December	134,600					
17							

5. Based on the examples of the common functions above, try to perform calculations for the average, maximum and minimum sales by quarters for the 12 months i.e. split the 12 months into four quarters and perform the calculations. In your own words, try to explain the deviations. Your final answer should be as illustrated.
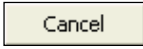
	A	B	C	D	E	F	G	H
1	Sales by Month							
2	Northern Region							
3								
4	Month	Sales ($)						
5	January	131,500		Total Sales:	1,513,100			
6	February	148,600		Average Sales:	126,092			
7	March	133,100		Total Count:	12			
8	April	125,500		Maximum Sales:	148,600			
9	May	100,500		Minimum Sales:	100,500			
10	June	102,600						
11	July	114,700		1st Quarter Sales			2nd Quarter Sales	
12	August	122,000		Average Sales:	137,733		Average Sales:	109,533
13	September	126,000		Maximum Sales:	148,600		Maximum Sales:	125,500
14	October	132,800		Minimum Sales:	131,500		Minimum Sales:	100,500
15	November	141,200						
16	December	134,600		3rd Quarter Sales			4th Quarter Sales	
17				Average Sales:	120,900		Average Sales:	136,200
18				Maximum Sales:	126,000		Maximum Sales:	141,200
19				Minimum Sales:	114,700		Minimum Sales:	132,800
20								

Authored by: Palani Murugappan

Central tendency of data

The **MODE** and **MEDIAN** are statistical functions too and in-built in Excel. To invoke either of these functions, click on the **Insert** Menu and select **Function**. The **Insert Function** dialog box is opened. Click to select the **Statistical** category as illustrated.
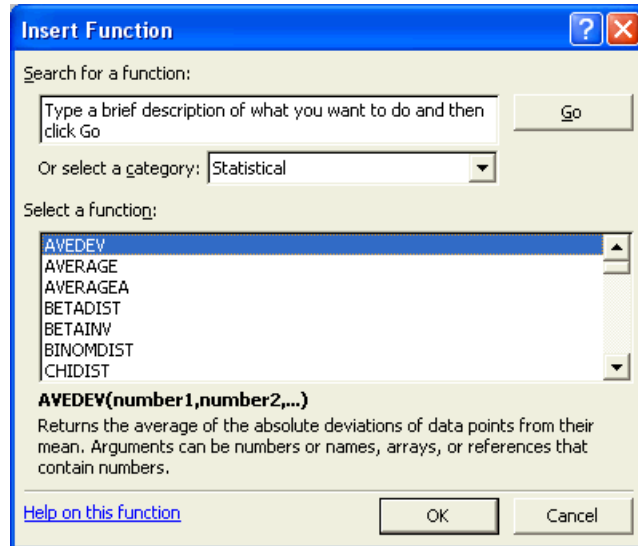
You can then click the **Cancel** button to close the above dialog box.

Assume you have the following set of data in Excel.

	A	B	C	D	E
1	50.5		Average:	48.45	
2	47.6		Geomean:	48.25244	
3	55.3		Harmean:	48.04919	
4	39.9		Mode:	#N/A	
5	43.6		Median:	48.25	
6	51.5				
7	48.9				
8	47.1				
9	53.3				
10	46.8				
11					

To view the formulas involved for the calculations of **AVERAGE**, **GEOMEAN**, **HARMEAN**, **MODE**, and **MEDIAN**, press the **Ctrl** + ~ keys. You should observe the respective formulas as follows.

Authored by: Palani Murugappan

	A	B	C	D	E
1	50.5		Average:	=AVERAGE(A1:A10)	
2	47.6		Geomean:	=GEOMEAN(A1:A10)	
3	55.3		Harmean:	=HARMEAN(A1:A10)	
4	39.9		Mode:	=MODE(A1:A10)	
5	43.6		Median:	=MEDIAN(A1:A10)	
6	51.5				
7	48.9				
8	47.1		Formula Auditing		
9	53.3				
10	46.8				
11					

Do not attempt to change the width of columns above. When the formulas are displayed (by pressing the **Ctrl** + ~ keys), the width of the columns are expanded automatically. Note that all of the formulas include the range **A1** to **A10**.

Press the **Ctrl** + ~ keys once again to display the original set of data. Note that the combination of **Ctrl** + ~ keys can be toggled to display the data and formulas at any time.

	A	B	C	D	E
1	50.5		Average:	48.45	
2	47.6		Geomean:	48.25244	
3	55.3		Harmean:	48.04919	
4	39.9		Mode:	#N/A	
5	43.6		Median:	48.25	
6	51.5				
7	48.9				
8	47.1				
9	53.3				
10	46.8				
11					

Let us analyze the data. Firstly, the average value for the set of observed data is calculated to be **48.45**. This value is the **arithmetic mean** of the values given.

The **geometric mean** for the set of data is calculated to be **48.25244**. The geometric mean is a specialized measure that is used to average proportional increases (the actual calculations can be quite tedious!).

The **harmonic mean** for the set of data is **48.04919**. The harmonic mean is another specialized measure of location and is used in particular circumstances only such as when the data set consists of a set of rates e.g. speed (km/h), productivity (output/hour), etc. It is calculated based on the reciprocal of the mean of the reciprocals of the item values.

Authored by: Palani Murugappan

The formula is as follows.

Harmonic mean (Hm) $= \dfrac{n}{\sum \dfrac{1}{x_i}}$

Where n is the total number of observed values and x_i is the observed value.

Note that care should be exercised when using the harmonic mean. For example, if the rates measured are averaged over constant numerator units, the harmonic mean should be used.

However, if the rates measured are averaged over constant denominator units, the arithmetic mean should be used.

To illustrate the above, let us consider the following production example. Suppose the production manager of a factory with two production lines wants to know the rate produced based on time and quantity. If the productivity is measured over a period of a month (i.e. same time – thus, the denominator is constant), the arithmetic mean is used to average the rate. However, if the production is measured based on quantity produced of say 10,000 items (i.e. the same quality – thus the numerator is constant), the harmonic mean would be more appropriate to average the rate.

Another example is as follows. Suppose you have two similar cars at home and want to measure the average petrol consumption. If both the cars were tested over a distance of 400 kilometers to give the two rates of 8 and 11 kilometers per liter (km/l), then the numerator unit (kilometer) is a uniform amount for both. Thus, the harmonic mean is calculated as follows:

Average petrol consumption (**Harmonic mean**) $= \dfrac{2}{\dfrac{1}{8} + \dfrac{1}{11}}$

which gives a value of **9.26 km/l**.

If both the cars are each filled with 40 liters of petrol and tested until the petrol runs out, then the denominator (liters of petrol) is the uniform amount and the arithmetic mean is more appropriate to be used. Thus, the arithmetic mean is calculated as follows:

Average Petrol consumption (**Arithmetic mean**) $= (8 + 11) / 2$

which gives a value of **9.5 km/l**.

How do you now interpret the above results?

The interpretation of the difference between the two averages is that the arithmetic mean of 9.5 km/l is based on consumption of 40 liters of petrol while the harmonic mean of 9.26 km/l is based on consumption over a distance of 400 kilometers.

Note that for all sets of data, the following relationship is always true:

Arithmetic mean is always greater than geometric mean which is always greater than harmonic mean i.e.

Arithmetic mean > Geometric mean > Harmonic mean

Coming back to the mode and median values, notice that **median** value is **48.25**. This implies that 50 percent of the observed data is greater than the median value and the balance 50 percent is less than the median value.

You can check this yourself. Observe that the values that are less than the median value (should be 5 values) are **47.6**, **39.9**, **43.6**, **47.1**, and **46.8** while the vales that are greater than the median value are **50.5**, **55.3**, **51.5**, **48.9** and **53.3**.

The **MODE** value returns an error of **#N/A**. This is so as all of the observed values are unique. The **MODE** should return the most commonly occurring value. In this case, as all the values are different, it returns the error message of **#N/A**.

To test if the **MODE** actually works, change the value of the 10th observation to the value of the 1st observation i.e. **50.5** in cell **A10**. This repeated value is then displayed as the **MODE** value as illustrated together with the changes for the other values.

A10		▼	fx 50.5	
	A	B	C	D	E
1	50.5		Average:	48.82	
2	47.6		Geomean:	48.621	
3	55.3		Harmean:	48.41337	
4	39.9		Mode:	50.5	
5	43.6		Median:	49.7	
6	51.5				
7	48.9				
8	47.1				
9	53.3				
10	50.5				
11					

Note that in the above illustration, the relationship where
Average mean > Geometric mean > Harmonic mean holds!