# MICROSOFT EXCEL

BASIC FORMATTING

- To create a new workbook: **[Start – All Programs – Microsoft Office - Microsoft Excel 2010]**
- To rename a sheet(1):
    - Select the sheet whose tab you want to rename (the selected worksheet is active and the current sheet name appears in bold letters).
    - Double-click the sheet tab. (You also can right-click the sheet tab and select **Rename** on its shortcut menu. The current name on the sheet tab appears selected.)
    - Replace the current name on the sheet tab by typing a new sheet name. (When you begin typing a new name, the previous name disappears.)
    - Press Enter.

- Other option used to rename a sheet (2): **[Home – Cells – Format – Organize Sheets – Rename Sheet]**
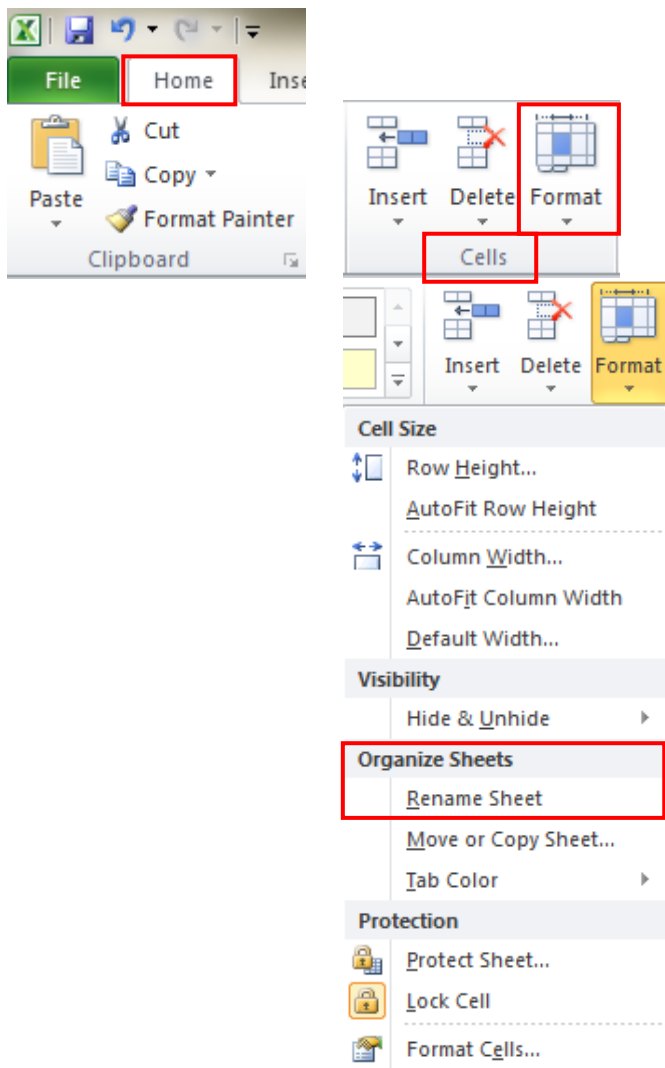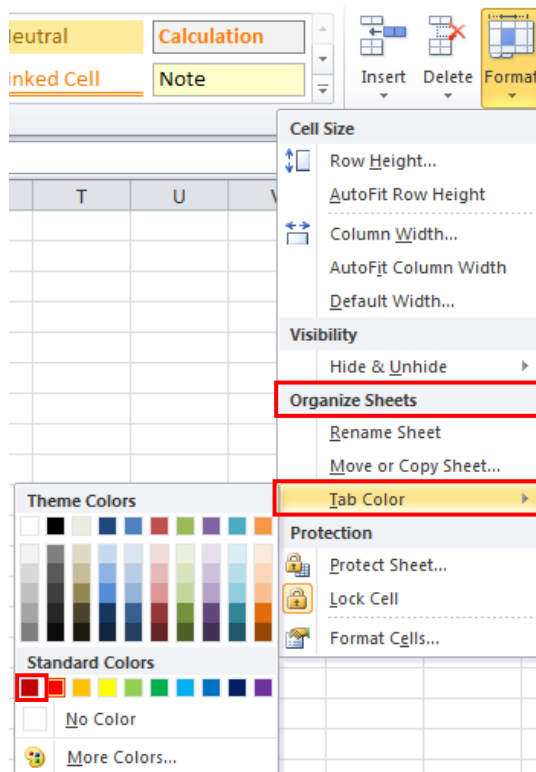

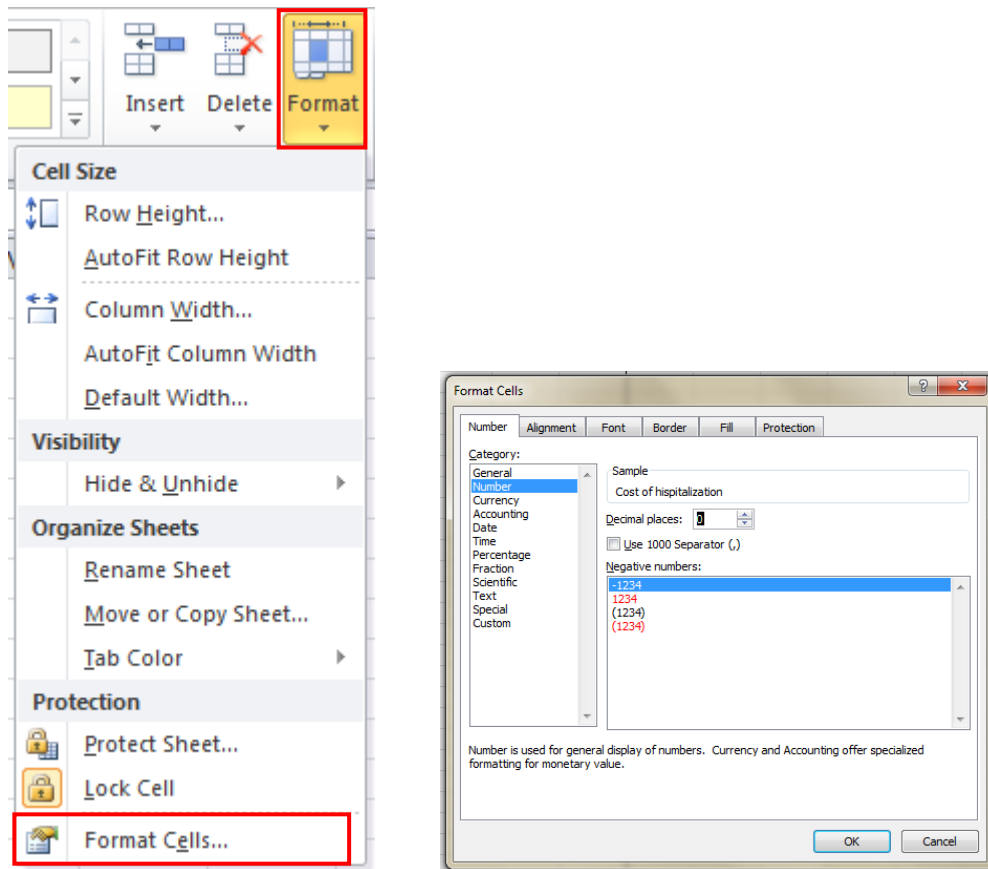- To insert a background: select the sheet and **[Page Layout – Background]**
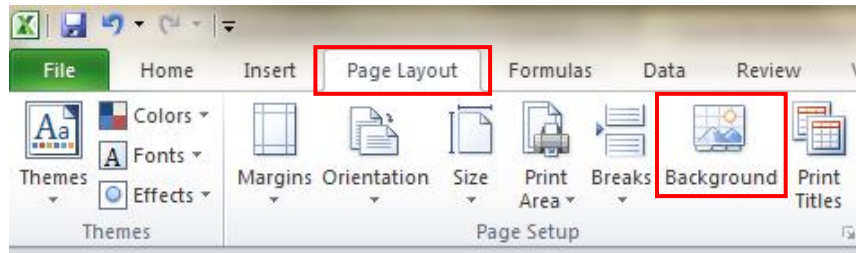
- To hide a sheet: select the sheet and **[Home – Cells – Format – Visibility – Hide&Unhide – Hide Sheet]**
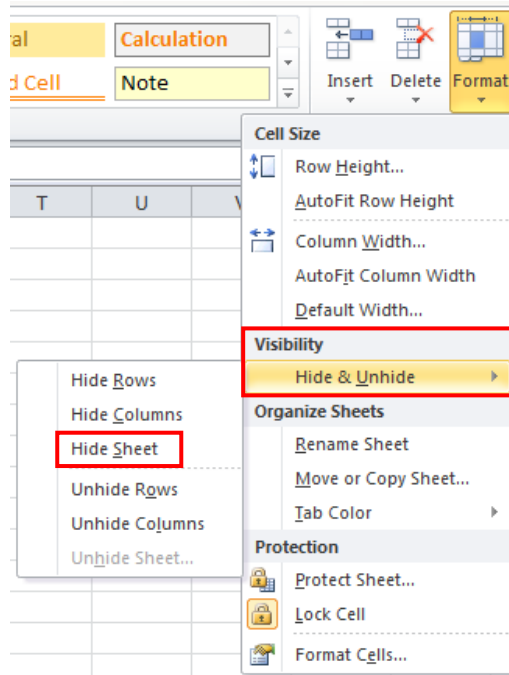
- To insert the tab to a sheet: select the sheet named <u>Data</u> and **[Home – Cells – Format – Organize Sheets – Tab color]** and choose the red color.

- To formatting a cell/column/row: **[Home – Cells – Format – Format Cells…]**

Category	Example	What to do?
Number	256	Write the number in the cell **[Format Cells …Number - Number]** N.B. Aligned implicit to the right of the cell.
Date	12-Oct-06	**[Format Cells … Number - Date]** Press CTRL+; (semicolon) to enter today's date.
Time	12:00 PM	**[Format Cells … Number - Time]** To enter a time based on the 12-hour clock, type the time + a space + "a"/"p" (a = AM – ante meridian – before midday; p = PM – after midday) ; To enter the current time, press CTRL+SHIFT+: (colon).
Text	Your name	Put some text hear (e.g. your first name) **[Format Cells – Number - Text]**. N.B. The text is aligned implicit to the left of the cell.
Scientific	0.00000058	Enter a number less than 1 **[Format Cells … - Number - Scientific]**

- To delete a sheet:
    - Right click on the sheet named and choose Delete option
    - **[Home – Cells – Delete – Delete Sheet]**
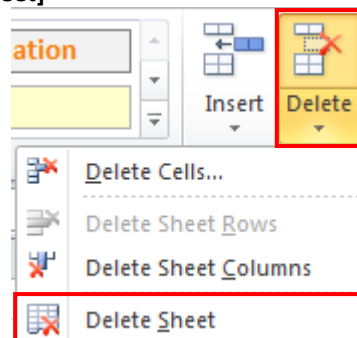
FORMULAS & FUNCTIONS

Predefined Functions

Functions are predefined formulas that perform calculations by using specific values, called arguments, in a particular order or structure. Functions can be used to perform simple or complex calculations.

Any function has the following structure: equal sign (=) followed by the function name, an opening parenthesis, the arguments for the function separated by commas, and a closing parenthesis. Arguments can be numbers, text, logical values such as TRUE or FALSE, arrays, error values such as #N/A, or cell references. The argument you designate must produce a valid value for that function. Arguments can also be constants, formulas, or other functions.

The main types of operators are:

1. Arithmetic operators: addition (+), subtraction (-), negation (-), multiplication (* - asterisk), division (/ - forward slash), percent (%), exponentiation (^ - caret).

2. Comparison operators: equal sign (=), greater than sign (>), less than sign (<), greater than or equal to sign (>=), less than or equal to sign (<=), not equal to sign (<>).

3. Text operator: ampersand (&).

4. Reference operators: colon (: - range operator - produces a reference to all cells between two references, including the two references), comma (, - union operator - combines multiple references into one reference), space (intersection operator, which produces one reference to cells common to the two references).

Formulas calculate values in a specific order. The equal sign tells Excel that the succeeding characters constitute a formula. Following the equal sign there are the elements to be calculated (the operands), which are separated by calculation operators. Excel calculates the formula from left to right, according to a specific order for each operator in the formula: reference operators (colon, single space, and comma), negation, percent, exponentiation, multiplication and division, addition and subtraction, concatenation and comparison. To change the order use round brackets to identify explain Excel which part to calculate first.

The most useful functions in medical statistics are presented in Table 1.

Click on an empty cell. To open the function wizard: [Formulas – Insert Function] and select the desired function by click OK (Figure 1). The shortcut is **Shift+F3** or click on the INSERT FUNCTION icon next to the formula bar (Figure 2).

Table 1. Statistics functions.

Function	Statistical parameter
Descriptive statistics	
=SUM(*data_range*)	Total sample
=COUNT(*data range*)	Sample size (n)
=AVERAGE(*data_range*)[a]	Arithmetic mean
=MODE(*data_range*)	Mode
=MEDIAN(*data_range*)	Median
=MAX(*data_range*)-MIN(*data_range*)	Range
=VARP(*data_range*)	Population variance
=VAR(*data_range*)	Sample variance
=STDEVP(*data_range*)	Population standard deviation
=STDEV(*data_range*)	Sample standard deviation
=CONFIDENCE(0.05,STDEV(*data_range*),COUNT(*data_range*))	Return the value that must be extracted and added from the mean to find the lower an upper bounds of confidence interval
=quartile(*data_range,x*)	Quartiles – where 1≤x≤4,Q1 = minimum, Q2 = median, Q4 = maximum
=quartile(data_range, 3)-quartile(data_range, 1)	Interquartile range
=SKEW(*data_range*)	Skewness (if data are normal distributed the skewness is 0)
=KURT(*data_range*)	Kurtosis – this function return the excess kurtosis (if data are normal distributed the KURT must return 0; a negative value indicate a platykurtotic shape while a positive value is indicative of leptokurtotic distributions)
=frequency(*data_range,bins_range*)	Counts how many of the values in the data occur within each interval
Measures of associations	
=CORREL(*Y data_range, X data_range*)	Compute the Pearson correlation coefficient between Y and X
=SLOPE(*Y data_range, X data_range*)	Compute the slope (b) for Y = a + b·X
=INTERCEPT(*Y data_range, X data_range*)	Compute the intercept (a) for Y = a + b·X
Functions that produce p-values	
=TDIST(*t,df,*2)	Two sided p-value for t-statistics (df=degrees of freedom)
=2*(1-NORMSDIST(ABS(z)))	P-value for two-tailed z-test (z value need to be previously calculated)
=CHIDIST(X^2, *df*)	Chi-Square distribution (X^2 = the Chi-Square value, df = degrees of freedom)
=CHITEST(*observed_frequency, expected_frequency*)	P-value for testing observed vs. expected frequencies.
=FDIST(*F, df1, df2*)	P-value for F distribution (df1 = degrees of freedom for numerator; df2 = degrees of freedom for numerator)
Other useful functions	
=SQRT(*number*)	Finds the square root of a number
=IF(*logical_expression, results-if-true, result-if-false*)	Returns one value if a specified condition evaluates to true, or another value if it evaluates to false
=COUNTIF(*data_range, criterion*)	Returns the absolute frequency of cases that accomplished the imposed criterion
=SUMIF(*data_range, criterion, reference*)	Sums values that accomplished a given criterion

data_range = the array where you have data in Excel (for example A2:A101)
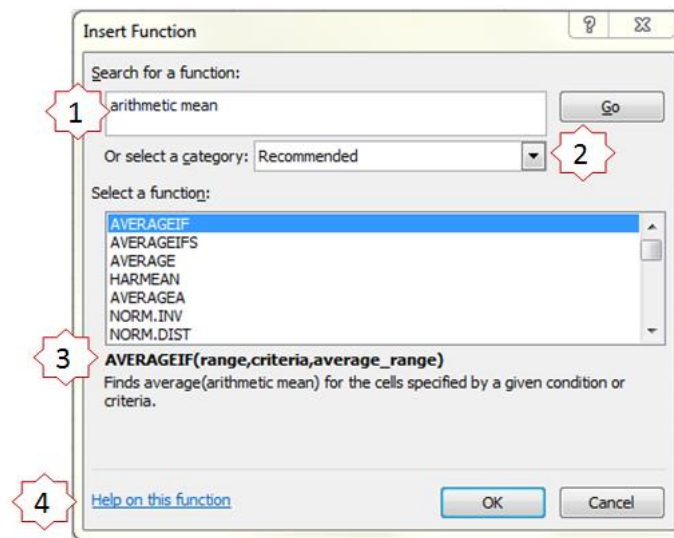
Figure 1. Insert Function window.

The wizard window allows searching a function using some keyword (1) such as "arithmetic mean" (Figure 1); the wizard will recommend us some functions related to the used keywords. If we did not have any idea about the keywords, the name of the function could be choose using the drop down list (2) and we will find all statistical function in the list of *Statistical*. If we want to use some of the functions recently used we could select the *Most Recently Used* category – this is a list of the last used 10 functions. As a function was selected, a brief description is available (3) while a full description of the function could be found on the Internet (4).

Once a function has been selected, the Function Arguments dialog box will be available (Figure 2).
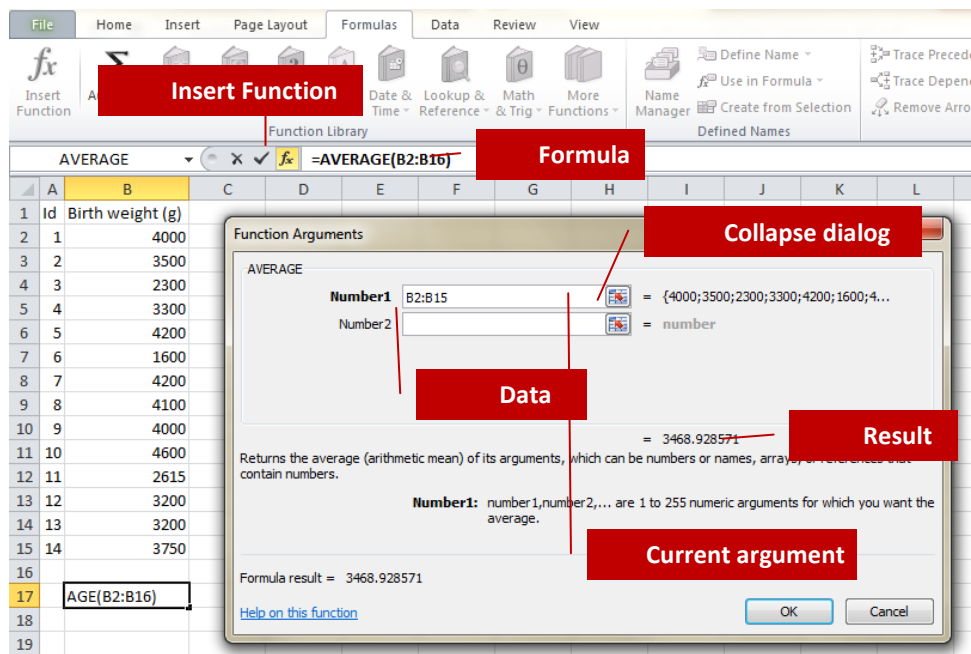

Figure 2. Function arguments dialog box.

User Defined Formulas

Three types of cell reference can be used when a formula is defined in Excel:

1. Relative reference will change column and row as formula is copied to other cells.

2. Absolute reference is the one that does not change when the formula is copied to other cells. If you refer to a cell in a certain row and column (for example $A$1) and then copy the formula elsewhere, it will still refer to exactly the same cell. Absolute reference could be also applied to range (e.g. $A$1:$D$1). An absolute address is defined by using the $ symbol

3. Mixed reference is the one that is half absolute and half relative (e.g. A$1 or $A1).

Relative reference	**A1**
Row absolute	**A$1**
Column absolute	**$A1**
Column and row absolute	**$A$1**
Relative reference & Row absolute	**C5*C$1**
Relative reference & Column absolute	**C5*$A1**
Relative reference & Column and row absolute	**C5*$A$1**

Any user-define formula must start with equal sign "=" followed by relative and/or absolute cell reference!!!

IF Function

One particular case of a logic function that could be very useful when you deal with medical data is the IF function. This function is very useful whenever different criteria are applied on a set of variables to return different answers.

The syntax of the function is as follows: = IF(*logical_test*, *value_if_true*, *value_if_false*), where *logical_test* is the test performed on a value or expression and could comprise the following operators: = (equal to), > (higher than), < (smaller than), ≥ (higher than or equal to), ≤ (smaller than or equal to), <> (not equal to). An example with two quantitative criteria is presented in Figure 3 (Q: The patient has hypertension? – is considered that a patient has hypertension if SBP ≥140 mmHg or DBP≥90 mmHg) while an example with a combination of qualitative and quantitative criteria is presented in Figure 4 (Q: The patient has a cardiovascular risk? A person has a cardiac risk if the *age of the person is greater than 30 year AND the person smoke AND the person has hypertension AND the person has glycemia higher or equal to 100 mg/dl*.). In each example an operator is used; OR will return the results if one of criterion is accomplished while AND will return a true value if all criterion ware accomplished. In the first example (Figure 3), a patient that has a SBP ≥ 140 mmHg and a DBP < 90 mmHg is consider at having hypertension (the same result is obtained if SBP < 140 mmHg and DBP ≥ 90 mmHg or SBP ≥ 140 mmHg and DBP ≥ 90 mmHg). In the second example (Figure 4), a patient will be considering of having a cardiac risk if his/her age is higher than 30 years, he/she is a smoker, he/she has hypertension and he/she has a value of glycemia ≥ 100 mg/dl. Thus, all criteria must be accomplished to be considered of having cardiac risk. If one criterion is not accomplished the person is considered not

having a cardiac risk (such as, a person with age > 30 years, non-smoker, with hypertension and with glycemia higher than or equal to 100 mg/dl).

Figure 3. IF function by example: two criteria

Figure 4. IF function by example: four criteria.