Microsoft EXCEL

Data Analysis Using Microsoft Excel

Bijay Lal Pradhan

Data analysis using Microsoft Excel

Table of Contents

Data analysis using Microsoft Excel

Introduction to Statistics

Statistics may be defined as the science of collection, organization presentation analysis and interpretation of numerical data from the logical analysis.

1.Collection of Data

It is the first step, and this is the foundation upon which the entire data set. Careful planning is essential before collecting the data. There are different methods of collection of data such as census, sampling, primary, secondary, etc., and the investigator should make use of correct method.

2.Organization of data

The mass data collected cannot conclude anything. So, the collected data should be condensed into suitable form and format. The arrangement and categorization will be done in this process.

3.Presentation of data

The collected data should be presented in a suitable, concise form for further analysis. The collected data may be presented in the form of tabular or diagrammatic or graphic form.

4.Analysis of data

The data presented should be carefully analyzed for making inference from the presented data   such as measures of central tendencies, dispersion, correlation, regression etc.,

5.Interpretation of data

The final step is drawing conclusion from the data collected. A valid conclusion must be drawn based on analysis. A high degree of skill and experience is necessary for the interpretation.

Data type

Measurement consists of counting the number of units or parts of units displayed by objects and phenomena. Measurement is a process of assigning numbers or symbols to any facts or objects or products or items according to some rule. It is a tool by which individuals are distinguished on the variables of area under study.
There are three types of data measurement

Nominal data

It is the simplest type of data, also known as categorical data. It is lowest level of measurement. It is simply a system of assigning number or the symbols to objects or events to distinguish one from another or in order to label them. The symbols or the numbers have no numerical meaning. The arithmetic operations cannot be used for these numerals. The orders of the symbol have no mathematical meaning.
For example, gender, occupation, religion are measured in nominal data. If we use data 1 for male programmer and 2 for female programmer for measuring the gender of programmer, then 1 and 2 have no numeric meaning. It is used to distinguish male and

female. In this case the number of a set of objects is not comparable to the other set. Mathematical operations such as addition, subtractions, multiplication, and division cannot be performed for the variable having nominal data. Frequency count is possible in this case so that percentage, mode can be obtained, and chi square test can be performed for test of significance for variable having nominal data.

Ordinal data

The second and the lowest level of ordered data is the ordinal data. It is the quantification of items by *ranking*. In this data, the numerals are arranged in some order but the gaps between the positions of the numerals are not made equal. It is used to rate preference of respondents. It indicates relative extent to which object possess certain characteristic. It represents qualitative values in ascending or descending order. The rank orders represent ordinal data and mostly useful in scaling the qualitative phenomena.

For example, qualification levels, preference to different statistical software, preference to different made of Laptops are measured in ordinal data. If we are going to study the computer literacy and we took the qualification level of people as primary, lower secondary, secondary, higher secondary, bachelor, master and Ph. D then we can use 1, 2, 3, 4, 5, 6 , 7 to represent primary , lower secondary , secondary , higher secondary , bachelor , master and Ph.D. respectively in ascending order. Mathematical operations such as addition, subtraction, multiplication, division cannot be performed for variable having ordinal data. Frequency count is possible and due to ranking partition values can be determined for variable having ordinal data. In this case median, mode, rank correlation also can be obtained.

Scale data

In addition to ordering the data, this data uses equidistant units to measure the difference between scores It assumes data have equal intervals. It is most powerful data of measurement. It possesses the characteristics of nominal, ordinal data. It can be expressed as relative level of measurement.

For example, Numbers on the data indicates the actual amount of property being measured. The ratio involved in the scale data possesses measure property and it facilitates comparison. Mathematical operations like addition, subtraction, multiplication and division all can be performed.

It represents actual number of variables and it is used to measure the physical dimensions. Some examples of scale data variables are disk space, amount expenses in IT department in different year, age etc. Arithmetic mean, Geometric mean, Harmonic mean, coefficient of variation along with all other measures can be obtained for variable having scale data. All statistical techniques can be applied to the variable present in scale data.

The following is the properties of categories, rank, equal interval of three data.

Level of measurement	Property		
	Categories	Ranks	Equal intervals
Nominal	Yes	No	No
Ordinal	Yes	Yes	No
Scale	Yes	Yes	Yes

Data analysis using Microsoft Excel

Variable

Variable is characteristic which can take more than one value. It may be characteristics of persons, things, events, groups, objects, feelings or any other category to be measured. It can take many values. Examples of variable are age of computer, sex of programmer, income of computer operator, job satisfaction, attitude of employee etc. If anyone is interested to find average age of CSIT first year second semester students, then the collection of information on variable age should be done. In this case value of variable age is expressed in years. To gather information on age one has to ask only one question "how old you are?" to the students of CSIT second semester students.

A Data provides the information about the individual. When the data taken by a variable is numeric then it is called numerical data. It is also called quantitative data. It can be divided into discrete and continuous data.

Discrete data is the data which can take countable values or whole numbers. For example, discrete variable, number of students in classes, number of rooms in houses, Number of vehicles in offices etc., then it takes only the whole numbers. So, the data taken by these variables are discrete and the respective variable is said to be discrete variable.

Continuous variable is the variable which can take all possible values i.e. whole numbers and fractions (real numbers). Examples of continuous variable are temperature recorded, height of students, wages of programmer, sales of Laptop etc. Here the data taken by these variables is within certain range. For example, wage of programmer is measured in Rupees as 5000 – 10000, 10000 – 20000, 20000 – 30000, etc then the data here collected are continuous in nature.

When the data taken by a variable is non-numeric then the variable is called qualitative variable. It is also called categorical variable. Examples of qualitative variable are religion of students, gender of sales man, education level of business executives etc.

Statistics and Technology

Statistics respond appropriately to the new demands of research and development in various areas of experimental sciences. With the advances in computer technology, many innovative techniques for use in applied statistics have recently been developed. In medicine, agriculture, business and government, these applications have solved important problems. Performing calculations almost at the speed of light, the computer has become one of the most useful research tools in modern times. Computers are ideally suited for data analysis concerning large research projects. Researchers are essentially concerned with huge storage of data, their faster retrieval when required and processing of data with the aid of various techniques. In all these operations, computers are of great help. Their use, apart expediting the research work, has reduced human drudgery and added to the quality of research activity.

The computers can perform many statistical calculations easily and quickly. Computation of means, standard deviations, correlation coefficients, 't' tests, analysis of variance, analysis of covariance, multiple regression, and various nonparametric analyses are just a

5

Data analysis using Microsoft Excel

few of the programs and subprograms that can be solve very easily using computer. Similarly using computer, linear programming, multivariate analysis, Monte Carlo simulation etc. are also can be done very easily. Software packages are readily available for the various simple and complicated analytical and quantitative techniques of which researchers generally make use of. The only work a researcher has to do is to feed in the data he/she gathered after loading the operating system and particular software package on the computer.

The output, or to say the result, will be ready within seconds or minutes depending upon the quantum of work. The storage facility of computers provides statistician the immense help to use the stored data whenever it required. Innumerable data can be processed and analyzed with greater ease and speed. Moreover, the results obtained are generally correct and reliable. Not only this, even the design, online data collection, pictorial graphing and report are being developed with the help of computers.

Hence, researchers should be given computer education and be trained in the line so that they can use computers for their research work.

Researchers interested in developing skills in computer data analysis, must be aware of the following steps:

(i) data organization and coding;

(ii) storing the data in the computer;

(iii) selection of appropriate statistical measures/techniques;

(iv) selection of appropriate software package;

(v) execution of the computer program.

First of all, researcher must pay attention toward data organization and coding prior to the input stage of data analysis. If data are not properly organized, the researcher may face difficulty while analyzing their meaning later on. For this purpose, the data must be coded. Categorical data need to be given a number to represent them. Once the data is coded, it is ready to be stored in the computer. Input devices may be used for the purpose. After this, the researcher must decide the appropriate statistical measure(s) he will use to analyze the data. Researcher will also have to select the appropriate program to be used. MS Excel, SPSS, SAS, STATA, R, Python etc. are the special statistical packaged program whereas Microsoft Excel can be used for simple statistical analysis. SPSS, SAS, STATA are the windows-based user-friendly program where as R, Python are the freeware where we have to write code to run the mathematical and statistical calculation. Among this all MS Excel is easily accessible and user-friendly software which mostly we are familiar with it. So, this data analysis workshop program is designed with MS Excel software.

Microsoft Excel

Microsoft Excel helps you to organize, attractively present and analyze data.

Data analysis using Microsoft Excel

A spreadsheet is the computer equivalent of a paper ledger sheet. It consists of a grid made from columns and rows. It is an environment that can make number manipulation easy and somewhat painless.

The statistics that goes on behind the scenes on the paper ledger can be overwhelming. If you change the any amount, you will have to start the calculation all over again (from scratch). The nice thing about using a computer and spreadsheet is that you can experiment with numbers without having to RE-DO all the calculations.

NO erasers! NO new formulas! NO calculators!

Excel has many applications:
- Sorting and organizing data
- Creating visual representations of the data
- Addition, Subtraction, Division, Multiplication, percentage of Cells
- Statistical analysis
  - Average (Mean)
  - Median
  - Quartile
  - Standard deviation
  - Estimation
  - Test of hypothesis (parametric and non-parametric)
  - Correlation and regressions…
- Matrix Operations
  - Addition/Subtraction
  - Multiplying
  - Inverse
- Optimization
  - Linear programming
  - Transportation
  - Assignments
- And many more…

Data analysis using Microsoft Excel

Introduction to Excel

Microsoft Excel helps you to organize, attractively present and analyze data.

A spreadsheet is the computer equivalent of a paper ledger sheet. It consists of a grid made from columns and rows. It is an environment that can make number manipulation easy and somewhat painless.

The statistics that goes on behind the scenes on the paper ledger can be overwhelming. If you change the any amount, you will have to start the calculation all over again (from scratch). The nice thing about using a computer and spreadsheet is that you can experiment with numbers without having to RE-DO all the calculations.

NO erasers! NO new formulas! NO calculators!

Excel has many applications:
- Sorting and organizing data
- Creating visual representations of the data
- Addition, Subtraction, Division, Multiplication, percentage of Cells
- Statistical analysis
    - Average (Mean)
    - Median
    - Quartile
    - Standard deviation
    - Estimation
    - Test of hypothesis (parametric and non-parametric)
    - Correlation and regressions…
- Matrix Operations
    - Addition/Subtraction
    - Multiplying
    - Inverse
- Optimization
    - Linear programming
    - Transportation
    - Assignments
- And many more…

Data analysis using Microsoft Excel

Opening Excel and Inputting Data

Opening a Document

New Document: Start → Programs → Microsoft Office → Excel

Saved Document: → File → Open then select the document you would
like to open (then click open)

Entering in Data

What you see on the screen is a new Excel Document. Each rectangle is a
Cell which is arranged in rows and columns each having a name. The first
cell (upper left-hand corner of the document) is A1, moving one cell to the
right is B1, and moving down one from A1 is A2.

To type in data (either number, words, formulas) simply click on a cell
and begin typing. When finished you can either click on a new cell to
enter more data. Or, to move one right press Tab and pressing Enter brings
you back to the first column entered but one row down. The Arrow Keys
will also move you from one cell to the next.

If you can not see all your data on the screen simply select the Row or
Column by clicking on the letter (A, B, C…) or number (1, 2, 3…) in gray
and then go to
    Format → Row → Auto Fit Section
                            Or
    Format → Column → Auto Fit Section

Sorting and Organizing Data

You may sort data in ascending, descending or alphabetical order.
Highlight the data.

Go to:
        Data→*Sort*
A window will open with several options for sorting the data
In the *Sort by* dialogue box, use the drop down menu to highlight the variable you
want the data sorted by (2003 ELA Scaled Score for example). Click on
Ascending or *Descending*.

The data will sort by that category. In other words, if you sorted by *Ascending*
2003 ELA Scaled Scores, the spreadsheet will now start with the lowest score and
progress to the highest score, and the other columns will correspond to the
appropriate score (i.e., all of the columns will not be sorted *ascending*; the
integrity of the data will be maintained)

9

Data analysis using Microsoft Excel

Graphing Data

You may create different types of graphs to visually represent your data. Perhaps the most pertinent type of graph you will be creating for this project is a *scatter-plot.* To make a scatter-plot

Scatter Plot

You may create different types of graphs to visually represent your data. Perhaps the most pertinent type of graph you will be creating for this project is a *scatter-plot.* To make a scatter-plot

Highlight the data you need for the scatter-plot (don't worry about highlighting extra data)

Go to:
> *Insert→Chart*

1. A window will appear asking what type of chart you would like to create Select *XY (Scatter)* from the menu on the left

2. When you select the *Scatter-plot* option the right side of the window will change to show *Chart Sub-Types*. Select the first one (where Excel will not connect the points) and click *Next >*

3. Click on the tab *Series* (top of window)

4. To select the values to be used on the x-axis, click on the blue and red box to the right of the *X Values* blank. You will now see your original spreadsheet. Click and drag over the appropriate values. Then press *Enter* or click on the box on the right.

5. To select the values for the y-axis, follow the same procedure as above except in the *Y Values* box.

6. You should see a sample of your graph in the window (make sure it makes sense).

    Click *Next >*

7. In the proceeding *Chart Options* box you may do a variety of things such as assign labels to the axes (under the *Titles* tab) or create a legend under the *Legend* tab

Data analysis using Microsoft Excel

8. Click *Next >* when you are done customizing the chart.

9. The last dialogue box asks if you would like the chart placed on a separate sheet (literally, a separate page) or as an object on the same page. Select one of the two options and click *Finish*.

Adding Trendlines

1. Click on your existing Chart

2. Select *Add Trendline*

   Chart→Add Trendline

3. Select an appropriate option from the *Trend/Regression Type* dialogue box (linear, exponential, polynomial, etc.)

4. Click on the *Options* tab

5. Check the boxes next to *Display equation on chart* and *Display r - squared value on chart* if not already checked (this will show the equation for the generated regression line)

6. The trendline should now be displayed on the graph (you may double-click it to change properties like color)

7. The equation for the line should also be displayed on the chart. Note that it may need to be dragged to an area of the chart where it is visible (it may be buried behind the actual data points)

Formulas for +,-,*,/ of Cells

Adding/Subtracting Two Cells

Click on a empty cell (where you want the output) type in
=Name of 1st Cell + Name of 2nd Cell
Example

	A	B	C	
1	8	9	=A1+B2	17
2	10	11		
3	7	6		

To add Column A and B for the other rows simply click on C1 so there is a black box around the cell then bring your curser to the lower right-hand

Data analysis using Microsoft Excel

corner when your curser turns in to a + sign, click, hold and drag the curser so it highlights C2 and C3 and release.

Adding/Subtracting Multiple Cells

Type in your two matrices

Example

	A	B	C	D	E	F	G
1	1	2	3		1	3	5
2	4	5	6		2	4	6
3	7	8	9		7	9	11
4							
5							
6							
7							

Click on an empty cell and type in the addition equation for A1 and E1

Example

	A	B	C	D	E	F	G
1	1	2	3		1	3	5
2	4	5	6		2	4	6
3	7	8	9		7	9	11
4							
5	=A1+E1						
6							
7							

Highlight A5 and click on the bottom right-hand corner and drag it to the right two and down two. When you let go you will see:

	A	B	C	D	E	F	G
1	1	2	3		1	3	5
2	4	5	6		2	4	6
3	7	8	9		7	9	11
4							
5	2	5	8				
6	6	9	12				
7	14	17	20				

Data analysis using Microsoft Excel

Dividing/Multiplying Individual Cells

Is done the same as adding/subtracting cells

Basic Statistical Analysis

Excel can compute a variety of basic statistics about a set of data.

Average (Mean) of a Set of Data

1. Click the cell where you would like the mean to be displayed

2. In the formula bar at the top of the document, type:
    =AVERAGE(starting cell:ending cell),

    Example:

    =AVERAGE(D2:D357)

    (this would compute the average of the set of data starting
    in cell D2 and ending in D357)
    Or
    You may type =AVERAGE(then highlight the set of cells
    you want the average of)

Median

1. Click the cell where you would like the median to be displayed

2. In the formula bar at the top of the document, type:
    =MEDIAN(starting cell: ending cell),

    Example:
    =MEDIAN(D2:D357)
    Or
    You may type =MEDIAN(then highlight the set of cells
    you want the average of)

Note: If there is an even number of numbers in the set, then MEDIAN
calculates the average of the two numbers in the middle

If an array or reference argument contains text, logical values or
empty cells, those values are ignored; however, cells with the value
zero are included

13

Data analysis using Microsoft Excel

Quartile

1. Click the cell where you would like the quartile to be displayed
2. In the formula bar at the top of the document type:
    =QUARTILE(starting cell: ending cell, quart)

If quart equals	QUARTILE returns
0	Minimum value
1	First quartile (25th percentile)
2	Median value (50th percentile)
3	Third quartile (75th percentile)
4	Maximum value

Standard Deviation

1. Click the cell where you would like the standard deviation to be displayed
2. In the formula bar at the top of the document type:
    =STDEV(starting cell: ending cell),

    Example:
    =STDEV(D2:D357)
    Or
    You may type =STDEV then simply highlight the relevant cells

NOTE: Excel uses the following formula to compute STDEV:

$$\sqrt{\frac{\sum(x-\bar{x})^2}{(n-1)}}$$ (using the "unbiased" or n-1 method)

If you want the standard error calculation to not be based on (n-1) and simply have n in the denominator, use the STDEVP function (input syntax same as STDEV)

t-Test

1. Click the cell where you would like the t-test to be displayed
2. In the formula bar at the top of the document type:
    = TTEST(array1,array2,tails,type)

14

Data analysis using Microsoft Excel

Where:

Array 1 is the first data set (Selected by highlighting)
Array 2 is the second data set (Selected by highlighting)
Tails specifies the number of distribution tails (1 or 2)
Type is the kind of t-Test to perform

If type equals	This test is performed
1	Paired
2	Two-sample equal variance (homoscedastic)
3	Two-sample unequal variance (heteroscedastic)

Covariance

Covariance can be calculated easily and on a large scale using the capabilities of an Excel spreadsheet. First, create a chart to compare the different stocks:

Covariance Matrix

	ADM	IBM	KO	MAY	XOM
ADM					
IBM			☺		
KO					
MAY					
XOM					

For example, the ☺ will represent the calculated covariance of KO and IBM stocks, or the degree to which their rates of return move together over the investigated period.

Next, the covariance must be calculated for each cell in the table.
As an example, for cell ☺, KO vs. IBM, follow the steps below.

1. Select the cell
2. Click the function (*f*x) button
3. Select the function COVAR, click OK
4. A box will appear into which you must enter two arrays
5. For Array 1, click the button at the end of the text field, then choose the entire column of daily returns for KO, rows 5-255. Press Enter.

Data analysis using Microsoft Excel

6. For Array 2, use the same process to select the
   entire column of IBM daily returns. Press Enter
7. Click OK

Repeat this process for each cell. The covariance function is
communitive, and therefore it does not matter which order the arrays are
selected in. Therefore, IBM vs ADM will have the same covariance value
as ADM vs IBM, and will not need to be calculated twice.

Following this procedure, the following covariance matrix was developed:

Covariance Matrix

	ADM	IBM	KO	MAY	XOM
ADM	0.00054433	0.00009065	0.00015905	0.00016825	0.00001369
IBM	0.00009065	0.00089061	0.00000014	0.00010558	0.00000771
KO	0.00015905	0.00000014	0.00071840	0.00017362	0.00010556
MAY	0.00016825	0.00010558	0.00017362	0.00081997	0.00005726
XOM	0.00001369	0.00000771	0.00010556	0.00005726	0.00039556

Matrix (Array) Operations

Matrix operations such as multiplying, finding the inverse, and finding the
determinant have to be down using arrays.

Multiplying Matrices

Remarks

- The number of columns in array1 must be the same as the
  number of rows in array2, and both arrays must contain only
  numbers.

- Array1 and array2 can be given as cell ranges, array constants,
  or references.

- If any cells are empty or contain text, or if the number of
  columns in array1 is different from the number of rows in
  array2, MMULT returns the #VALUE! error value.

- The matrix product array a of two arrays b and c is:

Data analysis using Microsoft Excel

$$a_{ij} = \sum_{k=1}^{n} b_{ik} c_{kj}$$

where i is the row number, and j is the column number.

- Formulas that return arrays must be entered as array formulas.

Steps

1. Enter in the Matrices
2. Click on an empty box and input the equation
   =MMULT(array1,array2) and press enter
3. Click and highlight the section of cells the same size as the desired matrix.
4. Press F2
5. Press Ctrl+Shift+Enter

Example

1. Enter in desired matrix

	A	B	C	D	E	F	G
1	1	2	3		1	3	5
2	4	5	6		2	4	6
3	7	8	9		7	9	11
4							
5							
6							
7							

2. Click on A5 and type in
   =MMULT(A1:C3,E1:G3) press enter

3. Ignore the number 40. Click on A5 and drag curser to C7 and release.
4. Press F2
5. Press Ctrl+Shift+Enter at the same time and release.

	A	B	C	D	E	F	G
1	1	2	3		1	3	5
2	4	5	6		2	4	6
3	7	8	9		7	9	11
4							
5	40	58	76				
6	51	75	99				

Data analysis using Microsoft Excel

7	84	126	168				

Inverse Matrix

Remarks

- The size of the array must not exceed 52 columns by 52 rows. If it does, the function returns a #VALUE! error.

- Array can be given as a cell range, such as A1:C3; as an array constant, such as {1,2,3;4,5,6;7,8,9}; or as a name for either of these.

- If any cells in array are empty or contain text, MINVERSE returns the #VALUE! error value.

- MINVERSE also returns the #VALUE! error value if array does not have an equal number of rows and columns.

- Formulas that return arrays must be entered as array formulas.
- Inverse matrices, like determinants, are generally used for solving systems of mathematical equations involving several variables. The product of a matrix and its inverse is the identity matrix— the square array in which the diagonal values equal 1, and all other values equal 0.
- As an example of how a two-row, two-column matrix is calculated, suppose that the range A1:B2 contains the letters a, b, c, and d that represent any four numbers. The following table shows the inverse of the matrix A1:B2.

Steps

1. Enter Matrix
2. Click on empty box and input the equation
        =MINVERSE(array1)
3. Click and highlight the section of cells the same size as the desired matrix.
4. Press F2
5. Press Ctrl+Shift+Enter

Example

1. Enter Matrix
2.

	A	B	C	D
1	1	2	1	
2	3	4	-1	

3	0	2	0	
4				
5				
6				
7				

3. Click on A5 and input the equation
    =MINVERSE(A1:C3) press enter
4. Ignore the number 0.25. Click on A5 and drag curser to C7 and release.
5. Press F2
6. Press Ctrl+Shift+Enter at the same time and release.

	A	B	C	D
1	1	2	1	
2	3	4	-1	
3	0	2	0	
4				
5	0.25	0.25	-0.75	
6	0	0	0.5	
7	0.75	-0.25	-0.25	

Determinant Matrix

Remarks

- Array can be given as a cell range, for example, A1:C3; as an array constant, such as {1,2,3;4,5,6;7,8,9}; or as a name to either of these.

- If any cells in array are empty or contain text, MDETERM returns the #VALUE! error value.

- MDETERM also returns #VALUE! if array does not have an equal number of rows and columns.

- The matrix determinant is a number derived from the values in array. For a three-row, three-column array, A1:C3, the determinant is defined as:

    MDETERM(A1:C3) equals
    A1*(B2*C3-B3*C2) + A2*(B3*C1-B1*C3) + A3*(B1*C2-B2*C1)

- Matrix determinants are generally used for solving systems of mathematical equations that involve several variables.

Data analysis using Microsoft Excel

- MDETERM is calculated with an accuracy of approximately 16 digits, which may lead to a small numeric error when the calculation is not complete. For example, the determinant of a singular matrix may differ from zero by 1E-16.

Steps

1. Enter Matrix
2. Click on empty box and input the equation
   =MDETERM(array1)
3. Click and highlight the section of cells the same size as the desired matrix.
4. Press F2
5. Press Ctrl+Shift+Enter

Example

1. Enter Matrix
2.

	A	B	C	D
1	7	8	1	
2	3	6	2	
3	9	5	4	
4				
5				
6				
7				

3. Click on A5 and input the equation
   =MDETERM(A1:C3) press enter
4. 107 is the output which is the determinant of the above matrix

Pivot Tables

Pivot Tables can be used to analyze tables of data.   Note that all examples in this section use the file Gr7 MAT7 Gr10 MCAS 2003 Que.xls supplied with the MCAS project.

Using a Pivot Table to Tally Columns:

Column I lists the performance level for the MCAS math scores (i.e. F, NI, P, A). If you want to find out how many students were in each performance level category, use the following steps to build the pivot table:

1. Select Column I

Data analysis using Microsoft Excel

2. From the Data pull down menu, select Pivot Table
3. Click Next through steps 1 and 2.
4. On step 3 choose Layout.
5. Drag the GR 10 MAT button to the Row area and again to the Data area. In the data area, it should say "Count of GR 10 MAT"
6. Click OK and Finish. In a new worksheet, you will have a table that lists the scores down the left, and for each score, the number of students that received that score.

Tallying Data by Two Categories

If you want a table of the students broken down by both performance level category and by gender, follow these steps:

1. Select Columns C through I
2. From the Data pull down menu, select Pivot Table, and click Next through steps 1 and 2.
3. Choose Layout on step 3.
4. Drag GR 10 MAT (the last one) to the Row and Data areas as before. But this time, also drag Sex to the Column area.
5. Click OK and Finish. In a new worksheet, you have a two dimensional table that tells how the students performed by gender.

Making a Graph Directly from the Pivot Table

1. With the Pivot Table from the previous example still displayed, click on the Chart Wizard button from the Pivot Table toolbar. This button looks like a bar chart.
2. A stacked bar chart will appear. If you'd like a different chart type, click Chart Type from the Chart Toolbar and select the type you prefer.

Removing Pivot Table Rows or Columns

Sometimes entries show up in the pivot table that you do not want to be displayed. For example, if you only want to present the information for students that passed the MCAS, you can hide the Failing row. In this example, the Blank row will also be hidden.

1. Starting from a pivot table generated in either of the previous examples, click on the down arrow button in the gray area labeled GR 10 MATH 2003 Performance Level.
2. A pop-up will appear showing all of the categories for this field. Uncheck the boxes that you do not want to include (e.g. Failing and blank).   The subtotals and totals will be updated with these counts removed.

Data analysis using Microsoft Excel

Using Pivot Tables to Display Calculated Data

The pivot table can perform calculations on the data, rather than just counting how many records fall into a category. For example, you may want to find the mean score broken down by gender and performance level.

1. Select Columns C through I.
2. From the Data pull down menu, select Pivot Table, and click Next through steps 1 and 2.
3. Choose Layout on step 3.
4. Drag GR 10 MAT (the last one) to the Row area. Drag Sex to the Column area.
5. Draw the first GR 10 MAT button to the Data area. Double click on the button in the data area that says "Count of GR 10 MAT."   You will be presented with a pop-up that lists other options for the data. Choose Average.
6. Click OK and Finish. In a new worksheet, you have a two dimensional table that tells the mean MCAS score by gender and performance level.