Data Analysis techniques allow professionals like engineers, social scientists and economists to extract meaningful information from a (typically) vast amount of data.

Excel is widely available, and provides useful functions/features for data analysis.

Some of these functions/features include:

- Histogram
- Descriptive Statistics
- Correlation
- Fitting Equations to Data
- Matrix Calculations
- Financial functions
- Look-up functions
- What-if analysis
- Testing/Debugging/Auditing Spreadsheets
- Iterative Solutions: Solving Equations
- Iterative Solutions: Optimization Problems
- Some Advanced Functions/Features
- Matrix Calculations
- Using Charts

See Topics 3, 4, 5 and 6 in the textbook.

Analysis Toolpak in Excel

- Analysis Toolpak is an add-in package for Excel. You need to activate it using Excel Options/Add-Ins. Make sure that both the Analysis ToolPak and Solver Add-in are in the "Active Application" list, otherwise click Manage Add-ins.

- For more info see "Enable or disable add-ins in Office programs" in "Excel Help".

The Analysis Toolpak provides tools to help users to quickly and easily carry out reasonably complex data analysis tasks.

- It provides support for many popular statistical and engineering data analysis tasks.

- Note: Excel’s statistics algorithms are not sufficiently robust for use by professional statisticians (but they’re OK for normal purposes)

Descriptive Statistics

- Excel provides many predefined statistical functions to calculate useful information like: mean, max, min, median, standard deviation, etc.

- Often a user wants to calculate most commonly used statistical functions (like the functions described above) for a given data set in order to get some description of the data set.

- Excel provides a tool called "Descriptive Statistics" that calculates such commonly used statistical functions for a given data set and produces a useful report.

- Alternatively, a user can apply each function individually and calculate the required information. Obviously this may prove very time consuming and tedious!
**Correlation**

- **Correlation** is a statistical measure referring to the strength of linear relationship between two or more variables.
- It can vary from $-1$ (perfect negative correlation) through 0 (no correlation) to $+1$ (perfect positive correlation).
- Excel provides a tool called "Correlation" that can calculate correlations between two or more variables.
- Alternatively, a user can plot a chart for two or more variables and try to visually identify possible correlations between variables.
- *Demonstration ……*

---

**Fitting Equations to Data**

- The method of least squares tries to fit a curve (including a straight line) through an aggregate of data. It tries to find the relationship among variables and express the relationship as an equation.
- Two approaches in Excel:
  - Using "Trendlines" on charts
  - Using the "Regression" tool
- If we do not need a lot of information about the regression, we can use Trendlines on charts. They provide the basic equation and $R^2$ value, and show the regression line superimposed on the data values.
- The $R^2$ value (varies from 0 to 1) is an indicator of how well the model fits the data. $R^2 = 1$ indicates that the regression line (curve) is a perfect fit to the data.
- *See Topic 4 in the textbook.*
- *Demonstration ……*

---

**Fitting Equations to Data**

- Calibrating an old device:
  - The Analysis Toolpak offers the tool called "Regression" to compute regression analysis.
  - The "Regression" tool provides lots of statistical values! and most likely you may not be able to understand most of them ;-) that's OK for now! In future when you learn more about regression analysis, you will be able to use these statistical values more effectively in your data analysis.
  - *Demonstration ……*
Fitting Equations to Data

Functions in Excel:

- **FORECAST** Returns a value along a linear trend
- **GROWTH** Returns values along an exponential trend
- **TREND** Returns values along a linear trend
- Etc. Etc. ....

Solving Single Equations

- **Graphical Solutions:**
  - One of the ways to find a possible root(s) of a given equation is to assign different values to a variable (say \( x \)), calculate the corresponding function values (say \( f(x) \)), and plot a graph for \( f(x) \) vs \( x \). Now try to see where the chart line intersects with the \( x \)-axis !
  - This approach is useful when we can guess a finite interval(s) within which to search for possible root values

- **See Topic 6 in the textbook.**

Solving Single Equations

\[ f(x) = 4x^3 + 12x^2 - 64x + 16 \]

Solving Single Equations

- **Using Goal Seek** (Tools - Goal Seek)
  - Goal Seek will try to find input values of an equation when the results are known. It starts with a provided initial guess for an input value, and uses iterative refinement to find a solution
  - If an initial guess is not close enough, Goal Seek may not be able to find a solution
  - Goal Seek tries a fixed number of iterations (guesses) and stops after that, even if the equation is not solved.
  - "Solver" tool is more powerful than "Goal Seek" tool.
  - Demonstration ......

Finding Optimum Solutions

- Excel's Solver (Tools - Solver) tool allows users to solve constrained nonlinear optimization problems.
- First make sure that Solver is installed. If not, select Tools - Add-Ins and install it.
- Often we want to maximize or minimize a value of a variable or a function. Such a function is called an objective function. For example, we may want to maximize our profit OR say minimize cost.
- In case we cannot find the best solution, often we are also interested in "optimal" solutions.
- In Solver, we provide the following:
  - Optimization function
  - Set of input variables (that could be changed)
  - Set of constraints to be satisfied
- See Topic 6 in the textbook.
Finding Optimum Solutions

• Demonstration...

(The above example is from the book: Microsoft Office System Inside Out, 2003 Edition by Michael J. Young and Michael Halvorson.)

Finding Optimum Solutions

• Options for Solver:

Solving Optimization Problems

• Find the values of $x$ and $y$ that maximize

$$f(x, y) = \sin(x) \cdot \cos(y)$$

Matrix Operations

• Excel provides all the standard matrix operations like: add, multiply, inverse, transpose, determinant, etc

• Excel uses the term arrays to refer to a collection of values that should be kept together.

• In Excel, both “vector” and “matrix” are considered as arrays.

• We can provide a name to an array using Insert/Name/Define. This may make your spreadsheet easier to read/understand. However, note that we do not have to name arrays in order to use matrix functions.

• Need to use Ctrl-Shift-Enter to enter arrays

• Basic Matrix Operations: add, scalar multiplication,

Matrix Operations

• Matrix (Linear Algebra) Functions

  – MMULT(M1, M2)
    Returns the matrix product of two matrices M1 and M2.

  – MINVERSE(M1)
    Returns the inverse matrix for matrix M1.

  – TRANSPOSE(M)
    Transposes matrix M

  – MDETERM(M)
    Returns the determinant of matrix M

• See Topic 3 in the textbook.

• Also see Microsoft Excel Help for more on the above Matrix operations

• Demonstration...

Solving Simultaneous Equations

• Using Matrix operations:
Solving Equations in Excel Using Solver

- In Solver we can specify
  - constraints on the independent variables.
  - For example, X >= 0
  - convergence criteria
  - etc...
- Simultaneous Equations
  - Solver could be used to solve linear as well as nonlinear simultaneous equations.
  - One way to do this is to force the following function (\( Y \)) to zero:
    \[ Y = f_1^2 + f_2^2 + \ldots + f_n^2 \]
    where \( f_1, f_2, \ldots, f_n \) represent the equations

Financial Functions

Excel offers a wide range of financial functions. For example:

- \( PMT(i, n, P) \)
  - Returns the periodic (e.g., monthly) payment for an n-payment loan of \( P \) dollars at interest rate \( i \).
- \( FV(i, n, A) \)
  - Returns the future value of a series of \( n \) payments of \( A \) dollars each at interest rate \( i \).
- \( PV(i, n, A) \)
  - Returns the present value of a series of \( n \) payments of \( A \) dollars each at interest rate \( i \).
  - Etc.. Etc..
- See pages 90-91 in the textbook. Also see Microsoft Excel Help for more on the above and other Financial functions

Lookup Functions

- \( VLOOKUP \) (lookup_value, table_array, column_index_num, range_lookup)
  - Searches for a value in the leftmost column of a table, and then returns a value in the same row from a column (whose index number you provide) in the table.
- See also \( HLOOKUP \)
- Demonstration ....
- See pages 84-86 in the textbook, and also Microsoft Excel Help for more on the above and other lookup functions

More on what-if analysis

- Goal Seek: we discussed this already
- Solver: we discussed this already
- Data tables:
  - A data table is a range of cells that shows how changing certain values in your formulas affects the results of the formulas.
  - Data tables provide a short cut for calculating multiple versions in one operation and a way to view and compare the results of all of the different variations together on your worksheet.
  - See Microsoft Excel Help for more details on how to create and use data tables.

More on what-if analysis

- Scenario:
  - A scenario contains a set of values that Microsoft Excel saves and can substitute automatically in your worksheet.
  - You can use scenarios to forecast the outcome of a worksheet model.
  - You can create and save different groups of values on a worksheet and then switch to any of these new scenarios to view different results.
  - See Microsoft Excel Help for more details on how to create and use scenarios.
  - Data tab – What-If Analysis – Scenario Manager

Testing/Debugging/Auditing Spreadsheets

- In spreadsheets, computational steps are often scattered across a wide range of cells and even across many worksheets.
- It is very easy to make mistakes while developing a solution in Excel. Therefore, it is essential to thoroughly test your solutions.

$24-million spreadsheet “clerical error”

June 03, 2003 TORONTO (Reuters) - TransAlta Corp. said on Tuesday it will take a $24 million charge to earnings after a bidding snafu landed it more U.S. power transmission hedging contracts than it bargained for, at higher prices than it wanted to pay.

[...] the company’s computer spreadsheet contained mismatched bids for the contracts, it said. "It was literally a cut-and-paste error," an Excel spreadsheet that we did not detect when we did our final sorting and ranking bids prior to submission," TransAlta chief executive Steve Snyder said in a conference call. "I am clearly disappointed over this event. The important thing is to learn from it, which we've done."
Testing/Debugging/Auditing Spreadsheets

- Spreadsheet mistakes – more news stories available on the class web page!
- Also see the class web page for selected papers and articles on how to design and develop spreadsheet solutions.
- From "How do you know your spreadsheet is right?: Principles, Techniques and Practice of Spreadsheet Style" by Philip L. Bewig, July 28, 2005 (available at http://www.eusprig.org/hdykysir.pdf):
  - A missing minus sign caused Fidelity’s Magellan Fund to overstate projected earnings by $2.6 billion (yes, billion) and miss a promised dividend.
  - Falsely-linked spreadsheets permitted fraud totaling $700 million at the Allied Irish Bank.
  - Voting officials reported spreadsheet irregularities in New Mexico and South Africa.

Testing/Debugging/Auditing Spreadsheets

- So…… Test, Test, Test. …….. your solutions ……!!
- Test your solution for obvious mistakes, like:
  - invalid or missing data in a cell
  - formula is replaced a constant in a cell
  - incorrect cell references in a formula (due to incorrect usage of relative/mixed/absolute addressing while copy and past actions)
  - Etc .. Etc ..

Excel Provides some debugging tools:

- Debugging by Using Cell Selection
  - Select Home/Editing/Find and select Go To Special
  - You can now locate cells with blanks, formulas, constants, etc.
  - You can locate cells that are “different” in a given range
  - You can also locate Precedents and Dependents of a given cell

- Test, Test, Test…!!