CSE 123
Introduction to Computing

Lecture 4

Linear Regression/Curve Fitting and Statistics Functions/Iteration, Analysis Toolpak, Goal Seek, Pivot Tables, External Data Sources

SPRING 2012

Assist. Prof. A. Evren Tugtas

Marmara Üniversitesi

CSE 123
Introduction to Computing

Linear Regression/Curve Fitting and Analysis Toolpak

SPRING 2012

Assist. Prof. A. Evren Tugtas

Marmara Üniversitesi

Linear Regression

- "Linear" term in linear regression means that the equation used to fit the data points must be linear in the coefficients.
- "Regression" part involves fitting the coefficients that produce the "best fit" of the regression model
- Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data.

Linear Regression

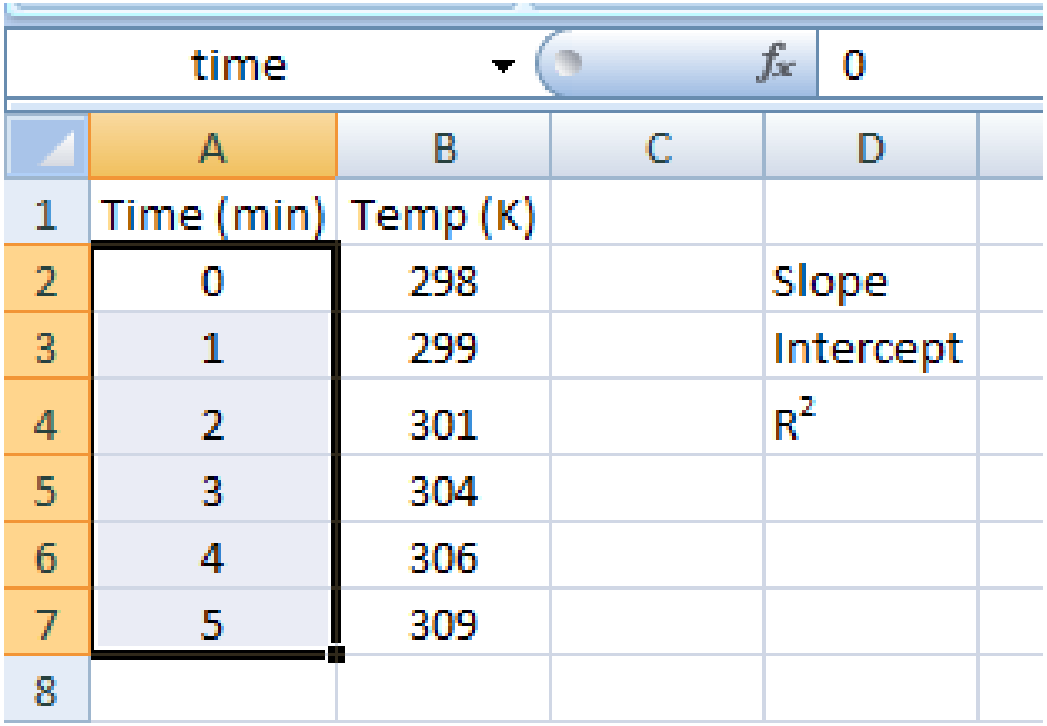
- Built-in regression functions
- Adding trendlines to graphs
- Using regression analysis package

Excel's Built-in Regression Functions

- SLOPE () → calculates the slope
- INTERCEPT () → finds the intercept of the best fit straight line
- RSQ() → Calculates the coefficient of determination or R^2

Regression

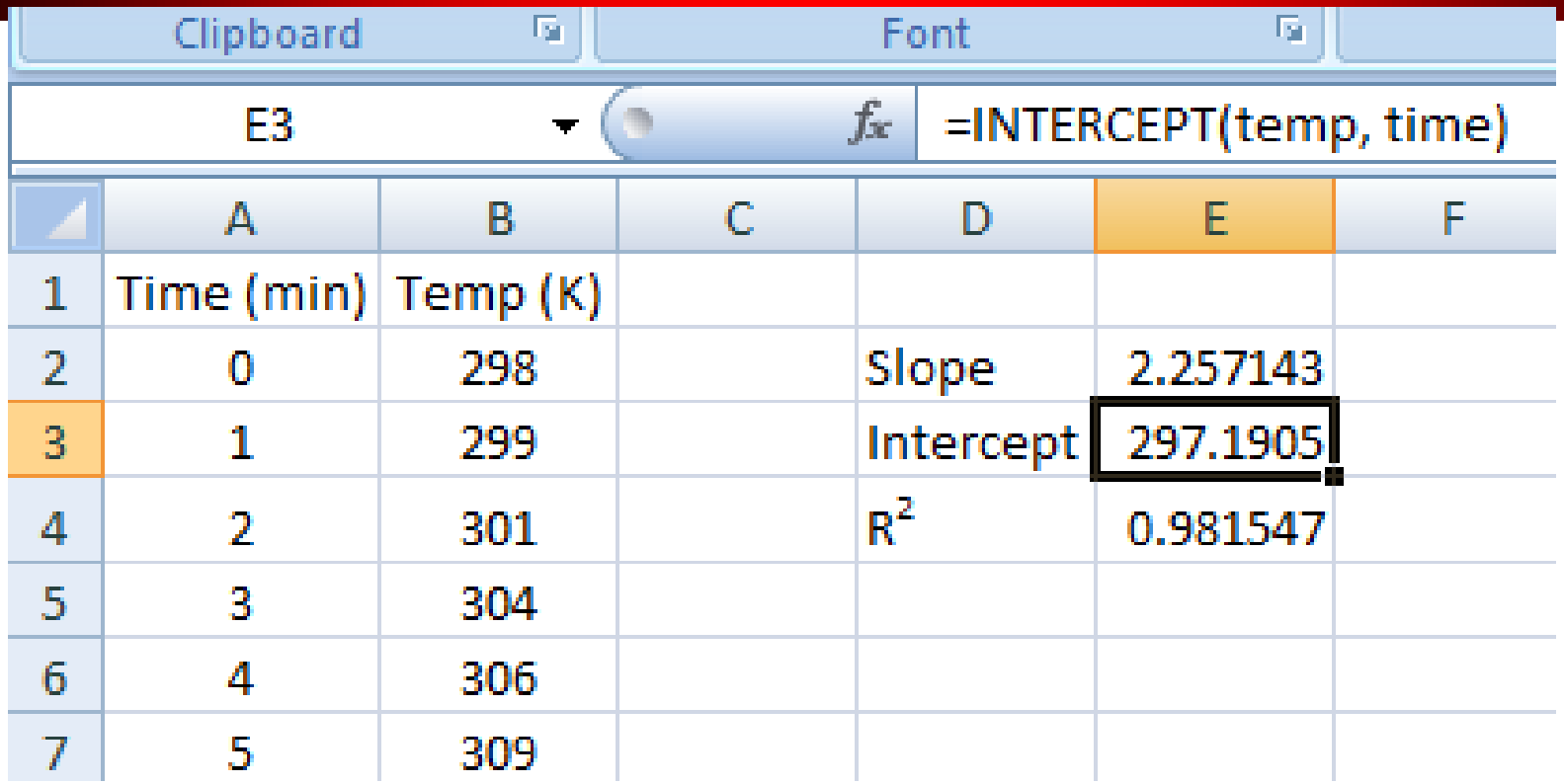
- Give names to cell ranges



time | fx | 0

	A	B	C	D
1	Time (min)	Temp (K)		
2	0	298		Slope
3	1	299		Intercept
4	2	301		R^2
5	3	304		
6	4	306		
7	5	309		
8				

Regression



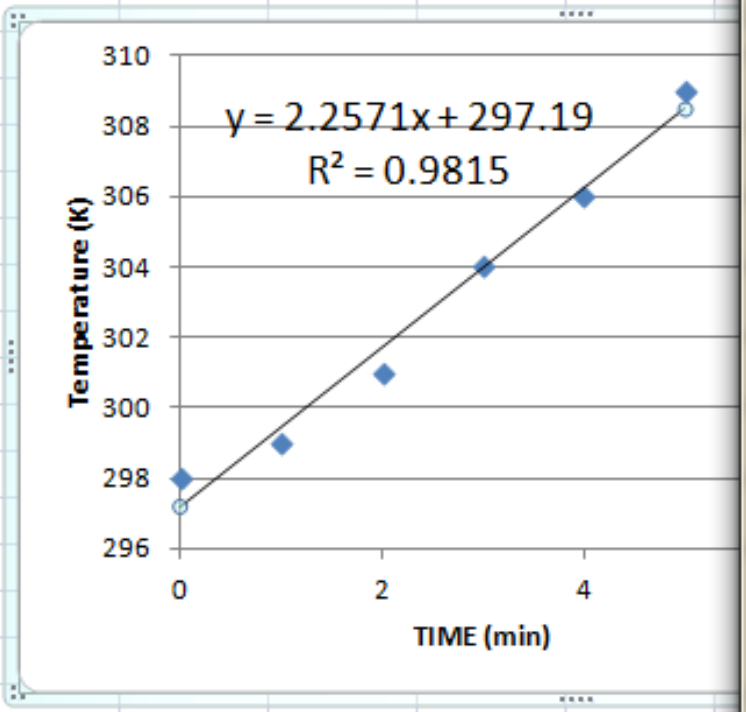
	A	B	C	D	E	F
1	Time (min)	Temp (K)				
2	0	298		Slope	2.257143	
3	1	299		Intercept	297.1905	
4	2	301		R^2	0.981547	
5	3	304				
6	4	306				
7	5	309				

=SLOPE(y, x)

=INTERCEPT (y, x)

=R^2 (y,x)

	A	B	C	D	E
1	Time (min)	Temp (K)			
2	0	298		Slope	2.257143
3	1	299		Intercept	297.1905
4	2	301		R^2	0.981547
5	3	304			
6	4	306			
7	5	309			



Format Trendline

Trendline Options
Line Color
Line Style
Shadow

Trendline Options

Trend/Regression Type
- Exponential
- Linear
- Logarithmic
- Polynomial Order: 2
- Power
- Moving Average Period: 2

Trendline Name
- Automatic : Linear (Series1)
- Custom:

Forecast
- Forward: 0.0 periods
- Backward: 0.0 periods

Set Intercept = 0.0
Display Equation on chart
Display R-squared value on chart

Close

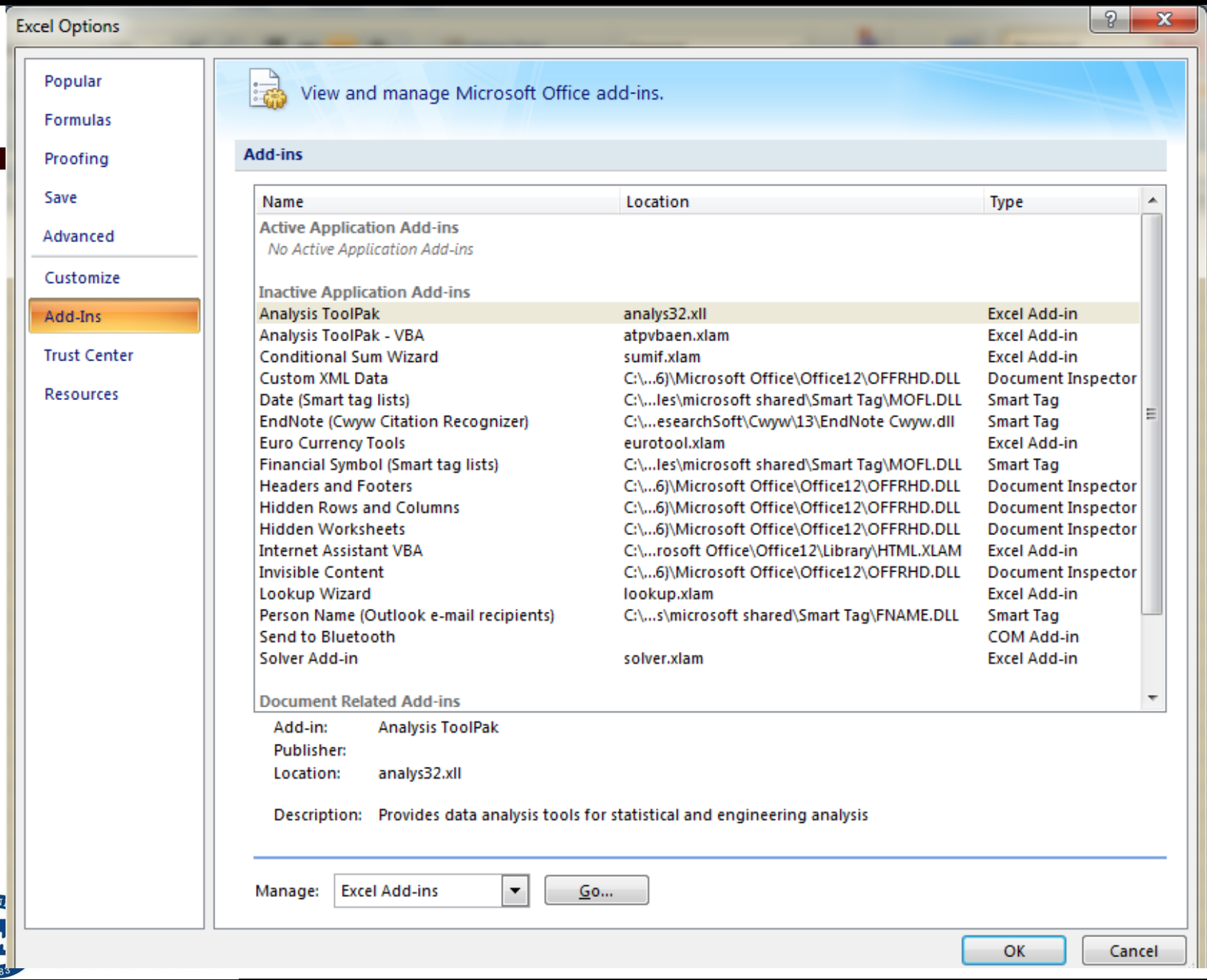
Linear Regression Using Excel's Regression Analysis Package

There are two reasons to use Excel's Regression Package

1. If the regression model is not available as a trendline
2. If you want to know the details of the regression process

Linear Regression Using Excel's Regression Analysis Package

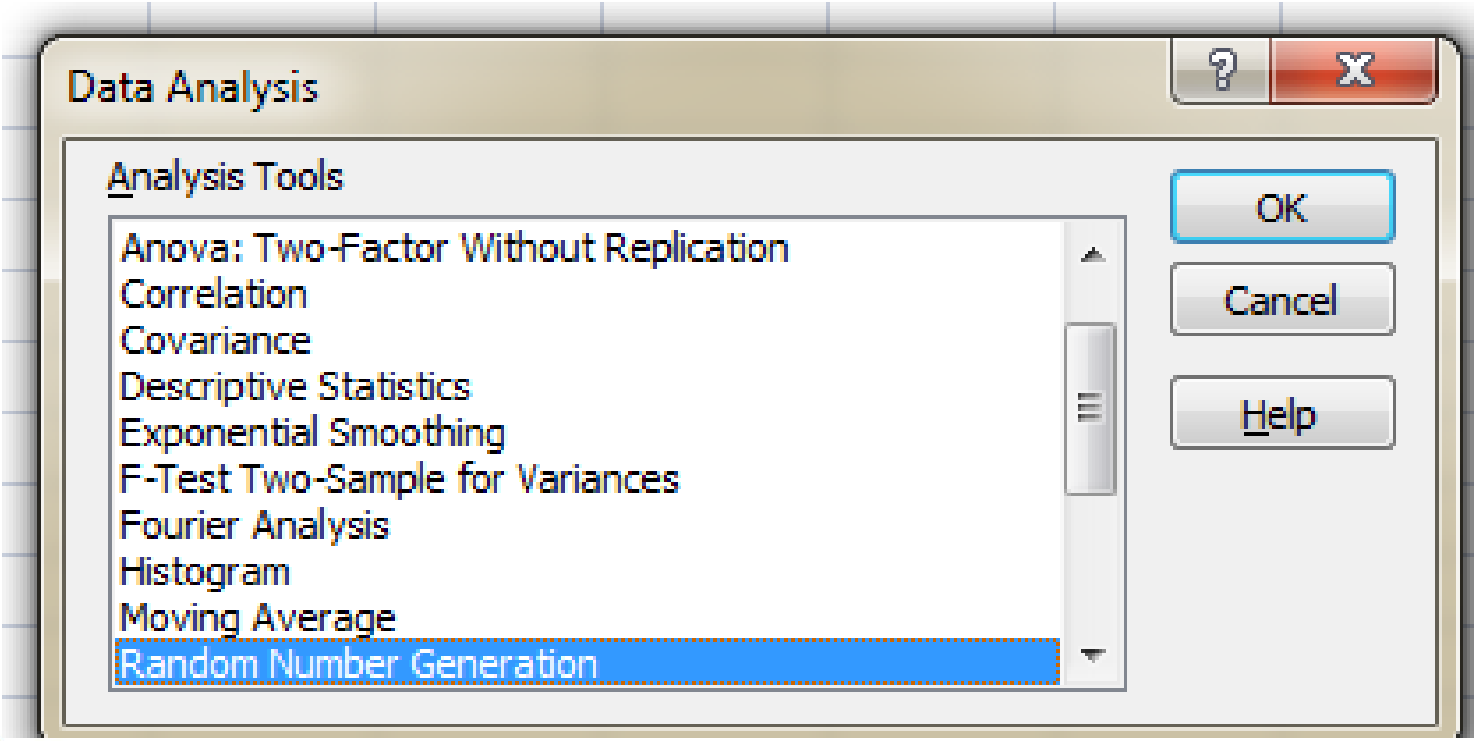
- Excel's Regression Analysis Package is part of the Excel's Analysis Toolpak Add-In
- Office/Excel Option/Add-Ins


Excel Options
Popular
Formulas
Proofing
Save
Advanced
Customize
Add-Ins
Trust Center
Resources
View and manage Microsoft Office add-ins.
Add-ins
Name
Location
Type
Active Application Add-ins
No Active Application Add-ins
Inactive Application Add-ins
Analysis ToolPak analys32.xll Excel Add-in
Analysis ToolPak - VBA atpvbaen.xlam Excel Add-in
Conditional Sum Wizard sumif.xlam Excel Add-in
Custom XML Data C:\...6)\Microsoft Office\Office12\OFFRHD.DLL Document Inspector
Date (Smart tag lists) C:\...les\microsoft shared\Smart Tag\MOFL.DLL Smart Tag
EndNote (Cwyw Citation Recognizer) C:\...esearchSoft\Cwyw\13\EndNote Cwyw.dll Smart Tag
Euro Currency Tools eurotool.xlam Excel Add-in
Financial Symbol (Smart tag lists) C:\...les\microsoft shared\Smart Tag\MOFL.DLL Smart Tag
Headers and Footers C:\...6)\Microsoft Office\Office12\OFFRHD.DLL Document Inspector
Hidden Rows and Columns C:\...6)\Microsoft Office\Office12\OFFRHD.DLL Document Inspector
Hidden Worksheets C:\...6)\Microsoft Office\Office12\OFFRHD.DLL Document Inspector
Internet Assistant VBA C:\...rosoft Office\Office12\Library\HTML.XLAM Excel Add-in
Invisible Content C:\...6)\Microsoft Office\Office12\OFFRHD.DLL Document Inspector
Lookup Wizard lookup.xlam Excel Add-in
Person Name (Outlook e-mail recipients) C:\...s\microsoft shared\Smart Tag\FNAME.DLL Smart Tag
Send to Bluetooth COM Add-in
Solver Add-in solver.xlam Excel Add-in
Document Related Add-ins
Add-in: Analysis ToolPak
Publisher:
Location: analys32.xll
Description: Provides data analysis tools for statistical and engineering analysis
Manage: Excel Add-ins
Go...
OK
Cancel

Linear Regression Using Excel's Regression Analysis Package

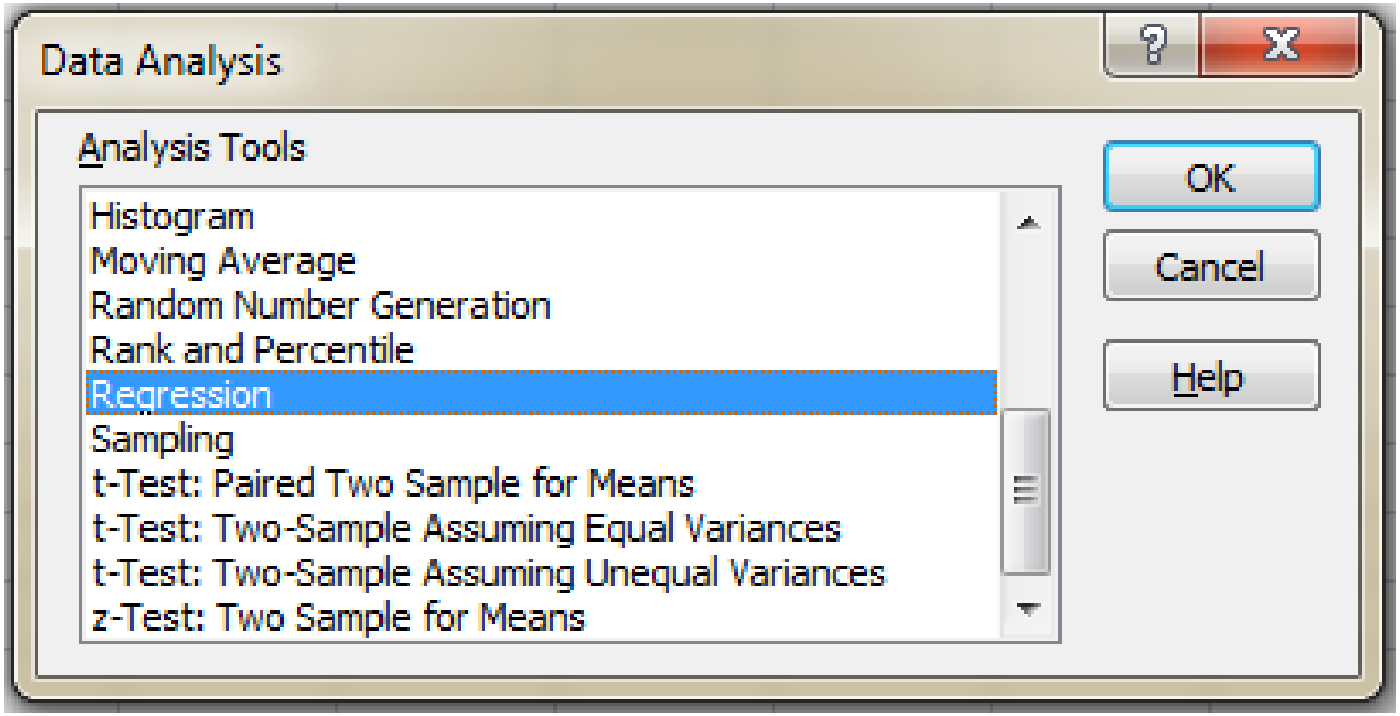
- The Add-Ins panel lists
  - Active application Add-Ins
  - Inactive application Add-Ins
- If the Analysis toolpak is in the Inactive list then,
  - Click on the Analysis toolpak
  - Click the Go button
  - Check the box labeled Analysis toolpak
  - Click the OK button
  - Data Analysis button will appear in the ribbon

- Click on the data analysis button
- Choose regression from the list

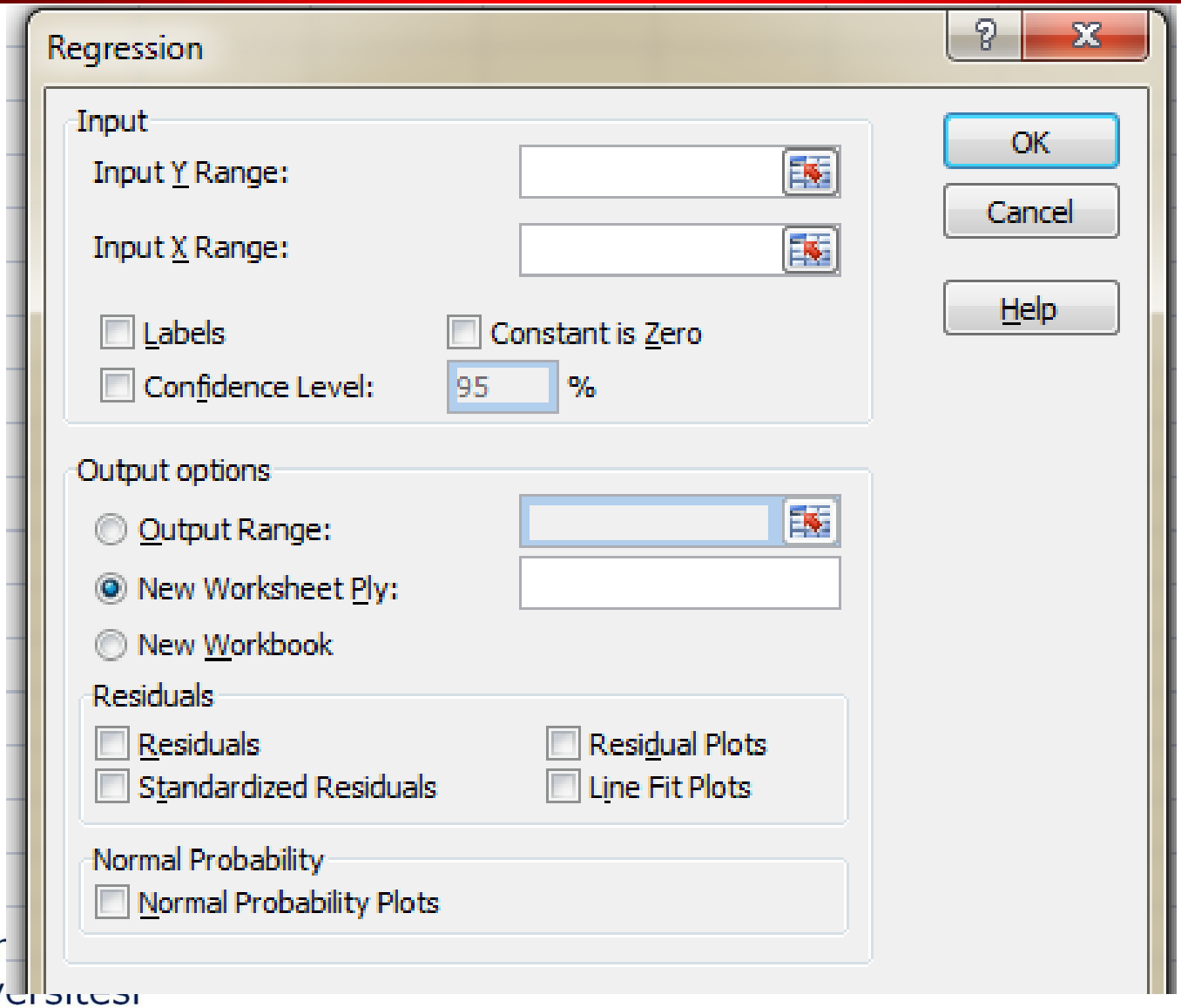


Simple Regression Using Analysis Toolpak

- Select "Regression" from Analysis toolpak

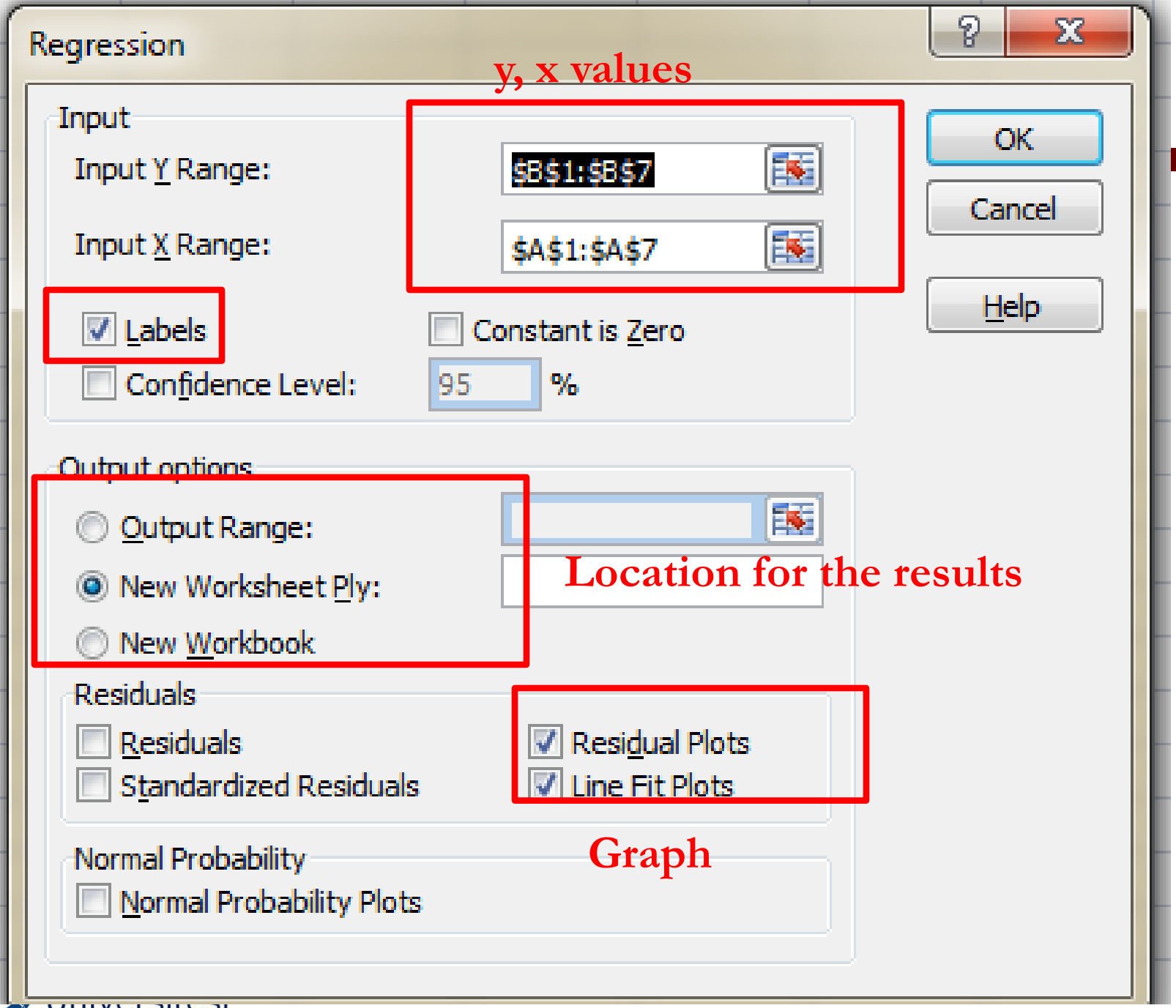


Simple Regression Using Analysis Toolpak



Steps of Simple Regression Using Analysis Toolpak

1) Tell excel to find the y values
2) Tell Excel to find the x values
3) Choose location for the results
4) Indicate you want your results to be graphed
5) Perform the regression


Regression
y, x values
Input
Input Y Range:
$B$1:$B$7
Input X Range:
$A$1:$A$7
OK
Cancel
Help
Labels
Constant is Zero
Confidence Level:
95
%
Output options
Output Range:
New Worksheet Ply:
New Workbook
Location for the results
Residuals
Residuals
Standardized Residuals
Residual Plots
Line Fit Plots
Graph
Normal Probability
Normal Probability Plots

SUMMARY OUTPUT

Regression Statistics	
Multiple F	0.99073
R Square	0.981547
Adjusted I	0.976933
Standard E	0.647339
Observati	6

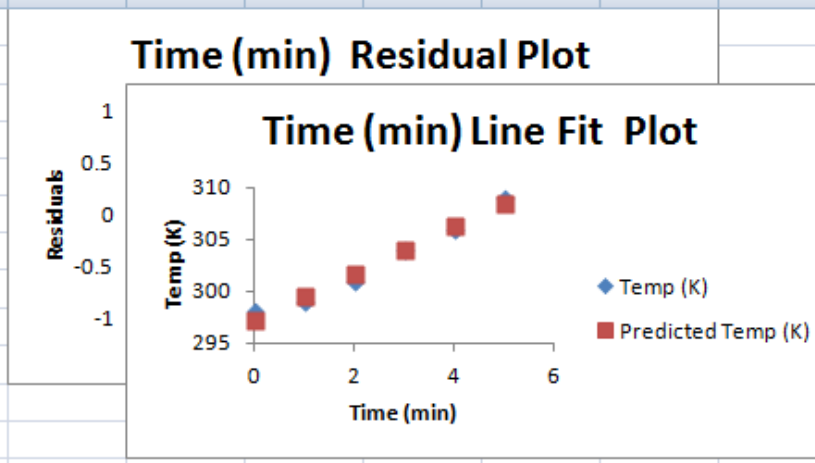
ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	1	89.15714	89.15714	212.7614	0.000128
Residual	4	1.67619	0.419048		
Total	5	90.83333			

	Coefficients	*ndard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	297.1905	0.468509	634.3318	3.71E-11	295.8897	298.4913	295.8897	298.4913
Time (min	2.257143	0.154744	14.58634	0.000128	1.827506	2.68678	1.827506	2.68678

RESIDUAL OUTPUT

)bservatio	*icted Temp*	*Residuals*
1	297.1905	0.809524
2	299.4476	-0.44762
3	301.7048	-0.70476
4	303.9619	0.038095
5	306.219	-0.21905
6	308.4762	0.52381



CSE 123
Introduction to Computing

Statistics Functions/Iteration, Goal Seek, Solver

SPRING 2012

Assist. Prof. A. Evren Tugtas

Marmara Üniversitesi

Statistics Functions

- =AVERAGE( )
- Variance and standard deviation provide information about the spread of the values.
- Small variance/Standard Deviation: data are closely clustered around the mean
- Large variance/Standard Deviation: wide range of values in the data set
- Standard deviation = SQRT(variance)

Statistics Functions

- Standard deviation: =STDEV( )
- Variance: =VAR( )
- Analysis toolpak can be used to perform other statistical analysis such as t-test and ANOVA

Iteration

- Several methods have been used to find iterative solutions;
  - Excel's Goal Seek Feature
  - Excel's Solver

Excel's Goal Seek

- Goal seek feature allows you to solve problems backwards.
- Examples:
- Velocity in the pipe should not exceed 3 m/s; what pipe diameter is required to keep this velocity?
- To use Goal Seek, you need to know the answer you want.

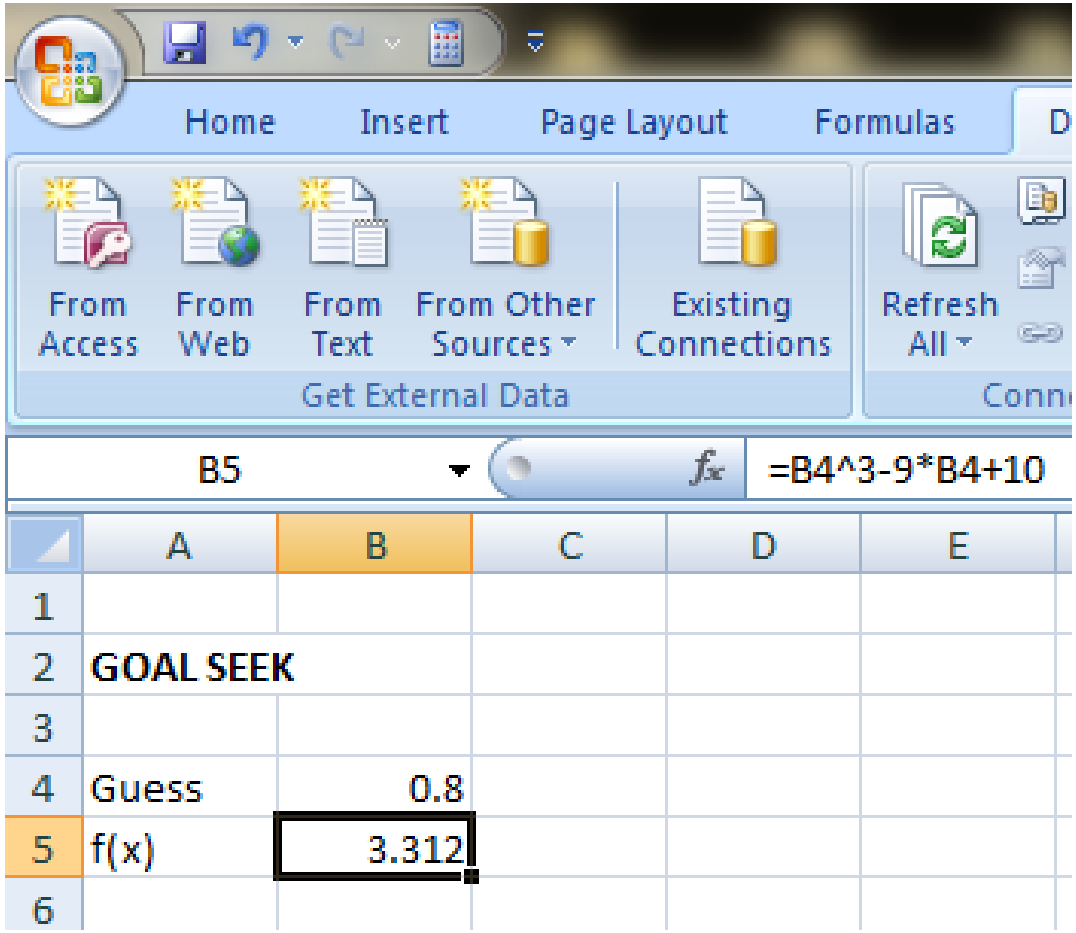
Excel's Goal Seek

- To use Goal Seek, you need a cell to hold guess value and another cell to contain a formula that computes the results
- To open Goal Seek;

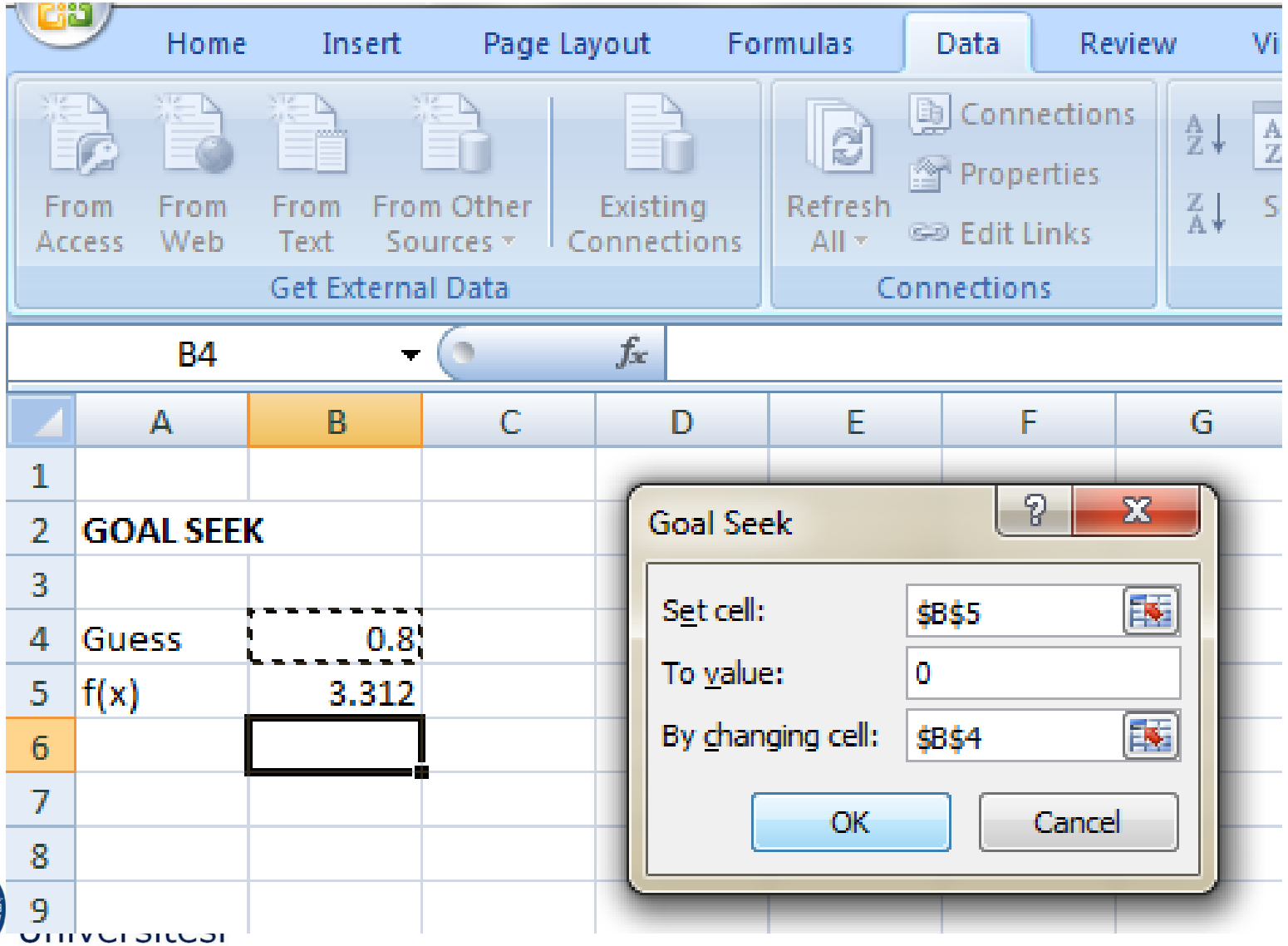
Data/Data Tools/What-If Analysis/Goal Seek

Excel's Goal Seek - Example

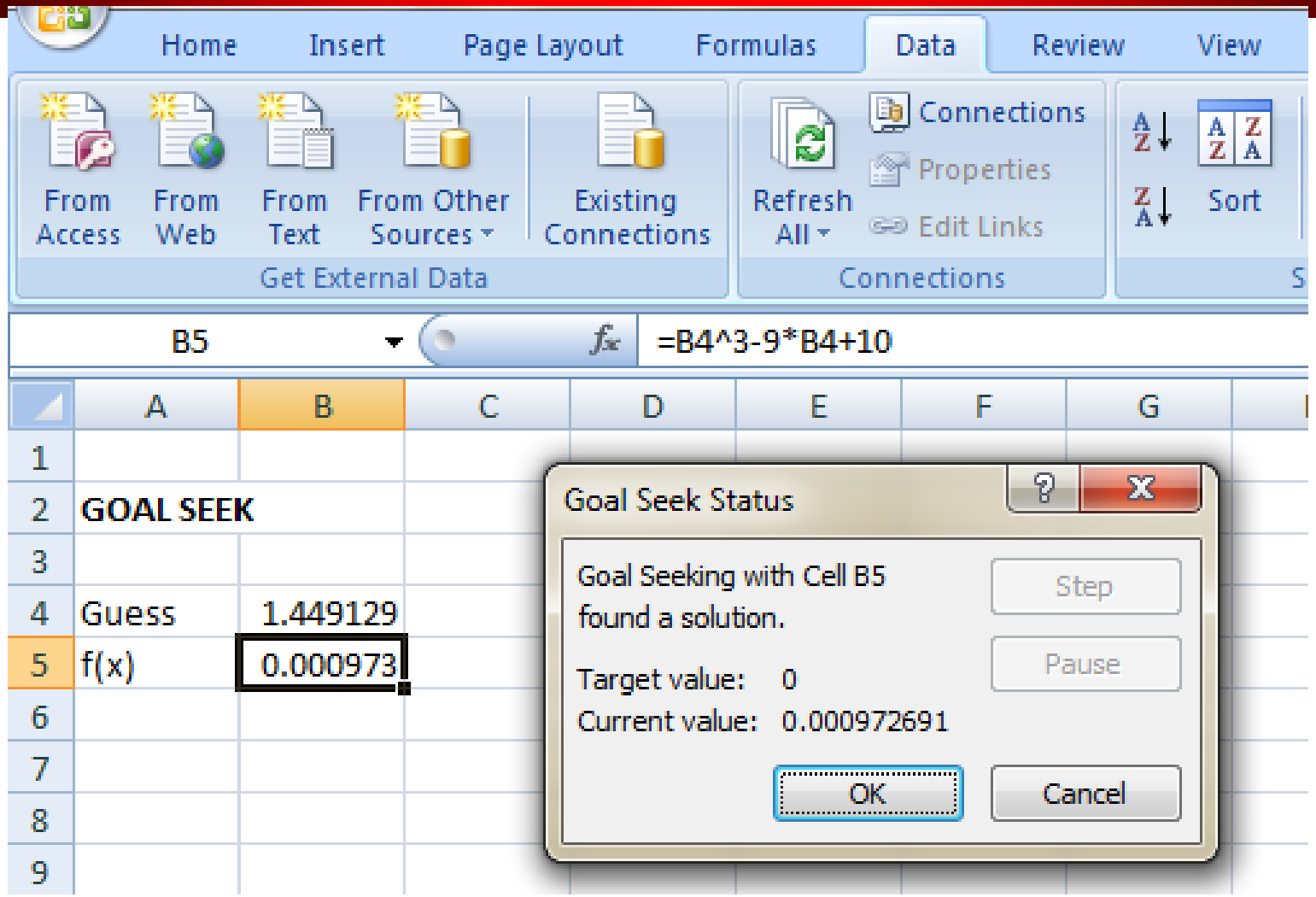
$$f(x) = x^3 - 9x + 10$$



Excel's Goal Seek - Example



Excel's Goal Seek - Example



Excel's Solver

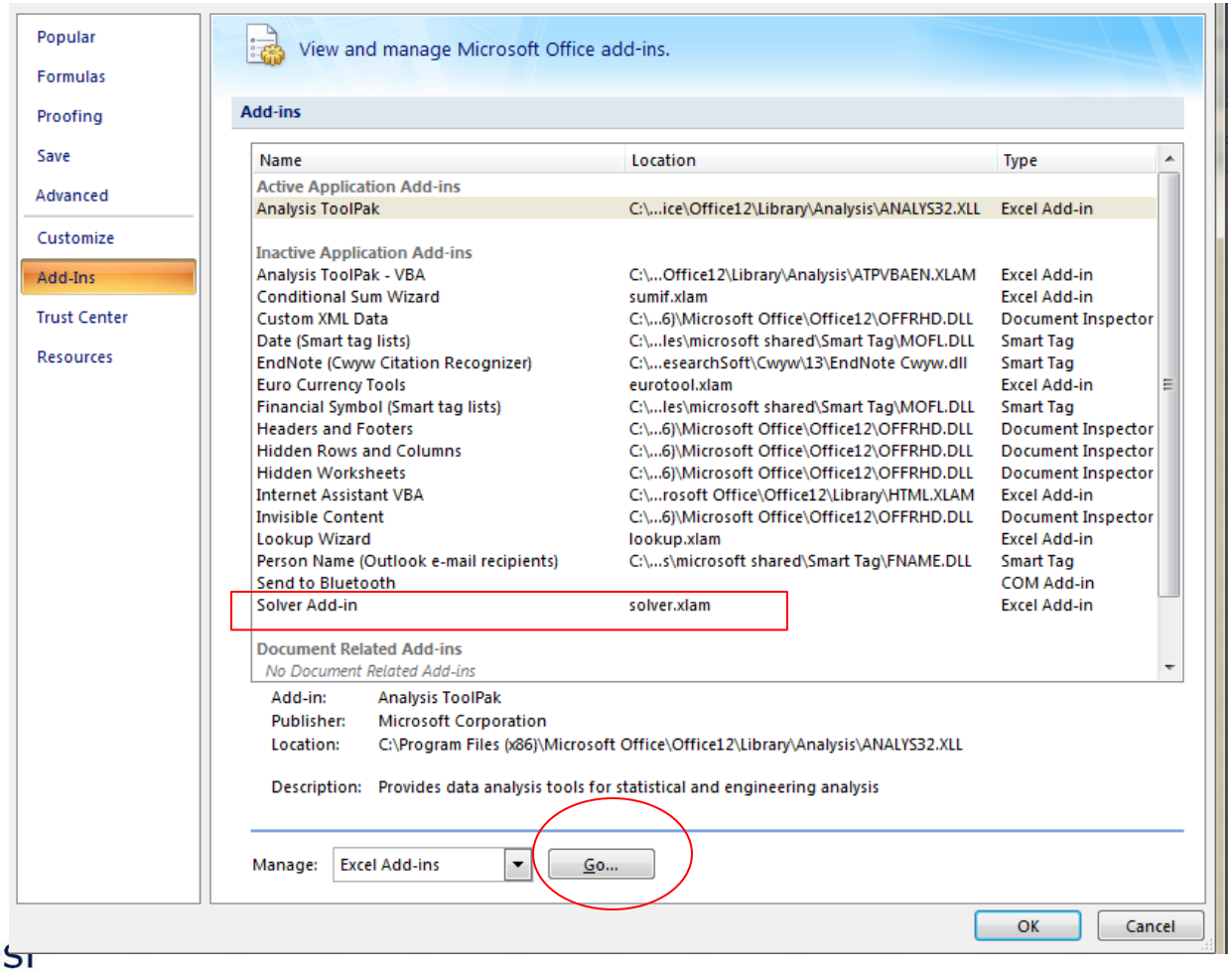
- Excel provides a very powerful iteration tool called Solver.
- Solver is an Add-In that installed but not activated by default when Excel is installed.
- If solver is already activated, you would see button labeled "Solver" at

  Data/Data Analysis/Solver

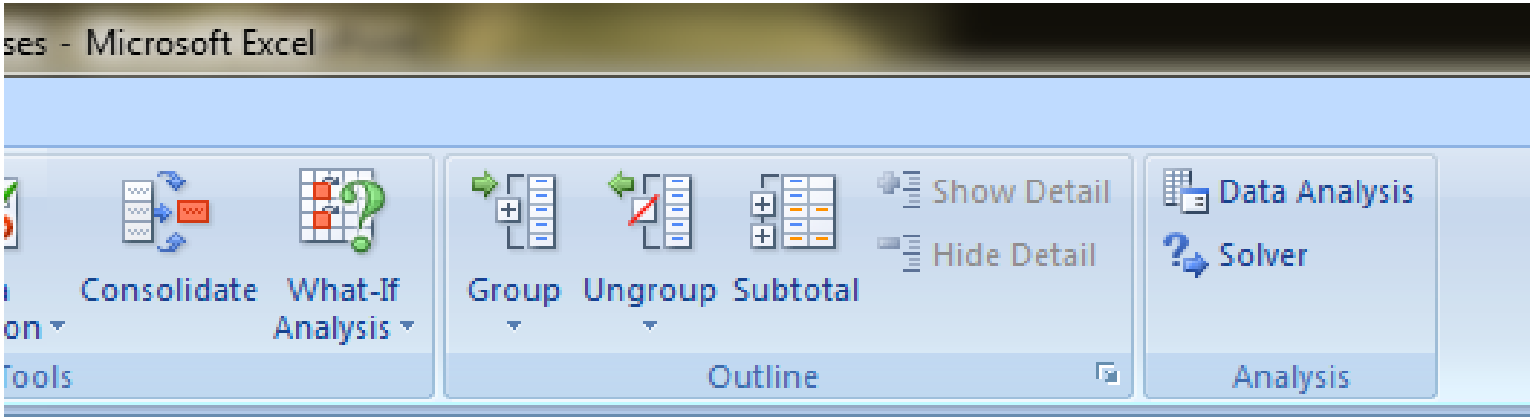
Excel's Solver

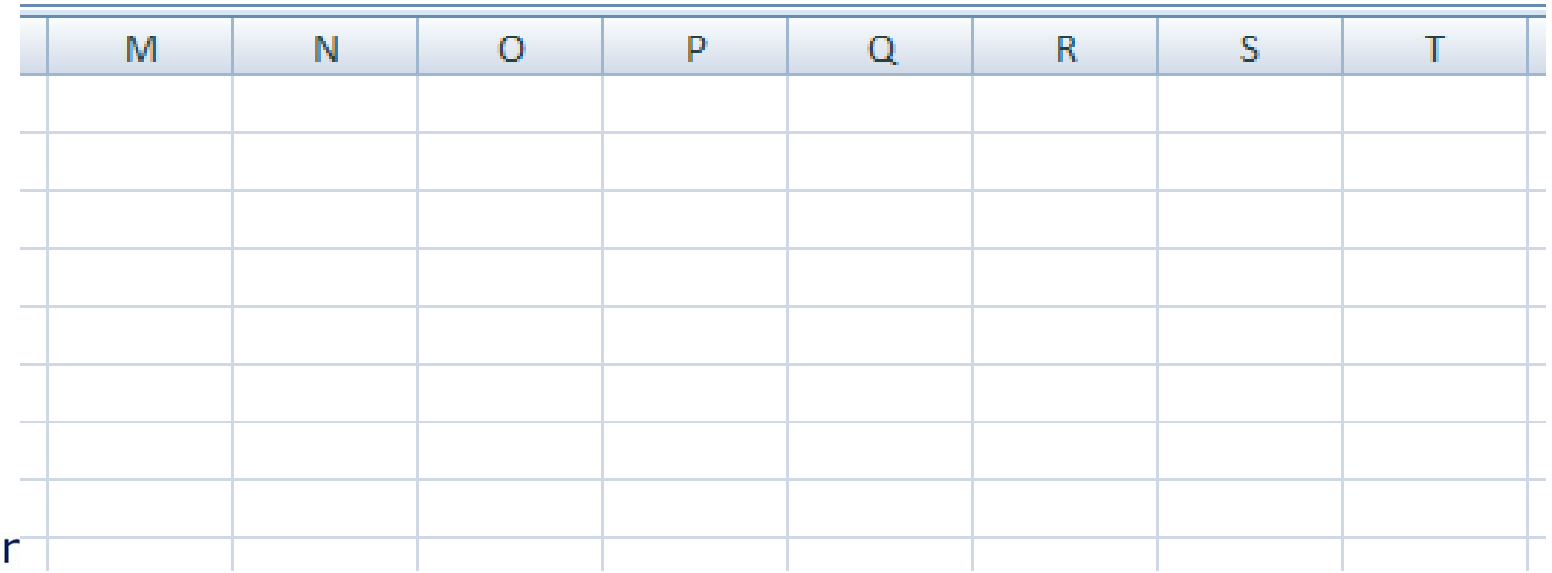
If solver is not activated

Office/Excel Options/Add-Ins/Solver Add-In/Go



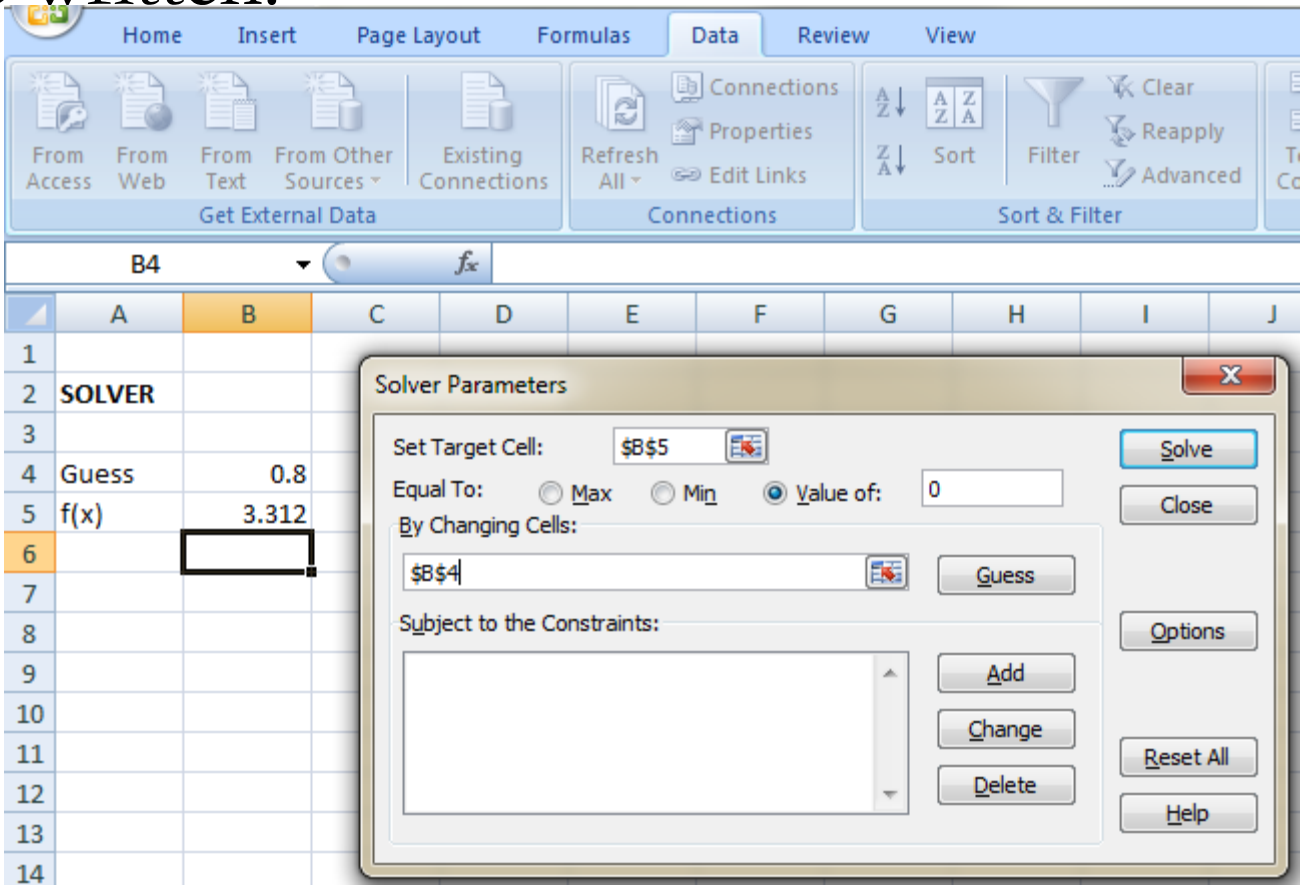
Excel's Solver



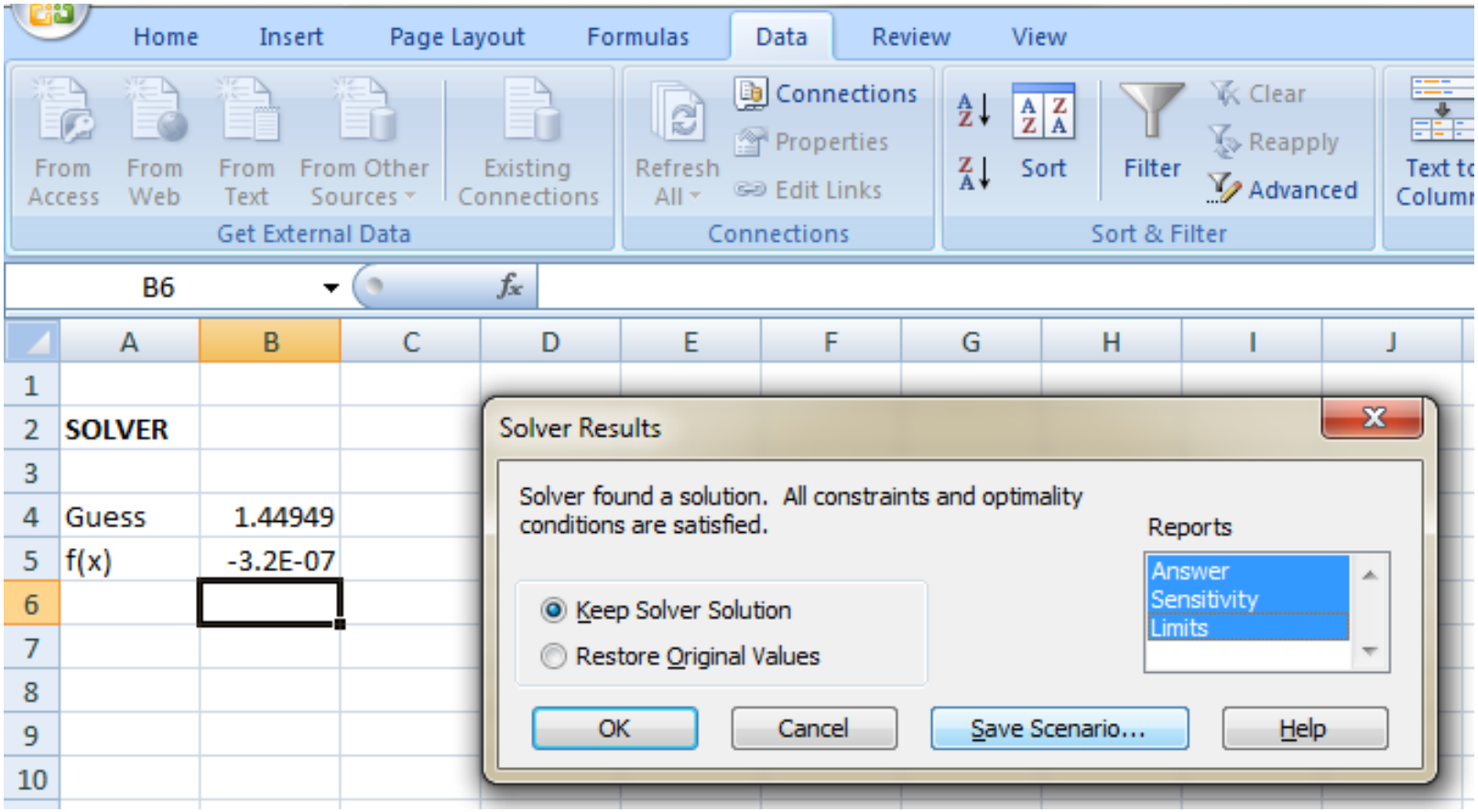


Excel's Solver

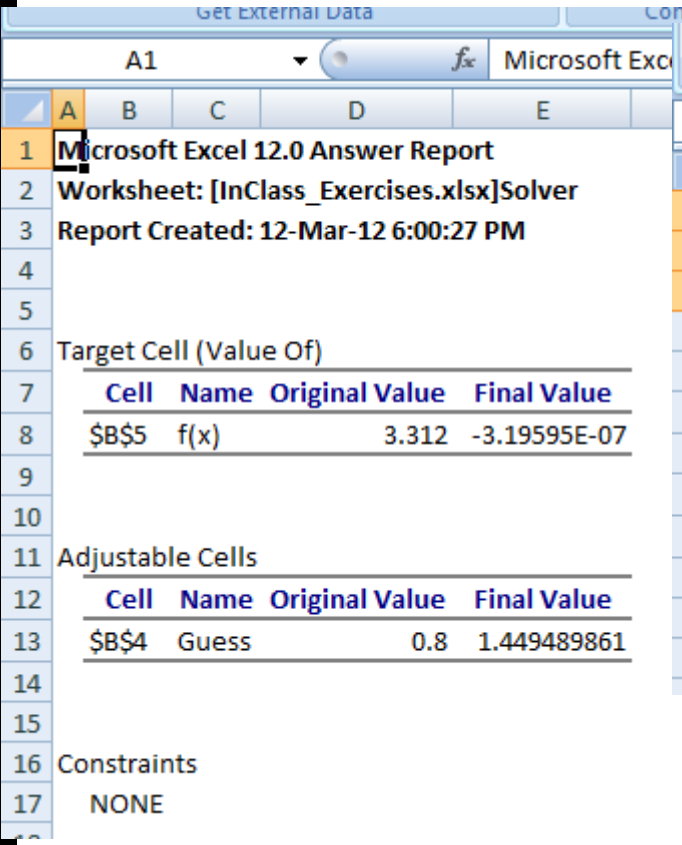
- Similar to Goal Seek, Excel requires the equation to be written.



Excel's Solver



Excel's Solver



Microsoft Excel 12.0 Answer Report
Worksheet: [InClass_Exercises.xlsx]Solver
Report Created: 12-Mar-12 6:00:27 PM

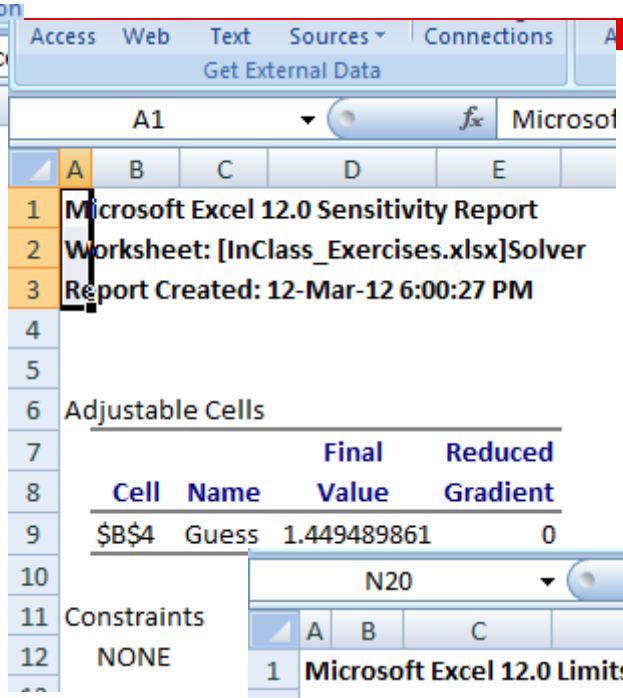
Target Cell (Value Of)

Cell	Name	Original Value	Final Value
$B$5	f(x)	3.312	-3.19595E-07

Adjustable Cells

Cell	Name	Original Value	Final Value
$B$4	Guess	0.8	1.449489861

Constraints
NONE

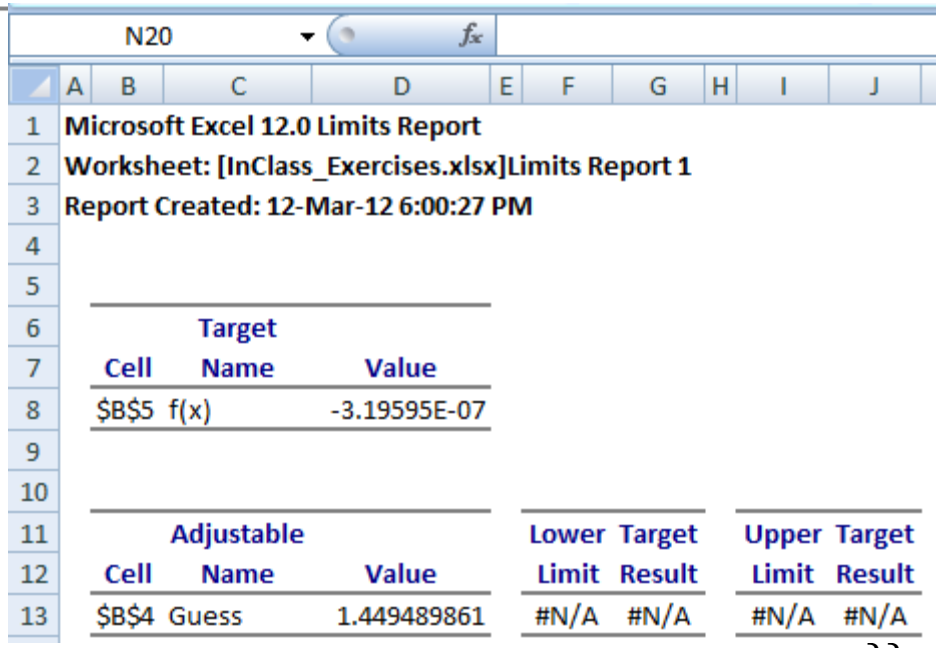


Microsoft Excel 12.0 Sensitivity Report
Worksheet: [InClass_Exercises.xlsx]Solver
Report Created: 12-Mar-12 6:00:27 PM

Adjustable Cells

Cell	Name	Final Value	Reduced Gradient
$B$4	Guess	1.449489861	0

Constraints
NONE

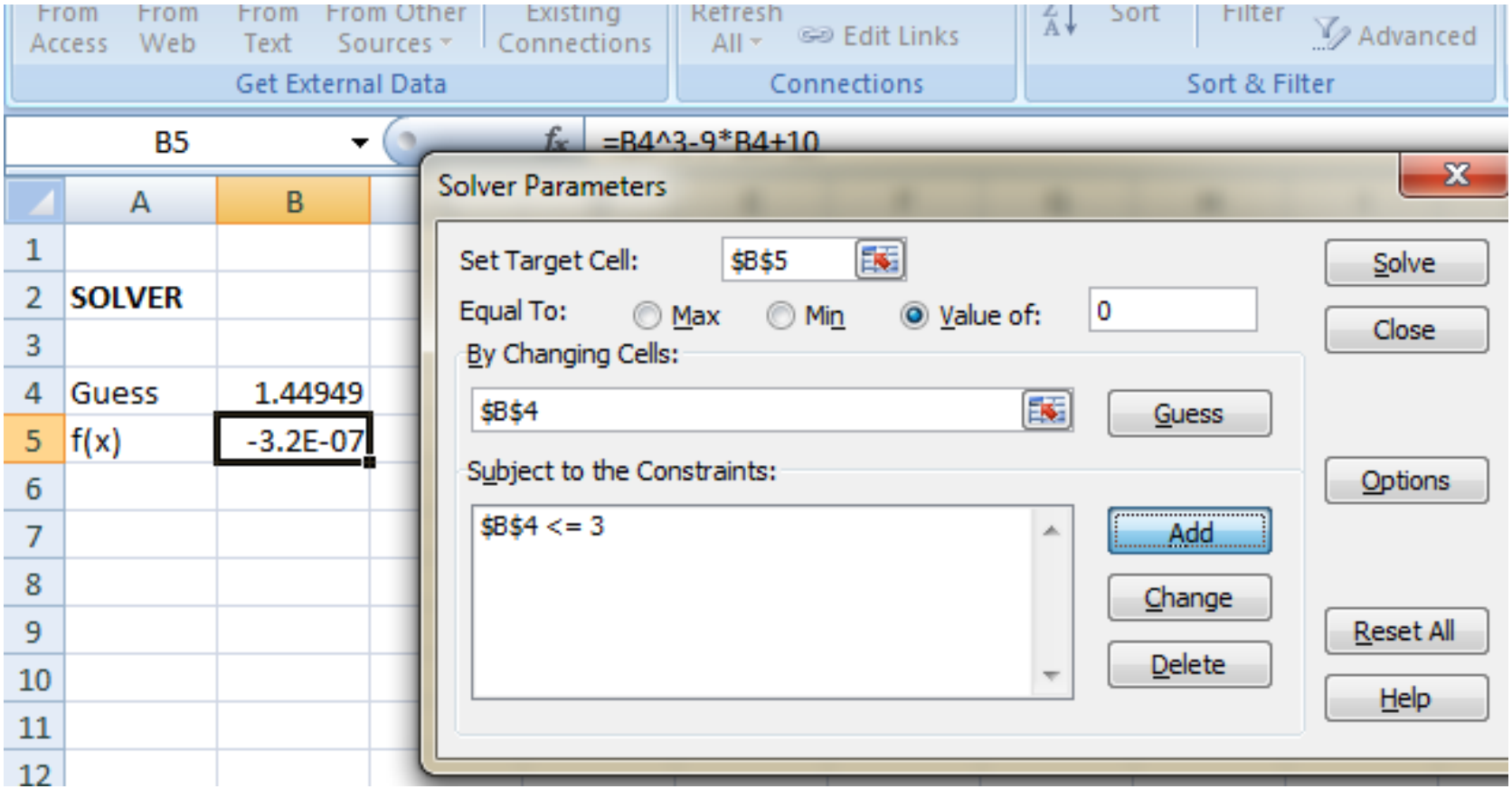


Microsoft Excel 12.0 Limits Report
Worksheet: [InClass_Exercises.xlsx]Limits Report 1
Report Created: 12-Mar-12 6:00:27 PM

	Target					
Cell	Name	Value	Adjustable	Lower Limit	Target Result	Upper Limit	Target Result
$B$5	f(x)	-3.19595E-07					
			Cell	Name	Value		
			$B$4	Guess	1.449489861	#N/A	#N/A	#N/A	#N/A

Excel's Solver

- You can add constraints to the Solver



CSE 123
Introduction to Computing

Pivot Tables

SPRING 2012

Assist. Prof. A. Evren Tugtas

Marmara Üniversitesi

Pivot Tables

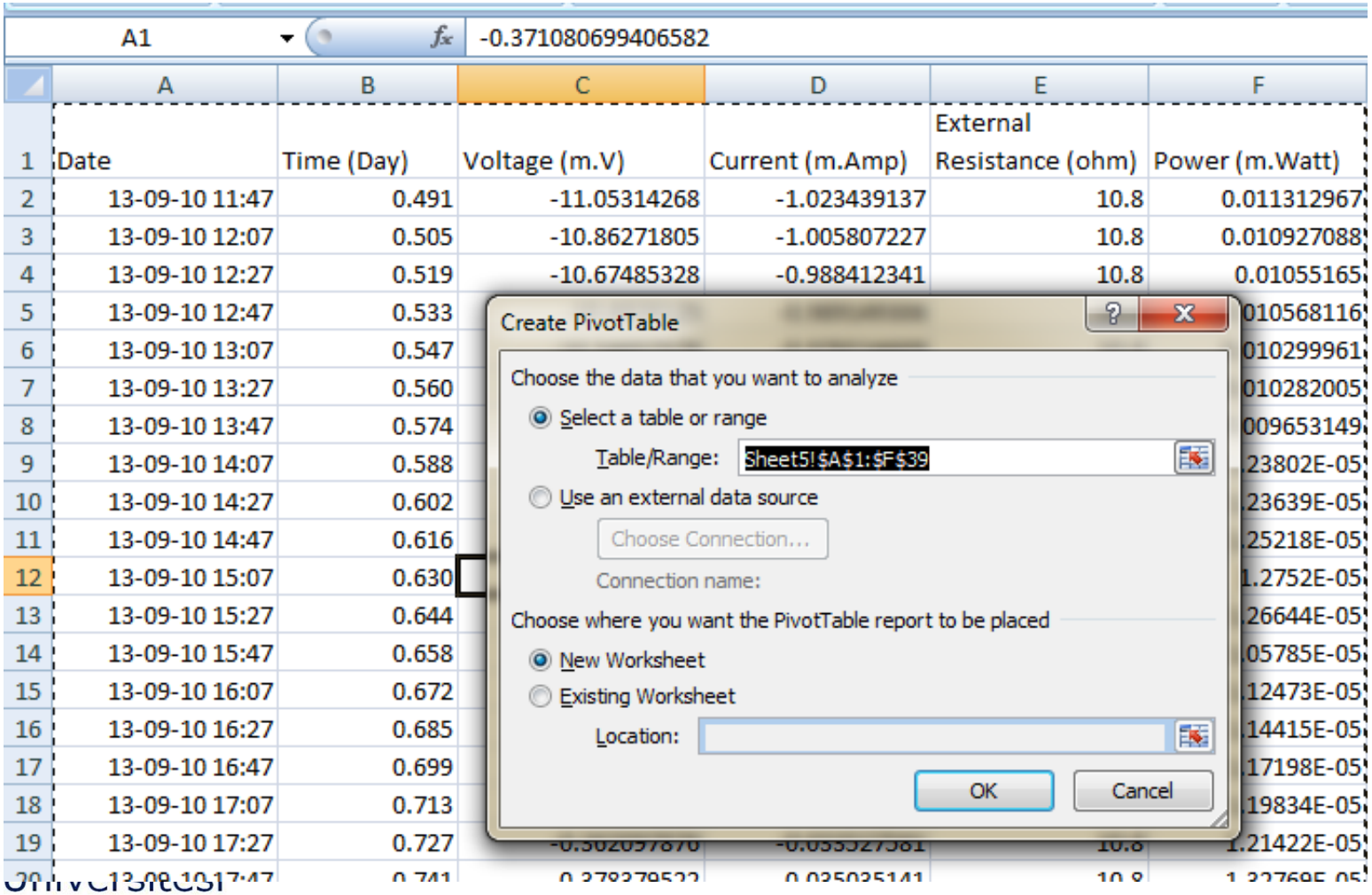
- It is a summary table for large data sets
- Pivot tables can be easily modified to get different views of a table
- You need to have a large data set with number of columns and raws

Pivot Tables

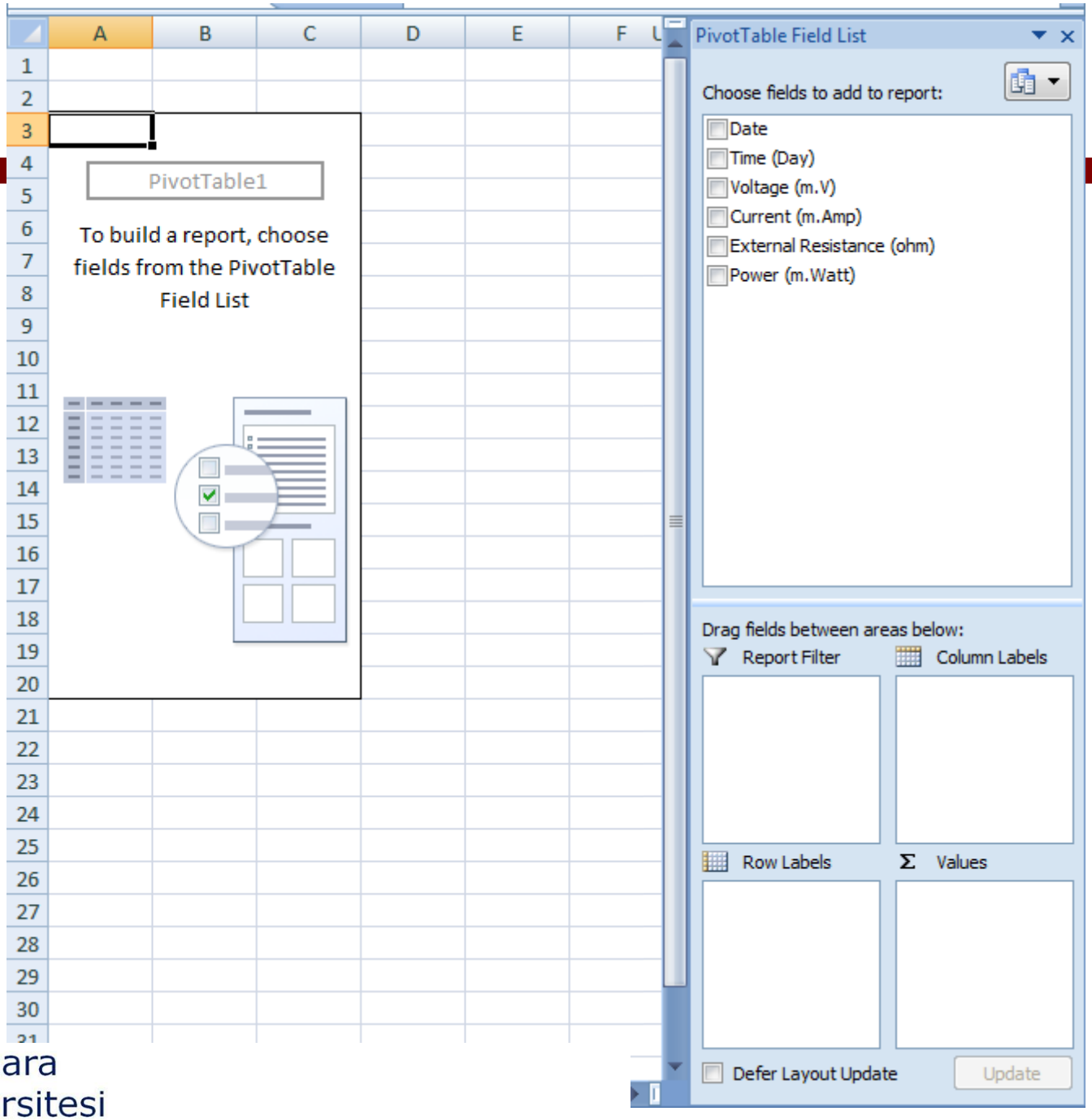
	A	B	C	D	E	F
1	Date	Time (Day)	Voltage (m.V)	Current (m.Amp)	External Resistance (ohm)	Power (m.Watt)
2	13-09-10 11:47	0.491	-11.05314268	-1.023439137	10.8	0.011312967
3	13-09-10 12:07	0.505	-10.86271805	-1.005807227	10.8	0.010927088
4	13-09-10 12:27	0.519	-10.67485328	-0.988412341	10.8	0.01055165
5	13-09-10 12:47	0.533	-10.6828125	-0.989149306	10.8	0.010568116
6	13-09-10 13:07	0.547	-10.54657378	-0.976534609	10.8	0.010299961
7	13-09-10 13:27	0.560	-10.53746612	-0.975691307	10.8	0.010282005
8	13-09-10 13:47	0.574	-9.970020286	-0.923150026	10.8	0.009653149
9	13-09-10 14:07	0.588	-0.365612061	-0.033852969	10.8	1.23802E-05
10	13-09-10 14:27	0.602	-0.365376	-0.033831111	10.8	1.23639E-05
11	13-09-10 14:47	0.616	-0.367713998	-0.034047592	10.8	1.25218E-05
12	13-09-10 15:07	0.630	-0.371080699	-0.034359324	10.8	1.2752E-05
13	13-09-10 15:27	0.644	-0.369211687	-0.034186267	10.8	1.26644E-05
14	13-09-10 15:47	0.658	-0.337944325	-0.031291141	10.8	1.05785E-05
15	13-09-10 16:07	0.672	-0.348478493	-0.032266527	10.8	1.12473E-05
16	13-09-10 16:27	0.685	-0.351403779	-0.032537387	10.8	1.14415E-05
17	13-09-10 16:47	0.699	-0.355738074	-0.032938711	10.8	1.17198E-05
18	13-09-10 17:07	0.713	-0.359724273	-0.033307803	10.8	1.19834E-05
19	13-09-10 17:27	0.727	-0.362097876	-0.033527581	10.8	1.21422E-05
20	13-09-10 17:47	0.741	-0.378379522	-0.035035141	10.8	1.32769E-05
21	13-09-10 18:07	0.755	-0.391726779	-0.036270998	10.8	1.42109E-05
22	13-09-10 18:27	0.769	-0.393030412	-0.036391705	10.8	1.43067E-05
23	13-09-10 18:47	0.783	-0.391551663	-0.036254784	10.8	1.4198E-05
24	13-09-10 19:07	0.797	-0.387779429	-0.035905503	10.8	1.39259E-05
25	13-09-10 19:27	0.810	-0.388928269	-0.036011877	10.8	1.4007E-05
26	13-09-10 19:47	0.824	-0.386216771	-0.035760812	10.8	1.38146E-05

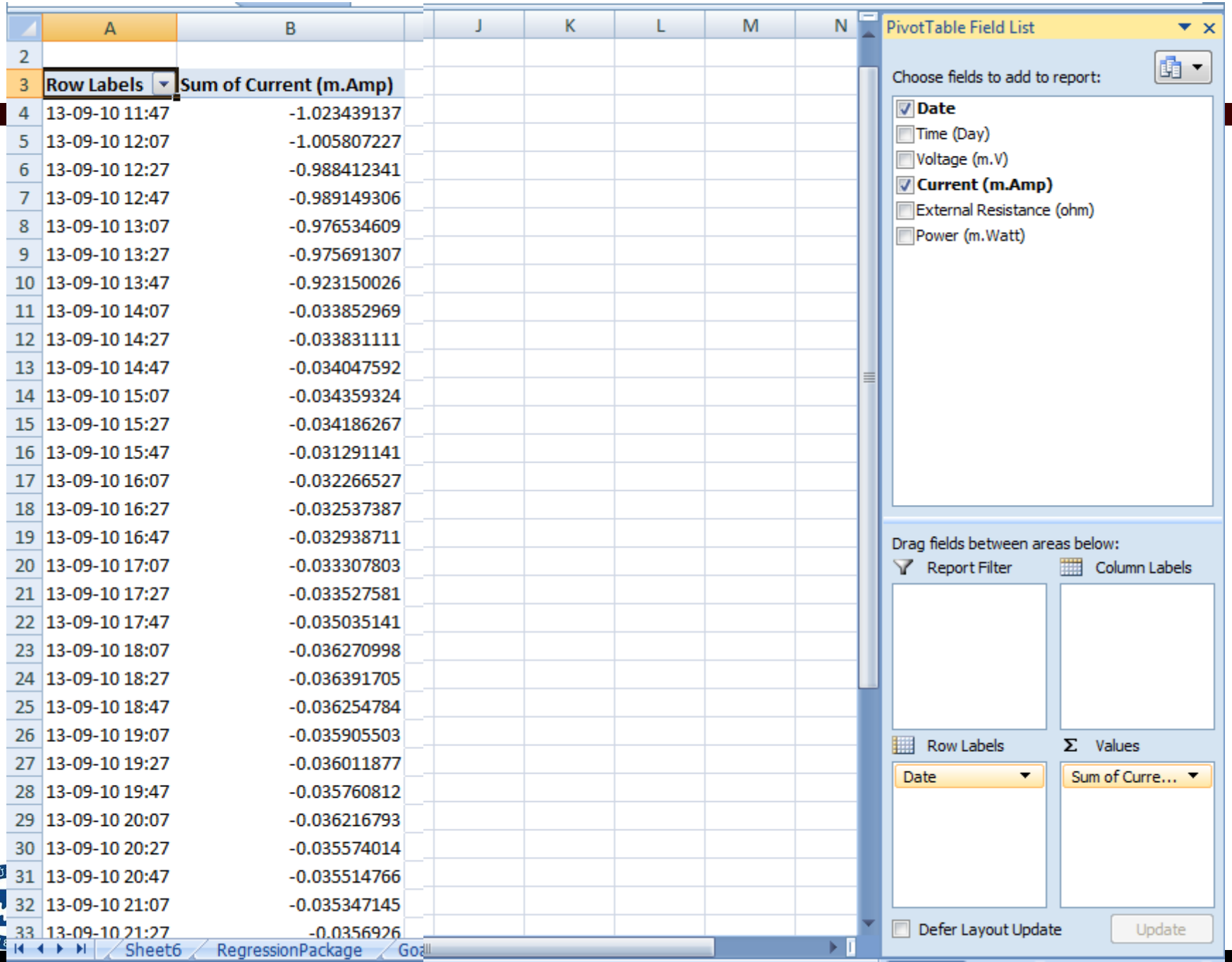
Pivot Table

- Insert/Pivot Table



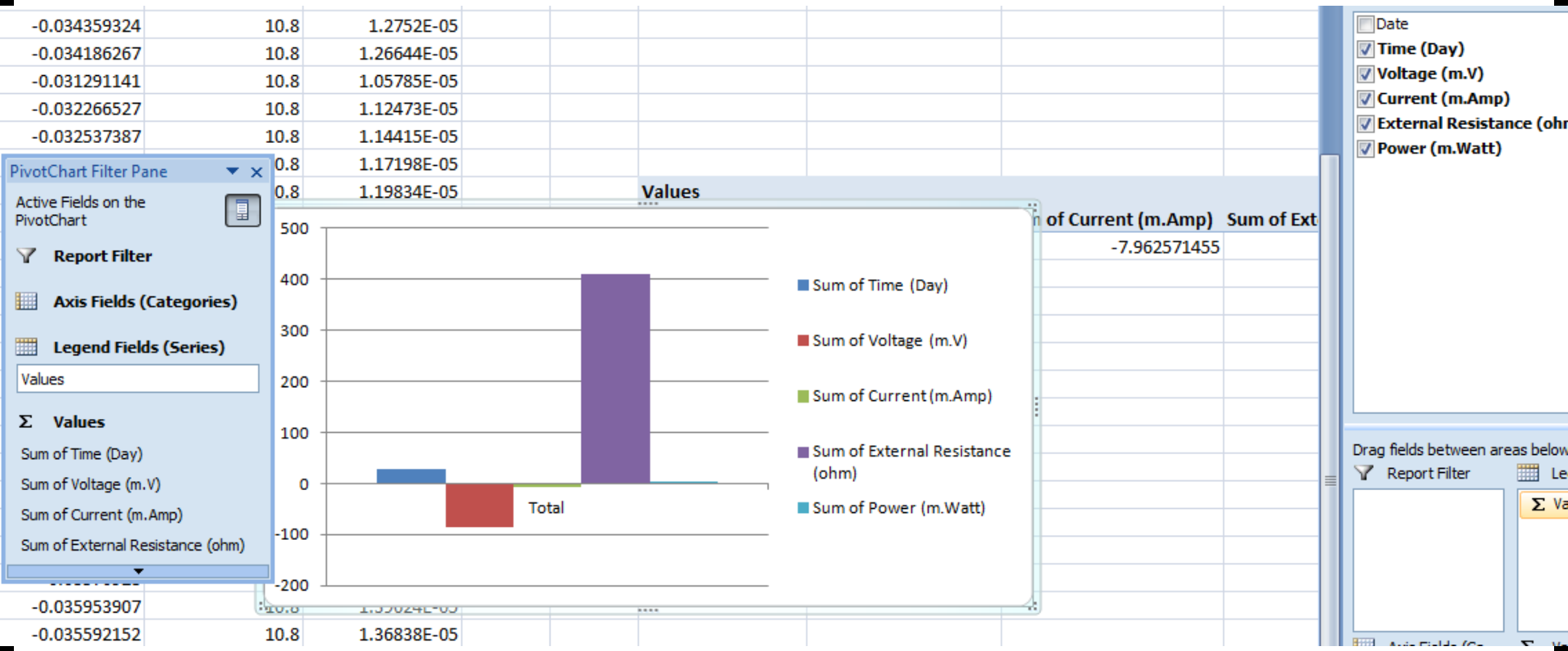
	A	B	C	D	E	F
1	Date	Time (Day)	Voltage (m.V)	Current (m.Amp)	External Resistance (ohm)	Power (m.Watt)
2	13-09-10 11:47	0.491	-11.05314268	-1.023439137	10.8	0.011312967
3	13-09-10 12:07	0.505	-10.86271805	-1.005807227	10.8	0.010927088
4	13-09-10 12:27	0.519	-10.67485328	-0.988412341	10.8	0.01055165
5	13-09-10 12:47	0.533				010568116
6	13-09-10 13:07	0.547				010299961
7	13-09-10 13:27	0.560				010282005
8	13-09-10 13:47	0.574				009653149
9	13-09-10 14:07	0.588				23802E-05
10	13-09-10 14:27	0.602				23639E-05
11	13-09-10 14:47	0.616				25218E-05
12	13-09-10 15:07	0.630				1.2752E-05
13	13-09-10 15:27	0.644				26644E-05
14	13-09-10 15:47	0.658				05785E-05
15	13-09-10 16:07	0.672				12473E-05
16	13-09-10 16:27	0.685				14415E-05
17	13-09-10 16:47	0.699				17198E-05
18	13-09-10 17:07	0.713				19834E-05
19	13-09-10 17:27	0.727				1.21422E-05
20	13-09-10 17:47	0.741	0.378379522	0.035035141	10.8	1.32769E-05


PivotTable1
To build a report, choose fields from the PivotTable Field List
PivotTable Field List
Choose fields to add to report:
Date
Time (Day)
Voltage (m.V)
Current (m.Amp)
External Resistance (ohm)
Power (m.Watt)
Drag fields between areas below:
Report Filter
Column Labels
Row Labels
Values
Defer Layout Update
Update



Row Labels	Sum of Current (m.Amp)
13-09-10 11:47	-1.023439137
13-09-10 12:07	-1.005807227
13-09-10 12:27	-0.988412341
13-09-10 12:47	-0.989149306
13-09-10 13:07	-0.976534609
13-09-10 13:27	-0.975691307
13-09-10 13:47	-0.923150026
13-09-10 14:07	-0.033852969
13-09-10 14:27	-0.033831111
13-09-10 14:47	-0.034047592
13-09-10 15:07	-0.034359324
13-09-10 15:27	-0.034186267
13-09-10 15:47	-0.031291141
13-09-10 16:07	-0.032266527
13-09-10 16:27	-0.032537387
13-09-10 16:47	-0.032938711
13-09-10 17:07	-0.033307803
13-09-10 17:27	-0.033527581
13-09-10 17:47	-0.035035141
13-09-10 18:07	-0.036270998
13-09-10 18:27	-0.036391705
13-09-10 18:47	-0.036254784
13-09-10 19:07	-0.035905503
13-09-10 19:27	-0.036011877
13-09-10 19:47	-0.035760812
13-09-10 20:07	-0.036216793
13-09-10 20:27	-0.035574014
13-09-10 20:47	-0.035514766
13-09-10 21:07	-0.035347145
13-09-10 21:27	-0.0356926

PIVOT CHARTS



CSE 123
Introduction to Computing

External Data Sources

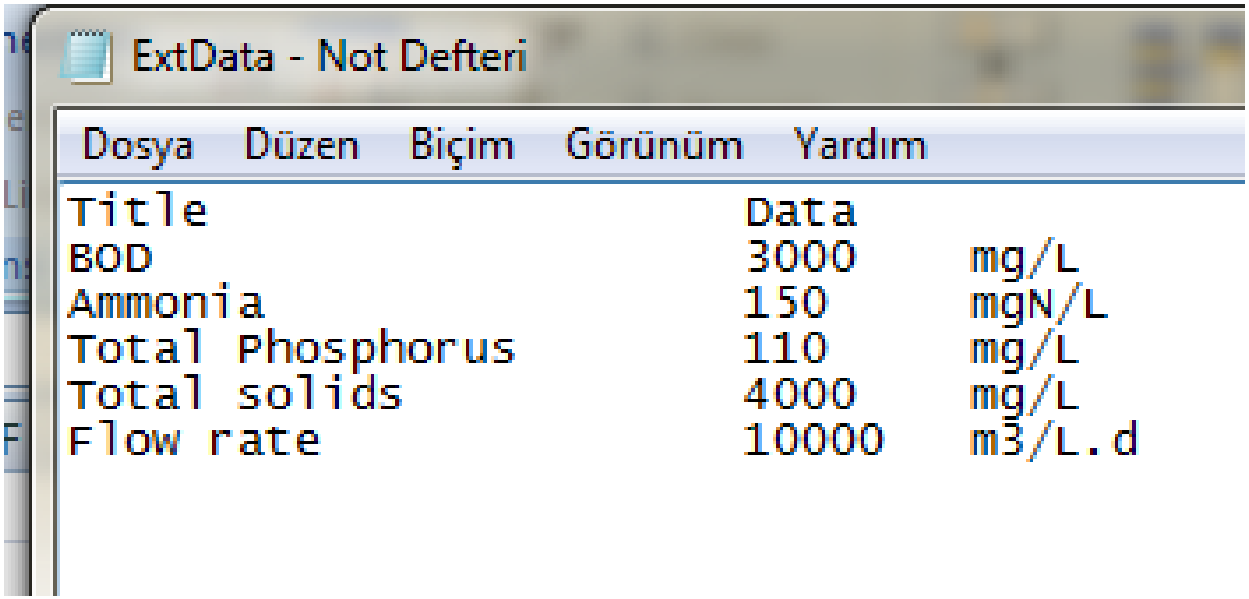
SPRING 2012

Assist. Prof. A. Evren Tugtas

Marmara Üniversitesi

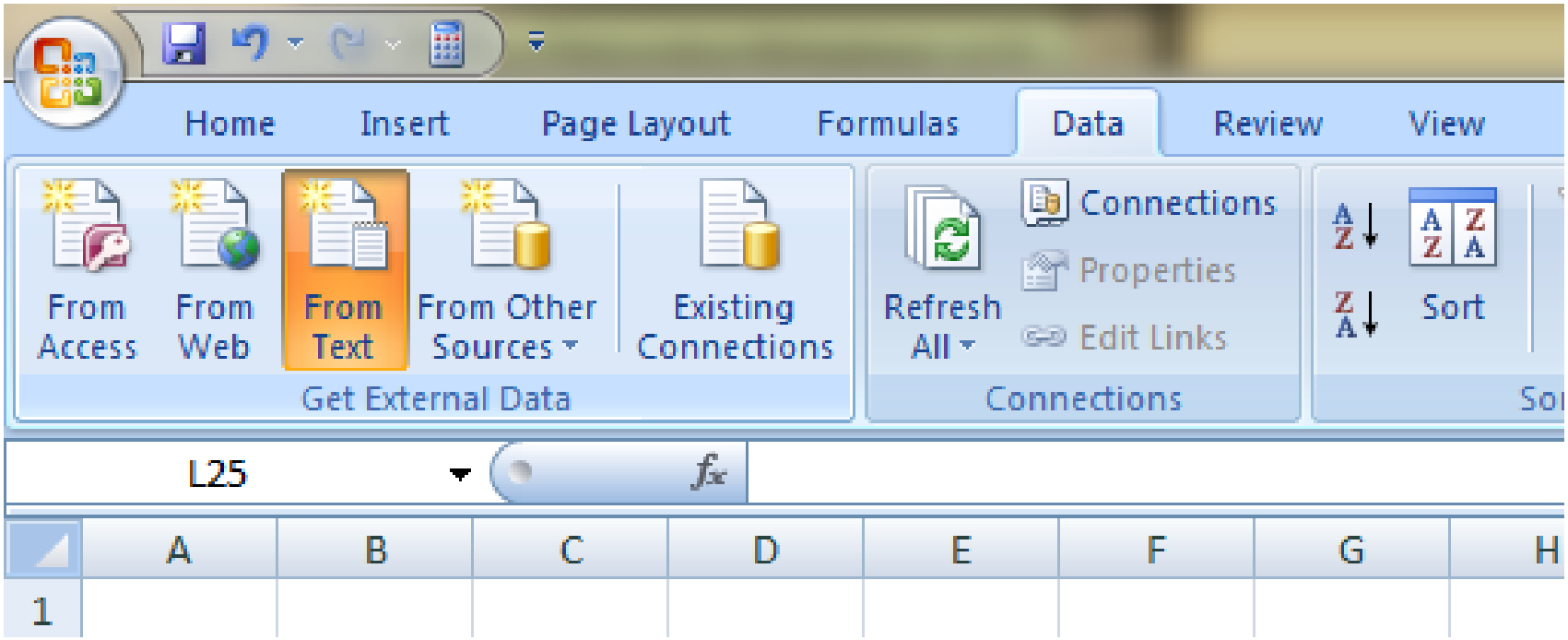
External Data Sources

- Data file used as an input to an Excel Worksheet
- Excel is updated whenever the external data changes
- Create a text file (Use tab to limit data)

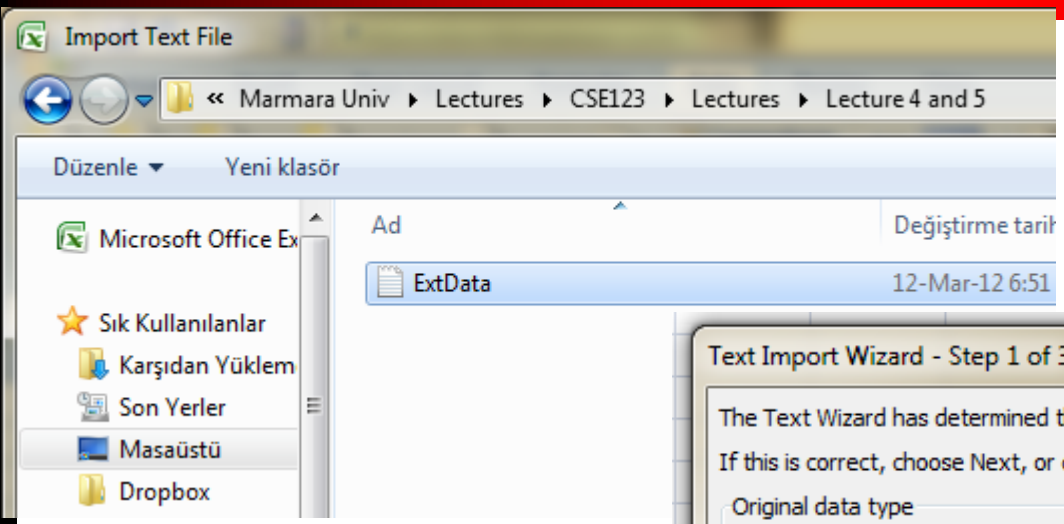


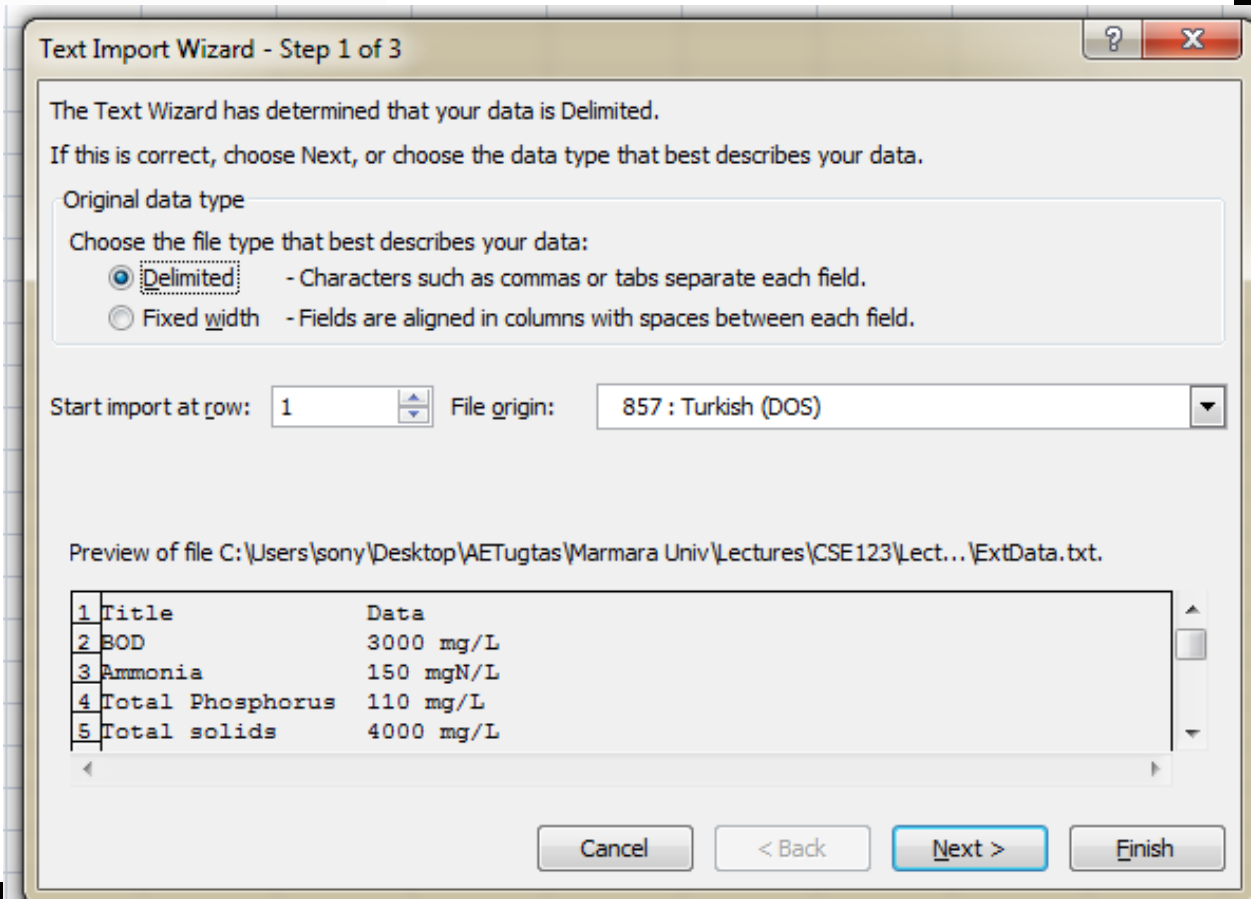
External Data Sources

- Data/Get External Data/From Text



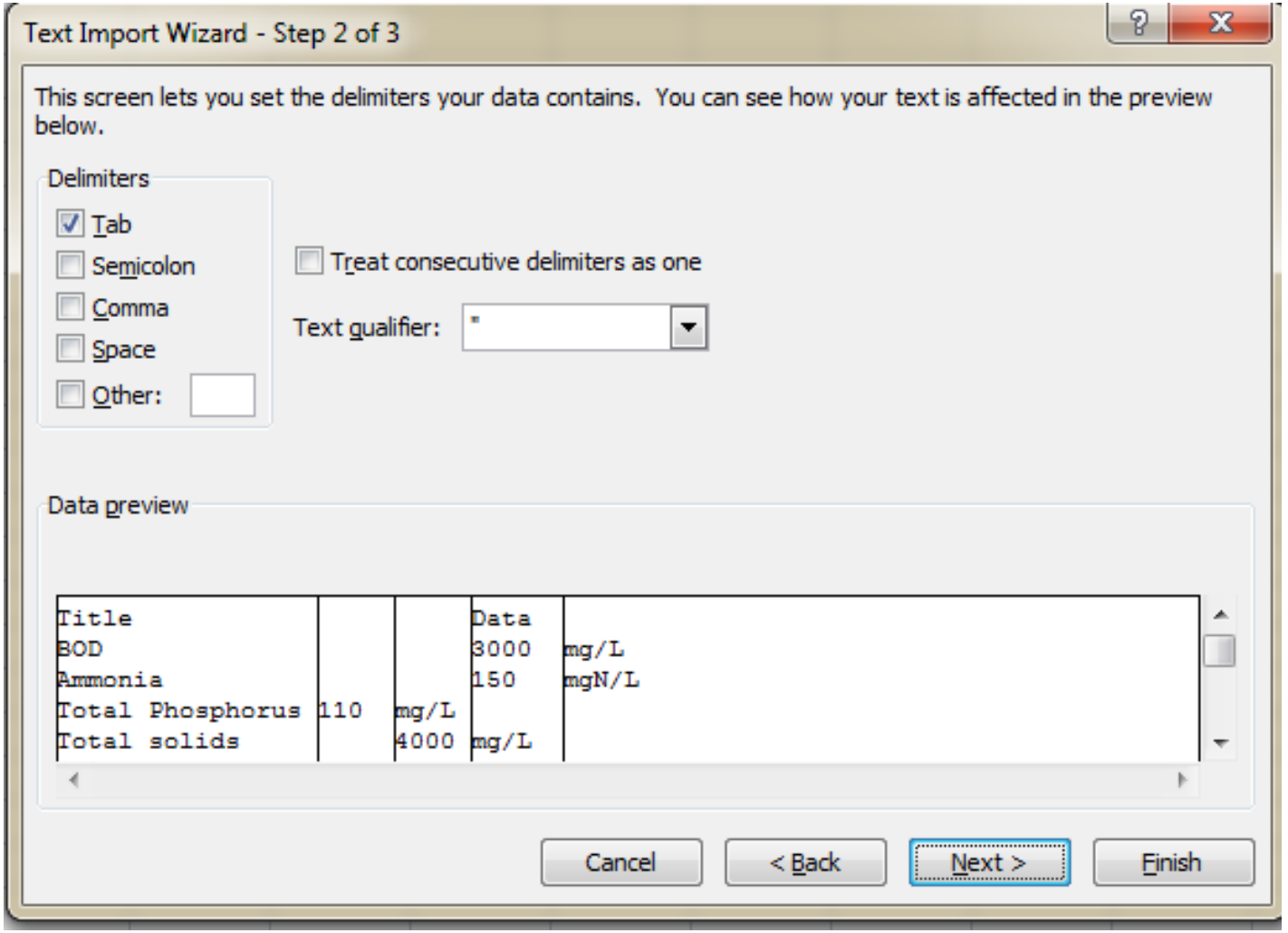
External Data Sources



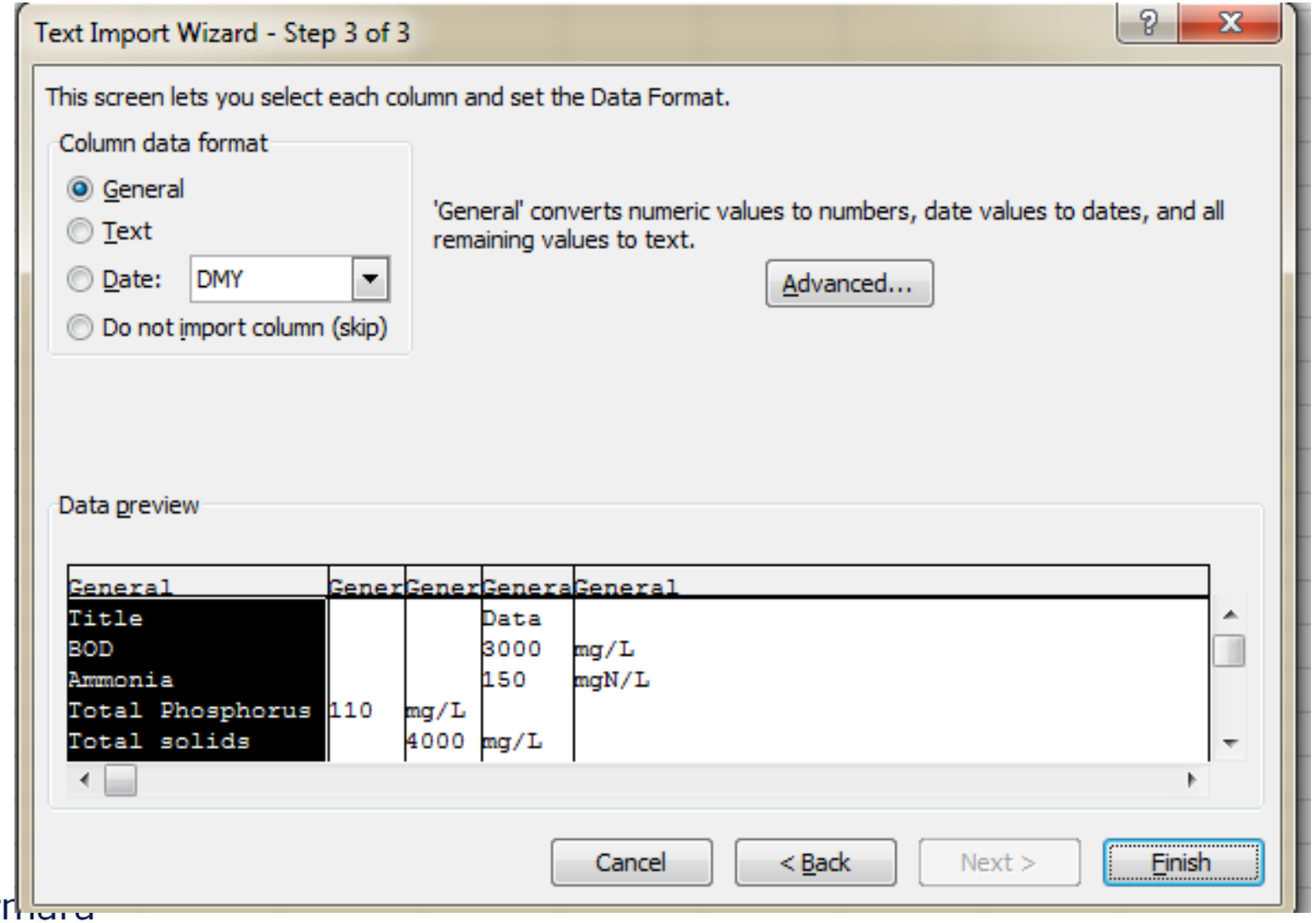


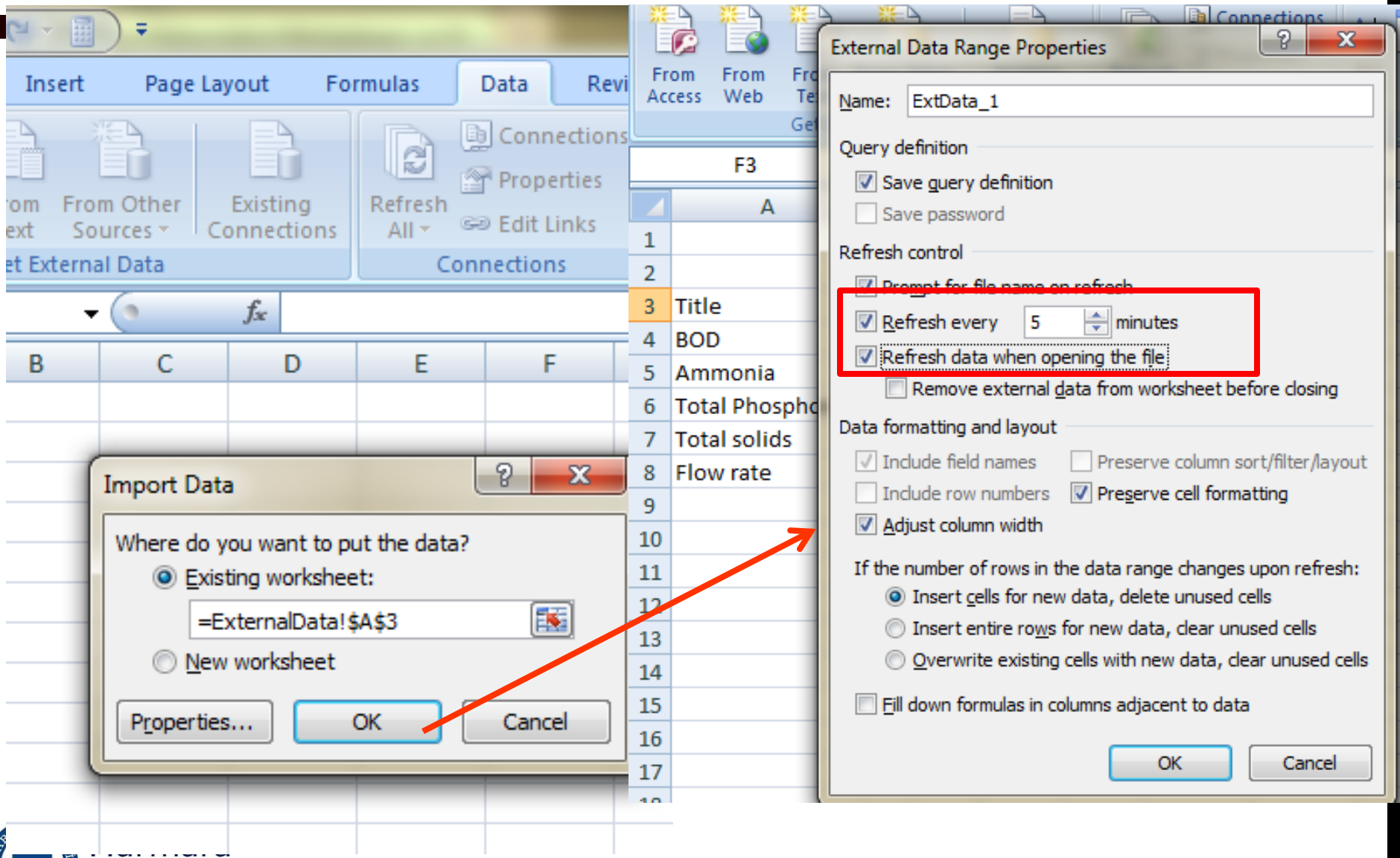
Marmara Üniversitesi

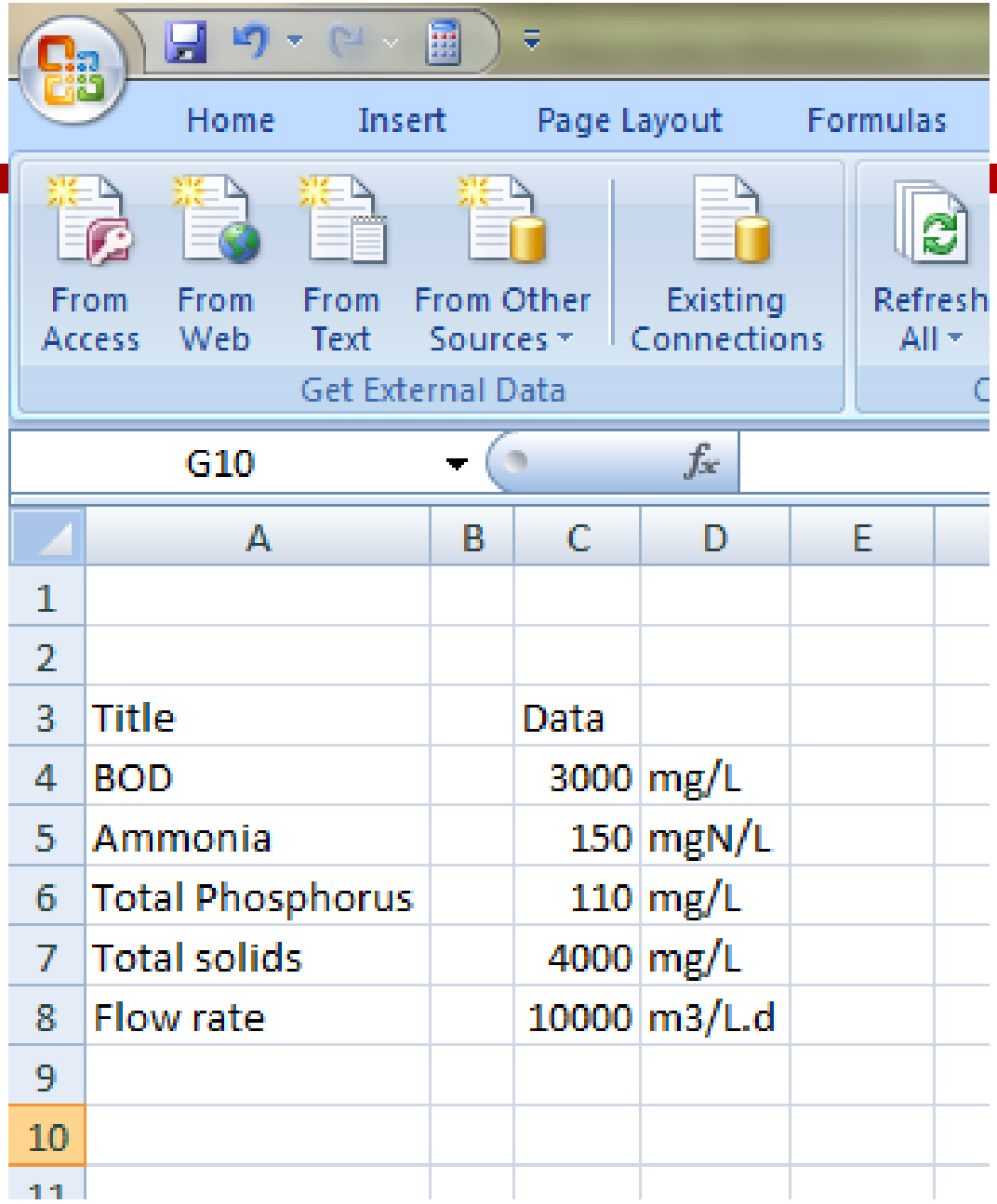
External Data



External Data




Insert
Page Layout
Formulas
Data
From Access
From Web
From Other Sources
Existing Connections
Refresh All
Connections
Properties
Edit Links
Get External Data
F3
Title
BOD
Ammonia
Total Phospho
Total solids
Flow rate
Import Data
Where do you want to put the data?
Existing worksheet:
=ExternalData!$A$3
New worksheet
Properties...
OK
Cancel
External Data Range Properties
Name: ExtData_1
Query definition
Save query definition
Save password
Refresh control
Refresh every 5 minutes
Refresh data when opening the file
Remove external data from worksheet before closing
Data formatting and layout
Include field names
Preserve column sort/filter/layout
Include row numbers
Preserve cell formatting
Adjust column width
If the number of rows in the data range changes upon refresh:
Insert cells for new data, delete unused cells
Insert entire rows for new data, clear unused cells
Overwrite existing cells with new data, clear unused cells
Fill down formulas in columns adjacent to data
OK
Cancel



	A	B	C	D	E
1					
2					
3	Title		Data		
4	BOD		3000	mg/L	
5	Ammonia		150	mgN/L	
6	Total Phosphorus		110	mg/L	
7	Total solids		4000	mg/L	
8	Flow rate		10000	m3/L.d	
9					
10					