

Prakarn Unachak Department of Computer Science Faculty of Science Chiang Mai University

Outline

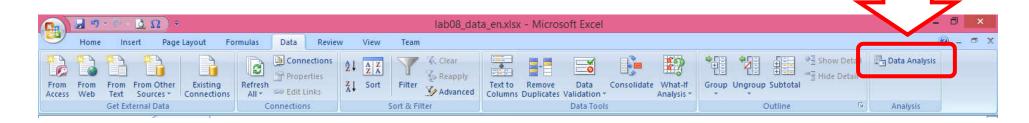
- Installing Data Analysis ToolPak
- Correlation
- Histogram
- What-If Analysis (Goal Seek)
- Linear Regression

Installing Data Analysis ToolPak

- 1. Check whether if *Data Analysis* command already appears under *Data* tab.
- If the command does not appear yet:
- 2. Click the office button.
- 3. Click *Excel Options*.
- 4. Select Add-ins.
- 5. Change Manage: option to *Excel Add-ins*.
- 6. Click Go.

Installing Data Analysis ToolPak

- 7. Under Add-Ins dialog box, make sure the Analysis ToolPak checkbox is checked.
- 8. Click OK (a few times).



Installing Data Analysis ToolPak (cont.)

N man	Recent Documents				-				Witnesser	NAMES OF THE OWNER		and a state		
New	1 lab08 data en.xisx	_⊨ 2 ↓	7 4	k Clear	*	-	3	è 🌇			P Sho		Data Analysis	
2	2 data mailmerge 02 en.xlsx	1000	Card Cillion	Reapply	Text to			olidate What-If		Ungroup Sub	THO	e Detan	~	
Open	3 lab07_scoresheet_fin.xlsx			Advanced	Columns D	uplicates Valio		Analysis *	1 . The				(1)	
	1 States and a state of States and	-	Sort & Filter			Da	ta Tools			Outlin	ne	19		
<u>S</u> ave	4 lab07_scoresheet_en.xlsx												$\mathbf{\vee}$	
	5 lab07_scoresheet_fin.xlsx	-	G	Н	I	J	К	la La la	М	N	0	Р	0	R
Save <u>A</u> s 🔸	<u>6</u> lab07_scoresheet_mid.xlsx	-1=1	Y2			-		-						
No. of the second se	7 lab07_scoresheet_mid.xlsx	-l= E-47	-1.700E+01											
Ъ.	8 lab07_scoresheet_en.xlsx	-j=i E-02	-1.662E+01											
Print >	9 lab06_data_mailmerge_en_example.xlsx	-i= -01	-1.554E+01											
NØ	lab06_data_mailmerge_en.xlsx	+00	-1.390E+01											
Prepare +	data_mailmerge_02_en.xlsx	-⊫ +00	-1.187E+01											
	data mailmerge en.xlsx	+00	-9.607E+00											
sen <u>d</u> >		+00	-7.189E+00											
<u> 8</u>		+01	-4.629E+00			1								
		+01	-1.899E+00											
Publish 🔸		+01	1.005E+00											
	-	+01	4.000E+00											
Close		+01	6.888E+00											
		+01	9.371E+00											
	Excel Options 🗙 I	Exit Excel +01	1.113E+01											
_0 0425	0.5878 -0.8000	-8 477F±00	1 100F±01											

Installing Data Analysis ToolPak (cont.)

	Excel	Options	?	×	
Popular Formulas	View and manage Microsoft Office	add-ins.		Add-Ins	? ×
Proofing	Add-ins			Add-Ins available:	
Save	Name Active Application Add-ins	Location	Туре	Analysis ToolPak Analysis ToolPak Analysis ToolPak - VBA	ОК
Customiza	Analysis ToolPak Team Foundation Add-in	C:\ice\Office12\Libran\/Analysis\ANALYS32.XLL "C:\ation Server\12.0\x86\TFSOfficeAdd-in.dll"	Excel Add-in COM Add-in	Co Col Sum Wizard	8
Add-Ins ITUST Center Reso	Inactive Application Add-ins Analysis ToolPak - VBA Conditional Sum Wizard Custom XML Data Euro Currency Tools Financial Symbol (Smart tag lists) Headers and Footers Hidden Rows and Columns Hidden Worksheets Internet Assistant VBA Invisible Content Lookup Wizard Person Name (Outlook e-mail recipients) Solver Add-in Document Related Add-ins	atpvbaen.xlam sumif.xlam C:\6}\Microsoft Office\Office12\OFFRHD.DLL eurotool.xlam C:\6\Microsoft Shared\Smart Tag\MOFL.DLL C:\6\Microsoft Office\Office12\OFFRHD.DLL C:\6\Microsoft Office\Office12\OFFRHD.DLL C:\6\Microsoft Office\Office12\Library\HTML.XLAM C:\6\Microsoft Office\Office12\Library\HTML.XLAM C:\6\Microsoft Office\Office12\OFFRHD.DLL lookup.xlam C:\s\Microsoft Shared\Smart Tag\FNAME.DLL solver.xlam	Excel Add-in Excel Add-in Document Inspector Excel Add-in Smart Tag Document Inspector Document Inspector Excel Add-in Document Inspector Excel Add-in Smart Tag Excel Add-in	Internet Assistant VBA Lookup Wizard Solver Add-in	Browse
	No Document Related Add-ins Disabled Application Add-ins Add-in: Analysis ToolPak Publisher: Microsoft Corporation	oft Office\Office12\Librany\Analysis\ANALYS32.XLL or statistical and engineering analysis		Analysis ToolPak Provides data analysis tools fo engineering analy	
		0			
	(5)	6	OK Cance		



CORRELATION

Correlation

- Correlation tells you how close two set of data related to each other. The value is between -1 and 1.
- If the two data sets have positive correlation, when one increase/decrease, the other will go to the same direction.
- If the two data sets have negative correlation when one increase/decrease, the other will go to the other direction.
- If the correlation is zero, there is no relation between two data sets. Whether the one decrease/increase will not predict the other.

Calculating Correlation

	Correlation	? ×	Į
Input	1	ОК	
Input Range: Grouped By:		Cancel	
Grouped by:	 <u>Columns</u> <u>Rows</u> 	Help	2
Labels in First Row			
Output options			3.
Output Range:			
New Worksheet Ply:			
O New Workbook			

- Select Data \rightarrow Data Analysis \rightarrow Correlation
- Select Input Range
- Select Options
 - Grouped By: Columns/Rows.
 - Check if first row of data is a label.
- 4. Select Output Range.

Correlation Results

	A	В	С	D
1	W	Х	Y	Z
2	1	1	10	1
3	2	1.5	9	1.5
4	3	2	8	2
5	4	2.5	7	2.5
6	5	3	6	3
7	6	3.5	5	3
8	7	4	4	3.5
9	8	4.5	3	4
10	9	5	2	4.5
11	10	5.5	1	5

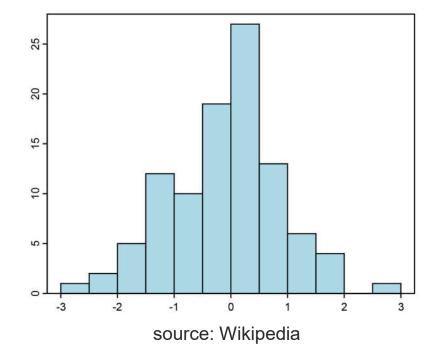
Correlation			?	×
Input Input Range	::	\$A\$1:\$D\$11	ОК	
Grouped By) <u>C</u> olumns Rows	Cancel	
Labels i		<u>Rows</u>	Help	
Output optio				
Output	Range: orksheet <u>P</u> ly:	\$E\$1	5	
E	F	G	Н	1
	F W	G X	H Y	l Z
		576	285	l Z
E	w	576	285	l Z
E	w	576	285	I Z



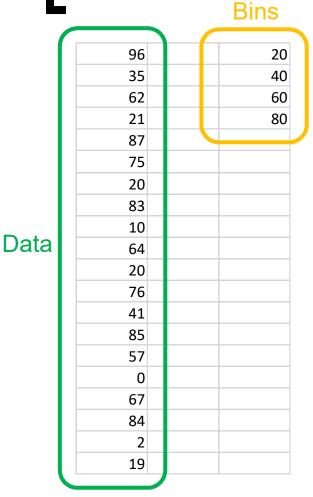
HISTOGRAM

Histogram

- A histogram is a way to display data by grouping each data point into ranges (bins) of their values
- Good for showing data distribution



Histogram (cont.)



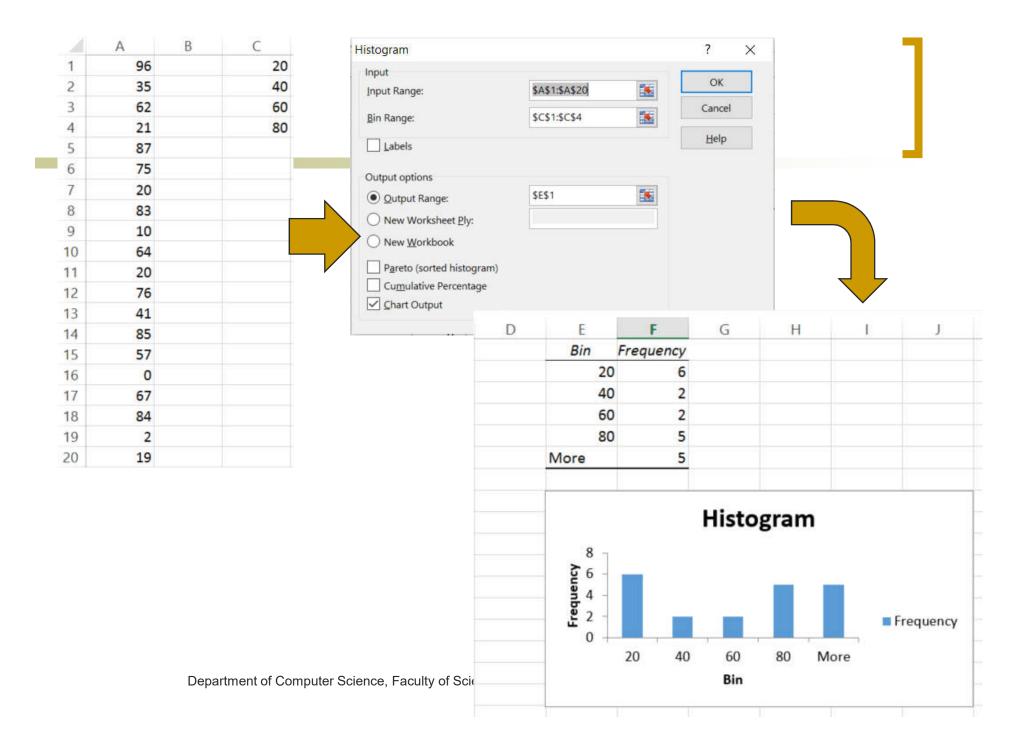
- To perform Histogram on Excel, you need data and bins.
- Bins are cells whose values used to determine which group a data point will go to.

Calculating Histogram

	Histogram		? ×
Input	1999		ОК
Input Range:	1		
Bin Range:		1	Cancel
Labels			Help
Output options Output Range: New Worksheet Ply: New Workbook Pareto (sorted histogram Cumulative Percentage Chart Output)		

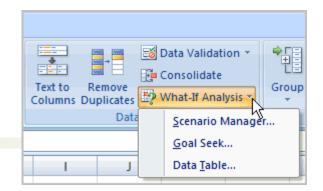
- 1. Select $Data \rightarrow Data$ Analysis \rightarrow Histogram
- 2. Select Input Range
- 3. Select Bin Range
- 4. Select Output Range
- 5. Check other options
 - Chart Output if you want charted version of histogram.

6. Click OK





WHAT-IF ANALYSIS (GOAL SEEK)

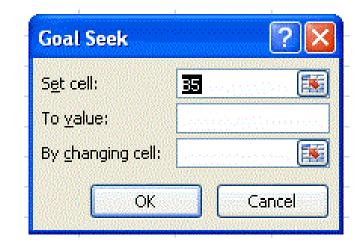


- Data \rightarrow Data Tools \rightarrow What-If Analysis
- Allow you to see the effect of different values on a (group of) formulas
- For this course, we will focus on Goal Seek
 - What should the input value be to get the result I need?

What-if Analysis

Goal Seek

- First, you need to set the formula on your spreadsheet
- Then you can use goal seek
 - Set cell: is the cell where the target output value is
 - To value: is the target output value
 - By changing cell: the input value



The answer will be here

Goal Seek Example

- You want to solve $x^2 16.5x + 35 = 0$
- Let Cell B1 be x. Set the value to 0.
- Let Cell A2:A4 contain the coefficients, 1, -16.5 and 35
- Let B2 = B1^2, B3 = B1, and B3 = 1
- Let C2 = A2*B2, C3 = A3*B3, C4 = A4*B4
- Let C5 = SUM(C2:C4)

Goal Seek (cont.) Start goal seek.

- - Set...
 - Set cell: to \$C\$5
 - To value: to 0

Goal Seek	? 🔀
S <u>e</u> t cell:	35 and a second iso
To <u>v</u> alue:	
By <u>c</u> hanging cell:	
ОК	Cancel

- And By changing cell: to \$B\$1
- Click OK. The answer should be at cell B1.
- You can try changing B2 to 20, and run goal seek again.



LINEAR REGRESSION

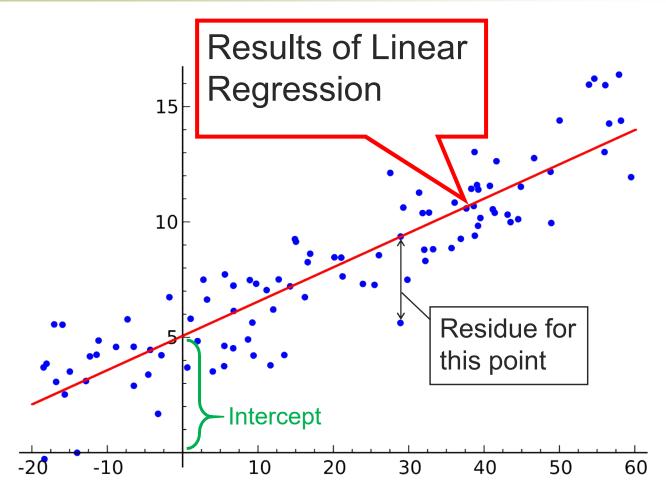
Linear Regression

- Linear regression is an analysis technique used to derive a relationship between dependent variable and one or more independent (explanatory) variables.
- Can be use both to explain data, and predict values of data in the future.

Linear Regression (cont.)

- Linear regression results will be in this form: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$
- Y is dependent variable.
- $X_1..X_n$ are independent or explanatory variables.
- β_0 is an intercept or constant values of the equations.
- $\beta_1 \dots \beta_n$ are coefficients of each independent variables.
- s is an error term or residue. This is not a constant.

Linear Regression, in Graph



Department of Computer Science, Faculty of Science, Chiang Mai University

Linear Regression (cont.)

	Regression	? ×
Input		ОК
Input <u>Y</u> Range:		
Input <u>X</u> Range:	1.	Cancel
Labels	Constant is Zero	Help
Confidence Level:	95 %	
Output options		
O Qutput Range:	ES.	
• New Worksheet Ply:		
O New Workbook		
Residuals		
Residuals	Residual Plots	
Standardized Residuals		

- 1. Select $Data \rightarrow Data$ Analysis \rightarrow Regression
- 2. Select Y (dependent variable) Range
- 3. Select X (independent variable) Range
- 4. Select Output Range.
- 5. Select options.
- 6. Click OK

		А	В	С	R	egression				?	×	-	
1	X1		X2	Y		Input					01/		
2	VI	0			-	Input <u>Y</u> Range:		\$C\$1:\$C\$2	1		ОК		
3		0.62	1.7		•	Input <u>X</u> Range:		\$A\$1:\$B\$21	1		Cancel		
4		1.47	3.01	11.036		-	_				<u>H</u> elp		
					-	Labels		Constant is <u>Z</u> er	0				
5		2.04				Con <u>f</u> idence	e Level: 9	5 %					
6		2.11	5.69	14.925		Output options							
7		2.16	6.74	16.087		Output options	ide:	\$E\$1					
8		2.19	8.28	17.725		New Works							
9		2.27	9.82	19.395		O New Work		1.1					
10		2.82	11	21.734		Residuals					1		
11		3.28	12.25	23.862		<u>R</u> esiduals			ual Plots				
12		3.45	13.3	25.272		Standardize	ed Residuals	Line F	it Plots				
13		4.13	14.45	27.809		Normal Probab	State of the second				:		
14		4.9	15.81	30.692		<u>N</u> ormal Pro	bability Plots				1		
15		5.33	16.81	32.505	-								
16		5.83	17.93	34.6	G SUMMARY OUTPUT	Н		J	-	L	Μ	N	0
17		6.76			SUMMART OUTPUT					7			
18		7.55			Regression St	tatistics							
					Multiple R	0.99999831							
19		8.36			R Square	0.99999661							
20		8.81	24.29	46.95	Adjusted R Square Standard Error	0.99999621 0.02599426							
21		9.69	25.83	50.238	Observations	20							
					ANOVA								
					ANOVA	df	SS	MS	F	Significance F			
					Regression	2	3391.344686			3.1883E-47			
					Residual	17	0.011486927	0.0006757					
					Total	19	3391.356173						
						Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Upper 95.0%
			Departme	nt of Compute	r Intercept	5.04734708	0.011855951						5.07236095
					X1	1.97983667	0.009738943		3.2E-30				2.00038405
					X2	1.00667981	0.003559745	282.795514	1.2E-32	0.9991694	1.01419021	0.99917	1.01419021

Reading the Results

G		Н	I.		J	K	L	М	N	Ο			
SUMMARY O	OUTPUT												
Regro	ession St	atistics		•	R-squ	are, a	djusted F	R-squar	e				
Multiple R		0.99999831			measi	ire ac	curacy of	fmodel	as				
R Square		0.99999661			a who	lo (Th	curacy of ne entire	onitatio	n				
Adjusted R S	Square	0.99999621				10. (11 r to 1 i		equalio					
Standard Eri	ror	0.02599426		•	Close		s better.						
Observation	าร	20											
ANOVA								• + (Stat D				
		df	SS		MS	F	Significance F	• l-v	 t-Stat, P-value measure accuracy 				
Regression		2	3391.34	4686	1695.67234	2509499	3.1883E-47	m					
Residual		17	0.01148	6927	0.0006757						-)		
Total		19	3391.35	6173				TO	r that	one			
								Va	ariable	2			
		Coefficients	Standard I	Error	t Stat	P-value	Lower 95%			-			
Intercept		5.04734708	0.01185	5951	425.722653	1.1E-35	5.02255521	• Th	ne hig	her the t-	-		
X1		1.97983667	0.00973	8943	203.290713	3.2E-30	1.9592893	C I	tat the	o hottor			
X2		1.00667981	0.00355	9745									
							• T	ne Iow	/er the P-	-			
Coofficiers	11-------------	a a a b a va	windala		interee	. 1							
Coefficien	it ior	r each variable + intercept						Va	alue, t	he better			
The regre	SSAC	lequation	n is Y	Y =	1 98*X	1 + 1	01*X2				۷1		
	0000	- oquutt		•	1100 //	· · ·							
+ 5.05													

Compare the Results

- You can then compute estimated
- $\hat{Y} = 1.98X_1 + 1.01X_2 + 5.05$

and compare the difference

	44.398
	46.95
Department of Computer Science, Faculty of Science, Chiang Mai University	50.238

Y	Y۸	diffY
5.049	5.05	0.02%
7.981	7.9946	0.17%
11.036	11.0007	0.32%
13.163	13.1898	0.20%
14.925	14.9747	0.33%
16.087	16.1342	0.29%
17.725	17.749	0.14%
19.395	19.4628	0.35%
21.734	21.7436	0.04%
23.862	23.9169	0.23%
25.272	25.314	0.17%
27.809	27.8219	0.05%
30.692	30.7201	0.09%
32.505	32.5815	0.24%
34.6	34.7027	0.30%
38.212	38.2712	0.15%
41.083	41.1888	0.26%
44.398	44.4591	0.14%
46.95	47.0267	0.16%
50.238	50.3245	0.17%