# How to get started.

This is a tutorial for you to work through at the computer at your own pace. Please make sure that you have a floppy disk which will also be supplied by Andrew. You can now proceed following the tutorial. If you have any questions or problems let Andrew know and he will help out.

### **1.** Switch on the PC.

The power button is on the right hand side of the PC System Unit at the front. "Booting Up" takes about 60 seconds.

After booting up the PC will display a log on box. Simply leave the password blank and hit enter.

### 2. Access Excel from Windows 95.

Excel is accessed either by double clicking on the Excel Icon or through the Start button / Programs options. Let me know if you have not used Windows 95 or Windows 3.1 before and I will happily give you a personalized introduction.

### 3. Disk Drives.

The LAN has several "logical disk drives". A: is the floppy disk as on your PC, but there are also C: through Z:

For this session, you will use the floppy disk supplied which contains the data files that you will need to follow the tutorials. When opening or saving files make sure that drive A: is selected.

# Using EXCEL for Statistical Analysis, a brief Tutorial.

I am making the assumption that you followed the general instructions and that you are now sitting in front of Windows 95 or Windows NT

If not, get me to come over.

So in this session we are going to:-

- 1. Learn how to install the Data Analysis Tools available in Excel.
- 2. Learn how to import data in other formats.
- 3. Learn how to download data from the World Wide Web.
- 4. Do correlation analysis on data on your floppy disk.
- 5. Perform a linear regression on data on your floppy disk.
- 6. Perform a multiple regression on data on your floppy disk.
- 7. If time allows, consider how to use the other statistical functions that are available.

If Excel is completely new to you, let me know and I will come over and give you an introduction.

# 1. Install the Data Analysis Tools available in Excel.

The following works for Excel version 5 and version 7. If you have an earlier version then the following may work, but I cannot guarantee it. On the SOM LAN Excel has the Tools installed, but for your own machine these are the procedures. Try them on the SOM LAN, anyway.

🗙 Microsoft E	xcel - Book1											_ 8 ×
🕙 <u>F</u> ile <u>E</u> dit	⊻iew <u>I</u> nserl	t F <u>o</u> rmat	<u>T</u> ools <u>D</u> ata	<u>W</u> indow	<u>H</u> elp							_ 8 ×
Arial	ائل ا ت	<mark>∦ ₪</mark>	<u>S</u> pelling Audi <u>t</u> ing Loo <u>k</u> Up F	F7 }eference	,   _	<u>2</u> ↓ Z↓ 3 \$6 9	∭ <b>∭</b> ∦,   %	<b></b> 8 <b>,</b> :::  <u> </u>	- - -	<b>₹</b>		
A1	•		<u>A</u> utoCorre	ot	1							
A 1 2 3	B	С	<u>G</u> oal Seek S <u>c</u> enarios Sol <u>v</u> er			F	G	H	1	J	К	<u>*</u>
4			<u>P</u> rotection		<b>₽</b> ¢							
5			Add-Ins		_							
6 7 8 9			<u>M</u> acro <u>R</u> ecord M Assi <u>gn</u> Ma	acro	•							
10												
11												
12												
13												
14												
15												
10												
18												
19												
20												
21												
22												
23												
III I P P Sh	eet1 / Sheet	2 / Sheet3	/ Sheet4 / S	heet5 / Sh	eet6 /	Sheet7 /	Sheet8	•				
Installs or unit	nstalls add-i	ns										
🚮 Start 🔍 🔍	Exploring - E:\M	18   <b>137</b> M	licrosoft Word -	🛛 🔀 Mie	crosof	t Exc	🕜 Printir	ng page 2	1		4:00	∮ 9:35 AM

Select the "Tools" Menu and the "Add-Ins" Option as shown below.

If the last item on the tools menu is "Data Analysis", then the tools are already installed.

Set the check boxes

- Analysis ToolPak,
- Analysis ToolPak VBA

as shown on the following page.

Microsoft Excel - Book1	
Arial	
A1 💌	Add-Ins
A     B     C       1	Add-Ins Available:   I   J   K   L     ✓ Analysis ToolPak   Cancel   I<
14	
16	
17	
19	
20	
21	
22	
21	
For Help on dialog acttings	3 / Sheet4 / Sheet5 / Sheet5 / Sheet7 / Sheet8   ◀
ror neip on dialog settings, press	
📴 Start   🔍 Exploring - E:\MB   👿	Microsoft Word   📉 Microsoft Exc 🕼 Printing page 2   🔤 9:35 AM

Then press OK.

At home, Excel may ask you to insert disks from the Excel or Office disks that you used to load Excel in the first place, so have these handy.

Now when you select the Tools pull down menu, the last item in the menu will be Data Analysis as you can see on the following page.

Microsoft Excel - Book1 File Edit View Insert Format	Tools Data Window H	Alline Constant Const
	Spelling F7 Auditing F Look Up Reference	
	- ·	
AI	AutoLorrect	
A B C 1 2 3	<u>G</u> oal Seek S <u>o</u> enarios Sol <u>v</u> er	
4 5	Protection ► Add-Ins	
6 7 8 9	Macro <u>B</u> ecord Macro Assi <u>gn</u> Macro	
10	Options	
11	Data Analysis	
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
H + H Sheet1 / Sheet2 / Sheet3	/ Sheet4 / Sheet5 / Sheet	t6 / Sheet7 / Sheet8   ∢
Select an analytic procedure		
🔀 Start 🔍 Exploring - E:\MB 🕅 🕅	ficrosoft Word 🔀 Micro	osoft Exc 🕜 Printing page 3

### 2. Import Data for analysis from files on the floppy disk.

The data that you are trying to analyze could come from a number of different sources. You may down load data from a government Web Site or from your company's accounting system or from a database package. Whichever source you select you will want to be able to import the data into Excel.

To import a file you simply open the file from within Excel using the File Open options on the pull down menus. You will then have the "open" window in front of you. Change the "Files of Type" to the type of file you want to import and then select your file ... in this case MULTIPLE.csv. "csv" stands for comma separated variables which describes the way in which the data is stored.

<b>∑</b> <i>Microsoft</i> E अ⊒ File Edit	xcel - Book2 View Insett	Format	Tools Data	Window	Help						
	a R V	∦ 🗈			f <sub>*</sub>	↓  <u>  % </u>  ⊗	4	▼ 3	<mark></mark>	1	
Arial					≕⊟ ∎⊟∎					-	
			BIIU		= [23] <b>-</b>	70 9 .00					
J24			D	F	- F					I K	
A	D	L	U	E	F	6	п			J N	<u>^</u>
2											
3	Оре	n								?	×
4	Loo	kin: 🚺	Excel		•		병 📰 🗰				
5	-		_		lana.			6-0 6-0 ·			-
5		MULTIPLE Drob14w7 c	.CSV							<u>U</u> pen	J
8	<u></u>	F1001407.0	.sv							Cancel	
9										Aduppood	
10											J
11										I <u>R</u> ead Only	
12											
13											
14											
15											
17											
18	Find	i files that m	atch these crit	ena:		T				The d Marco	1
19	File	name:				rext or proper	rty: ]			Find Now	
20	Files	s of type: 🚺	ext Files (*.pm	: *.txt; *.csv)	-	Last modified:	any time		•	Ne <u>w</u> Search	
21	2.6	e(s) found									- III
22					1	1		1	1		
23											
Image: Sheet1 / Sheet2 / Sheet3 / Sheet5 / Sheet5 / Sheet6 / Sheet7 / Sheet8   Image: Sheet3 / Sheet3 / Sheet4 / Sheet5 / Sheet5 / Sheet3 / Sheet3   Image: Sheet3 / S											
For Help on dialog settings, press F1  Sum=0											
🚮 Start 🛛 🕅	🕱 Start 🕎 Microsoft Word - Statistics 🔀 Microsoft Excel - Book2										

Click on Open when you have highlighted the file that you want or double click on the filename. In the case of the "multiple.csv" file the data is read without any problems by Excel. In other cases a "Wizard" may prompt you with questions which will allow it to put the data into sensible columns.

There are a number of different formats which are commonly used to store data and pass it around and you may encounter:

- Text files where columns of data are separated by commas or tabs or surrounded by quotation marks or with fixed column widths (fixed record length).
- Dbase files from DBaseII, III, or IV or from FoxPro or Clipper. (Usually named as \*.dbf).

To practice, import multiple.csv; multiple.dbf and multiple.txt from the floppy disk.

### 3. Getting Data From a Web Site.

The SOM PCLabs have been undergoing a major upgrade this summer and when I tried this yesterday Netscape was not working on the PC that I tried so you may have to try this from at home.

Invoke Netscape or your favorite browser and call up the following page:

http://www-unix.oit.umass.edu/~adhall/andrew.htm

Double click on "Multiple.csv" and you should see the following:

🙀 Netscape - [http://www-unix.oit.umass.edu/~adhall/MULTIPLE.csv]	_ 8 ×
<u>File Edit View Go Bookmarks Options Directory Window H</u> elp	and the second second
Image: Second	
Location: http://www-unix.oit.umass.edu/~adhall/MULTIPLE.csv	
What's New? What's Cool? Destinations Net Search People Software	
House_No,Selling_Pr,House_Size,House_Age,Location,OwnersSal 1,104307,1737,11,3,2500 2,196519,2744,10,5,40000 3,135515,1894,7,2,120000 4,224496,3311,8,3,100000 5,154898,2334,14,3,55000 6,144957,1816,4,5,45000 7,205938,2584,5,4,23000 8,173194,2013,5,3,42000 9,162475,2404,8,0,27000 10,113892,1573,4,0,29500 11,107622,1919,12,0,43500 12,221972,2698,0,4,54000 13,159906,2788,10,1,29500 14,120049,1612,12,2,57700 15,191532,2725,4,5,47500 16,126075,2005,12,1,35500 17,156356,2182,15,0,42500 18,141735,1973,5,0,35600 19,159674,2512,14,2,120000 20,150529,2536,12,3,150000 21,268092,3417,2,4,37500 22,184895,2290,5,3,52000	
Document: Done	<b>⊥</b> □ ⊠?

Go to the File pull down and select Save as.

Fill in the save as options as follows:

Save As					?	х
Save jn:	Sty Floppy (A: )	-	£	Ċ	8-6- 5-5- 6-6-	
900110962	2.txt					
🚵 fred.html						1
mgtss.html						
	CSV Cdbf					
	txt					
I						
File <u>n</u> ame:	MULTIPLE.csv				<u>S</u> ave	
Save as <u>type</u> :	All Files (*.*)		•		Cancel	

and select save using the  $\underline{S}$  ave button.

Netscape will save a copy of the file on your floppy disk.

Even if what you saw on the screen was meaningless, when you save it, it may nevertheless be meaningful. A good example is the Multiple.dbf file which cannot be read by Netscape's editor but when saved will be readable by Excel.

# 4. Do correlation analysis on data on the floppy disk.

First load up the Worksheet corel.wks from the floppy disk in the data directory A:\data

I have named two ranges to make selection of ranges easier. These are:-

Named Ranges	The Range
All_Data	=Business_Failures!\$A\$1:\$H\$14
All_Numeric_Data	=Business_Failures!\$B\$1:\$H\$14

Then you need to access the Data Analysis tools which you do using Tools and the last option in the pull down menu. As you can see in the following panel.

X /	Aicresoft	Excel - (	Corel.xls	Taala Data	At Good and and	l e le								
	<u>r</u> ne <u>E</u> d		Insert Format		<u>w</u> indow <u>F</u>									
			V 🔨 🖁 🖻	Auditina	•		<u>iii 🧶 🚽</u>	100%	≤ <u>\$</u> <b>\?</b>					
Aria	1		▼ 10 ▼	Loo <u>k</u> Up F	eference	<b>5 5</b>	<b>%</b> , <b>*</b> .0	※ = - (소	T					
	D16	-		AutoCorrect	:t									
		A	B	Ecol Sook		E	F	G	Н	1	J			
1	Year		Bus_Fail F	Scenarios		hemploy	Ret_Sales	House_Sts	C_P_Index					
2		1968	0.3	Solver		3.8	6.8	22.0	3.6	i 📃				
3		1969	-5.1	001_01	1997 (C. 1997)	4.5	6.0	20.0	4.0	)				
4		1970	24.3	<u>P</u> rotection	+	4.4	6.9	6.9	4.6	i				
5		1971	4.0	Add-Ins		5.7	2.3	-9.5	3.3	}				
6		1972	1.2	Macro		6.2	8.9	22.6	2.9	)				
7		1973	-4.8	Becord Ma		6.2	11.3	7.0	4.8	}				
8		1974	-4.9	Assign Ma	210	5.5	12.6	7.4	7.6	i				
9		1975	5.1	1.001311 11.0		5.3	16.9	-17.3	10.8	}				
10		1976	-4.9	Options		6.9	14.5	4.2	10.8	}				
11		1977	40.0	<u>D</u> ata Analy	isis	7.1	10.8	-7.3	7.5	5				
12		1978	40.4	2.8	10.0	8.1	8.4	-10.1	8.0	)				
13		1979	2.7	3.5	8.8	8.4	11.7	-7.3	9.0	)				
14		1980	15.8	3.2	11.0	0.5	11.9	-13.4	8.8	}				
15														
16														
17														
18														
19														
20														
21														
22														
23														
		Business	<b>Failures</b> Rar	nge_Names / S	heet3 🖌 Sheet	4 / Sheet5 ,	/ Sheet6							
Sel	ect an a	nalytic p	rocedure											
	🚯 Start 💓 Microsoft Word - Section 2 🔯 Microsoft Excel - Cor 🤄 🖗 11:13 AM													

Now select Correlation as on the following screen:-

X/ 80	<i>ficrosoft</i> Exe Filo Edit A	cel - C <i>licu</i>	Corel.xls	ormat	Toola Da	ta Mindau	Help									
	cne <u>c</u> un y		Insert F	unnau Les L	<u>1001s D</u> a								느미스			
		فالا	. <b>↓</b>			ο Γαι Σ		<u>k</u> 🧕 🚽	100%							
Aria			• 10	-	BIU		E 🗃 🕏 🤅	<b>∕, , </b> .00 ,	※ == (1	- To -						
	D16	-														
	А		В		С	D	E	F	G	Н		I	J			
1	Year		Bus_Fai	I R	eal_D_P	Wage_Sals	Unemploy	Ret_Sales	House_Sts	C_P_Index						
2		1968		0.3	3.5	10.9	3.8	6.8	22.0		3.6					
3		1969	-	5.1	5.5	8.7	4.5	6.0	20.0		4.0					
4		1970	2	4.3	6.2	12.1	4.4	6.9	6.9		4.6					
5		1971		4.0	2.2	8.4	5.7	2.3	-9.5		3.3					
7		1972		1.2	5.9	9.9	2.0 6.2	0.9	22.0		4.9					
8		1974		4.0 .4 9	7.8	15.8	5.5	12.6	7.0		7.6					
9		1975		5.1	4.7	19.4	5.3	16.9	-17.3		10.8					
10		1976	-	4.9	0.7	16.4	6.9	14.5	4.2		10.8					
11		1977	4	0.0	5.3	15.4	7.1	10.8	-7.3		7.5					
12		1978	4	0.4	2.8	10.0	Data Ana	ysis					×			
13		1979		2.7	3.5	8.8	Analysis 1	l ools	Borte de la de la Bo	202. 2015	ya 🗠	. 2912 199.				
14		1980	1	5.8	3.2	11.0	Anova: S	ingle Eactor		1993) 1993		OK				
15							Anova: T	wo-Factor With	Replication			Cancel				
16							Anova: T	wo-Factor With	out Replication							
1/							Covarian	ce				<u>H</u> elp				
10							Descripti	ve Statistics								
20							F-Test T	wo-Sample for V	ariances							
21							Fourier A	Fourier Analysis								
22																
23																
24	▶ ▶ Busi	iness_	Failures /	Rang	ge_Names /	Sheet3 / She	et4 / Sheet5 ,	Sheet6					• • •			
For	Help on dia	alog se	ettings, p	ress F	1			Sum=0	)							
	Start W Microsoft Word - Section2 K Microsoft Excel - Cor															

X / S)	<i>flicrosof</i> File Ec	Excel-( it View	Corel.xl Insert	s Forma	t Tools	Data	Window	Help					_ 8 ×
	   🚅   📮	1 <i>4</i> 0	HBC	X 🖻				f≈ A↓ Z↓	<b>I S S</b>	100%	- 😨 📢		
Aria			- 10	) 🔻	BI	<b>v</b>			×	이 프 -			
	B20	-						سلسبي لتتقلب		السر استحصب المت			
		A	E	3	С		D	E	F	G	Н	1	J
1	Year		Bus_F	ail	Real_D_F	, ∖v	Vage_Sals	Unemploy	Ret_Sales	House_Sts	C_P_Index		
2		1968		0.3	3	3.5	10.9	3.8	6.8	22.0		3.6	
3		1969		-5.1	6	5.5	8.7	4.5	6.0	20.0		4.0	
4		1970		24.3	6	6.2	12.1	4.4	6.9	6.9		4.6	
5		1971		4.0	2	2.2	8.4	5.7	2.3	-9.5		3.3	
6		1972		1.2	6	5.9	9.9	Correlation	00			20	a –
1		1973		-4.8	÷	5.6 7.0	11.1	Jacob	ar a ar ar ig a b b	a da anti-	10 - 10 - 10 - 10 - 10 - 10 - 10 - 10 -	and a state of the	
8		1974		-4.9		.8	15.8	Input Danger		All Numerie	Data	OK	
9		1975		0.1 4 Q		+.7 1 7	19.4	inputriange.		Par_Numeric_		Cancel	
11		1970		40.0		33	15.	Grouped By:		<u> C</u> olumns			
12		1978		40.0		28	10.1			C <u>R</u> ows		<u>H</u> elp	
13		1979		27		15	81	🔽 Labels in	First Row				
14		1980		15.8		3.2	11.0	-Output option	。 				
15								C Output D					
16									anye.				
17								• New Wor	iksheet <u>Ply</u> :	Juorrei			
18						_		C New <u>W</u> or	rkbook				
19													
20						_							
21						-							
23													
21													
Ent	er	D 02111622	_i anule	<b>.</b> • \ • •	inge_ivante:	×	meeto X one	ette X prieeto ;	Sum=0	)			
	Start W Microsoft Word - Section2 K Microsoft Excel - Cor												

I suggest that you complete the options as shown above. Using named ranges makes things easier. If you do more correlations in this file and you want to keep the results then you should give a different "New Worksheet Ply" each time.

# If you want to look at a couple of columns, say Bus\_Fail and Unemploy, then you have to move them so that they are side by side as Excel will not allow you to define a split range in this instance.

#### The following is from Excel's help.

Correlation measures the relationship between two data sets that are scaled to be independent of the unit of measurement. The population correlation calculation returns the covariance of two data sets divided by the product of their standard deviations.

You can use the Correlation tool to determine whether two data sets move together; that is, whether large values of one set are associated with large values of the other (positive correlation), whether small values of one set are associated with large values of the other (negative correlation), or whether values in both sets are unrelated (correlation near zero).

#### Input Range

Type the reference for the range of worksheet data you want to analyze. The input range should contain two or more blocks of data arranged in columns or rows. If your input range includes row or column labels, you must select the Labels In First Column or Labels In First Row check box, or Microsoft Excel displays a message.

#### **Output Range**

Type the reference for the upper-left cell of the output table. Only half of the table is complete because correlation between two data sets is independent of the order in which the data sets are processed. Cells in the output table with matching row and column coordinates contain the value 1 because each data set correlates perfectly with itself. The output table is a square whose width and height are one larger than the number of cell ranges of data. If existing data is about to be overwritten in the output range, Microsoft Excel displays a message.

#### **New Worksheet Ply**

This option inserts a new ply into the workbook the current ply resides in and pastes the results into cell A1 of the new ply. Use the text box next to New Worksheet Ply to optionally name the new ply.

#### New Workbook

This option creates a new workbook, adds a new workbook ply in the new workbook, and pastes the results into cell A1 of the new ply.

#### **Grouped By**

Select either the Rows or Columns option button to indicate whether the data in the input range is arranged in rows or in columns.

#### Labels In First Row/Labels In First Column

If you select the Columns option button under Grouped By and the first row of your input range contains labels, select the Labels In First Row check box.

If you select the Rows options button under Grouped By and the first column of your input range contains labels, select the Labels In First Columns check box.

If your input range does not include labels, clear the Labels In First Row or Labels In First Column check box. Microsoft Excel then generates the appropriate data labels for the output table (Row 1, Row 2, Row 3, and so on; or Column 1, Column 2, Column 3, and so on).

# 5. Perform a linear regression on data on the floppy disk.

First load up the Worksheet simple.xls from the floppy disk in the data directory A:\data

Then you need to access the Data Analysis tools which you do using Tools and the last option in the pull down menu. As you can see in the following panel.

×л	flicresoft E:	xcel - Simple.xls										_ 8 ×
*	<u>F</u> ile <u>E</u> dit	⊻iew Insert Fo	irmat <u>T</u> ools	<u>D</u> ata <u>W</u> indow	<u>H</u> elp							_ 8 ×
		<b>€ <u></u>] 10</b>	Bpe Aud	lling F7 iting k Un Beference	,		<u>× v</u>	<u>م</u> ا 100%				
Pana	1		<u> </u>			<b>D</b> 70	<b>,</b> .0	0 <b>→</b> .0 →				
	G8	<b>_</b>	Auto	Correct	1					_	-	
	A	B	<u> <u> </u></u>	l Seek	E	F		G	H		J	<u> </u>
1	VVeek	Overhead L	.abor_ S <u>c</u> e	narios	64 —							
2		76890	Soly	er								
3		94001	Prot	ection								
4		77643	bb4	-Ins	<u> </u>							
6	5	84703		1								
7	6	77328	<u></u> ac	xro	-							
8	7	78516	<u>R</u> ec	ord Macro			Ē					
9	8	64865	A.ssi	gn Macro								
10	9	84365	<u>O</u> pti	ons								
11	10	70247	<u>D</u> ata	a Analysis								
12	11	62261	675	216								
13	12	73207	932	289								
14	13	79742	1325	377								
15	14	93115	1348	454								
16	15	57579	618	236								
17	16	90639	974	454								
18	17	88547	840	485								
19	18	68113	1124	207								
20	19	54160	565	222								
21	20	60144	552 1000	428								
22	21	69144	640	∠44 AEA								
23	22	. 02000 . <u></u>	1403	404								<b>•</b>
	I F F SIN	PLE Range Na	imes /					•				
Sel	ect an ana	lytic procedure										
	Start 😗 🕅	licrosoft Word - Sec	tio34 🔀 M	icrosoft Excel -	5im							🕅 11:28 AM

Now select a Regression Analysis ...

<u>×</u> /	<i>Microsoft</i> Ex	cel - Simple.xl	S								_ B ×
8	<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>I</u> nsert F	F <u>o</u> rmat <u>T</u> ools	<u>D</u> ata <u>W</u> indow	<u>H</u> elp						_ 8 ×
R		3 B 189 S			£ A	z     s   a	) 🚚 🔤				
					/* <u>2</u> *			°	8.4.		
Aria	al l	▼ 10	▼ B <i>I</i>	<u> </u>	클 🔤 🧕 9	<b>β%</b> ,	◆.0 .00 ↓ ↓ 00. ◆ 00.	<b>⊥</b> • (∆)•	7 -		
	G8	-									
	A	B	С	D	E	F	G	Н	I	J	К
1	Week	Overhead	Labor_Hrs	Machine_Hrs							
2	1	76890	1266	339							
3	2	75671	668	468							
4	3	84001	1474	264							
5	4	77643	1201	Regress	ion	1		1		× I	
6	5	84703	1028	megness			///	147	<u>19</u>		
7	6	77328	1209	Input						К	
8	7	78516	618	Input )	<u>r'</u> Range:		Overhead				
9	8	64865	721	< Incut?	/ Dongo:		Labor Lin		Ca	ncel	
10	9	84365	850	mpurz	<u>n anye.</u>		JLabor_Hrs				
11	10	70247	841		abels		° `onstant is Zer	0	<u><u> </u></u>	elp	
12	11	62261	675				.onstancis <u>∠</u> ci	о 			
13	12	73207	932		on <u>fi</u> dence Lev	vei lap	~				
14	13	79742	1325	-Output	options						
15	14	93115	1348	- Capat							
16	15	57579	618		itput Hange:						
17	16	90639	974	🕘 🔍 Ne	ew Workshee	t <u>Ply:</u>	OverheadLa	bor			
18	17	88547	840	O Ne	ew Workbook						
19	18	68113	1124	Beside	Jals						
20	19	54160	565	E Br	esiduals		🔽 Besidu	al Plots			
21	20	71695	552		andardized Bi	esiduals	I line Fit	Plots			
22	21	69144	1096								
23	22	62065	640	Norma	al Probability						
24		IPLE / Range N	lames /		ormal Probabil	lity Plots					
Ent	ter										
	Start 🕎 M	icrosoft Word - Se	ectio34 🔀 🕷	licrosoft Excel	- Sim					4	💔 11:30 AM

Now complete the form.

I suggest that you complete the options as shown above. I have named ranges Overhead, Labor\_Hrs and Machine\_Hrs which makes it easier to enter ranges. When you do the second regression of Overhead against Machine\_Hrs remember to use a different Worksheet name to the one given above.

You will need to change column widths and so on to really see what you have as a result.

The resulting equation is Total Overhead = \$49,615 + \$28.60 for every labor hour.

If you run Overhead against Machine\_Hrs you should get:

Total Overhead = \$48,783 + \$82.36 for every machine hour.

The following is from Excel's help.

#### Input Y Range

Type the reference for the range of dependent data you want to analyze. The dependent data should be typed in a single column. If you include a report label in the first row of the input y range, you must select the Labels check box, or Microsoft Excel displays a message.

#### Input X Range

Type the reference for the range of independent data you want to analyze. The data should be typed in a single column for a simple analysis.

#### **Constant Is Zero**

Select the Constant Is Zero check box to force the regression line to pass through the origin.

#### Labels

If the first row in the input range includes report labels, select the Labels check box. If your input range does not contain labels, clear the Labels check box. Microsoft Excel then generates the appropriate data labels for the summary output table.

#### **Confidence** Level

Select the Confidence Level check box if you want an additional level included in the summary output table. In the Confidence Level box, type an additional confidence level that you want Microsoft Excel to apply to the regression in addition to the default 95% confidence level.

#### **Output Range**

Type the reference for the upper-left cell of the range where you want the summary output table to appear. Allow at least seven columns for the summary output table.

The summary output table includes the following:

- Anova table
- Coefficients
- Standard error of y estimate
- r2 values
- Number of observations
- Standard error of coefficients

#### New Worksheet Ply

This option inserts a new ply into the workbook the current ply resides in and pastes the results into cell A1 of the new ply. Use the box next to New Worksheet Ply to optionally name the new ply.

#### New Workbook

This option creates a new workbook, adds a new workbook ply in the new workbook, and pastes the results into cell A1 of the new ply.

#### Residuals

Select the Residuals check box if you want to include residuals in the residuals output table.

#### **Standardized Residuals**

Select the Standardized Residuals check box if you want to include standardized residuals in the output table.

#### **Residual Plots**

Select the Residual Plots check box if you want Microsoft Excel to generate a chart for each independent variable versus the residual.

#### Line Fit Plot

Select the Line Fit Plot check box if you want Microsoft Excel to generate a chart for predicted values versus the observed values.

#### Normal Probability Plot

Select the Normal Probability Plot check box if you want Microsoft Excel to generate a chart plotting normal probability plots.

### 6. Perform a multiple regression on data on the floppy disk.

First load up the Worksheet multiple.xls from the floppy disk in the data directory A:\data

In Multiple.xls, I have also named ranges as follows:-

All_Independent	=MULTIPLE!\$C\$1:\$F\$36
Selling_Pr	=MULTIPLE!\$B\$1:\$B\$36

The main difference between simple and multiple regression analysis is that multiple regression involves more than one Independent Variable and hence you can specify more columns for the Independent Variable range. Up to 16 are available.

📉 М	licrosoft Exc	el - Multiple.xls									
🎦 <u>E</u>	jile <u>E</u> dit <u>V</u> i	ew <u>I</u> nsert F <u>o</u> rma	at <u>T</u> ools <u>D</u> ata	<u>W</u> indow <u>H</u> elp							۶×
Ы	alol a			$\sum f_x A$	1 Z I I S	1 🚳 🛺   100	)% 🔻	<b>⊘ №</b>			
								8			
Arial		▼ 10 ▼	B I U	트 프 프 프	\$ %	••0 •00 ••• 00 ••0	<u> </u>	• 🛅 •			
	A2	SUI	MMARY OUTP	JT							
	A	В	C	D	E	F	G	Н	Ι	J	
2	1	104307	1737	11	3	25000					
3	2	196519	2744	10	5	40000					
4	3	135515	1894	7	2	120000					
5	4	224496	3311	8	3	100000					
6	5	154898	2334	Personaian		55000					
7	6	144957	1816	negression	447.446	ac. 557 10 5 10	697 - 316 316	147 - M.C.			
8	7	205938	2584	-Input					ОК		
9	8	173194	2013	Input <u>Y</u> Range:		Selling_Pr					
10	9	162475	2404			A.H. 1. 1	l		ancel		
11	10	113892	1573	Input <u>A</u> Hange:		All_Indeper	ndent				
12	11	107622	1919	V Labels	г	- Constant is Ze	20		<u>H</u> elp		
13	12	221972	2698				310				
14	13	159906	2788	I Lontidence I	_evel	35 %		- 61			
15	14	120049	1612	-Output options							
16	15	191532	2725					-			
17	16	126075	2005	O <u>O</u> utput Rang	e:			_			
18	17	156356	2182	💿 New Worksh	ieet <u>Ply</u> :	Housing		1			
19	18	141735	1973	C New Workbo	ook						
20	19	159674	2512	Besiduals				-			
21	20	150529	2536			🔽 Besidi	ual Plots				
22	21	268092	3417	Standardized	Besiduals	I line F	it Plots				
23	22	184895	2290			ien ränen					
24	23	161715	2001	Normal Probabilit	у						
25		ng <b>\ MULTIPLE</b> \	Named Ranges	- 🔽 <u>N</u> ormal Proba	ability Plots						١Ē
Enter											
🙀 S	itart 😗 Mici	rosoft Word - Sectio	34 🔀 Micros	oft Excel - Mult					4	<b>(</b> ∈ 🕅 11:57	7 AM

The resulting equation is:

Selling Price = 39,555 + 60 per unit of House Size - 2,568 per unit of House Age + 3,401 per unit of House Location - 5 cents per unit of Owner's Salary.

Unfortunately if you want to perform a regression analysis of Selling\_Pr against both House\_Age and OwnersSal you have to move the columns around to make the columns House\_Age and OwnersSal contiguous (next to one another!).

## 7. Other statistical functions that are available.

The following table is a list of Excel's statistical functions (You should use the Help system in Excel to get descriptions of these functions and how to use them.).

For a simple linear regression you can establish an equation using:

INTERCEPT() and SLOPE(). RSQ() will give you the R-squared value. CORREL() will give you the Correlation Coefficient. STEYX() will give you the Standard Error of Estimation.

Statistical	Functions
AVEDEV	Returns the average of the absolute deviations of data points from their mean.
AVERAGE	Returns the average of its arguments.
BETADIST	Returns the cumulative beta probability density function.
BETAINV	Returns the inverse of the cumulative beta probability density function.
BINOMDIST	Returns the individual term binomial distribution probability.
CHIDIST	Returns the one-tailed probability of the chi-squared distribution.
CHIINV	Returns the inverse of the one-tailed probability of the chi-squared distribution.
CHITES	Returns the test for independence.
CONFIDENCE	Returns the confidence interval for a population mean.
CORREL	Returns the correlation coefficient between two data sets.
COUNT	Counts how many numbers are in the list of arguments.
COUNTA	Counts how many values are in the list of arguments.
COVAR	Returns covariance, the average of the products of paired deviations.
CRITBINOM	Returns the smallest value for which the cumulative binomial distribution is less
	than or equal to a criterion value.
DEVSQ	Returns the sum of squares of deviations.
EXPONDIST	Returns the exponential distribution.
FDIST	Returns the F probability distribution.
FINV	Returns the inverse of the F probability distribution.
FISHER	Returns the Fisher transformation.
FISHERINV	Returns the inverse of the Fisher transformation.
FORECAST	Returns a value along a linear trend.
FREQUENCY	Returns a frequency distribution as a vertical array.
FTEST	Returns the result of an F-test.
GAMMADIST	Returns the gamma distribution.
GAMMAINV	Returns the inverse of the gamma cumulative distribution.
GAMMALN	Returns the natural logarithm of the gamma function, $G(x)$ .
GEOMEAN	Returns the geometric mean.
GROWTH	Returns values along an exponential trend.
HARMEAN	Returns the harmonic mean.
HYPGEOMDIST	Returns the hypergeometric distribution.
INTERCEPT	Returns the intercept of the linear regression line.
KURT	Returns the kurtosis of a data set.
LARGE	Returns the k-th largest value in a data set.
LINEST	Returns the parameters of a linear trend.
LOGEST	Returns the parameters of an exponential trend.
LOGINV	Returns the inverse of the lognormal distribution.

LOGNORMDIST	Returns the cumulative lognormal distribution.
MAX	Returns the maximum value in a list of arguments.
MEDIAN	Returns the median of the given numbers.
MIN	Returns the minimum value in a list of arguments.
MODE	Returns the most common value in a data set.
NEGBINOMDIST	Returns the negative binomial distribution.
NORMDIST	Returns the normal cumulative distribution.
NORMINV	Returns the inverse of the normal cumulative distribution.
NORMSDIST	Returns the standard normal cumulative distribution.
NORMSINV	Returns the inverse of the standard normal cumulative distribution.
PEARSON	Returns the Pearson product moment correlation coefficient.
PERCENTILE	Returns the k-th percentile of values in a range.
PERCENTRANK	Returns the percentage rank of a value in a data set.
PERMUT	Returns the number of permutations for a given number of objects.
POISSON	Returns the Poisson distribution.
PROB	Returns the probability that values in a range are between two limits.
QUARTILE	Returns the quartile of a data set
RANK	Returns the rank of a number in a list of numbers.
RSQ	Returns the square of the Pearson product moment correlation coefficient.
SKEW	Returns the skewness of a distribution.
SLOPE	Returns the slope of the linear regression line.
SMALL	Returns the k-th smallest value in a data set.
STANDARDIZE	Returns a normalized value.
STDEV	Estimates standard deviation based on a sample.
STDEVP	Calculates standard deviation based on the entire population.
STEYX	Returns the standard error of the predicted y-value for each x in the regression
TDIST	Returns the Student's t-distribution.
TINV	Returns the inverse of the Student's t-distribution.
TREND	Returns values along a linear trend
TRIMMEAN	Returns the mean of the interior of a data set.
TTEST	Returns the probability associated with a Student's t-Test.
VAR	Estimates variance based on a sample.
VARP	Calculates variance based on the entire population.
WEIBULL	Returns the Weibull distribution.
ZTEST	Returns the two-tailed P-value of a z-test.