Statistics & Regression Tools
for Excel

Duke MBA Computer Preparation

DUKE THE FUQUA SCHOOL OF BUSINESS

We Educate *Thoughtful Business Leaders*
Worldwide

Excel Statistics & Regression Tools Review Focus

The focus of this assignment is on understanding the mechanics of running a regression analysis using Excel. In addition, Excel's descriptive statistics generator is briefly explored.

Excel includes a built-in regression tool in its Analysis ToolPak. Although the regression tool is relatively easy to use it has some limitations. For example, if you use multiple variables for the X range in the regression, the variables must be located next to one another in the worksheet. And, the regression outputs this tool generates are tightly packed into a worksheet making them relatively hard to read, especially if you choose to generate charts.

Skills

- The Data Analysis ToolPak Excel Regression
- Naming Ranges
- Excel Data Analysis ToolPak Descriptive Statistics

Files You Need

To complete the work for this assignment, download this file from the class "Review Files" web page: **StatsRegr.xlsx**.

Deliverable

Complete the Statistics and Regression Tools for Excel Review online quiz.

Notes

For your convenience the instructions in this assignment are divided into these sections:

Completing the Assignment: *Summary Descriptions*

Short descriptions of the assignment requirements, useful if you are already familiar with the core skills.

Completing the Assignment: *Detailed Descriptions*

More detailed descriptions of the assignment tasks and techniques, useful if you are unfamiliar with the assignment topics. You may find it helpful to have a good Excel reference guide on hand or to become familiar with Excel's online help.

A Skills Review Summary for this Assignment

Excel Statistics & Regression Tools Review
Table of Contents

Statistics and Regression Tools Review – *Summary Descriptions*

Task 1: Use Excel to Run a Multiple Regression & Create a Chart of Results

Use Excel's built-in regression tool to analyze whether a variety of information about price, advertising, income, etc. collected by the Mrs. Smyth's Gourmet Frozen Fruit Pie Company can predict company pie sales. Make the "Task 1–Pie Data" worksheet current. Then:

1. Run the regression using Excel's Analysis ToolPak regression option. Use the data in the **Unit Sales** column as the dependent, Y variable. Use the data in the **Price, Ads, Comp_Price, Income, Population,** and **Time Var** columns as the independent, X variables.
2. Choose to output "Residuals" and "Standardized Residuals" and send the output to a new worksheet. Name the new worksheet "Task 2". Position this sheet to the left of the "Task 2 – Pie Data" worksheet.
3. Create a line chart comparing Unit Sales from the "Task 2–Pie Data" worksheet with the "Predicted (Q)" output in Excel's "Residual Output" section of results. Locate the line chart to the right of the "Residual Output" section of Excel's regression results.

End of the Task 1 Summary Description

Task 2: Use Excel's Descriptive Statistics Generator to Analyze Data

Make the "Task 2-Auto Imports" worksheet current. This sheet contains data on imports into the U.S. passenger car market from 1971 to 1997, comparing Korean, Japanese, and German import sales with sales of passenger cars manufactured in North America.

1. Run Excel's Analysis ToolPak "Descriptive Statistics" analysis on the data located in this worksheet. Do not include the "Year" column in the analysis. Select column headers for the rest of the data and make sure the "Labels in first row" option is checked. Send the descriptive statistics report to a "New Worksheet Ply" and request all the available outputs: Summary statistics, Confidence Level for Mean, Kth Largest, and Kth Smallest. Leave the settings for the confidence level, Kth largest, and Kth smallest at the defaults.

Task 2 Intermediate Check

In the set of initial results, many columns are not wide enough to display their data completely. In addition, Excel generates a column of descriptive statistics labels (Mean, Standard Error, Median, etc.) for each country/region analyzed (Columns A, C, E, and G). This label redundancy makes the output more difficult to read.

Korea		Japan		Germany		North America	
Mean	101097.1	Mean	1459132	Mean	356176.2	Mean	7470701
Standard E	24773.69	Standard E	108339.5	Standard E	31866.44	Standard E	202393.6
Median	0	Median	1476038	Median	301123.5	Median	7191147
Mode	0	Mode	#N/A	Mode	#N/A	Mode	#N/A
Standard D	126321.5	Standard D	552425.2	Standard D	162487.6	Standard D	1032009
Sample Va	1.6E+10	Sample Va	3.05E+11	Sample Va	2.64E+10	Sample Va	1.07E+12
Kurtosis	0.134265	Kurtosis	-1.22913	Kurtosis	1.832032	Kurtosis	-0.47534
Skewness	0.990079	Skewness	-0.18657	Skewness	1.438401	Skewness	0.515567
Range	419099	Range	1803637	Range	601385	Range	3917204
Minimum	0	Minimum	578977	Minimum	186177	Minimum	5758586
Maximum	419099	Maximum	2382614	Maximum	787562	Maximum	9675790
Sum	2628525	Sum	37937442	Sum	9260582	Sum	1.94E+08
Count	26	Count	26	Count	26	Count	26
Largest(1)	419099	Largest(1)	2382614	Largest(1)	787562	Largest(1)	9675790
Smallest(1	0	Smallest(1	578977	Smallest(1	186177	Smallest(1	5758586
Confidence	51022.37	Confidence	223129.3	Confidence	65630.17	Confidence	416837.5

2. Edit the output so it has a *single* labels column at the left of the data and so that country/region names are above their data. Widen columns so all the data is visible and so your worksheet looks like the illustration below.

	Korea	Japan	Germany	North America
Mean	101097.1154	1459132.385	356176.2308	7470700.846
Standard Error	24773.69004	108339.4856	31866.44492	202393.6268
Median	0	1476037.5	301123.5	7191147
Mode	0	#N/A	#N/A	#N/A
Standard Deviation	126321.5289	552425.1513	162487.6245	1032009.053
Sample Variance	15957128675	3.05174E+11	26402228113	1.06504E+12
Kurtosis	0.134265053	-1.229126977	1.832032364	-0.475338779
Skewness	0.990079421	-0.186566776	1.438401394	0.515566924
Range	419099	1803637	601385	3917204
Minimum	0	578977	186177	5758586
Maximum	419099	2382614	787562	9675790
Sum	2628525	37937442	9260582	194238222
Count	26	26	26	26
Largest(1)	419099	2382614	787562	9675790
Smallest(1)	0	578977	186177	5758586
Confidence Level(95.0%)	51022.33399	223129.1911	65630.12587	416837.1852

End of the Task 2 Summary Description

End of the Statistics and Regression Tools Review Summary Description

Statistics & Regression Tools Review – *Detailed Descriptions*

Task 1: Run a Multiple Regression with Excel & Create a Chart of Results

Introduction
Use Excel's built-in regression tool to analyze whether information collected by the Mrs. Smyth's Gourmet Frozen Fruit Pie Company about price, advertising, competitors' pricing, etc. can predict company pie sales. Make the "Task 1–Pie Data" worksheet current and review the data. There are ten columns of data with forty-eight records.

The illustration below shows a partial view of the data.

City	Yr/Qtr	Row	Unit Sales	($) Price	($) Ads	($) Comp_Price	($) Income	Population	(T) Time Var
Atlanta, GA	2003-1	1	193,334	6.39	15,827	6.92	33,337	4,116,250	1
	2003-2	2	170,041	7.21	20,819	4.84	33,390	4,140,338	2
	2003-3	3	247,709	5.75	14,062	5.28	33,599	4,218,965	3
	2003-4	4	183,259	6.75	16,973	6.17	33,797	4,226,070	4
	2004-1	5	282,118	6.36	18,815	6.36	33,879	4,278,912	5
	2004-2	6	203,396	5.98	14,176	4.88	34,186	4,359,442	6
	2004-3	7	167,447	6.64	17,030	5.22	35,691	4,363,494	7
	2004-4	8	361,677	5.30	14,456	5.80	35,950	4,380,084	8
Chicago, IL -	2003-1	9	401,805	6.08	27,183	4.99	34,983	9,184,926	1
Gary, IN -	2003-2	10	412,312	6.13	27,572	6.13	35,804	9,237,683	2
Kenosha, WI	2003-3	11	321,972	7.24	34,367	5.82	35,898	9,254,182	3
	2003-4	12	445,236	6.08	26,895	6.05	36,113	9,272,758	4
	2004-1	13	479,713	6.40	30,539	5.37	36,252	9,300,401	5
	2004-2	14	459,379	6.00	26,679	4.86	36,449	9,322,168	6
	2004-3	15	444,040	5.96	26,607	5.29	37,327	9,323,331	7
	2004-4	16	376,046	7.21	32,760	4.89	37,841	9,348,725	8
Dallas-Fort	2003-1	17	255,203	6.55	19,880	6.97	34,870	5,294,645	1
Worth, TX	2003-2	18	270,881	6.11	19,151	6.25	35,464	5,335,816	2
	2003-3	19	330,271	5.62	15,743	6.03	35,972	5,386,134	3

For this regression operation:

>Use as the predictor variables (the independent or X variables) the data for **Price, Ads, Comp_Price, Income, Population,** and **Time Var**.

>Use as the predicted variable (the dependent or Y variable) the data for **Unit Sales**. Unit Sales holds actual sales data.

Steps to Follow When Running a Regression with Excel

1. Rearrange the Data as Necessary
With Excel's regression tool the independent X variables you use in your analysis must be *located together* in the worksheet. There must be no blank columns or columns-with-non-relevant data interrupting the range of X variables. The dependent Y variable need not be located adjacent to the X variables, but all Y variable values must be in a single unbroken column. These location specifications are required by Excel's built-in regression tool.

For the data in the "Task 1–Pie Data" worksheet, you need not rearrange the data. The X variables are together in Columns F through K and the Y variable values are all in Column E.

| A	B	C	D	E	F	G	H	I	J	K
	Mrs. Smyth's Gourmet Frozen Fruit Pie Regional Market Demand Data, 2003-1 to 2									
					($)	($)	($)	($)		(T)
	City	Yr/Qtr	Row	Unit Sales	Price	Ads	Comp_Price	Income	Population	Time Var
	Atlanta, GA	2003-1	1	193,334	6.39	15,827	6.92	33,337	4,116,250	1
		2003-2	2	170,041	7.21	20,819	4.84	33,390	4,140,338	2
		2003-3	3	247,709	5.75	14,062	5.28	33,599	4,218,965	3
		2003-4	4	183,259	6.75	16,973	6.17	33,797	4,226,070	4
		2004-1	5	282,118	6.36	18,815	6.36	33,879	4,278,912	5

Columns not used in the regression.
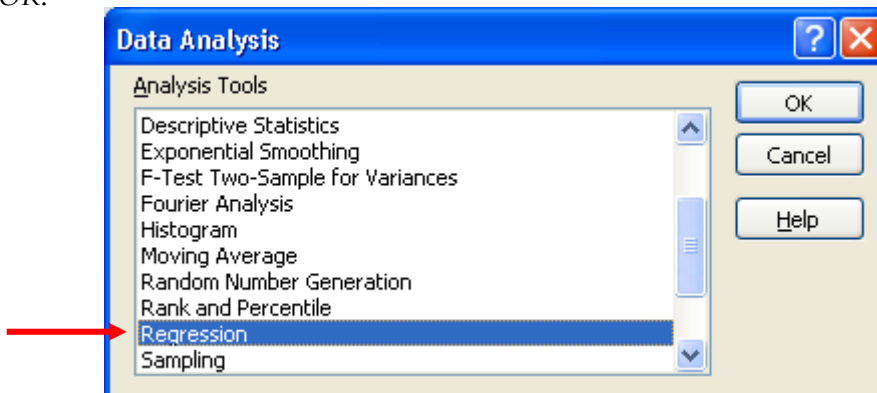
The dependent, Y variable.

Independent, X variables, located together in the worksheet.
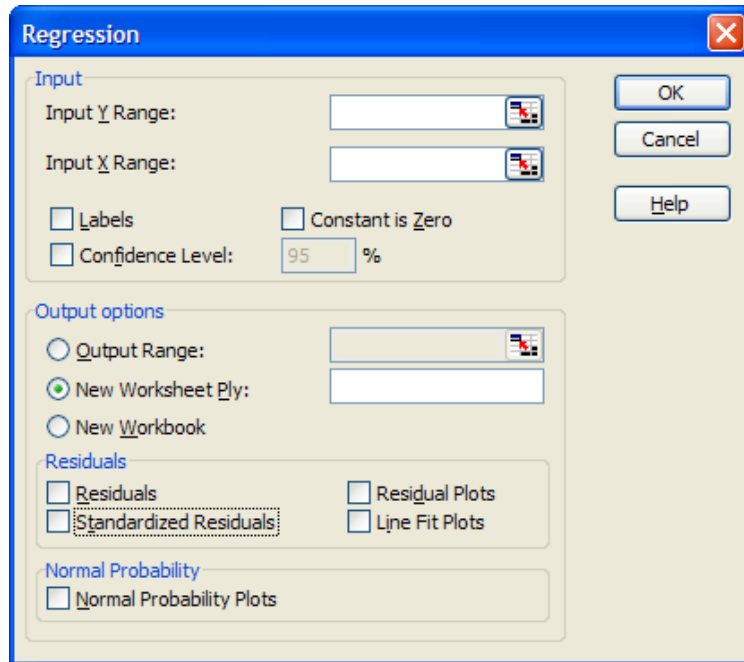
2. Run the Regression Analysis

In the "Data" tab on Excel's ribbon find the "Analysis" group and choose "Data Analysis".
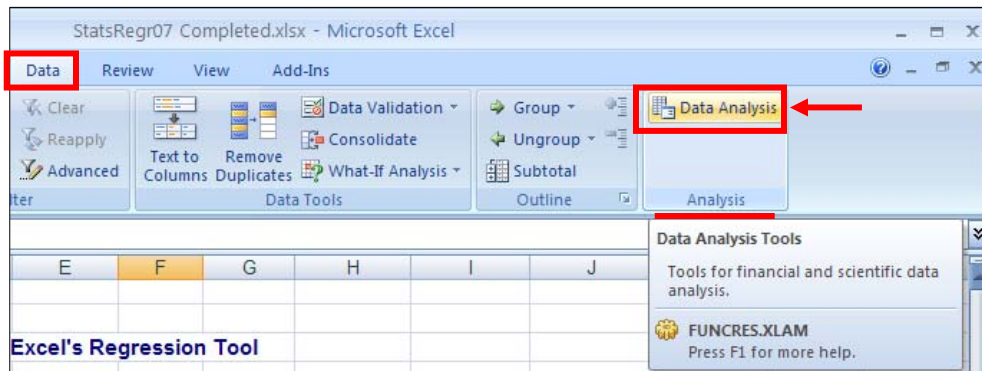
Excel opens its "Data Analysis" dialog. Scroll down and find the "Regression" option. Click *OK*.

The "Regression" dialog displays. This is the sole interface to Excel's regression tool. All your interaction with Excel's regression tool happens in this dialog.

Complete the "Regression" dialog box as follows:

- Specify "Unit Sales" as the "Input Y Range". We will be turning on the "Labels" option in this dialog, so include the column header cell (Cell E9) in your Unit Sales range specification[1]. This range is E9:E57.

- Specify the range of data in Columns F through K as the "Input X Range". Include only *one* row of column headers with your range specification. That is, include the labels in Row 9 but *not* those in Row 8. This range is F9:K57.

- In the "Regression" dialog box make sure the *Labels* box is checked. Excel will recognize that the first cell in each column of data is that column's label.

- Direct Excel to put its results on a new worksheet by choosing the option *New Worksheet Ply*.

- For outputs, select *Residuals,* and *Standardized Residuals.*

With data ranges and choices specified, your dialog should now look like the one below.

---

[1] In this selection operation be careful to select *data and header only* and not an entire column. Here's an easy way to select just the contiguous data from a column of data: Click the top cell. Depress the SHIFT key. Tap the End key and then tap the down arrow key. Excel selects all data down the column until it reaches the first empty cell in the column.
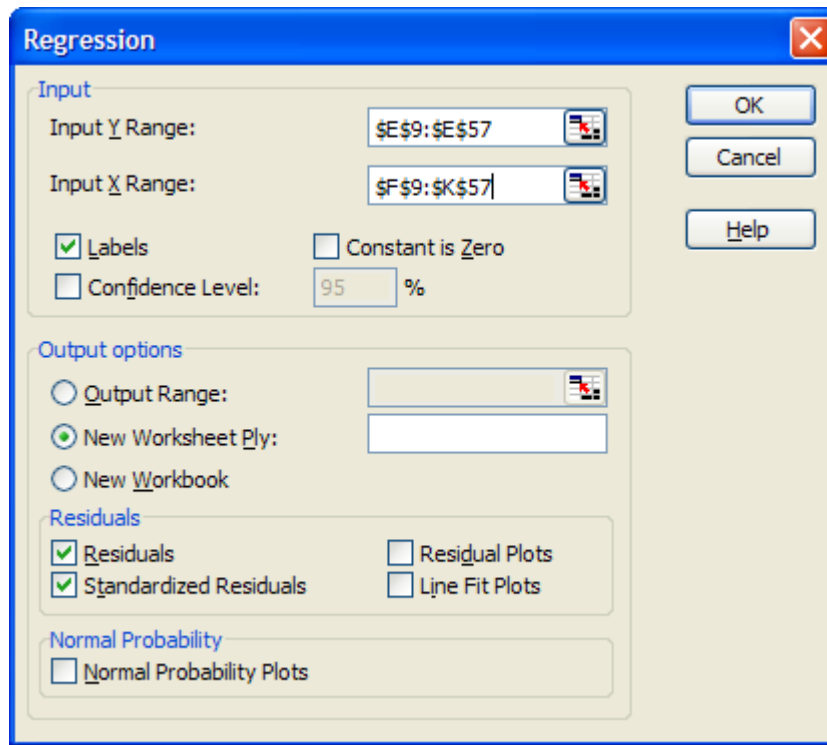
- Click *OK* to execute the regression.

The *New Worksheet Ply* default output option means that Excel locates the regression outputs on a new, separate worksheet that it creates and puts to the left of the worksheet that holds the original data.

Excel completes the regression analysis leaving a large range of the worksheet selected. Click any cell to turn off the range selection. Because the Regression tool generates many outputs on a single sheet, you may want to use Excel's Zoom Control option to reduce the new worksheet size in order to get a view of how the outputs are arranged.

Widen the columns on the new worksheet in order to see all the text results properly. A quick way to do this is to select the blank square at the intersection of the row numbers and column headers (to select the entire worksheet), and then double-click the dividing line between any two column headers. Excel widens each column so the widest entry in that column displays completely.

To rename this worksheet "Task 1" right-click the worksheet tab and enter the name in place of the default worksheet name.

Task 1 Intermediate Check

Selecting the *Residuals,* and *Standardized Residuals* options in the "Regression" dialog produces the output partially shown below.

In this view columns have been widened to show the results completely.

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.946559629							
5	R Square	0.895975132							
6	Adjusted R Square	0.880751981							
7	Standard Error	60699.8107							
8	Observations	48							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	6	1.30112E+12	2.16853E+11	58.8560875	1.39375E-18			
13	Residual	41	1.51063E+11	3684467019					
14	Total	47	1.45218E+12						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
17	Intercept	646957.8174	154146.8797	4.197021819	0.00014152	335651.8818	958263.753	335651.8818	958263.753
18	Price	-127442.8033	15112.59845	-8.432884903	1.699E-10	-157963.3147	-96922.2919	-157963.315	-96922.2919
19	Ads	5.352976863	1.113917711	4.805540671	2.0919E-05	3.103374433	7.602579293	3.103374433	7.602579293
20	Comp_Price	29337.34026	12388.5933	2.368092934	0.02267716	4318.0688	54356.61171	4318.0688	54356.61171
21	Income	0.344132846	3.186720717	0.107989647	0.91453065	-6.091580132	6.779845825	-6.09158013	6.779845825
22	Population	0.023965666	0.002349041	10.20232045	8.1243E-13	0.019221682	0.02870965	0.019221682	0.02870965
23	Time Var	4405.940135	4400.39741	1.001259596	0.32257427	-4480.842623	13292.72289	-4480.84262	13292.72289
24									
25									
26									
27	RESIDUAL OUTPUT								
28									
29	*Observation*	*Predicted Unit Sales*	*Residuals*	*Standard Residuals*					
30	1	234861.2336	-41527.23358	-0.732491627					
31	2	101059.9918	68981.00824	1.216743968					
32	3	270227.062	-22518.06199	-0.397192166					
33	4	189121.3616	-5862.361642	-0.103405174					
34	5	259958.8857	22159.11427	0.39086075					

Introduction / **Sheet1** / Task 2 - Pie Data / Task 3 - Auto Imports

Excel labels the output areas "SUMMARY OUTPUT", "ANOVA", and "RESIDUAL OUTPUT".

3. Create a Chart

Create a line chart comparing the actual **Unit Sales** in the "Task 1–Pie Data" worksheet with the **Predicted Unit Sales** produced by the regression analysis.

	A	B	C	D
26				
27	RESIDUAL OUTPUT			
28				
29	*Observation*	*Predicted Unit Sales*	*Residuals*	*Standard Residuals*
30	1	234861.2336	-41527.23358	-0.732491627
31	2	101059.9918	68981.00824	1.216743968
32	3	270227.062	-22518.06199	-0.397192166
33	4	189121.3616	-5862.361642	-0.103405174
34	5	259958.8857	22159.11427	0.39086075

Introduction / **Sheet1** / Task 2 - Pie Data / Task 3 - Aut

	C	D	E	F	G
8				($)	($)
9	Yr/Qtr	Row	Unit Sales	Price	Ads
10	2003-1	1	193,334	6.39	15,827
11	2003-2	2	170,041	7.21	20,819
12	2003-3	3	247,709	5.75	14,062
13	2003-4	4	183,259	6.75	16,973
14	2004-1	5	282,118	6.36	18,815
15	2004-2	6	203,396	5.98	14,176

Partial view of Excel's regression output showing the "Predicted Unit Sales".

Partial view of the "Task 1 - Pie Data" worksheet showing "Unit Sales".

The data needed to create this chart are located on two different worksheets. The **Unit Sales** data is on the original "Task 1–Pie Data" worksheet. The **Predicted Unit Sales** data – output by the regression analysis – is under the "Residual Output" header on the worksheet of outputs you just generated and now named "Task 1".

To create a chart using data from two different sheets:

1. Select the **Predicted Unit Sales** (Predicted Y) header cell and the data beneath it from the regression output. Click the "Insert" tab and from the "Charts" group select a line chart style from the "2-D Line" group.
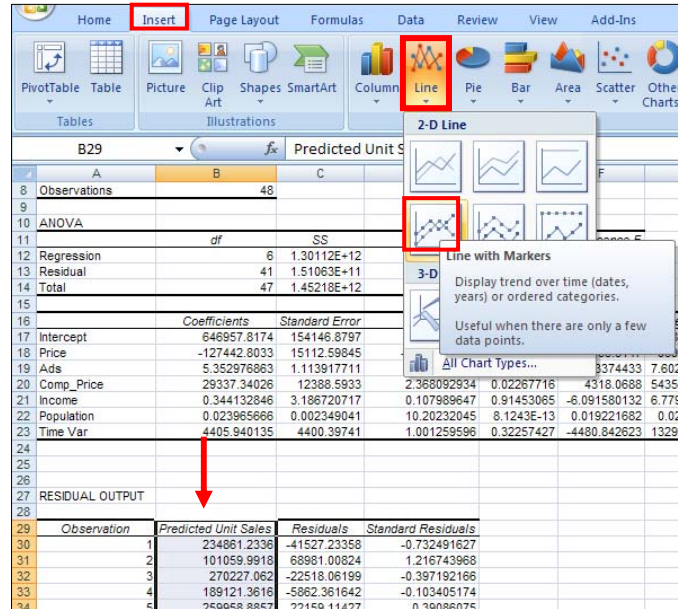
   Drag the chart to the right of the RESIDUAL OUTPUT section of the worksheet. The basic chart and its location in relation to the outputs on your "Task 1" worksheet should look like the illustration below.
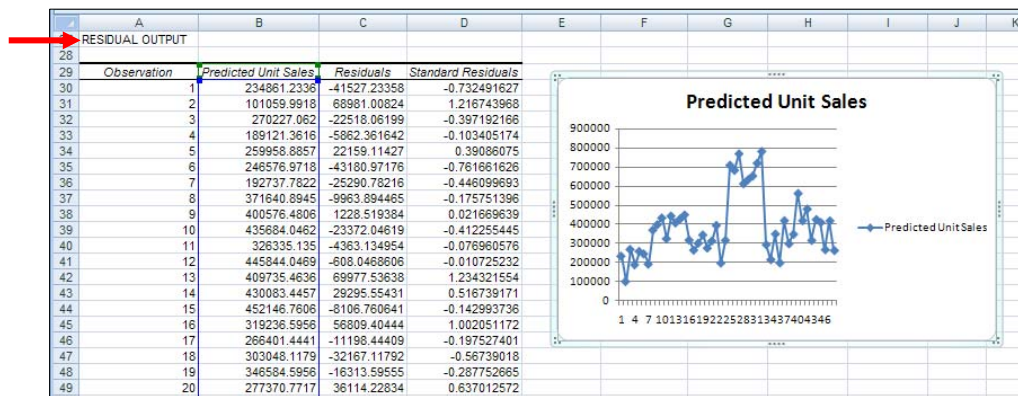
2. Make your "Task 1–Pie Data" worksheet current. Select the **Unit Sales** header cell and the data beneath it from this worksheet. Copy the selection to the Windows Clipboard using *Edit, Copy* (or CTRL+C).

3. Make your "Task 1" worksheet current. Select the outside edge of the line chart you just created. Small, empty circles and squares should appear around the edges of the chart to show it's properly selected.

4. With the chart selected, use the *Edit, Paste* commands (or CTRL+V) to copy the data from the Clipboard to the chart. Excel adds the data from the clipboard to the existing chart as a second line.

Task 1 Check
A simple line chart with sales data from the two worksheets should appear similar to the chart shown at right. You need not make any formatting changes to this cart. In the illustration here, it's a bit hard to see both lines, but the legend at right indicates that two data series are included in the chart.
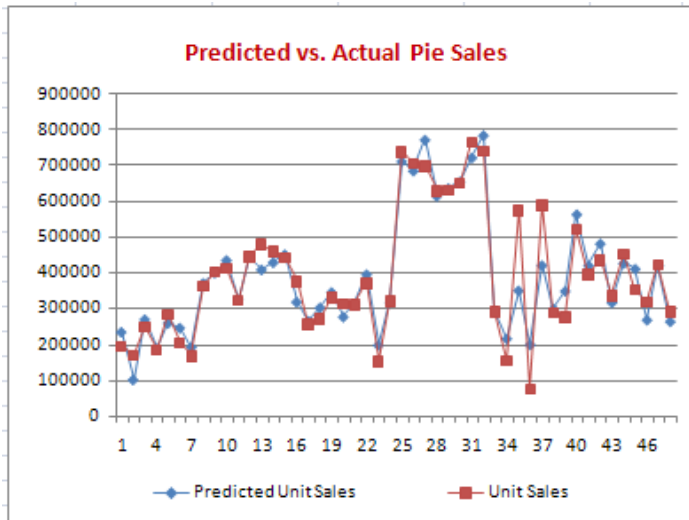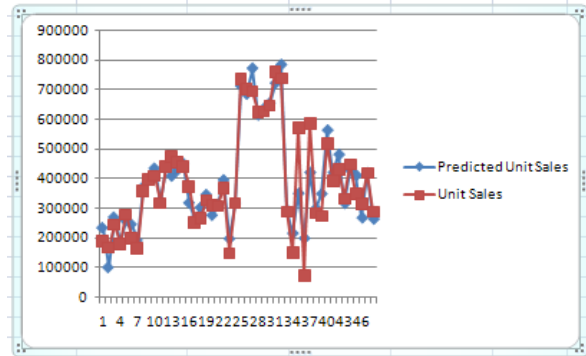
Optional Enhancement
At left is a chart modified to include a title, formatted X and Y axes, and a modified legend.

You need not reproduce these modifications in your own chart.

End of the Task 1 Detailed Description

Range Naming Note

You need not name the ranges of data you plan to use in an Excel analysis, but being able to work with named ranges can be helpful. When the data already has headers (as does the data we used in Task 1), a quick way to assign names to all the variables is to:

1) Select *all* the headers from the "Unit Sales" header to through "Time Var". Note that some of the data columns have 2-row headers, but Excel allows only the contents of a single cell to be a range name, so select the single row of headers closest to the data.

	A	B	C	D	E	F	G	H	I	J	K	
1												
2		**Statistics and Regression Tools**										
3		**Task 1: Run a Regression Using Excel's Regression Tool**										
4												
5												
6		**Mrs. Smyth's Gourmet Frozen Fruit Pie Regional Market Demand Data, 2003-1 to 2004-**										
7												
8						($)	($)	($)	($)		(T)	
9		City	Yr/Qtr	Row	Unit Sales	Price	Ads	Comp Price	Income	Population	Time Var	
10		Atlanta, GA	2003-1	1	193,334	6.39	15,827	6.92	33,337	4,116,250	1	
11			2003-2	2	170,041	7.21	20,819	4.84	33,390	4,140,338	2	
12			2003-3	3	247,709	5.75	14,062	5.28	33,599	4,218,965	3	
13			2003-4	4	183,259	6.75	16,973	6.17	33,797	4,226,070	4	

Depress and hold down the Shift key, tap and release the End key, and tap and release the down arrow key. Release the Shift key. Excel selects all the data beneath the selected headers for you.

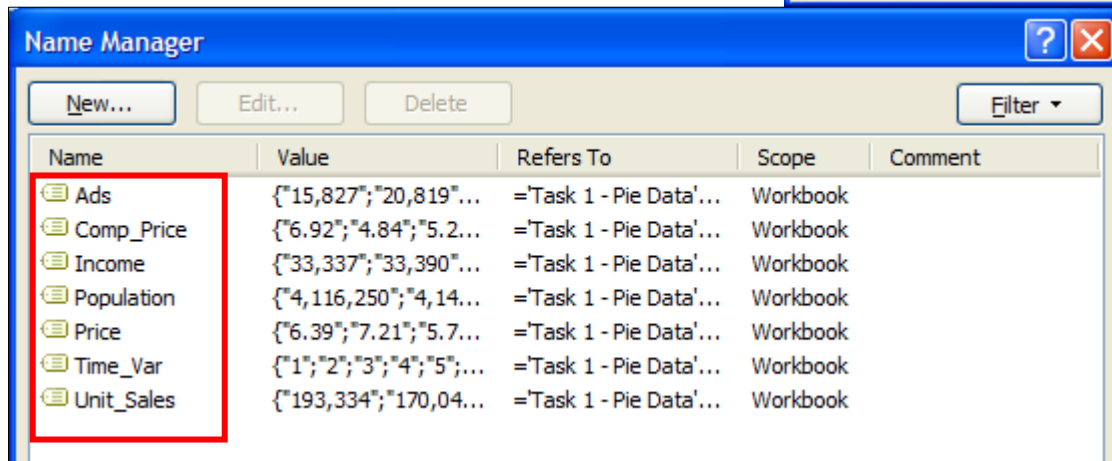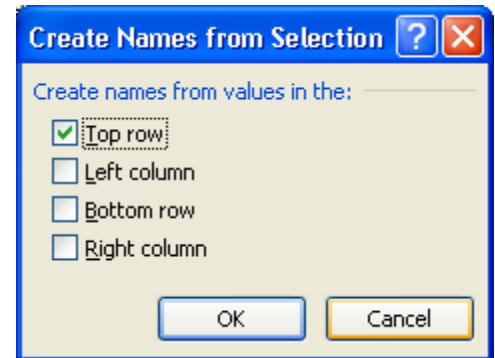|	City	Yr/Qtr	Row	Unit Sales	Price	Ads	Comp_Price	Income	Population	Time Var
8					($)	($)	($)	($)		(T)
9	City	Yr/Qtr	Row	Unit Sales	Price	Ads	Comp_Price	Income	Population	Time Var
10	Atlanta, GA	2003-1	1	193,334	6.39	15,827	6.92	33,337	4,116,250	1
11		2003-2	2	170,041	7.21	20,819	4.84	33,390	4,140,338	2
12		2003-3	3	247,709	5.75	14,062	5.28	33,599	4,218,965	3
13		2003-4	4	183,259	6.75	16,973	6.17	33,797	4,226,070	4
14		2004-1	5	282,118	6.36	18,815	6.36	33,879	4,278,912	5
15		2004-2	6	203,396	5.98	14,176	4.88	34,186	4,359,442	6

Partial view.

2) Next, click the "Formulas" tab and find the "Defined Names" group.

3) Choose the "Create from Selection" button. Excel opens a "Create Names from Selection" dialog.
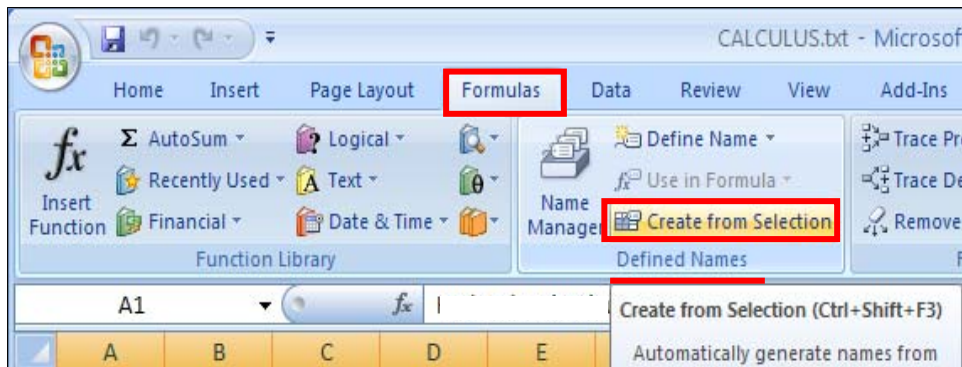
4) Choose the "Top row" option in this dialog and click *OK*.

Excel uses the text in the top row of each column of data as the range name for that column.

Choose the "Name Manager" button in the same "Defined Names" group to open a dialog that displays all the named ranges in your workbook. The names from the selected headers in the pie data should be listed in this dialog.

Create Names from Selection

Create names from values in the:
- ☑ Top row
- ☐ Left column
- ☐ Bottom row
- ☐ Right column

[OK] [Cancel]

Name Manager

[New...] [Edit...] [Delete]   [Filter ▾]

| Name	Value	Refers To	Scope	Comment
Ads	{"15,827";"20,819"...	='Task 1 - Pie Data'...	Workbook	
Comp_Price	{"6.92";"4.84";"5.2...	='Task 1 - Pie Data'...	Workbook	
Income	{"33,337";"33,390"...	='Task 1 - Pie Data'...	Workbook	
Population	{"4,116,250";"4,14...	='Task 1 - Pie Data'...	Workbook	
Price	{"6.39";"7.21";"5.7...	='Task 1 - Pie Data'...	Workbook	
Time_Var	{"1";"2";"3";"4";"5";...	='Task 1 - Pie Data'...	Workbook	
Unit_Sales	{"193,334";"170,04...	='Task 1 - Pie Data'...	Workbook	

You can now use any of these names in place of their corresponding range references to refer to those ranges. For example, the range E10:E57 can be referred to by the range name "Unit Sales".

Additional Notes
- A range named in this way (using create from selection) does *not* include the header cell that provides the name.
- If a header consists of two or more words separated by spaces, Excel automatically replaces the spaces with underscores in the range names.
- Two other ways to create a range name:
  - Use the "Define Name" option in the "Defined Names" group on the Formulas " tab.
 - Use the "name box" at the left-hand side of the formula bar.

Task 2: Use Excel's Descriptive Statistics Generator to Analyze Data

Excel has a wide variety of built-in functions and tools for performing statistical analyses. For example, Excel includes over 80 functions in its "Statistical" category of built-in functions. Clicking the *Paste Function* button on the Formula Toolbar displays an "Insert Function" dialog showing Excel's built-in functions.
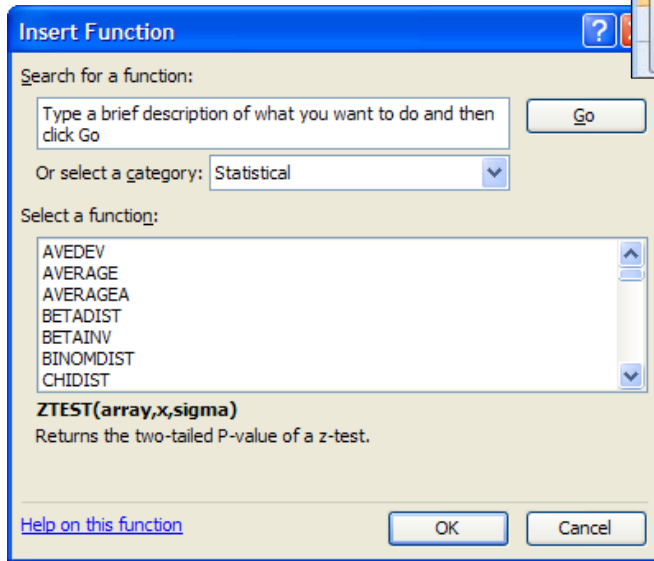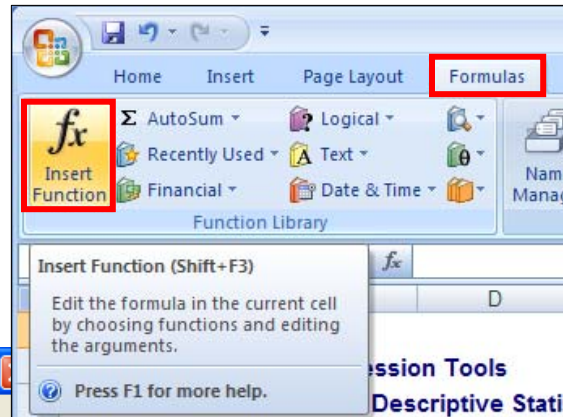
These functions in particular are often used for projections: FORECAST, TREND, GROWTH, LINEST, and LOGEST. In addition to the built-in functions, a number of statistical analysis tools besides regression are included in Excel's Analysis ToolPak add-in (*Tools, Data Analysis)*.

In this task you use the "Descriptive Statistics" data analysis tool to generate descriptive statistics about a set of car import data and rearrange the results to make the data more readable.
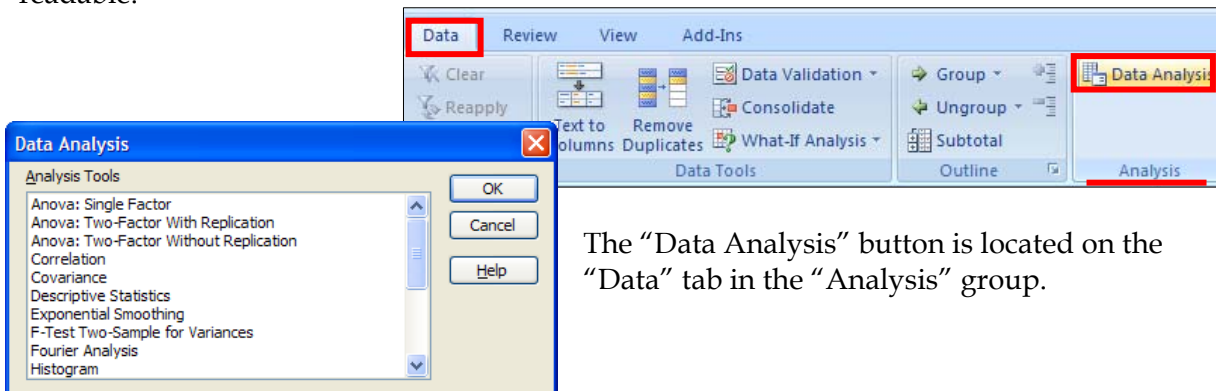
The "Data Analysis" button is located on the "Data" tab in the "Analysis" group.

Make the "Task 2 - Auto Imports" worksheet current. This worksheet contains data on imports into the U.S. passenger car market from 1971 to 1997, comparing Korean, Japanese, and German import sales with sales of passenger cars manufactured in North America.

Choose the "Data" tab and find the "Analysis" group. Click the "Data Analysis" button. Select the "Descriptive Statistics" option to open the dialog for that tool. Complete the "Descriptive Statistics" dialog as follows:

1. For "Input Range" specify all columns of data *except for* the "Year" column in the analysis. Do not include the "Year" data. Include column headers when you select the Input Range (the range **C9:F35**).
2. Check the "Labels in first row" option.
3. For "Output options" choose "Output Range" and specify a cell below the data (for example, Cell B38).
4. Request all the available outputs: Summary statistics, Confidence Level for Mean, Kth Largest, and Kth Smallest. (Leave the defaults for these last three output options as they are, at 95%, 1, and 1.)
5. Click *OK.*
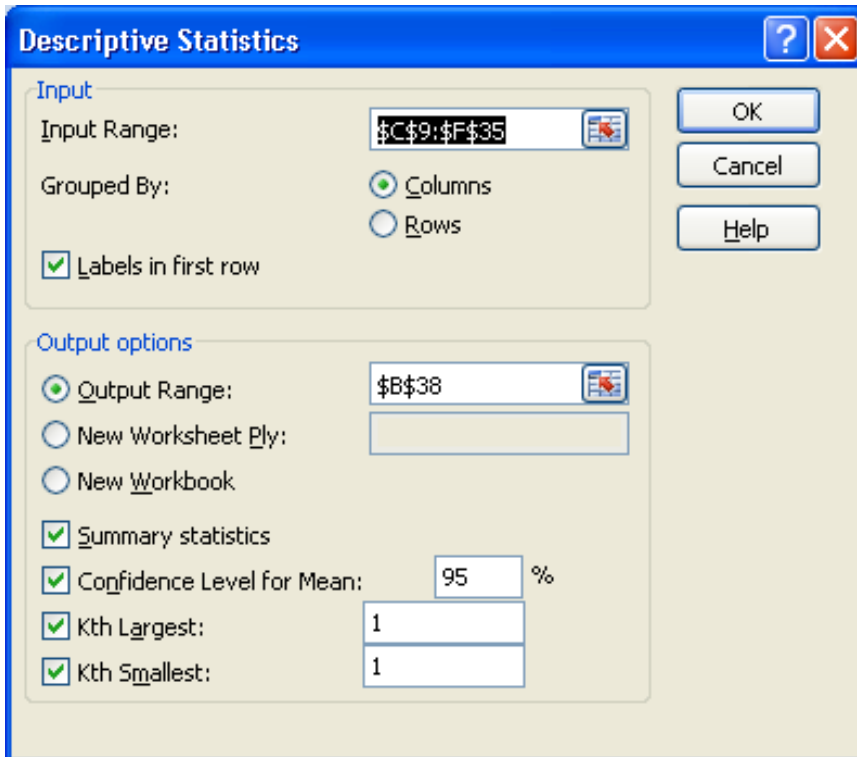
Excel automatically converts any cell references you enter in this dialog to absolute references.

Excel generates a complete set of descriptive statistics for the data that looks like the illustration that follows.

Task 2 Intermediate Check

The initial results remain selected; click any cell to deselect the range. In this "raw" output, many columns are not wide enough to display their data completely. In addition, Excel generates a column of descriptive statistics labels (Mean, Standard Error, Median, etc.) for each country/region analyzed (Columns B, D, F, and H). This label redundancy makes the output more difficult to read.

Columns with redundant labels

	A	B	C	D	E	F	G	H	I
31		1993	134,764	1,328,448	186,177	6,741,667			
32		1994	173,214	1,239,450	192,275	7,255,303			
33		1995	183,214	981,462	207,555	7,128,712			
34		1996	172,193	726,940	237,984	7,253,582			
35		1997	181,923	726,104	297,028	6,916,769			
36									
37									
38		*Korea*		*Japan*		*Germany*		*North America*	
39									
40		Mean	101097.1154	Mean	1459132.385	Mean	356176.2	Mean	7470701
41		Standard Err	24773.69004	Standard Error	108339.4856	Standard Error	31866.44	Standard E	202393.6
42		Median	0	Median	1476037.5	Median	301123.5	Median	7191147
43		Mode	0	Mode	#N/A	Mode	#N/A	Mode	#N/A
44		Standard De	126321.5289	Standard Deviati	552425.1513	Standard Deviation	162487.6	Standard D	1032009
45		Sample Vari	15957128675	Sample Varianc	3.05174E+11	Sample Variance	2.64E+10	Sample Va	1.07E+12
46		Kurtosis	0.134265053	Kurtosis	-1.229126977	Kurtosis	1.832032	Kurtosis	-0.47534
47		Skewness	0.990079421	Skewness	-0.186566776	Skewness	1.438401	Skewness	0.515567
48		Range	419099	Range	1803637	Range	601385	Range	3917204
49		Minimum	0	Minimum	578977	Minimum	186177	Minimum	5758586
50		Maximum	419099	Maximum	2382614	Maximum	787562	Maximum	9675790
51		Sum	2628525	Sum	37937442	Sum	9260582	Sum	1.94E+08
52		Count	26	Count	26	Count	26	Count	26
53		Largest(1)	419099	Largest(1)	2382614	Largest(1)	787562	Largest(1)	9675790
54		Smallest(1)	0	Smallest(1)	578977	Smallest(1)	186177	Smallest(1	5758586
55		Confidence L	51022.36931	Confidence Leve	223129.3456	Confidence Level(9	65630.17	Confidence	416837.5

Introduction / Task 2 / Task 2 - Pie Data / **Task 3 - Auto Imports**

Widen columns so all the data is visible and edit the output so it has a *single* labels column at the left of the data. Arrange the output so country/region names are above their data. Use Excel's delete and move options to rearrange the data.

	Korea	*Japan*	*Germany*	*North America*
Mean	101097.1154	1459132.385	356176.2308	7470700.846
Standard Err	24773.69004	108339.4856	31866.44492	202393.6268
Median	0	1476037.5	301123.5	7191147
Mode	0	#N/A	#N/A	#N/A
Standard De	126321.5289	552425.1513	162487.6245	1032009.053
Sample Vari	15957128675	3.05174E+11	26402228113	1.06504E+12
Kurtosis	0.134265053	-1.229126977	1.832032364	-0.475338779
Skewness	0.990079421	-0.186566776	1.438401394	0.515566924
Range	419099	1803637	601385	3917204
Minimum	0	578977	186177	5758586
Maximum	419099	2382614	787562	9675790
Sum	2628525	37937442	9260582	194238222
Count	26	26	26	26
Largest(1)	419099	2382614	787562	9675790
Smallest(1)	0	578977	186177	5758586
Confidence L	51022.36931	223129.3456	65630.17131	416837.4738

For example, select the data for Japan and drag it to the left.

Task 2 Check

With these changes complete, your output will look like the illustration below.

	A	B	C	D	E	F
37						
38			Korea	Japan	Germany	North America
39						
40		Mean	101097.1154	1459132.385	356176.2308	7470700.846
41		Standard Error	24773.69004	108339.4856	31866.44492	202393.6268
42		Median	0	1476037.5	301123.5	7191147
43		Mode	0	#N/A	#N/A	#N/A
44		Standard Deviation	126321.5289	552425.1513	162487.6245	1032009.053
45		Sample Variance	15957128675	3.05174E+11	26402228113	1.06504E+12
46		Kurtosis	0.134265053	-1.229126977	1.832032364	-0.475338779
47		Skewness	0.990079421	-0.186566776	1.438401394	0.515566924
48		Range	419099	1803637	601385	3917204
49		Minimum	0	578977	186177	5758586
50		Maximum	419099	2382614	787562	9675790
51		Sum	2628525	37937442	9260582	194238222
52		Count	26	26	26	26
53		Largest(1)	419099	2382614	787562	9675790
54		Smallest(1)	0	578977	186177	5758586
55		Confidence Level(95.0%)	51022.36931	223129.3456	65630.17131	416837.4738
56						

End of the Task 2 Detailed Description

End of the Statistics and Regression Tools Review Detailed Descriptions

Duke University
The Fuqua School of Business

Statistics & Regression Tools Review Skills Summary

Import a Text File into Excel

Use Excel's normal *File, Open* commands. A 3-step "Text Import Wizard" dialog opens. Complete the wizard steps to specify the data type, characteristics, and import formats. Also use the wizard to determine which columns of data to import and the import starting row. Once a text file is imported into Excel it remains a text file until converted to Excel format by clicking the Office button, then *Save As*, and then *Excel Workbook.*
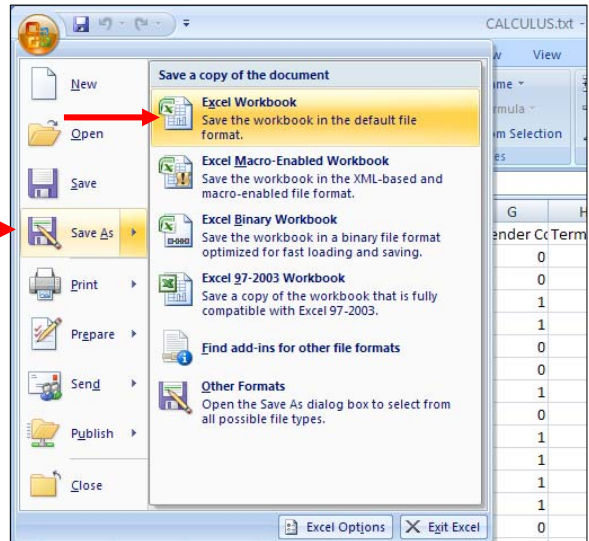
Excel's Analysis ToolPak

To Access

From the "Data" tab and the "Analysis" group, click the "Data Analysis" button to open the "Data Analysis" dialog.
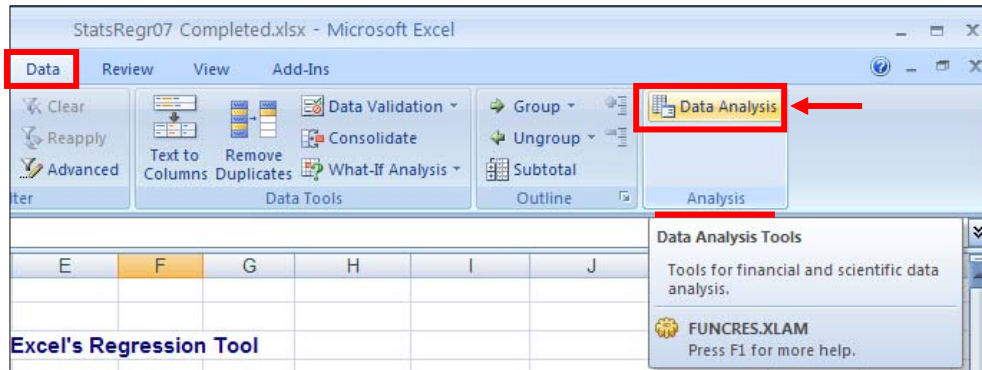
The Analysis ToolPak is a part of MS Excel. It's packaged as an add-in. It's available if you have loaded the add-in during your Excel installation and if you've turned it on as an available add-in.

Tools

Provides functions and interfaces for financial and scientific data analyses.

Operation

From the "Data Analysis" dialog select the analysis you want to perform and then complete the dialog that displays. The dialog prompts for input data, output specifications, and (in many cases) run-time options.

Run a Regression Analysis Using Excel's Built-In Regression Tool

Excel's built-in regression analysis tool is part of its Analysis ToolPak add-in.

1) Open the "Data Analysis" dialog (see the notes above.)
2) Choose *Regression* from the list of options in the "Data Analysis" dialog that displays. A "Regression" dialog opens.
3) Complete the "Regression" dialog, specifying inputs (the dependent (Y) variable and the independent (X) variables and outputs. Note that range of X variable values must be located together in the worksheet.
4) Choose *OK* to run. Excel generates regression results.
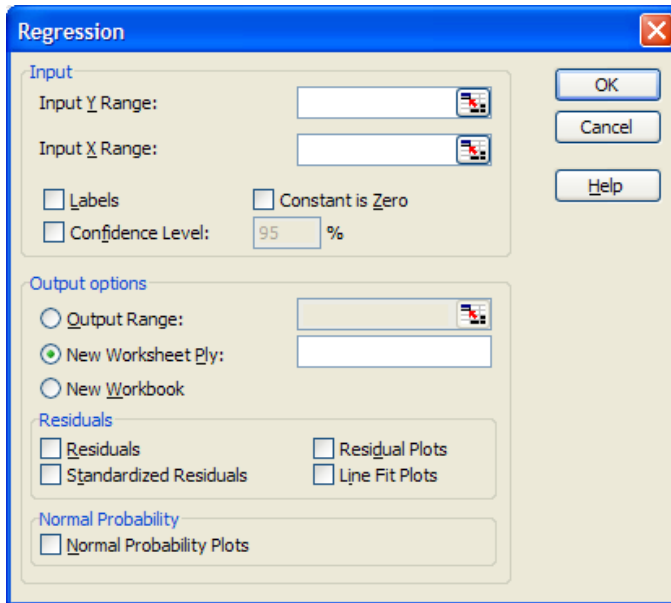
Chart Data from Different Worksheets

1) Create a chart using the data from one worksheet.
2) Select and copy the data from the other worksheet.
3) Go back to the initial worksheet-with-chart and click the chart's edge to select it.
4) *Edit, Paste* the copied data onto the chart.

Find a List of Excel's Statistical Functions

On the "Formulas" tab choose the "Function Library" group and click the "Insert Function" button. Choose "Statistical" from the category list to see a list of built-in functions that handle statistical operations.
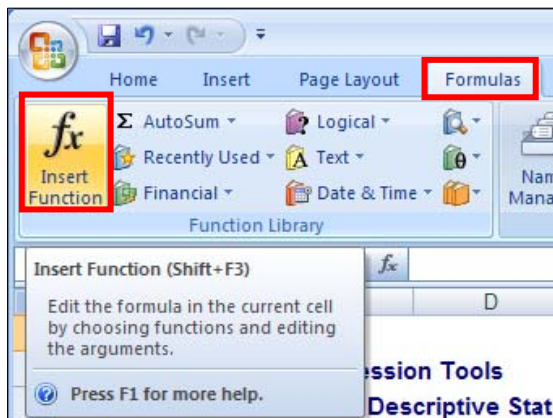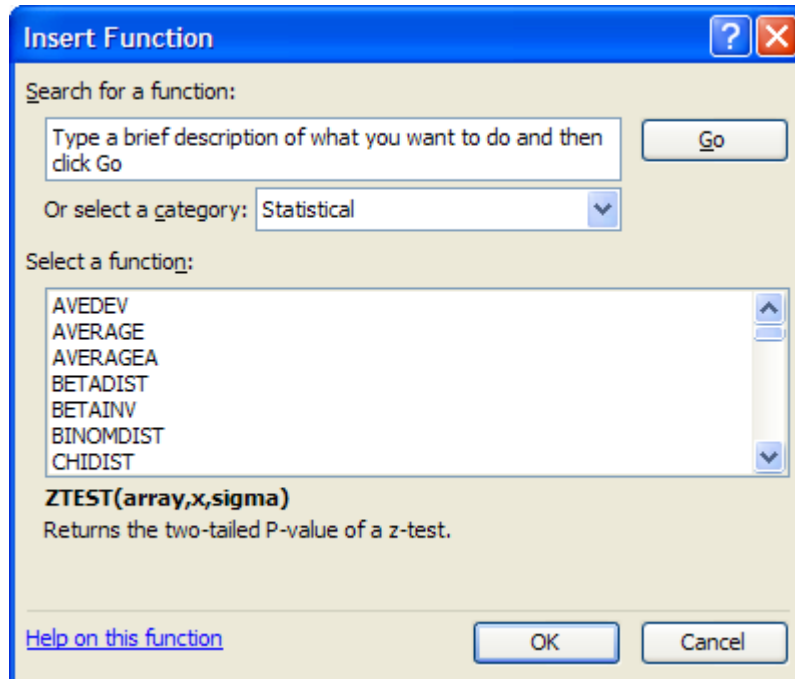
End of the Excel Statistics & Regression Tools Skills Review Summary