

Using Minitab for regression analysis, a brief tutorial.

USING MINITAB FOR REGRESSION ANALYSIS, A BRIEF TUTORIAL.

In this tutorial, you will learn:

1. About the “Basic Features” of Minitab.
2. How to import data from
 - Excel worksheets;
 - Comma-Separated Variable (CSV) formatted files, and
 - HTML Tables.
3. How to make sure the data is stored in the appropriate type, “Text” or “Numeric”.
4. How to do correlation analysis in Minitab.
5. How to do a multiple regression in Minitab.
6. How to calculate basic statistics for a column of data.
7. How to create dummy variables, from a column of data.

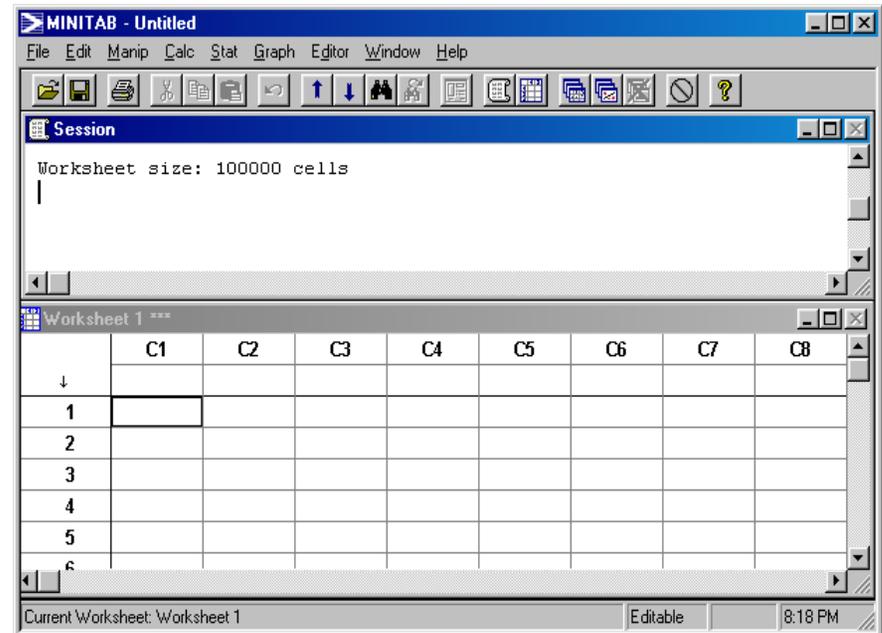
The graphics used in this document are from Minitab version 12. I checked out the ISOM Computer Labs, which have version 13. I didn't notice any significant differences in the menus. There may be small differences in the dialogues.

Minitab, basic features

The Minitab start-up windows

When you start up Minitab, what you will see is something like the image of a window on the right.

Note that Minitab has a set of pull down menus like every other Windows based program you have ever used. Mostly, you will use the “File”; “Manip”; “Stat” and “Window” pull down menus.

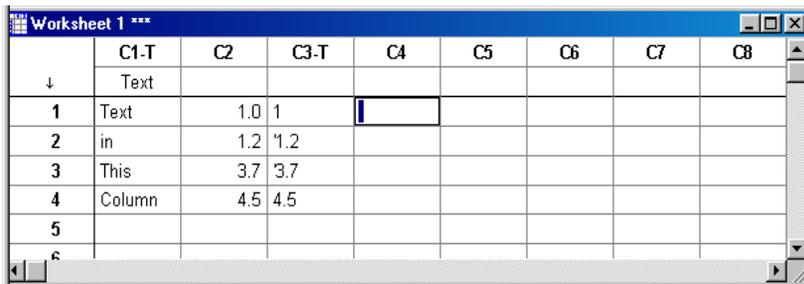


Minitab has four main window types, the “Session” window, which will display the results of your actions; the “Worksheet” window, which is where you will keep and manipulate your data; the “History” window that displays a list of commands that you have used in your session and the “Information”

Using Minitab for regression analysis, a brief tutorial.

window which displays a summary of each worksheet. Since the “Session” and “Worksheet” windows are the ones you will want to use first, these are generally displayed first when you open up Minitab.

The Worksheet window



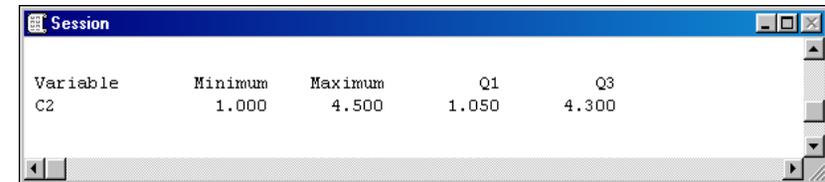
	C1-T	C2	C3-T	C4	C5	C6	C7	C8
↓	Text							
1	Text	1.0	1					
2	in	1.2	1.2					
3	This	3.7	3.7					
4	Column	4.5	4.5					
5								
6								

This is the window into which you will enter data in two main types, “Text” and “Numeric”. Note from the example above that text is identified in the Column names “C1-T” and “C3-T” by the “-T” component.

Often when we import data into worksheets we will see that *numeric* data is seen by Minitab as being *text* and when we try to perform operations on it that assume it is *numeric* they won’t work. For this reason in the third main section of this

document we will describe the way in which you can convert text columns to numeric columns.

The Session window



Variable	Minimum	Maximum	Q1	Q3
C2	1.000	4.500	1.050	4.300

I ran some “Basic Statistics” to put some results into the Session window to show how this window is used.

Other Windows

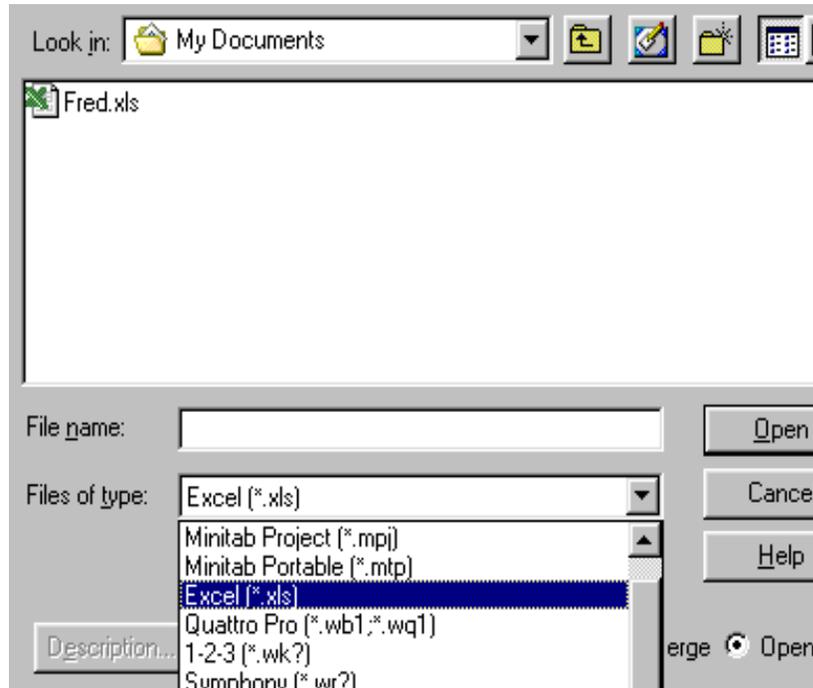
I do not think you will need to use the “History” or the “Information” windows, so I won’t describe them further here. Note, however, that Minitab has “help” which you can use to further inform yourself about these and the other windows that you may encounter in Minitab.

Importing Data to a Worksheet

Importing data from Excel

Using Minitab for regression analysis, a brief tutorial.

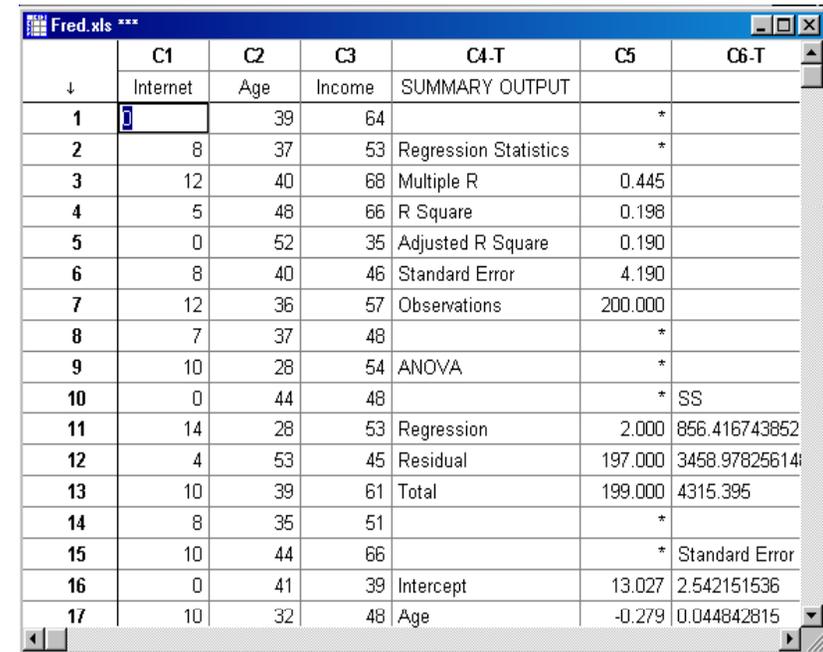
Use the “File” pull down menu and select the “Open Worksheet” option. This will bring up a dialogue to open up a file, which will look like the following:



Note that there are a number of different formats, which can be imported, essentially by opening them from Minitab.

I will select Fred.xls, but first note that Minitab will import the all of the Spreadsheets in an Excel Workbook, so make sure that the data you want is in the first Spreadsheet of the Excel

Workbook (it will be easier to locate!!). Also, make sure that you store the data in a simple table with column headings on the first row. Minitab will create a new “Worksheet” window containing the data from “Fred.xls”, which should look like the following.



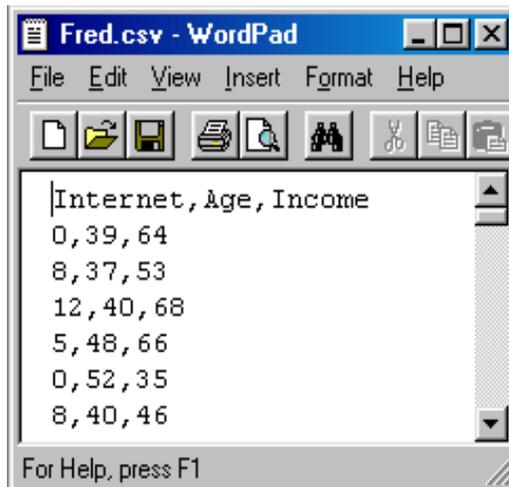
	C1	C2	C3	C4-T	C5	C6-T
↓	Internet	Age	Income	SUMMARY OUTPUT		
1		39	64		*	
2	8	37	53	Regression Statistics	*	
3	12	40	68	Multiple R	0.445	
4	5	48	66	R Square	0.198	
5	0	52	35	Adjusted R Square	0.190	
6	8	40	46	Standard Error	4.190	
7	12	36	57	Observations	200.000	
8	7	37	48		*	
9	10	28	54	ANOVA	*	
10	0	44	48		*	SS
11	14	28	53	Regression	2.000	856.416743852
12	4	53	45	Residual	197.000	3458.97825614
13	10	39	61	Total	199.000	4315.395
14	8	35	51		*	
15	10	44	66		*	Standard Error
16	0	41	39	Intercept	13.027	2.542151536
17	10	32	48	Age	-0.279	0.044842815

You can select columns and use “Edit”, “Delete” to remove them if you want to be tidy. Note that the column names “Internet”, “Age” and “Income” have now been created and can be used to reference the columns as well as the column

Using Minitab for regression analysis, a brief tutorial.

numbers “C1”, “C2” and “C3”. Also note that on this occasion the columns have come in as “Numeric” data, which is what we want if we are to analyze correlation and do regressions on the data.

Getting data from Comma Separated Variable (CSV) formatted files



A CSV file looks like the example on the left when it is opened up in WordPad. As the name suggests the variables are separated by commas and the data is in rows with one set of data to a row.

The easiest way to deal with CSV files is to import them into Excel, by opening the file in Excel. Then save the data as an Excel Workbook and follow the procedures for importing data from Excel.

Getting data from an HTML table on the Web

Open a new Workbook in Excel.

In your browser, go to a web page with some data in a table. Select the data using the mouse (right hand mouse button depressed at start and cursor moved to the end of the data you want to acquire). See bottom of page.

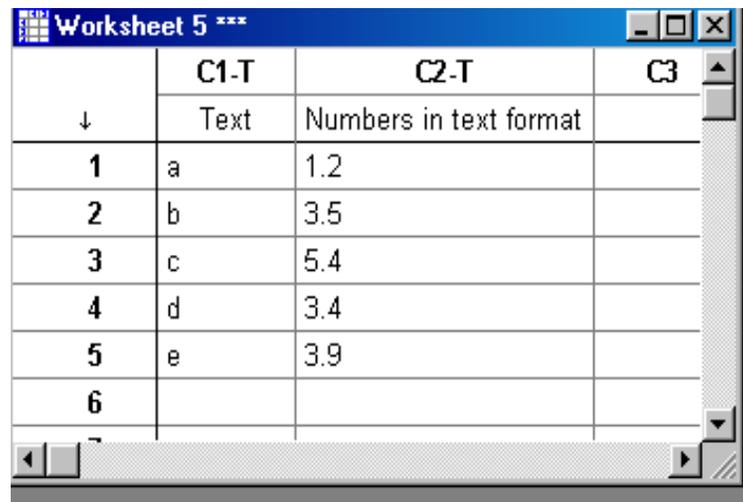
Do “ctrl-C” or “Edit”, “Copy” to copy the selected data to the clipboard. Return to Excel. Click into Row 1, Column A and either use “ctrl-V” or “Edit”, “Paste” to past in the data from the clipboard. Name your columns, if necessary. Close the Workbook and follow the procedures for importing data from Excel.

CONSTRUCTION	134.6	134.5	134.6	135.9	137.2	136.9	137.1	136.3	137.5	138.4	139.2	140.2
General Building Contractors	28.1	28.0	27.9	27.8	27.6	27.4	27.0	27.2	27.3	27.4	27.7	27.9
Heavy Construction, ex. Bldg.	15.8	15.7	15.7	16.3	16.6	16.8	16.8	16.8	16.7	16.5	16.5	16.6
Special Trade Contractors	90.7	90.8	91.0	91.8	93.0	92.7	93.3	92.3	93.5	94.5	95.0	95.7
MANUFACTURING	438.8	437.2	434.3	430.1	426.9	423.2	420.2	417.7	416.5	414.1	412.1	409.7
DURABLE GOODS	279.9	279.0	277.1	274.0	271.5	268.8	267.4	264.9	263.9	262.0	260.2	258.5
Stone, Clay & Glass	9.3	9.2	9.1	9.0	9.1	9.0	9.0	9.0	9.0	9.0	9.0	9.1

Using Minitab for regression analysis, a brief tutorial.

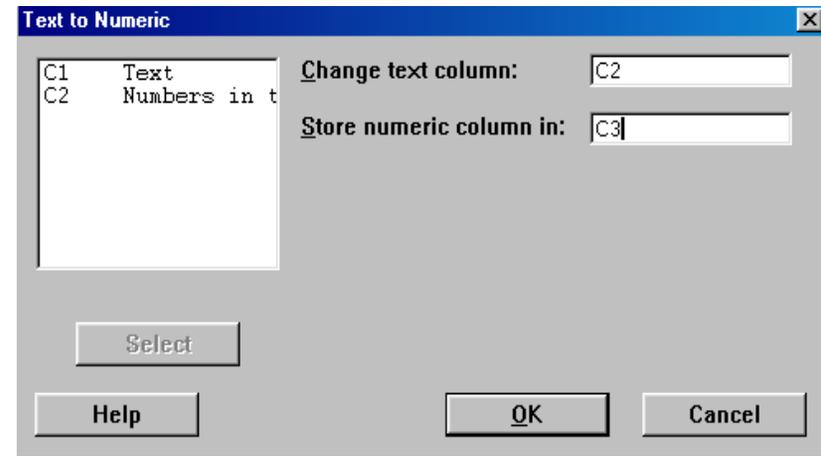
Converting Numbers in *Text* format to *Numeric* format

Assuming you have imported some data and Minitab has incorrectly given Text format to some numbers. Before you can work on the numbers you need to turn them back into Numeric format.



	C1-T	C2-T	C3
↓	Text	Numbers in text format	
1	a	1.2	
2	b	3.5	
3	c	5.4	
4	d	3.4	
5	e	3.9	
6			

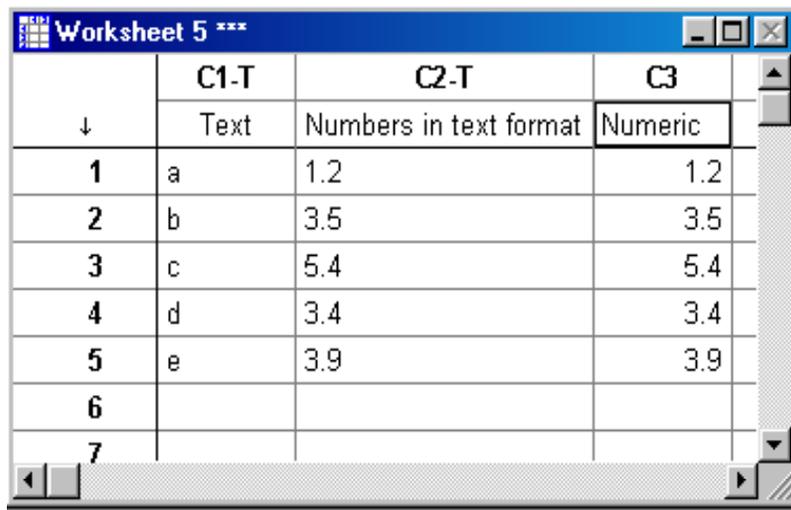
Use the “Manip” pull down menu option and its menu item “Change Data Type”. Select “Text to Numeric...” The following dialogue will appear:



Fill in the name of the column you want to convert to numeric in the top right text entry box. I selected “C2” before capturing the image given above. Also fill in the name of an empty column, such as “C3” in the lower right text entry box. You can overwrite a column, so beware! You could in this instances have chosen “C2” as the destination, assuming you did not want to keep the textual column!

Select OK and the worksheet will be changed to the following:

Using Minitab for regression analysis, a brief tutorial.



	C1-T	C2-T	C3
↓	Text	Numbers in text format	Numeric
1	a	1.2	1.2
2	b	3.5	3.5
3	c	5.4	5.4
4	d	3.4	3.4
5	e	3.9	3.9
6			
7			

Column “C3” is now ready for you to work on it as numbers. You may find that giving the column a meaningful name will help you later and you can do this by “clicking” in the cell under the column heading “C3” and entering something like “Numeric”, but more meaningful in the context of your problem!!!

Correlation Analysis in Minitab

We are interested in the results of correlation analysis in the context of a multiple regression analysis. Significantly correlated independent variables (Collinearity) may result in an R^2 value which is misleading. So it is good practice to check

the degree of correlation between the independent variables that you might include in your model.

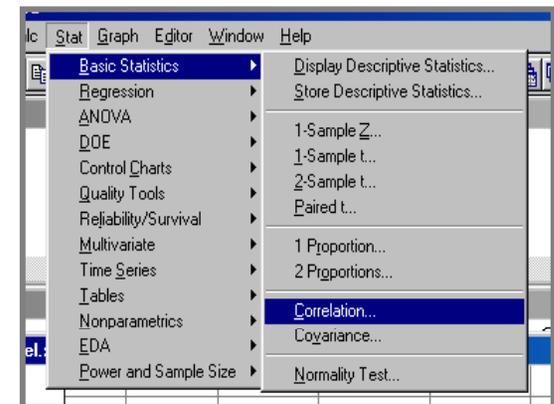
When I open up the file Corel.xls, I get the following Minitab worksheet:



↓	C1	C2	C3	C4	C5	C6	C7	C8
	Year	Bus_Fail	Real_D_P	Wage_Sals	Unemploy	Ret_Sales	House_Sts	C_P_Index
1	1968	0.3	3.5	10.9	3.8	6.8	22.0	3.6
2	1969	-5.1	5.5	8.7	4.5	6.0	20.0	4.0
3	1970	24.3	6.2	12.1	4.4	6.9	6.9	4.6
4	1971	4.0	2.2	8.4	5.7	2.3	-9.5	3.3
5	1972	1.2	5.9	9.9	6.2	8.9	22.6	2.9
6	1973	-4.8	5.6	11.3	6.2	11.3	7.0	4.8
7	1974	-4.9	7.8	15.8	5.5	12.6	7.4	7.6
8	1975	5.1	4.7	19.4	5.3	16.9	-17.3	10.8
9	1976	-4.9	0.7	16.4	6.9	14.5	4.2	10.8
10	1977	40.0	5.3	15.4	7.1	10.8	-7.3	7.5
11	1978	40.4	2.8	10.0	8.1	8.4	-10.1	8.0
12	1979	2.7	3.5	8.8	8.4	11.7	-7.3	9.0
13	1980	15.8	3.2	11.0	0.5	11.9	-13.4	8.8

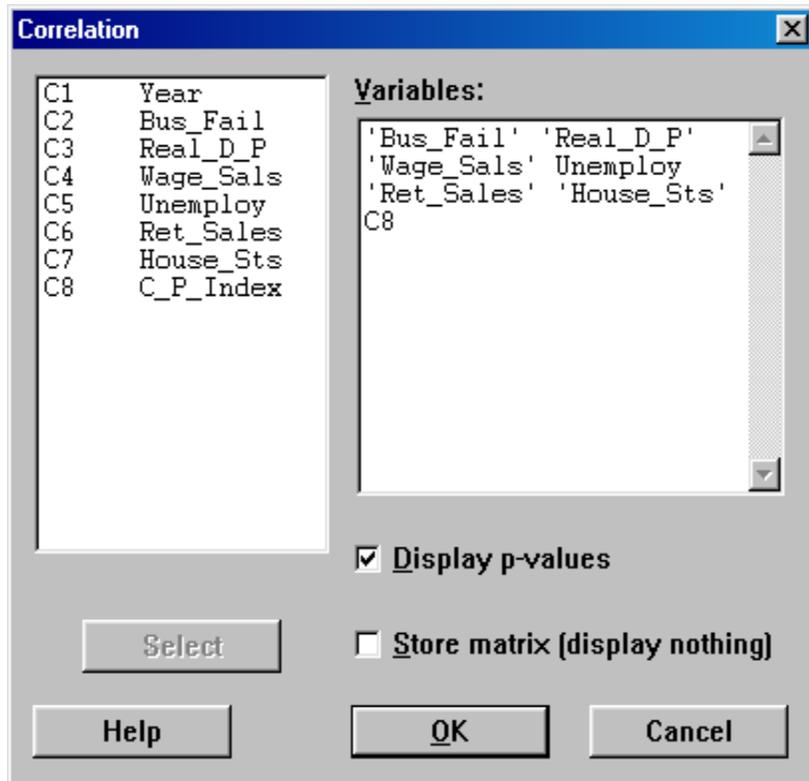
Using the “Stat” pulldown menu, Basic Statistics option and “Correlation” sub-option:

I get the “Correlation” dialogue on the next



Using Minitab for regression analysis, a brief tutorial.

page. I can either type in the “C2”, “C3” etc., for the variables I want to test the correlation between, or I can double click on the variable in the list on the left hand side of the dialogue.



When I have selected the variables of interest (year won't tell me much!!). I click on the “OK” button and Minitab will fill the bottom of the “Session” window with a correlation matrix such as the one shown at the top of the next column.

The screenshot shows the 'Session' window in Minitab displaying a correlation matrix. The matrix is upper triangular, with the diagonal elements all equal to 1.000. The off-diagonal elements represent the correlation coefficients between the selected variables. The variables are: Bus_Fail, Real_D_P, Wage_Sal, Unemploy, Ret_Sale, and House_St. The matrix is as follows:

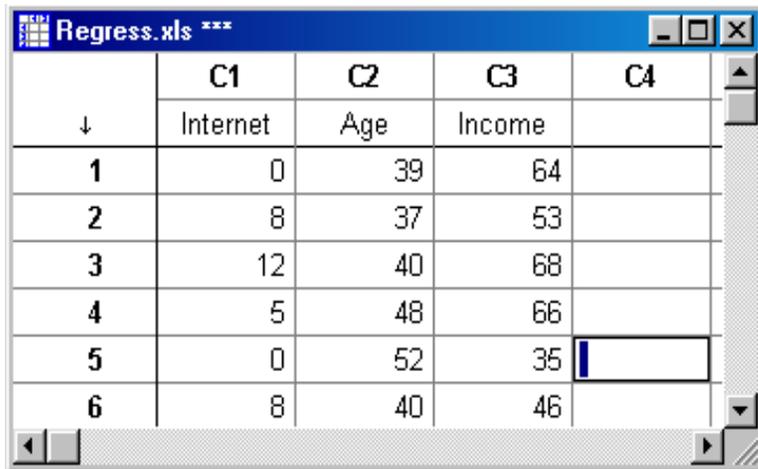
	Bus_Fail	Real_D_P	Wage_Sal	Unemploy	Ret_Sale	House_St
Bus_Fail	1.000					
Real_D_P	-0.051	1.000				
Wage_Sal	0.022	0.133	1.000			
Unemploy	0.140	-0.097	0.033	1.000		
Ret_Sale	-0.108	0.038	0.769	0.094	1.000	
House_St	-0.467	0.359	-0.279	-0.091	-0.342	1.000

Below the matrix, the text 'Cell Contents: Correlation' and 'P-Value' is displayed.

Using Minitab for regression analysis, a brief tutorial.

Multiple Regression in Minitab

Assume that using the data in the Multi.xls and that I have imported it into a worksheet which is shown below:

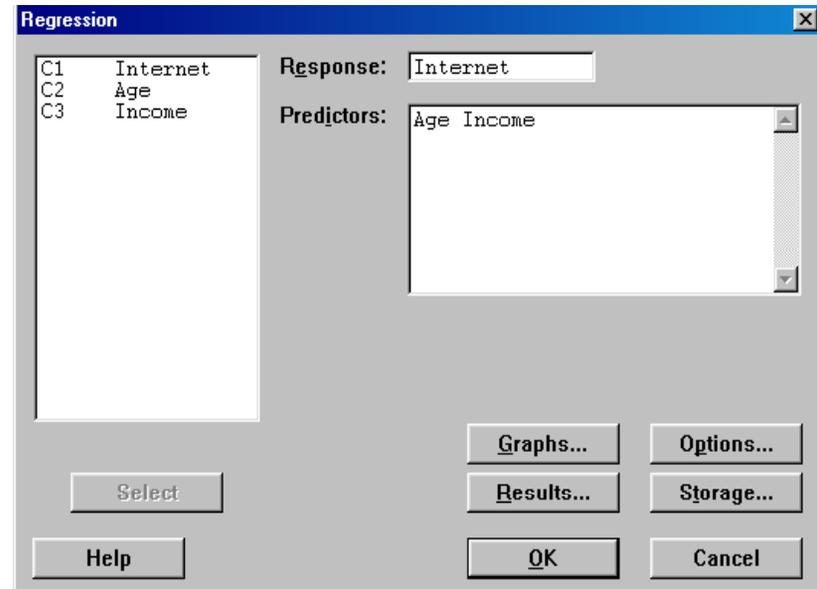


	C1	C2	C3	C4
↓	Internet	Age	Income	
1	0	39	64	
2	8	37	53	
3	12	40	68	
4	5	48	66	
5	0	52	35	
6	8	40	46	

There are 200 rows of data, but you get the idea!

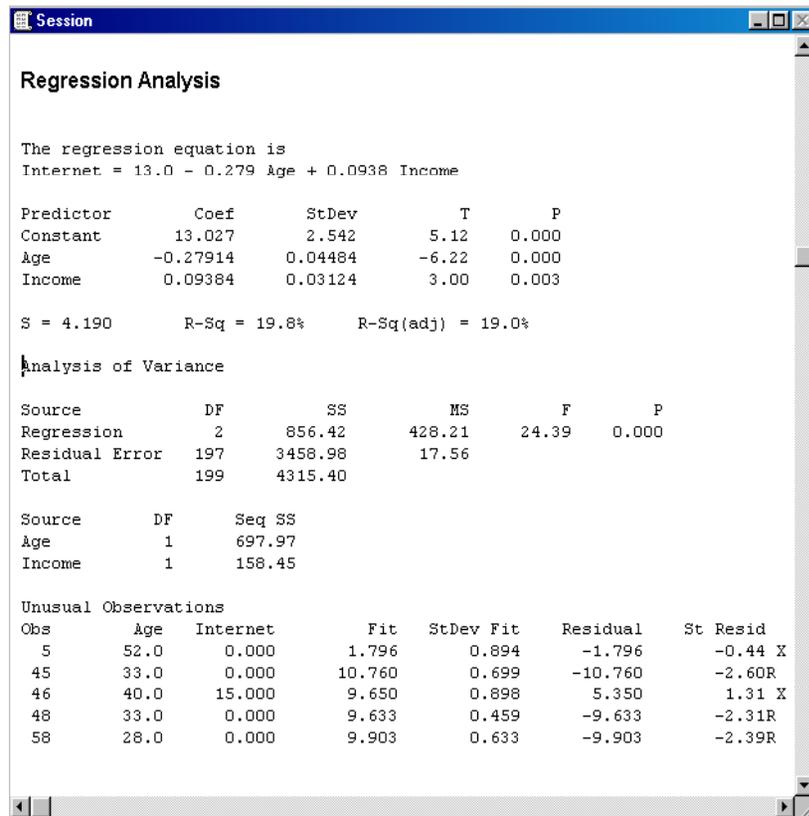
Using “Stat”, with the option “Regression” and the sub-option “Regression” you will get up the Regression dialogue, which looks like the image in the right hand column.

Select “Internet” as the Response and “Age” and “Income” as the Predictors.



Depress the “OK” button and the results of your analysis will be displayed in the “Session” window and will look as it does on the following page.

Using Minitab for regression analysis, a brief tutorial.



The regression equation is
Internet = 13.0 - 0.279 Age + 0.0938 Income

Predictor	Coef	StDev	T	P
Constant	13.027	2.542	5.12	0.000
Age	-0.27914	0.04484	-6.22	0.000
Income	0.09384	0.03124	3.00	0.003

S = 4.190 R-Sq = 19.8% R-Sq(adj) = 19.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	856.42	428.21	24.39	0.000
Residual Error	197	3458.98	17.56		
Total	199	4315.40			

Source	DF	Seq SS
Age	1	697.97
Income	1	158.45

Unusual Observations

Obs	Age	Internet	Fit	StDev Fit	Residual	St Resid
5	52.0	0.000	1.796	0.894	-1.796	-0.44 X
45	33.0	0.000	10.760	0.699	-10.760	-2.60R
46	40.0	15.000	9.650	0.898	5.350	1.31 X
48	33.0	0.000	9.633	0.459	-9.633	-2.31R
58	28.0	0.000	9.903	0.633	-9.903	-2.39R

If you want a different level of detail, click on “Results...” on the Regression dialogue and select a different level from the 4 levels available.

Also check out the “Options...” and “Graphs...” buttons for all the good things that you can specify. I don’t think you will need the ”Storage...” options, but if you do try the help menu.

Confidence Intervals on the Regression Coefficients

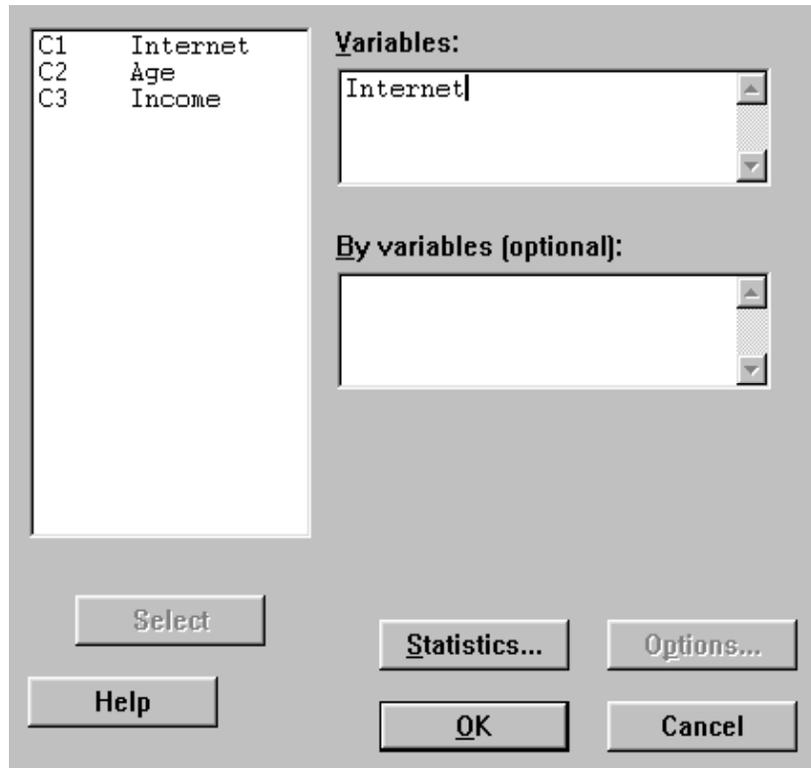
One thing that this analysis does not provide is Confidence Intervals on the Regression Coefficients. Note that in the “Session” window displayed on the left you have the “Coef” values and by their side you have the “StDev” (which may have been referred to as the “Standard Error” in other places). Use these values to calculate the confidence intervals for the three coefficients.

Calculating Basic Statistics for a Column of Data

Using the data in “Regress.xls”, select the “Stat” pull down menu, select the “Basic Statistics” option and either the “Display Descriptive Statistics” or the “Store Descriptive Statistics” sub-options. The former will print the statistics to the “Session” window and the latter will insert them in the worksheet.

“Display” will give you a predefined list of statistics, while “Store” has an option in its dialogue to select the statistics that you want to store. The “Store” dialogue set up to give statistics on the “Internet” column is show in the next image.

Using Minitab for regression analysis, a brief tutorial.



Note the “Statistics” button and the fact that you will need to select the Dialogue to select the statistics you want to store.

I ran the “Display” option for the “Internet” column and obtained the following in the “Session” window:

The screenshot shows the Minitab 'Session' window with the following descriptive statistics for the 'Internet' variable:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Internet	200	7.195	8.000	7.144	4.657	0.329

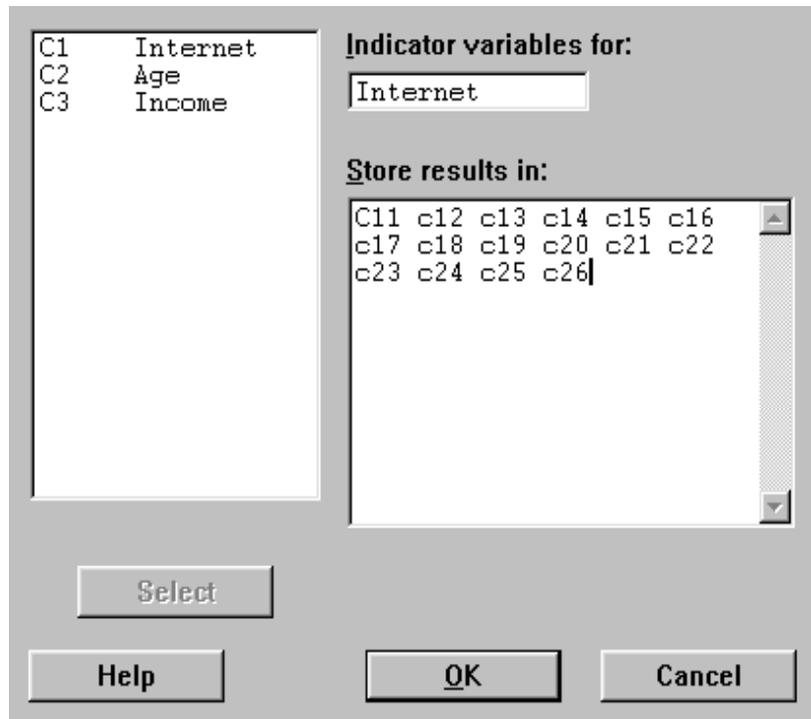
Variable	Minimum	Maximum	Q1	Q3
Internet	0.000	18.000	4.000	11.000

Note that the min is 0 and the max is 18.

Creating Dummy Variables

Again using the “Regress.xls” worksheet. To create dummy variables for the “Internet” variable. Select the “Calc” pull down menu and the “Make Indicator Variables” option. The following dialogue will appear. Select “Internet” as the variable for which to make Indicator Variables. I need nineteen columns for the results, so I select “Store results in:” as “c11 c12 c13 c14 c15 c16 c17 c18 c19 c20 c21 c22 c23 c24 c25 c26”. Note that this is 16 columns, because the “Internet” column only contains 16 distinct integers between 0 and 18 inclusive. Also note that I could not find a way to specify a range of columns between c11 and c26.

Using Minitab for regression analysis, a brief tutorial.



At this point I am beginning to think this operation might be easier in Excel.

Any way, when I click on OK, lo and behold I get dummy variables in columns 11 through 26...

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
1	0	39	64								1	0	0	0	0	0	0	0
2	8	37	53								0	0	0	0	0	0	1	0
3	12	40	68								0	0	0	0	0	0	0	0
4	5	48	66								0	0	0	1	0	0	0	0
5	0	52	35								1	0	0	0	0	0	0	0
6	8	40	46								0	0	0	0	0	0	1	0
7	12	36	57								0	0	0	0	0	0	0	0
8	7	37	48								0	0	0	0	0	1	0	0
9	10	28	54								0	0	0	0	0	0	0	0
10	0	44	48								1	0	0	0	0	0	0	0
11	14	28	53								0	0	0	0	0	0	0	0
12	4	53	45								0	0	1	0	0	0	0	0
13	10	39	61								0	0	0	0	0	0	0	0
14	8	35	51								0	0	0	0	0	0	1	0
15	10	44	66								0	0	0	0	0	0	0	0
16	0	41	39								1	0	0	0	0	0	0	0
17	10	32	48								0	0	0	0	0	0	0	0

Last words

The data for these worked examples is available from

<http://www-unix.oit.umass.edu/~adhall/MBA/minitab.html>

note the page is sensitive to case so type precisely as seen here.

Saving Minitab work in Minitab project form saves the data and the history of the actions you have taken, so if you create graphs and don't close them they will be there when you open up the project again.

