Computer Handout 01: Simple Regression in Excel

Diego Escobari

Econ 3341

The goal of this Computer Handout 01 is to use MS Excel to implement the following Ordinary Least Squares estimators:

$$b_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$b_1 = \overline{Y} - b_2\overline{X}$$

The file BEAUTY.DES contains a short description of the 17 variables included in the corresponding MS Excel file "BEAUTY.xls". In particular, in column 1 we have `wage` and in column 5 we have `exper` (experience).

To make computations easier, copy columns 1 and 5 (wage and exper) from the file BEAUTY.xls into a separate sheet or to a separate MS Excel file. Then label the two variables y = `wage` and x = `exper`.

To implement the formulas above, let's first calculate the average of x and the average of y. Just go to the end of these two columns and type =average(A2:A1261). Just make sure to select all the corresponding observations to calculate the average.

	A	B	C
1	wage	exper	
2	5.73	30	
3	4.28	28	
4	7.96	35	
5	11.57	38	
6	11.42	27	
7	3.91	20	
1259	3.29	35	
1260	2.31	15	
1261	1.92	24	
1262	=AVERAGE(A2:A1261)		
1263			

On column C, let's calculate the deviations from the mean (Xi-Xbar). Therefore on cell C2 type:

	A	B	C	D
1	wage	exper	(Xi-Xbar)	
2	5.73	30	=(B2-B1262)	
3	4.28	28		

To make sure you lock the formula of the average, make sure you include the dollar signs $ in the formula.

Double click on the bottom-right corner of cell C2. This will automatically copy the formula on C2 to all cells in column C calculating the values of (Xi-Xbar) for all the observations in the sample.

Do the same thing on column D for the deviations from the mean for the variable y: (Y-Ybar).

On column E, get the product of columns C and D. On column F get the square of column C. This can be done by writing the formulas =C2*D2 and =D2^2 on cells E2 and F2 respectively. Then just pull down the formulas by double-clicking on the lower-right corner of cells E2 and F2.

	A	B	C	D	E	F
1	wage	exper	(Xi-Xbar)	(Yi-Ybar)	(Xi-Xbar)(Yi-Ybar)	(Xi-Xbar)^2
2	5.73	30	11.7937	-0.57669	-6.801286092	=C2^2
3	4.28	28	9.79365	2.02669	19.84869879	

Once you pull down the formulas, you need to calculate $\Sigma(X_i\text{-}Xbar)(Y_i\text{-}Ybar)$ and $\Sigma(X_i\text{-}Xbar)^2$, which are the numerator and the denominator in the formula for b_1 above. In order to do this, type =SUM(E2:E1261) and =SUM(F2:F1261) on cells E1263 and F1263.

	A	B	C	D	E	F	G
1	wage	exper	(Xi-Xbar)	(Yi-Ybar)	(Xi-Xbar)(Yi-Ybar)	(Xi-Xbar)^2	
2	5.73	30	11.7937	-0.57669	-6.801286092	139.0902	
3	4.28	28	9.79365	-2.02669	-19.84869879	95.915596	
4	7.96	35	16.7937	1.65331	27.7651028	282.02671	
5	11.57	38	19.7937	5.26331	104.1801107	391.78861	
6	11.42	27	8.79365	5.11331	44.96465835	77.328294	
7	3.91	20	1.79365	-2.39669	-4.298825775	3.2171832	
1259	3.29	35	16.7937	-3.01669	-50.66124641	282.02671	
1260	2.31	15	-3.20635	-3.99669	12.81478534	10.280675	
1261	1.92	24	5.79365	-4.38669	-25.41495276	33.56639	
1262	6.30669	18.2063					
1263			Summation:		16470.87048	=SUM(F2:F1261)	
1264							

Once you have these calculations done, it is easy to get the values for b_0 and b_1 by just implementing the last step of the formulas above. You can do this in one of the cells.

For b_2: (notice that b_2 is named b1 in the Excel file)

	Summation:		16470.87048	180194.35
	b1:	=E1263/F1263		
	b0:			

And for b_1: (b_1 is names b0 in the Excel file)

1261	1.92	24	5.79365	-4.36669	-25.41495276	3.
1262	6.30669	18.2063				
1263			Summation:		16470.87048	18(
1264						
1265			b1:		0.091406143	
1266			b0:	=A1262-E1265*B1262		

This last equation gives us a value for $b_0 = 4.6425$

These two values can be summarized in the following formula:

```
wage = 4.6425 + 0.0914 * exper
```

The interpretation is as follow:

- An individual who has no experience (`exper` = `0`) will have an hourly wage of 4.64 dollars per hour.

- For the average individual in our sample, an additional year of experience increases the hourly wage by 0.0914 dollars. This is the same as an increase of 9.14 cents. This means that if you have 10 years more of experience you will be making 91.4 cents per hour more. Do you think that the magnitude of this coefficient is reasonable? Recall that this result assumes that everything else remains constant.

This can also be done using the MS Excel graphing tools. Just graph a simple scatter plot of `wage` on the vertical axis and `exper` on the horizontal axis, then under the "Design" menu go to the "Chart Layouts" section and pick Chart Layout number 9. You should then be getting the following graph:

Excel Linear Regression

$y = 0.0914x + 4.6425$
$R^2 = 0.0551$
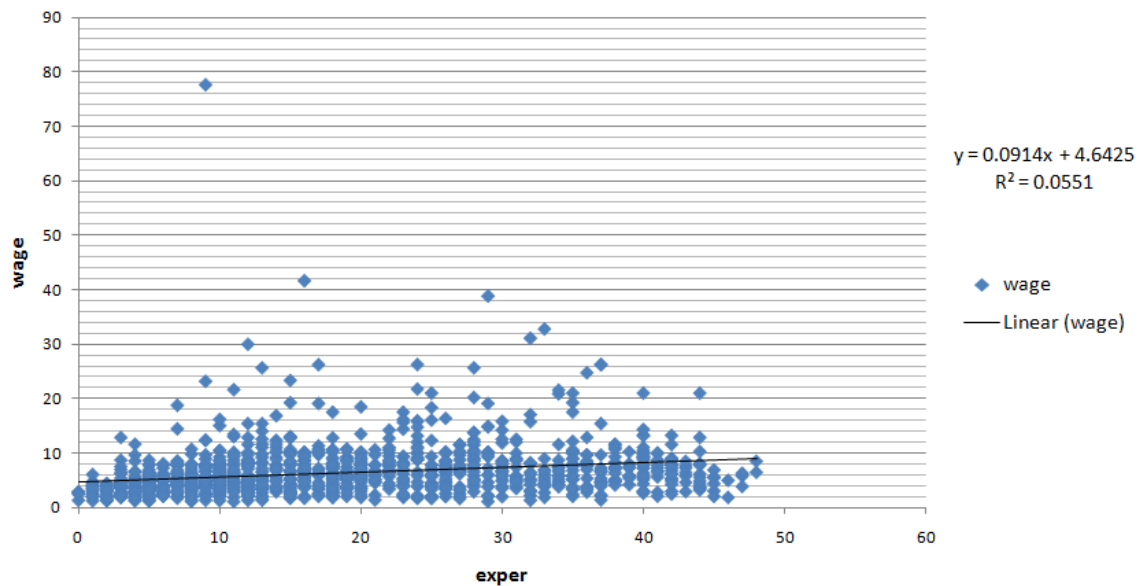
To use the Excel built-in procedure to run a regression you first need to make sure that the "Data Analysis" tool is active. Click the "Office Button" in the upper right corner of the screen. Then go to "Excel Options" then "Add-Ins" then select "Analysis ToolPak" and then click "Go." A new window will open and you will have to select "Analysis ToolPak" again and click "OK." This should activate the "Data Analysis" tool. In the newer versions of Excel you need to go to "File" and then "Options." The rest of the steps are the same as before.

Now, go "Data" in the main menu and on the right hand side you will find the icon "Data Analysis." Click it and then go to "Regression" and OK. You will have to select the input range for both variables Y and X and then a single cell for the output. Once you are done, you should be getting the following output:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.234632212
R Square	0.055052275
Adjusted R Sq	0.054301124
Standard Error	4.532333981
Observations	1260

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1505.538744	1505.538744	73.2905746	3.21658E-17
Residual	1258	25841.90055	20.54205132		
Total	1259	27347.4393			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.642518315	0.232574148	19.96145467	3.12383E-77	4.18624238	5.098794249	4.18624238	5.098794249
X Variable 1	0.091406143	0.010677051	8.56099145	3.21658E-17	0.070459355	0.112352932	0.070459355	0.112352932