Lab 9 – OLS Regressions I

This lab will cover how to perform a simple OLS regression, both by reproducing the formulas seen in class and by using Excel's Data Analysis option.

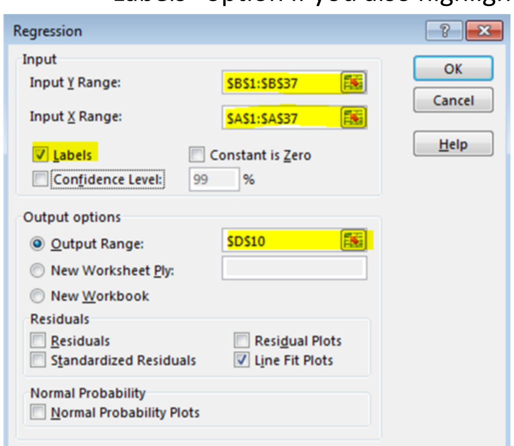
Lab 9 Quick View

- These OLS estimates can be calculated using the excel formulas learned up to this point:

$$\hat{\beta}_2 = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sum (X-\bar{X})^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

- Sample means such as $\bar{Y}$ can be calculated using the **=average(X,Y)** formula, generally at the end of the data
- Remember that placing a "$" before a letter or number in a formula makes sure that letter or number does not chance if that cell is copied.
- OLS regressions can be automatically calculated by opening the "Data" menu and selecting "Data Analysis".
    - When the "Data Analysis" box opens, select "Regression" and press "OK". Select both your Y (dependent) and X (independent) variables, clicking the "Labels" option if you also highlighted the variable labels.



- By clicking the "Confidence Level" box in the Regression window, you can choose a Confidence Interval percentage that will be shown for both β_1 and β_2.

Written by Lorne Priemaza

A) OLS Regressions Using Formulas

Recall the following formulas for OLS regressions seen in class. Notice that these can be calculated using the excel formulas learned up to this point:

$$\hat{\beta}_2 = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2}$$

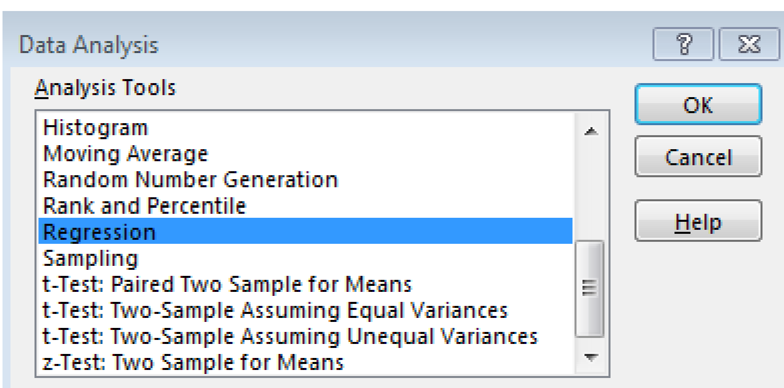
$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$$

- Sample means such as $\overline{Y}$ can be calculated using the **=average(X,Y)** formula, generally at the end of the data
- Deviation from means such as $(X - \overline{X})$ can be calculated using **=CellA-CellB**, where CellA is the X value that changes with each observation and CellB (the sample mean) is entered with $ signs so that when it is copied it always refers to the sample mean. (For example, F48 will always refer to cell F48).
- Through multiplication, subtraction and division, you can then calculate these formulas (see the practice lab below if needed).

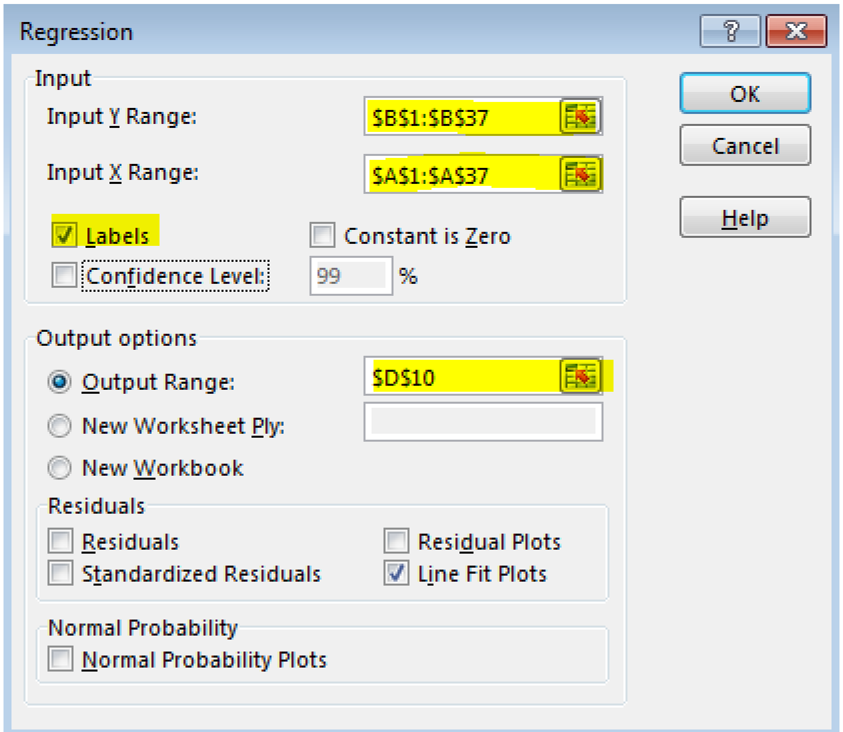
B) OLS Regressions Using Data Analysis

OLS regressions can be automatically calculated by opening the “Data” menu and selecting “Data Analysis”.

When the “Data Analysis” box opens, select “Regression” and press “OK”. Select both your Y (dependent) and X (independent) variables, clicking the “Labels” option if you also highlighted the variable labels.



Written by Lorne Priemaza



C) Confidence Interval Options

By clicking the “Confidence Level” box in the Regression window, you can choose a Confidence Interval percentage that will be shown for both β_1 and β_2. Some of the data shown may not be useful to you yet, but in the output:

R Square is the R^2 calculation for goodness of fit

The Intercept row gives data referring to $\hat{\beta}_1$ and **X** refers to $\hat{\beta}_2$.

Coefficients are the estimated $\hat{\beta}_1$ and $\hat{\beta}_2$.

Standard Error is the standard error for each estimate.

t Stat is the test t statistic of the hypothesis test that the estimate is equal to zero.

Upper and **Lower** percentages give upper and lower bounds of confidence intervals.

Written by Lorne Priemaza

Econ 299 Practice Lab 9

A) Download (highly recommended) or copy the following Data Sets into excel:

X	Y
15	36
16	34.6
35	46
42	53.4
26	41.6
18	38.4
21	34.5
17	39.3
41	67
38	45
38	51.6
36	61
20	45
20	37
40	50
20	44
29	49.2
41	58.8
38	59.2
33	52.4
27	42.9
25	50
28	43.6
20	46
18	43
20	36
15	40
37	59.3
30	45
39	48.4
18	38.4
35	54
41	53.7
29	44.3
29	54
18	36.6

Written by Lorne Priemaza

B) Move these data sets into columns B and C either by cut and pasting or by inserting a new column.

C) Under each data set, calculate the sample mean of each variable. For example, in cell B38, the sample mean is calculated as =AVERAGE(B2:B37). Put the appropriate label in column A.

D) In columns D and E, calculate each variables deviation from its mean. For example, cell D1 would have a title like "X-Xbar", and cell D2 would have the formula =B2-B$38, which then can be copied for each observation.

E) In column F we calculate the product $(X-\overline{X})(Y-\overline{Y})$ by multiplying columns D and E. For example, cell F2 would have the formula =D2*E2.

F) In column G we calculate the product $(X-\overline{X})^2$ by multiplying columns D and E. For example, cell G2 would have the formula =D2^2.

G) Underneath columns F and G, calculate the sum of each column. For example, cell F39 would have the formula =sum(F2,F37). Put the appropriate title in column A.

H) In Cell A41, put a title such as "Beta2hat" and calculate it in Cell B41 as =F39/G39.

I) In Cell A42, put a title such as "Beta1hat" and calculate it in Cell B42 as =C38-B41*B38.

J) Use the "Regression" option under "Data Analysis" to estimate the real function $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$.

Written by Lorne Priemaza

Econ 299 Practice Lab 9 Answers

A	B	C	D	E	F	G
	X	Y	x-xbar	y-ybar	(x-xbar)(Y-Ybar)	(X-Xbar)-squared
	15	36	-13.1389	-10.64444444	139.8561728	172.6304012
	16	34.6	-12.1389	-12.04444444	146.2061728	147.3526235
	35	46	6.861111	-0.644444444	-4.421604938	47.07484568
	42	53.4	13.86111	6.755555556	93.63950617	192.1304012
	26	41.6	-2.13889	-5.044444444	10.78950617	4.574845679
	18	38.4	-10.1389	-8.244444444	83.58950617	102.7970679
	21	34.5	-7.13889	-12.14444444	86.69783951	50.96373457
	17	39.3	-11.1389	-7.344444444	81.80895062	124.0748457
	41	67	12.86111	20.35555556	261.7950617	165.408179
	38	45	9.861111	-1.644444444	-16.21604938	97.24151235
	38	51.6	9.861111	4.955555556	48.86728395	97.24151235
	36	61	7.861111	14.35555556	112.8506173	61.7970679
	20	45	-8.13889	-1.644444444	13.38395062	66.24151235
	20	37	-8.13889	-9.644444444	78.49506173	66.24151235
	40	50	11.86111	3.355555556	39.80061728	140.6859568
	20	44	-8.13889	-2.644444444	21.52283951	66.24151235
	29	49.2	0.861111	2.555555556	2.200617284	0.741512346
	41	58.8	12.86111	12.15555556	156.3339506	165.408179
	38	59.2	9.861111	12.55555556	123.8117284	97.24151235
	33	52.4	4.861111	5.755555556	27.97839506	23.63040123
	27	42.9	-1.13889	-3.744444444	4.264506173	1.297067901
	25	50	-3.13889	3.355555556	-10.53271605	9.852623457
	28	43.6	-0.13889	-3.044444444	0.422839506	0.019290123
	20	46	-8.13889	-0.644444444	5.245061728	66.24151235
	18	43	-10.1389	-3.644444444	36.95061728	102.7970679
	20	36	-8.13889	-10.64444444	86.63395062	66.24151235
	15	40	-13.1389	-6.644444444	87.30061728	172.6304012
	37	59.3	8.861111	12.65555556	112.142284	78.51929012
	30	45	1.861111	-1.644444444	-3.060493827	3.463734568
	39	48.4	10.86111	1.755555556	19.06728395	117.9637346
	18	38.4	-10.1389	-8.244444444	83.58950617	102.7970679
	35	54	6.861111	7.355555556	50.46728395	47.07484568
	41	53.7	12.86111	7.055555556	90.74228395	165.408179
	29	44.3	0.861111	-2.344444444	-2.01882716	0.741512346
	29	54	0.861111	7.355555556	6.333950617	0.741512346
	18	36.6	-10.1389	-10.04444444	101.8395062	102.7970679

Written by Lorne Priemaza

38	Mean	28.13888889	46.64444444						
39	Sum					2178.377778	2928.305556		
40									
41	Beta2hat	0.74390385							
42	Beta1hat	25.71181666							
43									
44	SUMMARY OUTPUT								
45									
46	*Regression Statistics*								
47	Multiple R	0.820006853							
48	R Square	0.67241124							
49	Adjusted R Squ	0.662776276							
50	Standard Error	4.818729612							
51	Observations	36							
52									
53	ANOVA								
54		*df*	*SS*	*MS*	*F*	*Significance F*			
55	Regression	1	1620.503616	1620.504	69.78866469	9.42067E-10			
56	Residual	34	789.4852725	23.22016					
57	Total	35	2409.988889						
58									
59		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
60	Intercept	25.71181666	2.631273719	9.771624	2.10079E-11	20.36442509	31.05920822	20.3644251	31.05920822
61	X	0.74390385	0.089048039	8.353961	9.42067E-10	0.562936461	0.92487124	0.56293646	0.92487124

Written by Lorne Priemaza