Regression Analysis using Excel[1]

Simple regression

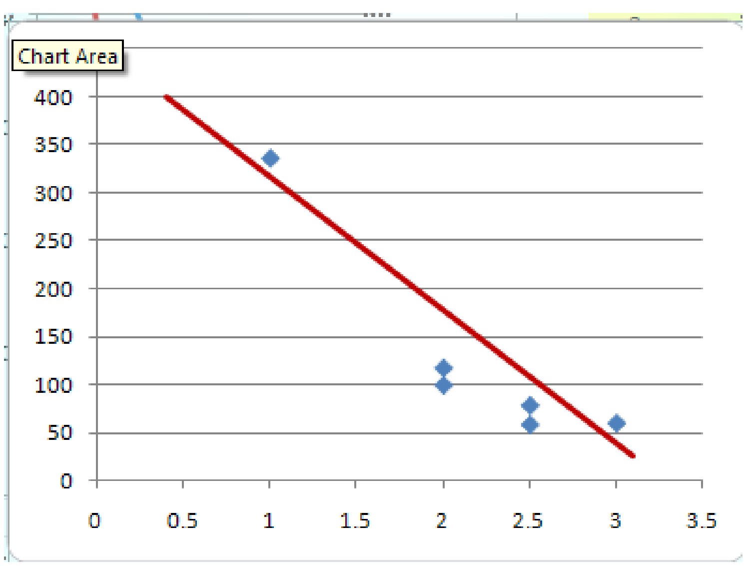
Use Solver and some simple utility function to compute some demands for commodity 1 as a function of income and prices. Here is a small sample

J	K	L	M
Results			
I	p1	p2	x1
420	1	4	336
300	2	4	100
330	2	5	117.8
290	3	5	60.4
320	2.5	4	78.8
270	2.5	3	58.9

We now try to fit a linear demand curve

$$x_1 = a - bp_1$$

The data points are depicted below.

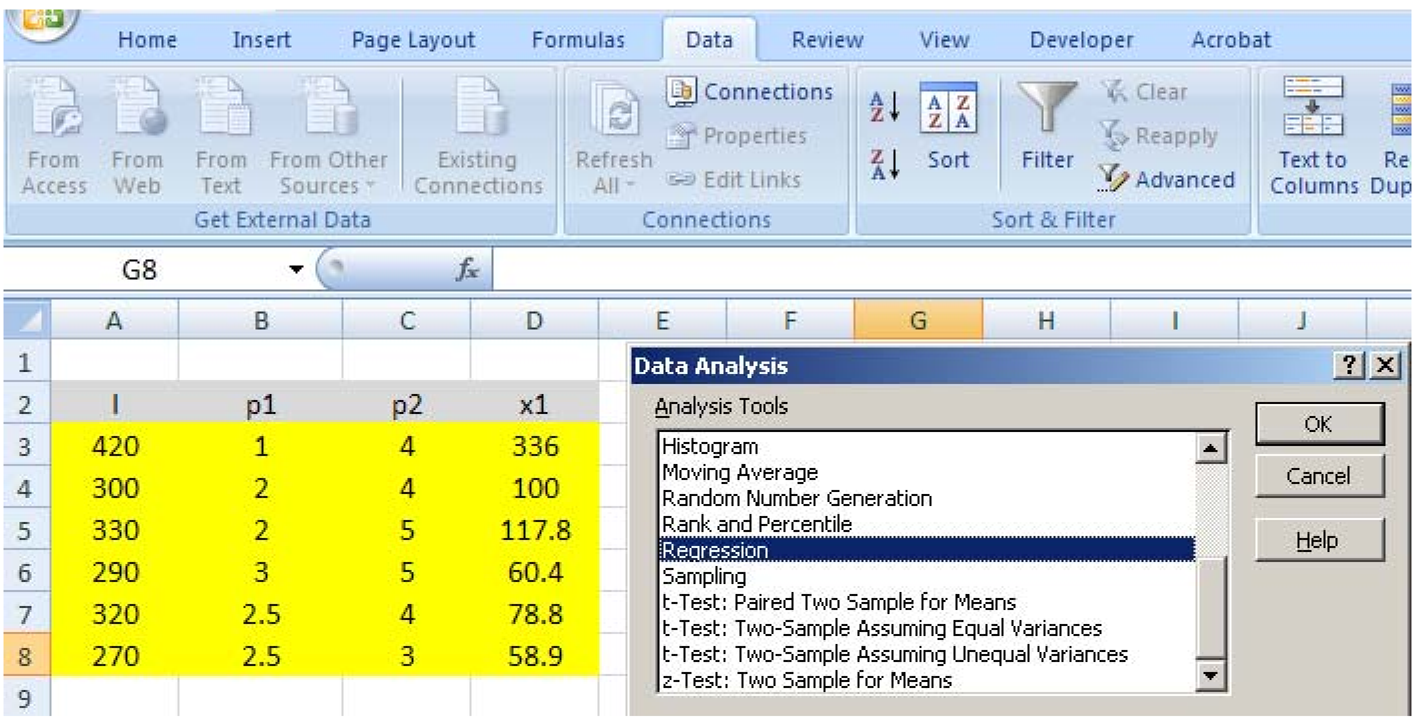


[1] These notes are almost identical to those in the Regression Analysis Slides

Clearly no line is going to pass through each point. Regression analysis starts with an initial guess as to the values of the parameters a and b. It then computes the vertical distance between the line and each dot and then sums the square of these distances.
The regression program then finds the parameters that minimize this sum of "squared errors."
The estimated parameter vector is then called the least squares estimate.

From the Home Ribbon, click on Data. On the far right you will see the Analysis Options

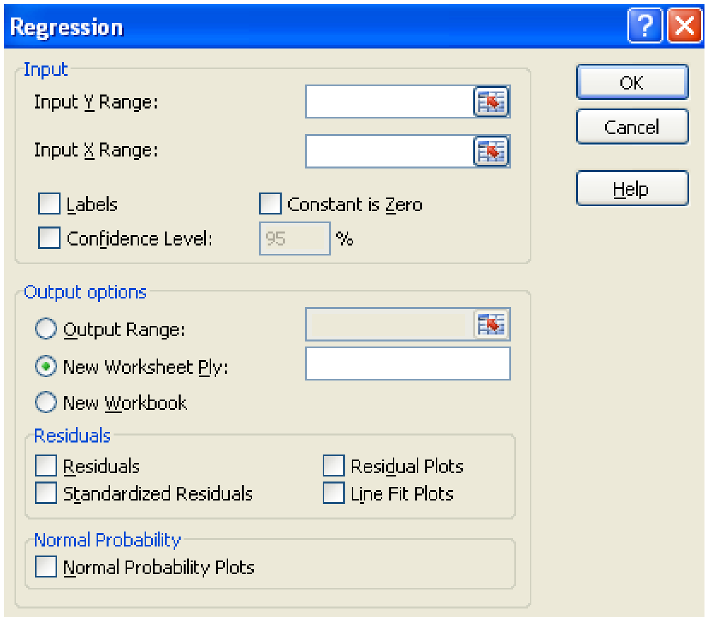
Click on Data Analysis and scroll down to Regression.



Click on OK.

You need to enter the <u>Y</u> (dependent variable data). This is the array of demands [D3:D8].

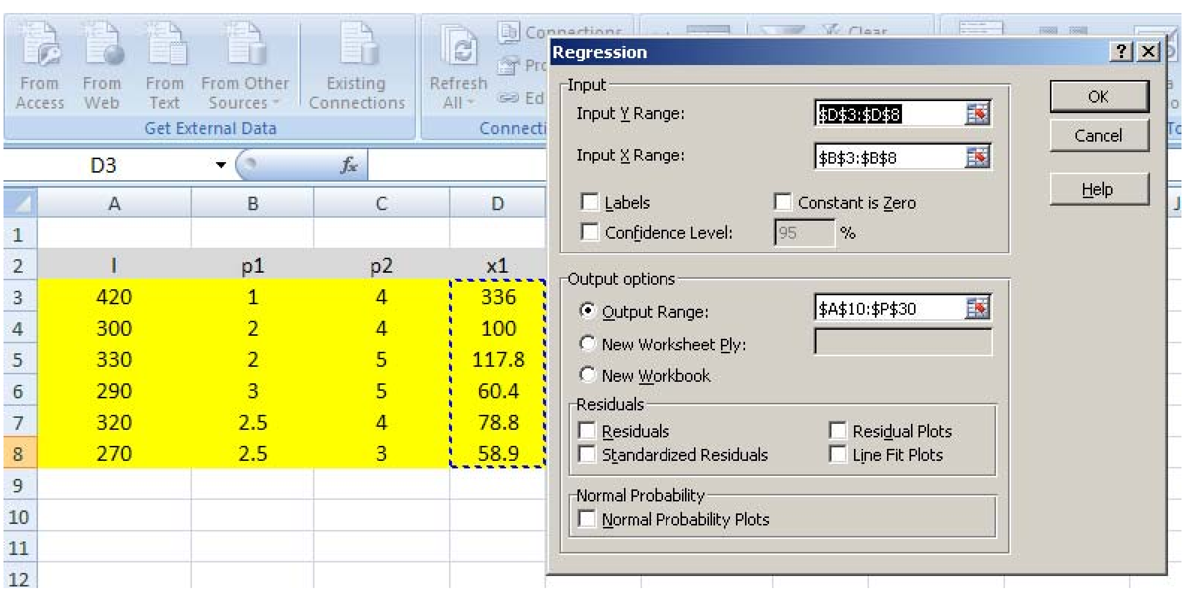
You also need to enter the <u>X</u> (independent variable array.) This is the array [B3:B8].



You do want a constant term so leave the "constant is zero" box blank.

Note that the default is for the regression results to appear on a new worksheet.

If would like to see them on the same worksheet you need to tell the program where to write the results. Om my spreadsheet everything is blank to the South-West of A10 so I choose the region A10:P30.



Click OK and the results will appear.

SUMMARY OUTPUT

10	SUMMARY OUTPUT								
11									
12	*Regression Statistics*								
13	Multiple R	0.919669578							
14	R Square	0.845792132							
15	Adjusted R Squa	0.807240165							
16	Standard Error	46.4117923							
17	Observations	6							
18									
19	ANOVA								
20		*df*	*SS*	*MS*	*F*	*Significance F*			
21	Regression	1	47257.83048	47257.83	21.93901	0.00942028			
22	Residual	4	8616.217857	2154.054					
23	Total	5	55874.04833						
24									
25		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
26	Intercept	433.6642857	68.50373356	6.33052	0.003187	243.46743	623.8611	243.46743	623.8611414
27	X Variable 1	-142.314286	30.38365019	-4.68391	0.00942	-226.6728226	-57.9557	-226.6728226	-57.9557489

We are interested in the coefficients. What the SUMMARY OUTPUT reveals is that the least squares estimate is

$$x_1 = 433 - 142 p_1$$

The R^2 is 0.85 indicating that we have “explained” 85% of the variation.

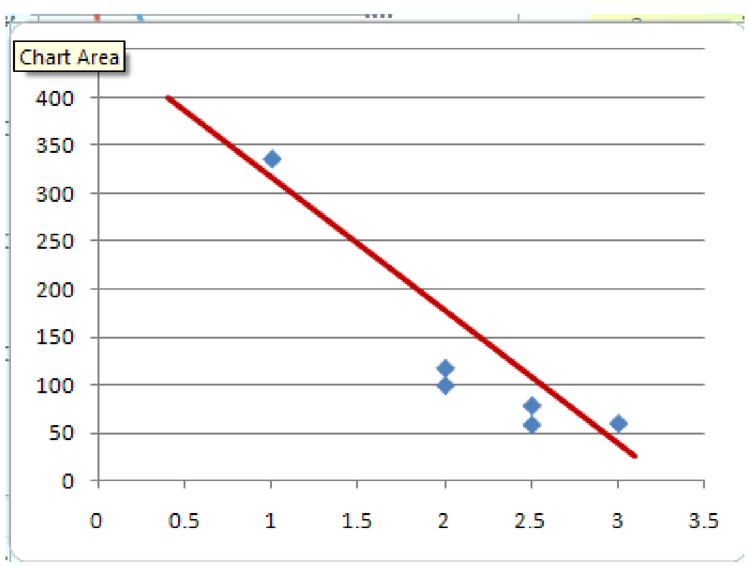
Drawing the chart

Using the estimated coefficients we can compute the estimated demand for each price.

F3 =B$26+B$27*B3

	A	B	C	D	E	F
1						Estimated
2	I	p1	p2	x1		x1
3	420	1	4	336		291.35
4	300	2	4	100		149.035714
5	330	2	5	117.8		149.035714
6	290	3	5	60.4		6.72142857
7	320	2.5	4	78.8		77.8785714
8	270	2.5	3	58.9		77.8785714
9						

We can then plot the estimated demand curve on the same chart as the actual scatter plot.



Multiple Regression

Simple regression leaves out the fact that the variation in x_1 results not just from the variation in p_1 but in the other price and income as well. The analysis proceeds as before but now the X-array is J3:L8. The new SUMMARY OUTPUT IS

1	SUMMARY OUTPUT			
2				
3	*Regression Statistics*			
4	Multiple R	0.97885635		
5	R Square	0.95815976		
6	Adjusted R S	0.89539941		
7	Standard Err	34.1890585		
8	Observation	6		
9				
10	ANOVA			
11		*df*	*SS*	*MS*
12	Regression	3	53536.26489	17845.42
13	Residual	2	2337.783445	1168.892
14	Total	5	55874.04833	
15				
16		*Coefficients*	*Standard Error*	*t Stat*
17	Intercept	-228.22899	290.2264944	-0.78638
18	X Variable 1	1.52306723	0.691991836	2.20099
19	X Variable 2	-38.214286	53.27979956	-0.71724
20	X Variable 3	-12.858403	24.04800874	-0.5347

Thus the least squares estimate is $x_1 = -228.2 + 1.5 * I - 38.2 * p_1 - 12.9 * p_2$.

We should not put too much faith in this result as there are only 6 observations and 4 explanatory variables (including the constant.)