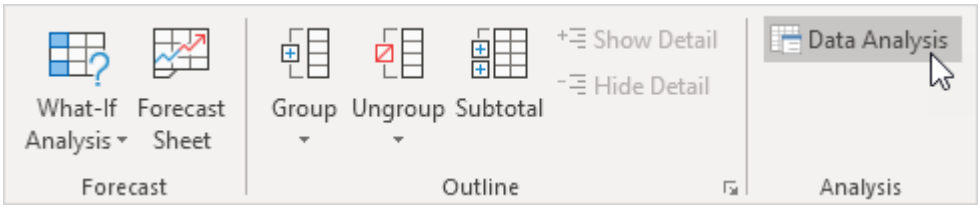
Practical work 7
Regression

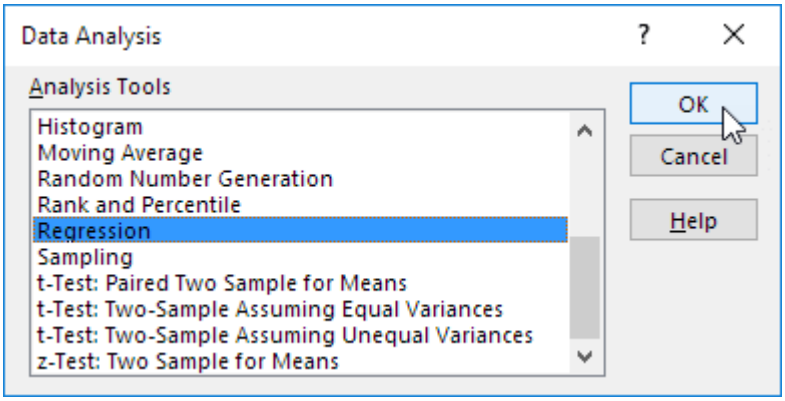
This example teaches you how to run a linear regression analysis in Excel and how to interpret the Summary Output.

	A	B	C	D
1	Quantity Sold	Price	Advertising	
2	8500	$2	$2,800	
3	4700	$5	$200	
4	5800	$3	$400	
5	7400	$2	$500	
6	6200	$5	$3,200	
7	7300	$3	$1,800	
8	5600	$4	$900	
9				

1. On the Data tab, in the Analysis group, click Data Analysis.



2. Select Regression and click OK.



3. Select the Y Range (A1:A8). This is the predictor variable (also called dependent variable).

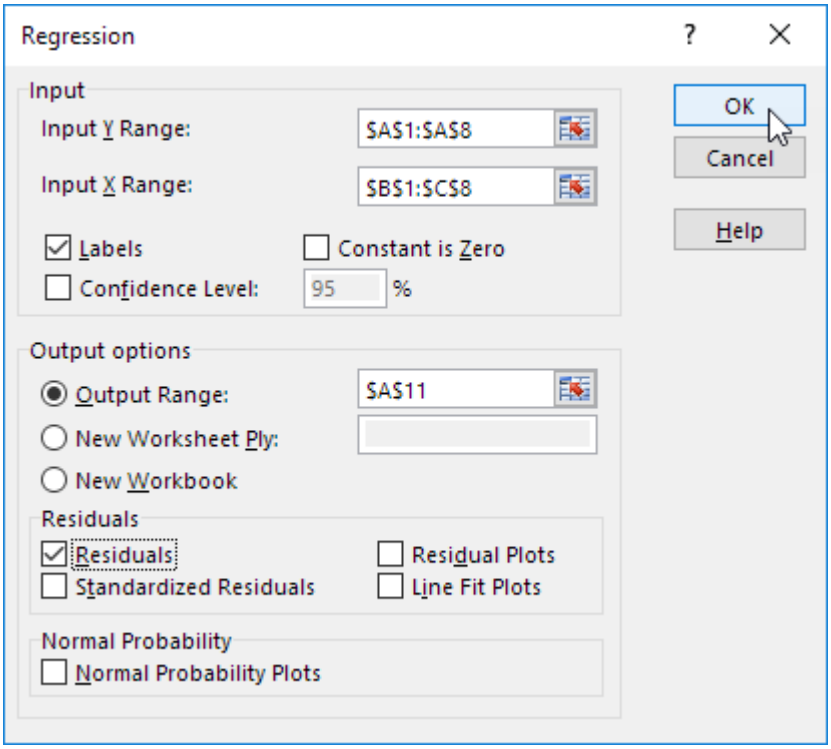
4. Select the X Range(B1:C8). These are the explanatory variables (also called independent variables). These columns must be adjacent to each other.

5. Check Labels.

6. Click in the Output Range box and select cell A11.

7. Check Residuals.

8. Click OK.



Excel produces the following Summary Output (rounded to 3 decimal places).

R Square

R Square equals 0.962, which is a very good fit. 96% of the variation in Quantity Sold is explained by the independent variables Price and Advertising. The closer to 1, the better the regression line (read on) fits the data.

11	SUMMARY OUTPUT	
12		
13	*Regression Statistics*	
14	Multiple R	0.981
15	R Square	0.962
16	Adjusted R Square	0.943
17	Standard Error	310.524
18	Observations	7
19		

Significance F and P-values

To check if your results are reliable (statistically significant), look at Significance F (0.001). If this value is less than 0.05, you're OK. If Significance F is greater than 0.05, it's probably better to stop using this set of independent variables. Delete a variable with a high P-value (greater than 0.05) and rerun the regression until Significance F drops below 0.05.

Most or all P-values should be below below 0.05. In our example this is the case. (0.000, 0.001 and 0.005).

20	ANOVA						
21		*df*	*SS*	*MS*	*F*	*Significance F*	
22	Regression	2	9694299.568	4847149.784	50.269	0.001	
23	Residual	4	385700.432	96425.108			
24	Total	6	10080000.000				
25							
26		*Coefficients*	*Std Error*	*t Stat*	*P-values*	*Lower 95%*	*Upper 95%*
27	Intercept	8536.214	386.912	22.062	0.000	7461.975	9610.453
28	Price	-835.722	99.653	-8.386	0.001	-1112.404	-559.041
29	Advertising	0.592	0.104	5.676	0.005	0.303	0.882
30							

Coefficients

The regression line is: y = Quantity Sold = 8536.214 -835.722 * Price + 0.592 * Advertising. In other words, for each unit increase in price, Quantity Sold decreases with 835.722 units. For each unit increase in Advertising, Quantity Sold increases with 0.592 units. This is valuable information.

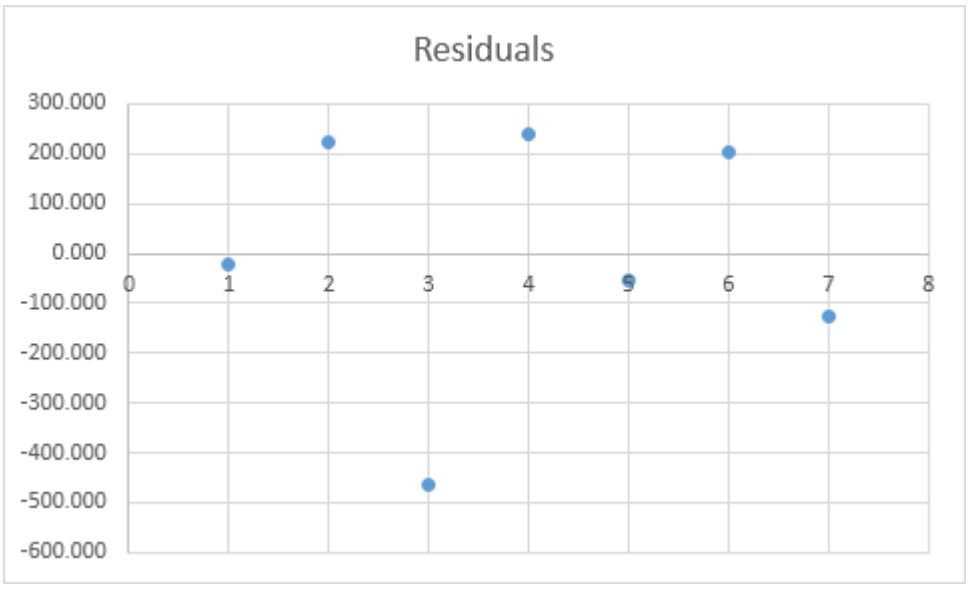
You can also use these coefficients to do a forecast. For example, if price equals $4 and Advertising equals $3000, you might be able to achieve a Quantity Sold of 8536.214 -835.722 * 4 + 0.592 * 3000 = 6970.

Residuals

The residuals show you how far away the actual data points are fom the predicted data points (using the equation). For example, the first data point equals 8500. Using the equation, the predicted data point equals 8536.214 -835.722 * 2 + 0.592 * 2800 = 8523.009, giving a residual of 8500 - 8523.009 = -23.009.

33	RESIDUAL OUTPUT		
34			
35	*Observation*	*Predicted Quantity Sold*	*Residuals*
36	1	8523.009	-23.009
37	2	4476.048	223.952
38	3	6265.938	-465.938
39	4	7160.883	239.117
40	5	6252.733	-52.733
41	6	7095.058	204.942
42	7	5726.330	-126.330
43			

You can also create a scatter plot of these residuals.


Residuals
300.000
200.000
100.000
0.000
-100.000
-200.000
-300.000
-400.000
-500.000
-600.000
0
1
2
3
4
5
6
7
8