Lab 5: Data Analysis in Excel

Analysis ToolPak (cont.)

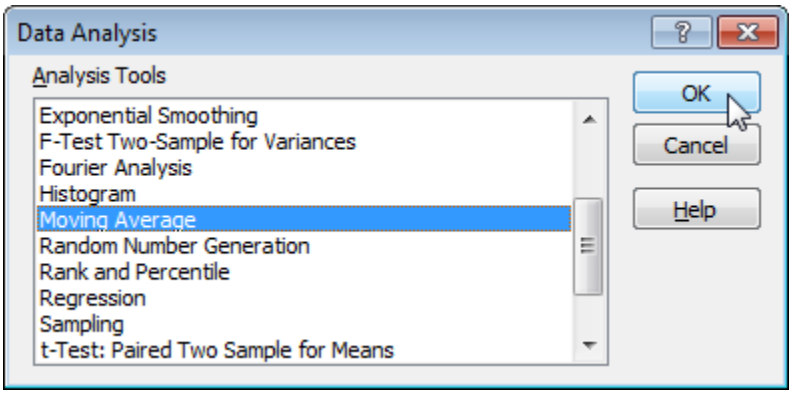
1. Moving average

moving average (also referred to as *rolling average*, *running average* or *moving mean*) can be defined as a series of averages for different subsets of the same data set.

Month	Temp. (°F)	Moving average
Jan	39	
Feb	42	
Mar	50	44
Apr	60	51
May	71	60
Jun	79	70
Jul	85	78
Aug	81	82
Sep	76	81

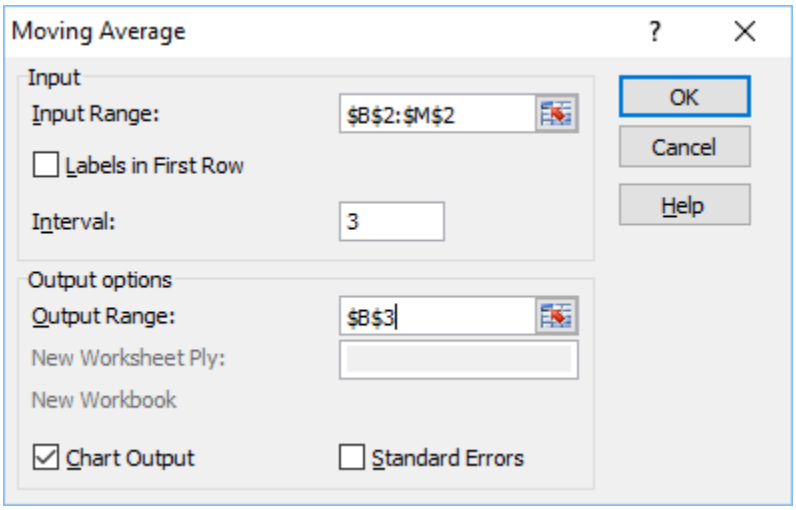
Using ToolPak in Excel

1. On the Data tab, click Data Analysis. Then select Moving Average and click OK.

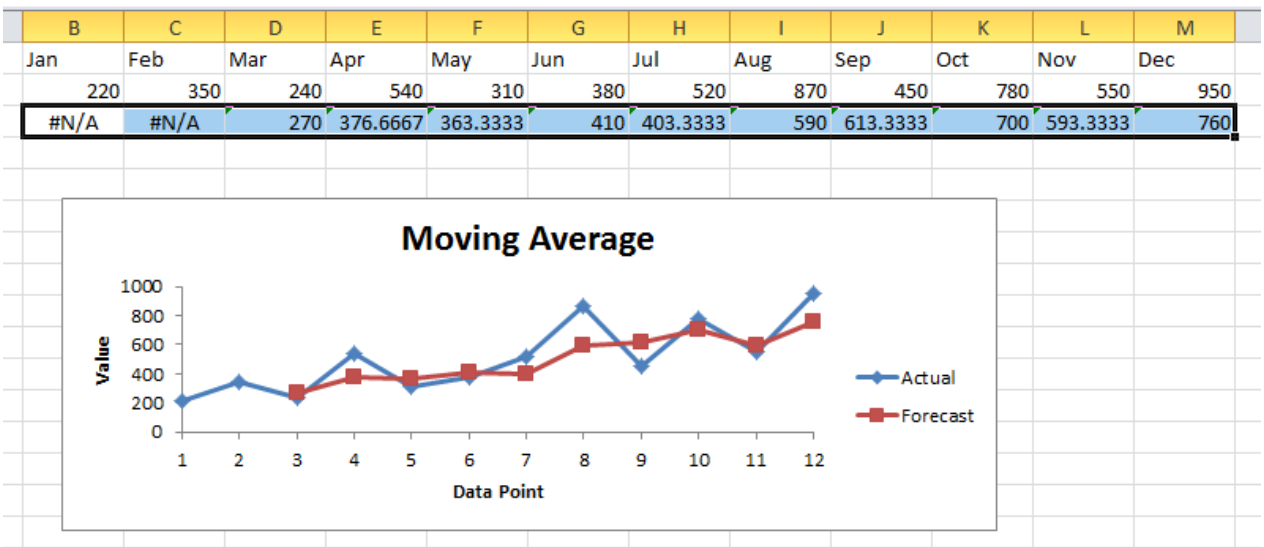


2. Click in the Input Range box and select the range of data input ➔ Click in the Interval box and type 3 (the subset you want to calculate average) ➔

Click in the Output Range box and select any cell ➔ Check Chart output ➔ Click OK.



3. Result



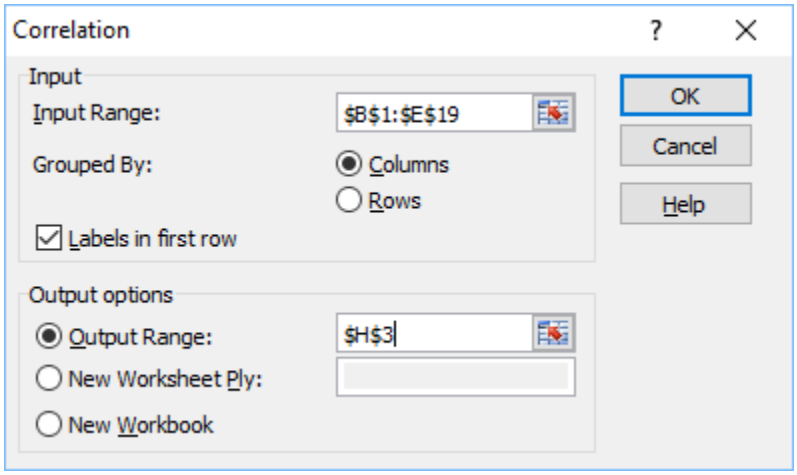
B	C	D	E	F	G	H	I	J	K	L	M
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
220	350	240	540	310	380	520	870	450	780	550	950
#N/A	#N/A	270	376.6667	363.3333	410	403.3333	590	613.3333	700	593.3333	760

2. Correlation

The correlation coefficient (a value between -1 and +1) tells you how strongly two variables are related to each other. We can use the CORREL function or the Analysis Toolpak add-in in Excel to find the correlation coefficient between two variables.

a) Using ToolPak

1. On the Data tab, click Data Analysis ➔ Select Correlation and click OK➔select the range, for example B1:E19 as the Input Range ➔ choose grouped by columns➔ select output range➔ OK.



2. Result

	Unemployment	*Inflation*	*residential Approva*	*nsumer Confidence*
Unemployment	1			
Inflation	0.183556548	1		
Presidential Approva	0.079018643	-0.463719926	1	
Consumer Confidenc	-0.103269917	-0.438373535	0.734298328	1

3. Conclusion
   - The correlation between Unemployment and Inflation (0.1835), Presidential Approval (0.079), Consumer Confidence (-0.103) are very weak.
   - The correlation between Inflation and Presidential Approval (-0.4637), Consumer Confidence (-0.4383) are weak.
   - There is a correlation between Presidential and Consumer Confidence (0.734).

b) Using CORREL Function

Click any cell to output ➔ type =CORREL and select 2 arrays of data

Unemployment	Inflation	Presidential App	Consumer Confidence
3.8	9.5	38	79.4
3.1	2.2	32	86.2
3.9	0.1	67	100.2
4.8	1.4	57	90.1
5.4	1.4	73	99.9
3.8	3.6	41	91.7
5.7	3.8	56	90.7
7.7	6.3	45	87.0
6.3	14.2	37	76.7
7.8	4.7	58	95.7
5.7	3.9	53	93.0
7.4	2.8	37	85.3
5.5	2.7	54	99.2
4.1	3.2	56	107.6
5.6	1.7	47	92.8
4.9	4.1	31	55.3
8.3	2.9	51	82.7
4.9	1.0	50	91.7

=CORREL(B2:B19,C2:C19)

3. Regression

Regression analysis is a statistical process for estimating the relationships between a dependent variable and one or more independent variables (or 'predictors'). The target is to find a **regression function** for prediction and forecasting.

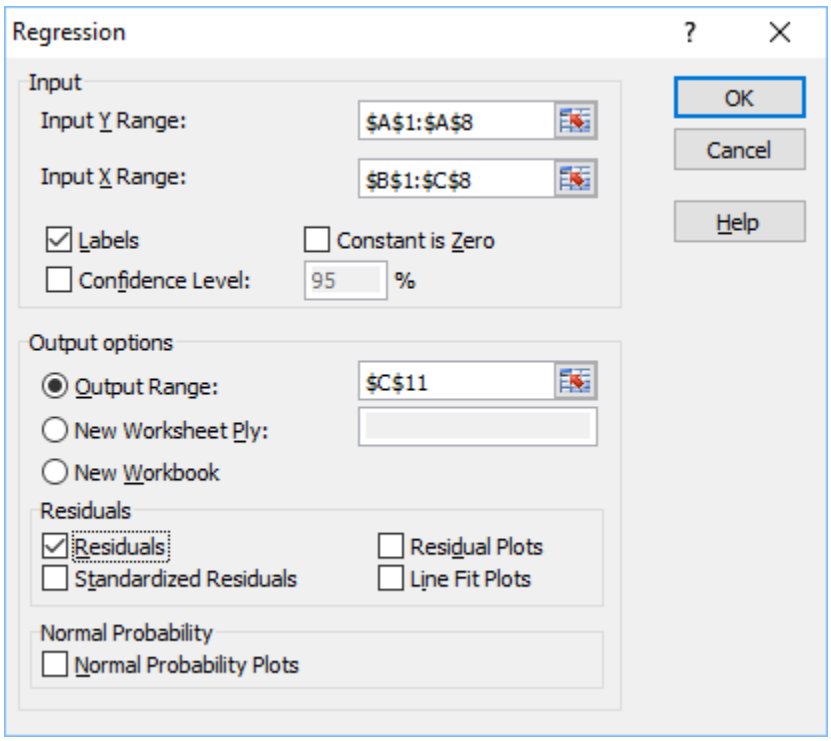
Given the following data

Quantity Sold	Price	Advertising
8500	$2	$2,800
4700	$5	$200
5800	$3	$400
7400	$2	$500
6200	$5	$3,200
7300	$3	$1,800
5600	$4	$900

The question is how to predict the quantity sold if we set the price to be $5 and advertising to be $1400?

To answer this question, we need to find a regression function based on the given data. Step-by-step using ToolPak to find this function as follows.

1. On the Data tab, click Data Analysis ➔ Select Regression and click OK➔ Select the Y Range (the predictor variable, here is quantity sold)➔Select the X Range (the explanatory variables or independent variables). These columns must be adjacent to each other (here are price and advertising).



2. Result, the tool generates three following tables

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.980681
R Square	0.961736
Adjusted R Square	0.942604
Standard Error	310.5239
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance*

					F
Regression	2	9694300	4847150	50.26854	0.001464
Residual	4	385700.4	96425.11		
Total	6	10080000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	8536.214	386.9117	22.06243	2.5E-05	7461.975	9610.453	7461.975	9610.453
Price	-835.722	99.65304	-8.38632	0.001106	-1112.4	-559.041	-1112.4	-559.041
Advertising	0.592228	0.104347	5.675579	0.004755	0.302515	0.881942	0.302515	0.881942

RESIDUAL OUTPUT

Observation	*Predicted Quantity Sold*	*Residuals*
1	8523.009	-23.009
2	4476.048	223.9522
3	6265.938	-465.938
4	7160.883	239.1166
5	6252.733	-52.7333
6	7095.058	204.9419
7	5726.33	-126.33

Conclusion

- R square = 0.961736 saying that 96% of the variation in Quantity Sold is explained by the independent variables Price and Advertising.
- Significance F = 0.001464 saying that the results are reliable (this value is less than 0.05*). 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.*
- Coefficients (Regression function): **Quantity Sold = 8536.214 - 835.722 * Price + 0.592228 * Advertising**
- The residuals show you how far away the actual data points are from the predicted data points (using the equation).