Leveraging maths and stats in Excel for deeper data insight

John Tennent
22 June 2023

Agenda

1. Setting up
   - Principles
   - Status Bar
   - Averages and Moving Averages using the Analysis Toolpak
2. Trends and Regression
   - Line of best fit
   - Trend, forecast, slope and intercept
3. Calendar functions
   - Useful date functions
   - Seasonalisation of data
4. Normal Distribution and probability
   - Sensitivity analysis
   - Standard deviations

Four Key Principles

Past does not predict the future

Trend lines cannot predict the future! Energy costs doubling in 2022 never showed up on any charts…

Unnecessarily searching for a single point forecast

Current inflation is 8.7% - what will it be next year? A range is easier to define and then explore implications

Avoid spurious accuracy

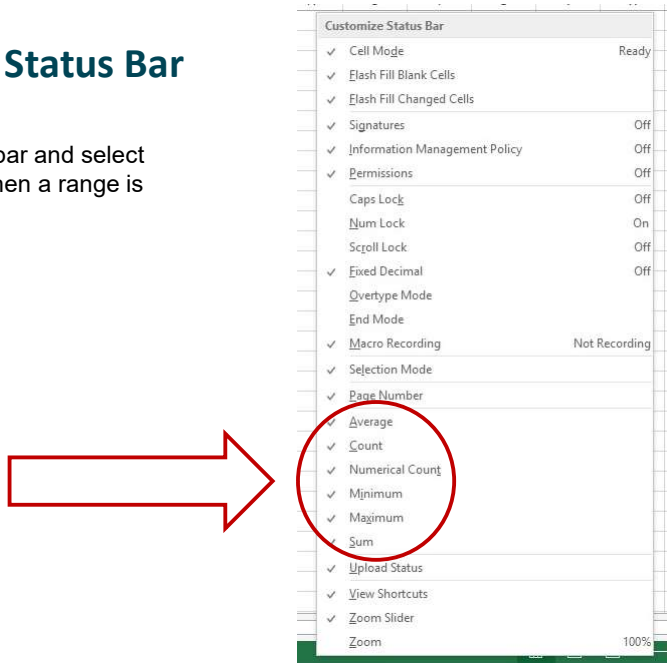
ASC Ehrenberg suggests you only need two digits to make a decision…

Only use maths functions you can explain to others

Avoid functions that decision makers don't understand

Customise your Status Bar

Right click on the status bar and select attributes to be shown when a range is highlighted



Average – Median – Mode

=**AVERAGE**(range) The same as **SUM**(range)/**COUNT**(range)

Beware zero values as they are included in the count and therefore lower the average.

Blank cells are excluded and keep the calculation valid

=**MEDIAN**(range)

When sorted in order it is the middle value. If there is an odd number of items then it an actual value. If there is an even number of items it the average of the middle two.

=**MODE**(range)

The most frequent value. If the range is just single instances of values the result is N/A. If there are two or more items with the most frequent value then the first of these in the range will be shown.

Moving Averages

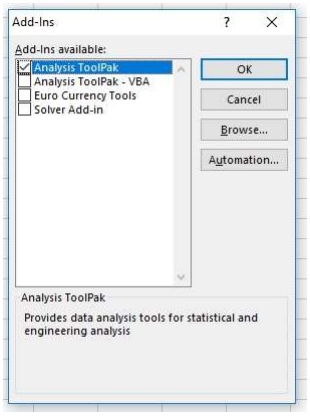
To smooth data it can be useful to have a moving average ie take the average of the last seven days on a continual basis

This can be done with a formulae but Excel has an Add-in that can do this and much more.

File – Options – Add-ins

Excel Add-ins – Go

Select Analysis Toolpak

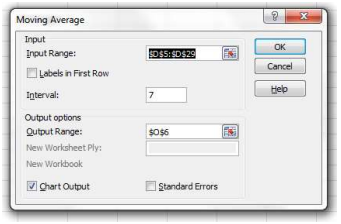


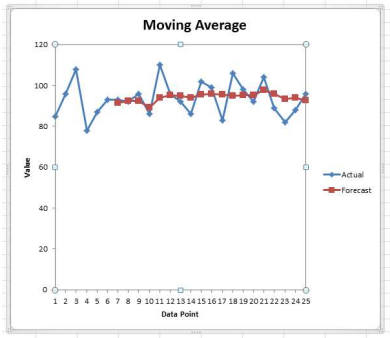
Moving Averages

Go to the far right of Data ribbon – Data Analysis – select Moving Averages

Set the range, interval and location for results

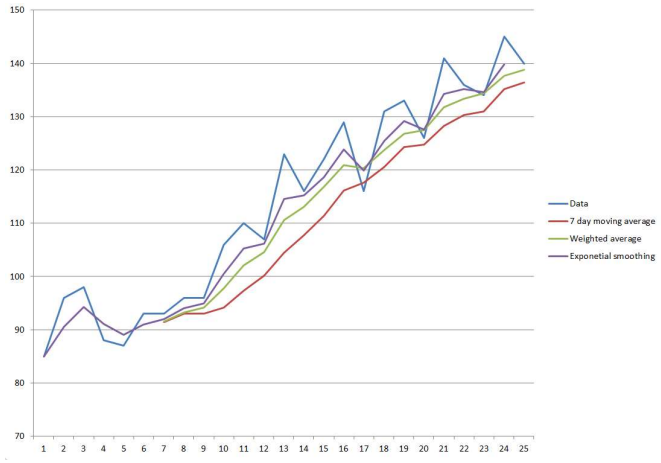
A graph can be drawn too.





Moving averages, weighted averages and exponential smoothing

These all smooth data, but are lagging the line of actual data.



Moving averages, weighted averages and exponential smoothing

Moving averages over a block of data – useful for smoothing uneven patterns

Weighted averages – these can be used to bias more recent results over older results and thus reveal growth better. The period and weightings are judgement

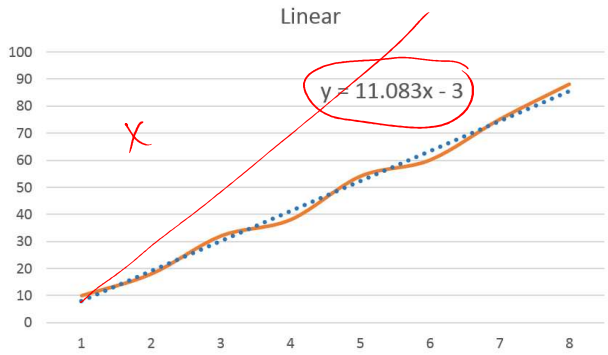
Exponential smoothing mixes current values in a proportion with past values. The proportion being a factor α. A value of 1 is the actual data and a value of 0 is the past data.

Formulae = current value * α + last calculated value * (1 – α)

Regression = Line of best fit

Right click and 'Add Trendline' - linear and Click 'Display Equation' on Chart that will enable you to create a formula to project month 9 onwards.

Beware the impact of outliers.



Functions for Forecasting

You can get the same results as Trendline with four functions. This enables you to automate forecasting without having to key in the equation.

Achieved with either **=FORECAST, =TREND** or **=SLOPE** and =**INTERCEPT** together.

	A	B	C	D	E	F	G	H
1	Trend Functions							
2								
3	1	10		=Forecast	96.75	=FORECAST(A11,B3:B10,A3:A10)		
4	2	18		=Trend	96.75	=TREND(B3:B10,A3:A10,A11)		
5	3	32						
6	4	38		=Slope	11.08333	=SLOPE(B3:B11,A3:A11)		
7	5	54		=Intercept	-3	=INTERCEPT(B3:B11,A3:A11)		
8	6	60						
9	7	75		To predict the value for month 9				
10	8	88			96.75	=(9 * 11.08333) - 3		
11	9							

52

Some useful calendar functions

- **=DATE**(year, month, day) generates an date in Excel format where 1 Jan 1900 is day 1
- **=EOMONTH**(Start date, months away) date of last day in month. Months away of zero is current month
- **=NETWORKDAYS**(Start date, End date, Bank holiday array) the number of working days (Mon to Fri) in date range
- **=WEEKDAY**(Date) the day of the week of date where 1 = Sun, 2 = Mon, 3 = Tue, 4 = Wed, 5 = Thu, 6 = Fri and 7 = Sat

Data Seasonalisation

Determine the activity levels to understand the performance month on month

Create seasonal index – beware Easter (March/April) and the month of Ramadan which moves forward 11 days each year.

Use the experience data to forecast the calendarised budget for future months

Data Seasonalisation

The budget for next year is £1,540,000 how should it be seasonalised / phased

	A	B	C	D	E
1	Data Seasonalisation				
2					
3			2020	2021	2022
4					
5	1	January	62,938.07	79,632.87	78,619.47
6	2	February	46,868.94	35,534.45	47,545.81
7	3	March	89,343.49	82,820.87	105,662.93
8	4	April	120,546.15	97,388.77	106,852.69
9	5	May	141,394.92	127,975.26	135,328.43
10	6	June	121,524.11	163,847.91	136,028.85
11	7	July	147,605.07	181,883.05	191,748.41
12	8	August	177,890.13	186,172.04	224,088.39
13	9	September	122,994.42	162,743.66	133,440.79
14	10	October	107,075.50	98,357.99	102,452.57
15	11	November	79,179.55	90,478.29	75,911.36
16	12	December	136,086.88	119,879.22	132,199.29
17			1,353,447.23	1,426,714.38	1,469,878.99

Data Seasonalisation

Find the proportion of the year that each month contributes and take an average of these
=AVERAGE(Range)

	A	B	G	H	I	J	K	L
1	Data Seasonalisation							
2								
3			**2020**	**2021**	**2022**		**Average**	**Round to 2 places**
4								
5	1	January	4.65%	5.58%	5.35%		5.19%	5.19%
6	2	February	3.46%	2.49%	3.23%		3.06%	3.06%
7	3	March	6.60%	5.81%	7.19%		6.53%	6.53%
8	4	April	8.91%	6.83%	7.27%		7.67%	7.67%
9	5	May	10.45%	8.97%	9.21%		9.54%	9.54%
10	6	June	8.98%	11.48%	9.25%		9.91%	9.91%
11	7	July	10.91%	12.75%	13.05%		12.23%	12.23%
12	8	August	13.14%	13.05%	15.25%		13.81%	13.81%
13	9	September	9.09%	11.41%	9.08%		9.86%	9.86%
14	10	October	7.91%	6.89%	6.97%		7.26%	7.26%
15	11	November	5.85%	6.34%	5.16%		5.79%	5.79%
16	12	December	10.05%	8.40%	8.99%		9.15%	9.15%
17			100.00%	100.00%	100.00%		100.00%	100.00%

Data Seasonalisation

Allocate the budget in this ratio

	A	B	N	O	P
1	Data Season		1,540,000		
2					
3			**2023 Budget (to nearest 100)**		
4					
5	1	January	79,900		
6	2	February	47,100		
7	3	March	100,600		
8	4	April	118,100		
9	5	May	146,900		
10	6	June	152,600		
11	7	July	188,300		
12	8	August	212,700		
13	9	September	151,800		
14	10	October	111,800		
15	11	November	89,200		
16	12	December	141,000		
17			1,540,000		
18					

=ROUND(L6*N$1,-2)
The -2 rounds to whole hundreds

=N1-**SUM**(N5:N15)
Use December to soak up any rounding to ensure the total is correct

Distribution Curves and Forecast Probability

- Business cases – how confident are you of the expected result?
- Monte Carlo analysis
- Normal distribution curves
- Valuation confidences

Distribution Curves and Forecast Probability

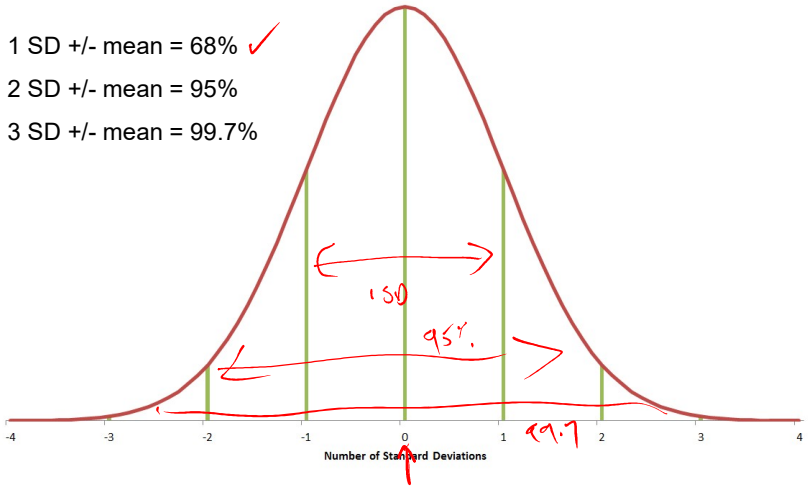
Basic	Expected outcome, best case, worse case and perhaps 10% up on costs and 10% down on revenue
% change	By how much does each assumption need to change before you would reject the project (NPV = 0)? Use Goal Seek to find answers (See 'Data' Ribbon and What if analysis)
Monte Carlo	Set assumption ranges and use the random number generator to find sets of assumptions to fire through the model (becomes dynamic not deterministic) Using normal distribution find the expected outcome and confidence intervals

Standard Deviations

1 SD +/- mean = 68%

2 SD +/- mean = 95%

3 SD +/- mean = 99.7%



Standard Deviations

Two types – population and sample

If you have the whole population (our milk example) use

=**STDEV.P**(range) same as **STDEVP**(range)

If you have a sample and want to estimate the population (average height of people) use

=**STDEV.S**(range) same as **STDEV**(range)

The difference between P and S is:

The sum of the differences between each value and the average squared is divided by n-1 for the sample making the Standard Deviation larger (less accurate). The larger the sample the nearer the two values become.

Thank you for attending

Please take the time to fill out our short survey