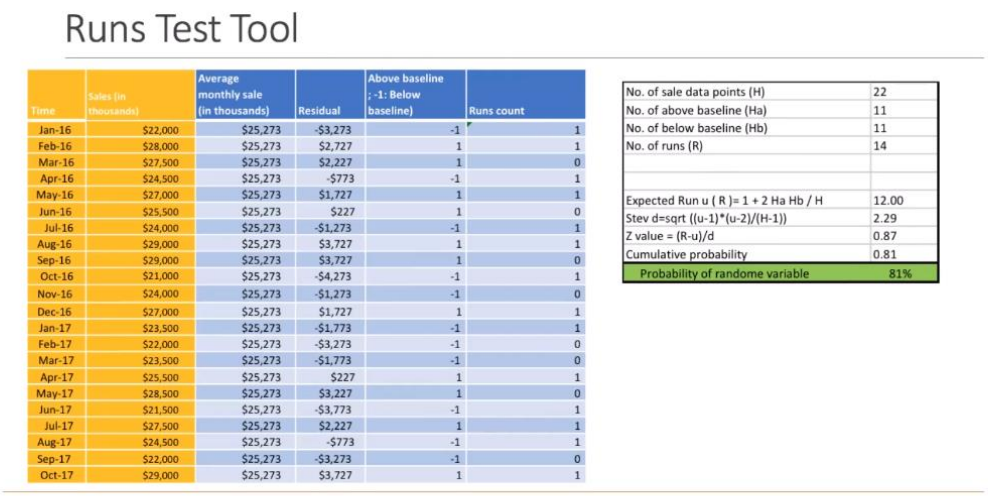
M6L2c. Random Noise vs Trend

Slide #1

Slide #2



Time	Sales (in thousands)	Average monthly sale (in thousands)	Residual	Above baseline ; -1: Below baseline)	Runs count
Jan-16	$22,000	$25,273	-$3,273	-1	1
Feb-16	$28,000	$25,273	$2,727	1	1
Mar-16	$27,500	$25,273	$2,227	1	0
Apr-16	$24,500	$25,273	-$773	-1	1
May-16	$27,000	$25,273	$1,727	1	1
Jun-16	$25,500	$25,273	$227	1	0
Jul-16	$24,000	$25,273	-$1,273	-1	1
Aug-16	$29,000	$25,273	$3,727	1	1
Sep-16	$29,000	$25,273	$3,727	1	0
Oct-16	$21,000	$25,273	-$4,273	-1	1
Nov-16	$24,000	$25,273	-$1,273	-1	0
Dec-16	$27,000	$25,273	$1,727	1	1
Jan-17	$23,500	$25,273	-$1,773	-1	1
Feb-17	$22,000	$25,273	-$3,273	-1	0
Mar-17	$23,500	$25,273	-$1,773	-1	0
Apr-17	$25,500	$25,273	$227	1	1
May-17	$28,500	$25,273	$3,227	1	0
Jun-17	$21,500	$25,273	-$3,773	-1	1
Jul-17	$27,500	$25,273	$2,227	1	1
Aug-17	$24,500	$25,273	-$773	-1	1
Sep-17	$22,000	$25,273	-$3,273	-1	0
Oct-17	$29,000	$25,273	$3,727	1	1

No. of sale data points (H)	22
No. of above baseline (Ha)	11
No. of below baseline (Hb)	11
No. of runs (R)	14
Expected Run u (R)= 1 + 2 Ha Hb / H	12.00
Stev d=sqrt ((u-1)*(u-2)/(H-1))	2.29
Z value = (R-u)/d	0.87
Cumulative probability	0.81
Probability of randome variable	81%

I prepared an Excel tool to help you do the runs analysis.

You can find this Excel tool on Canvas.

This slide shows the interface of this Excel tool.

It is a very simple tool that you can use.

Column A is the time and column B is the monthly sale, which is the test data set.

Column C is the baseline value, which is the average monthly sale, and you can use the Excel average formula to calculate the baseline value.

Column D is the difference between the actual sale and the baseline value.

Negative number means that the sales is lower than average.

The baseline value and the positive means that the sales is greater than the average.

Column E represents the sign of the difference.

One means the sales is above the baseline and negative one means that the sale is below the baseline.

Those two numbers are used for the number of runs calculation.

Column F is the count of runs.

If the signs of adjacent cells in column E change, it registers a run.

The calculation of the statistics analysis is shown in the right table.

The total number of data points is 22.

The number of data points above the baseline is 11.

And the number of data points below the baseline is also 11.

The number of runs is 14 and you can see the formula for those counts in the Excel cells.

The normal distribution analysis calculation is at the bottom of this table.

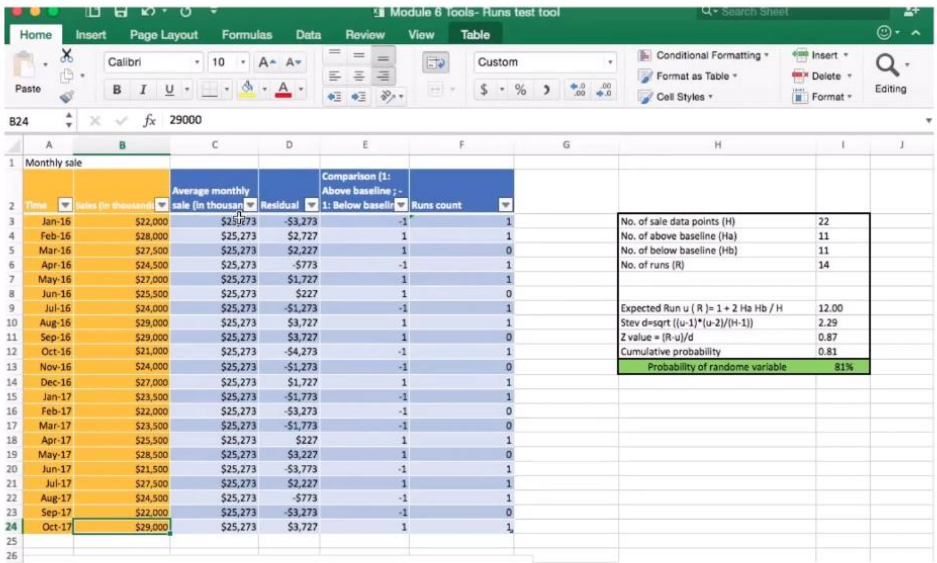
The mean of expected runs is 12.

The standard deviation is 2.29 and then the z value is 0.87.

The probability is 0.81.

So all of those numbers are calculated based on data in the table at the left.

Slide #3



Module 6 Tools- Runs test tool

Home | Insert | Page Layout | Formulas | Data | Review | View | Table

B24: 29000

	A	B	C	D	E	F
1	Monthly sale					
2	Time	Sales (in thousands)	Average monthly sale (in thousan	Residual	Comparison (1: Above baseline ; -1: Below baselin	Runs count
3	Jan-16	$22,000	$25,273	-$3,273	-1	1
4	Feb-16	$28,000	$25,273	$2,727	1	1
5	Mar-16	$27,500	$25,273	$2,227	1	0
6	Apr-16	$24,500	$25,273	-$773	-1	1
7	May-16	$27,000	$25,273	$1,727	1	1
8	Jun-16	$25,500	$25,273	$227	1	0
9	Jul-16	$24,000	$25,273	-$1,273	-1	1
10	Aug-16	$29,000	$25,273	$3,727	1	1
11	Sep-16	$29,000	$25,273	$3,727	1	0
12	Oct-16	$21,000	$25,273	-$4,273	-1	1
13	Nov-16	$24,000	$25,273	-$1,273	-1	0
14	Dec-16	$27,000	$25,273	$1,727	1	1
15	Jan-17	$23,500	$25,273	-$1,773	-1	1
16	Feb-17	$22,000	$25,273	-$3,273	-1	0
17	Mar-17	$23,500	$25,273	-$1,773	-1	0
18	Apr-17	$25,500	$25,273	$227	1	1
19	May-17	$28,500	$25,273	$3,227	1	0
20	Jun-17	$21,500	$25,273	-$3,773	-1	1
21	Jul-17	$27,500	$25,273	$2,227	1	1
22	Aug-17	$24,500	$25,273	-$773	-1	1
23	Sep-17	$22,000	$25,273	-$3,273	-1	0
24	Oct-17	$29,000	$25,273	$3,727	1	1
25						
26						

No. of sale data points (H)	22
No. of above baseline (Ha)	11
No. of below baseline (Hb)	11
No. of runs (R)	14
Expected Run u (R)= 1 + 2 Ha Hb / H	12.00
Stev d=sqrt ((u-1)*(u-2)/(H-1))	2.29
Z value = (R-u)/d	0.87
Cumulative probability	0.81
Probability of randome variable	81%

The video clip demonstrates how to build a runs test tool model in Microsoft Excel. Let me show you how to build and run a runs test model with Microsoft Excel.

First, you can have raw data in the first two columns.

And then we are going to calculate the baseline, the average monthly sale, based on the two-year sale data.

To calculate the baseline number, we use the average function. And this is the baseline of monthly sale.

After that, we are going to calculate the difference between the baseline and the actual sale. So we use the actual sale minus the baseline number to calculate the difference. Very simple.

After that, we are going to determine whether the data point is above the baseline or below the baseline.

If it is above the baseline, we give the number 1. If it is below the baseline, we give the number minus 1.

Then, we use the if function, which basically says if the residual is less than 0, then it is minus 1, it is below the baseline.

If it is above or equal to 0, we give a 1.

That means that it is above or equal to the baseline.

The next step is to calculate the run.

So we always start with number one.

We give one run to the beginning, then we use the what if function to calculate the number of runs.

For example, if two adjacent data points run across the baseline, we count it as one additional run.

We use the if function.

If two adjacent data points have different signs, then we count it as one additional run.

For the rest of the column, each cell will have the same formula, and you can see that where two adjacent points run across the baseline, we count one additional run.

We use this table to calculate the number of total data points.

We use the count function to do the calculation.

Also, we use countif to calculate the number of data points above the baseline and below the baseline.

We count the range of data, and if it's above the baseline, the number should be 1, so we count both the number of points above and below the baseline.

We also use the sum function to calculate the total number of runs.

After that, we're going to use the formula listed over here to calculate the statistical parameters.

The mean is calculated with this formula and the standard deviation is calculated with this formula.

I'm not going to repeat the formula.

So that is the calculation.

Then we calculate the z number based on the standard deviation and the number of runs.

So, the z value of this case is 0.87.

After that, we use the norm distribution function to calculate the cumulative probability based on the z value.

The cumulative probability basically tells us the probability of the random variable, that there is no correlation between the sale and time.

That is the probability.

That is represented over here.

There is 81% of chance that the monthly sale is not correlated to time.

That is what this model is.

You can use this model and you can change the sale to calculate whether there is a correlation between the sale and the time.

Feel free to use this model.

If you have more data, you can expand this and you can just use the formula to calculate the statistics parameters.

This is a very simple model.

You can find it on the Canvas website and feel free to use it.