Supply Chain Masterclass Series

3 Magic Numbers for Demystifying Demand Patterns

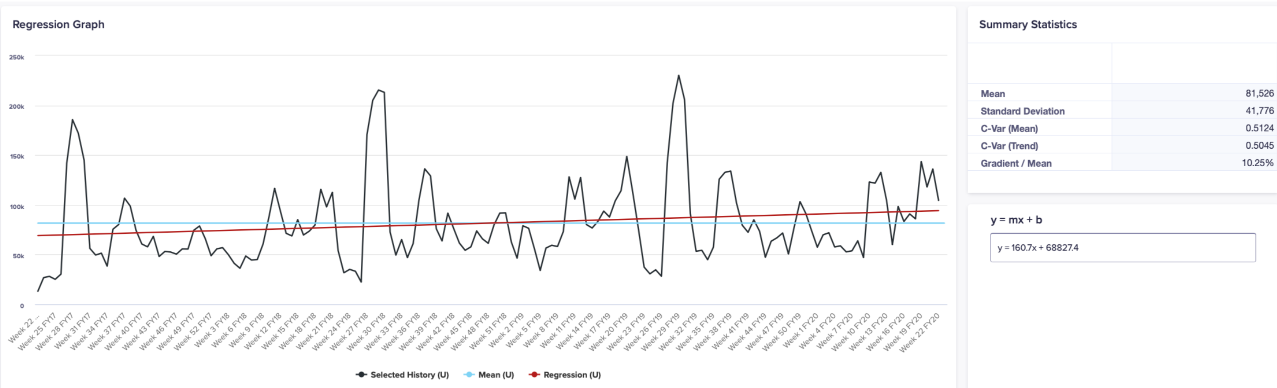
by Antony Lovell

The 3 Magic Numbers

Many of us find statistics daunting, so the purpose of this series it to demystify some of those powerful statistics for use by supply chain professionals. This paper, the first in our Supply Chain Masterclass Series, explains how to understand underlying demand patterns with three simple statistical calculations. In subsequent editions of the series, we will build on this to consider seasonality and promotions and the overall implications for history correction, buffer stock management and forecast performance management processes.

The Data Example

Let's begin by taking a look at the graph below. This is a screenshot from the Vuealta Demand Planning Application which is built on the Anaplan platform. The graph shows us the historical data in black. The mean is represented by the light blue line and the regression or 'best fit' line is shown in red. To the right we have a series of statistical values that we will delve into in order to understand the data pattern.



From a cursory glance at the graph, you can probably guess that the data is trending and seasonal. For the purposes of this paper, we will use the base measures of the mean, gradient and standard deviation, which can be applied to any dataset to broadly understand the principles. We will explore the additional considerations for trending and seasonal products in the next edition of this series.

01

The Mean as the Anchor

There are in fact several measures of the mid-point of a dataset (mean, median and mode) but here we will be using the mean. The table to the right of the graph, above, gives us various useful statistics. The first of these is the mean at 81,526. The mean is also drawn on the graph as the light blue line, visually representing how the mean serves as the central point of the dataset, or the anchor.

While the mean is probably the most used statistical measure in existence, it is arguably also one of the least descriptive. There is a real danger of relying on averages in supply chain management, so we consider the mean as nothing more than an anchor here and we need to supplement it with other descriptive statistics for trend and variability.

02

Regression for the Trend

Regression lines are incredibly useful for identifying trends in data. As scary as regression analysis sounds, the system does the calculations for us and we only need to extract the gradient from the regression equation, which is the 'm' in the regression equation y = mx + b. The box labelled 'y = mx + b' next to the graph shows the regression equation for this product of y = 160.7x + 68,827.4 and so we can extract that our gradient is 160.7. This effectively means that our straight-line regression is increasing by 160.7 units per week.

The regression line is shown in red on the graph and interestingly it highlights an upward trend that may not be spotted by the naked eye. In fact, the trend is very shallow on the graph so let's estimate the trend as an annual percentage. The regression line is increasing by 160.7 units per week which equates to 8,356 units per year. Dividing this annual increase by any point along the trendline will give us the annual percentage increase from that point. To gain the annual percentage at the 'anchor' we can divide the gradient by the mean and this is shown as 'Gradient / Mean' in the statistical summaries next to the graph, with a value of 10.25%.

Had we not looked at the statistics it would have been hard to identify the underlying trend of 10.25% due to the large seasonal peaks present in the data, which are masking the underlying trend. And we now have a way to estimate the underlying trend in any dataset which is particularly useful in highly variable data.

Technical Note
If the trend were declining then m would be negative.
For example y = -160.7x + 68,827.4. If there is no trend, then m would be 0.

Standard Deviation for the Variability

The fact that the variability in the data is strong enough to disguise a 10% trend shows why the regression analysis is important and also why we need a measure of variability. The standard deviation allows us to get a handle on the amplitude of the underlying variability from our anchor (the mean).

The statistical summaries next to the graph show that the system calculated standard deviation is 41,776. Assuming that we wanted to know how frequently historic demand would likely exceed a certain level (i.e. our planned inventory levels), it is useful to consider the normal distribution theory which again sounds daunting, but simply tells us that:

84% of values will be less than one standard deviation from the mean, or 123,302 in this example (81,526 + 41,776)

97.8% of values will be less than two standard deviations from the mean, or 165,078 in this example (81,526 + 83,552).

99.87% of values will be less than three standard deviations from the mean or 206,854 in this example (81,526 + 125,328).

Technical Note
These values are obtained by using the z values in a one-tailed normal distribution table. It is a one-tailed distribution because we are only interested in when the values are higher than the mean and could therefore cause a stockout.

A cursory glance at the graph confirms that a couple of data points exceed the value for three standard deviations (206,854) during the three years of weekly data (156 data points). The standard deviation is incredibly powerful in allowing us to understand the underlying variability in data and it enables us to put some boundaries around expected values. This has implications for history correction, safety stocks, and performance management.

But before we consider these in more detail in the next masterclass of the series, let us first consider making the standard deviation a relative value so that we can compare the variability of different products or different product ranges.

03

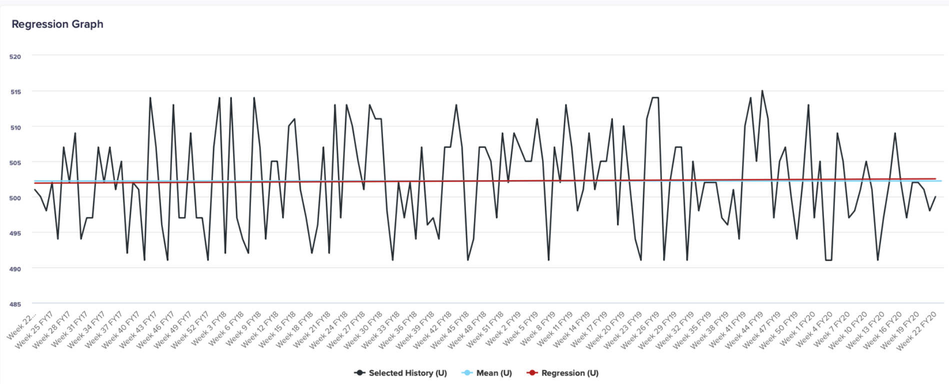
The Coefficient of Variability: A Relative Measure of Deviation

As powerful as the standard deviation is, it is not useful for comparing variability between products because it is an absolute value. A standard deviation of 100 would be very high if the underlying data had a mean of 5, but would be almost insignificant with a mean of 1,000. The coefficient of variability sounds scary, but it is simply the standard deviation divided by the mean. In this example it is therefore 41,776 by 81,526 which equals 0.5124. This value is the 'C-VAR (Mean)' in the summary statistics next to the graph.

Because the coefficient of variability is a relative measure, we can use it to understand inherent variability in products and product ranges and to be able to compare them to each other. What is high for the 'C-VAR (Mean)' will depend on your business and the underlying variability, but we typically draw the line for 'high variability' somewhere between 0.5 at the low end and 1.5 at the top end.

Additional Example

To further our development of the mean, gradient, and C-VAR let us now contrast the analysis with a product that is relatively stable.



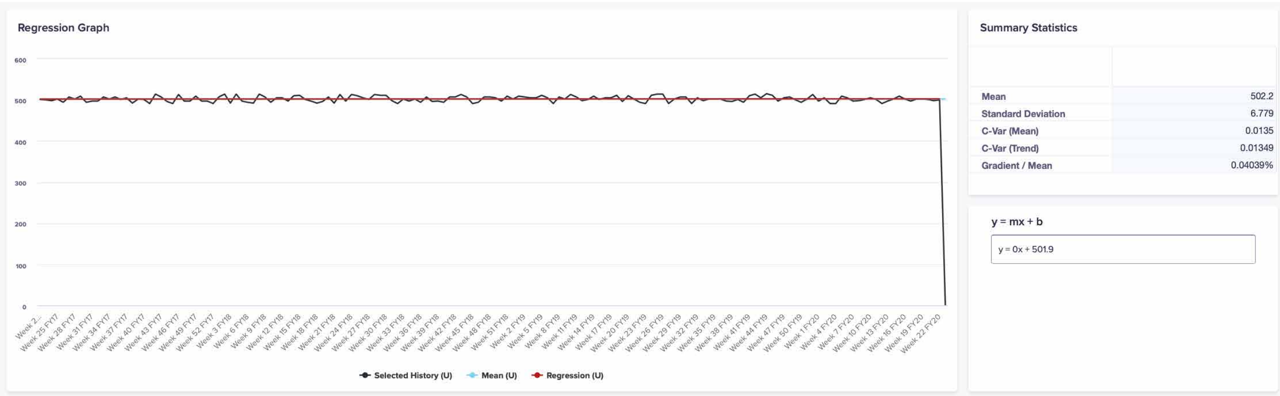
Summary Statistics

Mean	502.2
Standard Deviation	6.779
C-Var (Mean)	0.0135
C-Var (Trend)	0.01349
Gradient / Mean	0.04039%

y = mx + b

y = 0x + 501.9

In the example above, we have demand that is almost a straight line (note that the scale of the x-axis on the graph above is from 485 to 520). To highlight the lack of variability further, the same data is also plotted below, with the x-axis starting at 0.



If we now consider the summary statistical values generated by the system for this product, we see that the mean is 502.2, the standard deviation is 6.779, and we therefore have a coefficient of variability of 0.01349 (6.779 / 502.2). The regression equation is y = 0x + 501.9, so with a gradient of 0 we know that there is no trend. In summary we basically have a straight line with very little variability around the mean of 502.2.

Implications for Business Processes

The three values that we have used in this example are useful in understanding underlying demand patterns. Without even looking at a graph you can begin to visualize approximate data patterns, for instance by knowing that the mean is the anchor at 500, that you have an upward gradient of 25% and high variability of 0.5, it is possible to picture the graph in your mind and to compare products en masse.

We have identified that a product with a coefficient of variability of 0.5 is relatively more variable than a product with a coefficient of variability of 0.01. Therefore, we could conclude that it should be harder to forecast, has a greater need for history correction, and requires higher levels of buffer stocks to protect service levels. However, because the trend and seasonality in our first dataset is clearly forecastable, this would be erroneous. We would have overstated the 'random variability' and would be protecting against the expected. In the next edition of this masterclass series, we will enhance our techniques for managing trending and seasonal products with the introduction of forecast decomposition.

Our next Supply Chain Masterclass will focus on Managing Products with Trend or Seasonality: Decomposition Analysis.

CLICK HERE TO BE ONE OF THE FIRST TO READ IT

VUEALTA

Follow us
@vuealta

vuealta.com

London New York Stockholm Helsinki Rotterdam Singapore Sydney