Course Name - Operations and Revenue Analytics

Professor Name - Prof. Rajat Agrawal

Department Name - Department of Management Studies

Institute Name - IIT, Roorkee

Week - 01

Lecture - 03

Welcome friends! In our first two sessions of this particular course on Operations Analytics. We understood what the changing landscape of operations management is. We understood that the role of analytics is to improve the value addition process of operations management. We discussed how operations management is also going to become a strategic source for organizations. It is no longer a reactive or neutral activity, but it is positively contributing to achieving the objectives of the organization.

We also discussed a large number of applications ranging from PPC, forecasting, inventory management, quality management, and scheduling problems using analytics. We discussed how analytics can be helpful in all those places where you have the generation of data. Wherever you have uncertainty, please understand one thing that is underlying across the domains: wherever you have more uncertainty, data is going to help you make better decisions. In those areas where uncertainty is less, things are more or less on the usual path, or you can say a stable path. Using too much data, spending too much energy on analysis may not be advisable from an operations management point of view. So, for your organization, you should be very clear about the areas where the amount of uncertainty is very high, and therefore, more data analytics will help you handle the challenges coming from this element of uncertainty.

As we discussed in our very first session, there can be three levels of business analytics, starting with descriptive analytics, then predictive analytics, and ultimately prescriptive analytics, where we are looking for the best decision in a particular scenario. So, in this particular session, we are going to focus on descriptive analytics, where we are trying to

understand what has happened and why it has happened. So, it is more about using your past data in a very systematic, transparent, and fair manner so that the data arrangement itself shows you various interesting insights. We just concluded. A very popular tournament in our country, the IPL, has just concluded, and now the T20 World Cup is starting.

Now, in IPL, in the World Cup, a large number of analysts are available, and they are predicting the success or failure of different types of players and different teams. Now, in this World Cup, there are some new teams, for example, Canada, USA, etc. These are some new teams that are playing in this World Cup for the first time. Now, we do not have much data about them. So, it is very difficult to give any kind of analytical suggestions to these teams, and on the basis of no data available, other teams will also not be able to understand their strengths and weaknesses. While the teams that are already playing the game of cricket for very long periods, like India, Pakistan, England, South Africa, Australia, etc. A huge amount of data for every player is available, and a huge amount of data as a team is also available.

So, all that data can be used for descriptive analytics. So, descriptive analytics is basically based on the availability of a huge amount of data, historic data, and how you are arranging that historic data to get some interesting insights. That is what we are going to discuss in this particular session. So, we will be talking about what is basically descriptive analytics. Some use cases of descriptive analytics, and then whenever a set of data is available to you, these are the two very important things that are going to define the characteristics of that data.

One is the measure of central tendency; there can be different types of measures of central tendency, and we will be talking about them in the coming slides. The measures of variation: How much inconsistency is there in that data? A cricket player may hit a century in one match, and in another match, the same player may be out on a duck. There is another player who is consistently scoring, let us say, 40 runs in every match. So, a player who is scoring a century and a duck has a huge variation from 0 to 100.

And on the other hand, there is a player who is consistently scoring only 40 runs in every match. Though it is a hypothetical example, for our understanding purpose, you can consider that there is no variation in this particular case. So, the central tendency and the measure of variation are important characteristics of any data. If you want to specify the characteristics of a data set, you need these two measures. Those who have done any course on stats or any other courses related to analytics know that many times we used to ask, what is the mean demand and what is the sigma of the demand? So, the mean is basically the central tendency, and sigma is the variation.

So, that is how you define the data. We will be talking about some interesting central tendency and variation measures, and then we will give you some hands-on experience of getting the descriptive analytics outcome using a very simple tool, Microsoft Excel. So, let us first understand what descriptive analytics is, which involves the use of current and historical data to identify trends in the data and the relationships in the data, such as how the data is increasing, decreasing, or remaining static over a period of time. So, it provides a straightforward view of what has happened without delving deeper into the underlying causes. So, what is available to us in the form of some data, we just see the face value of the data.

Why are these numbers like this? That probably is not the interest of descriptive analytics. We just see the numbers. For example, this is the summer season, and every day the mercury is rising to 42, 43, 45 plus. So, I can calculate what the average temperature of the month of June in Roorkee is.

So, this is just one example of descriptive analytics. The other person will also be interested in why this temperature is rising so much. What are the reasons? That is not the scope of descriptive analytics at the moment. So, as I already mentioned, simple software like Excel, etc., is used for doing your descriptive analytics. We will show in the class how we can use Excel's power for getting different types of descriptive analytics answers. It is actually a foundation for various other serious types of data analytics approaches like predictive and prescriptive data analytics, which are only based on the data provided by descriptive analytics. So, you can say that for your predictive and prescriptive analytics, descriptive analytics is the input for the higher-order analytical tools.

Some of the popular use cases of descriptive analytics include use cases in the field of traffic management. So, organizations track engagement through social media analytics or web traffic. So, how about different intersections? For example, what type of number of two-wheelers, three-wheelers, and four-wheelers are coming in the morning, afternoon, evening, and night time? So, you can create different types of engagement reports for different crossings, which itself is a very interesting use case of descriptive analytics. So, your daily data is generated, and nowadays, all that data can easily be collated, and a huge amount of reports can be generated about your traffic flows from different types of intersections in your city.

Sales and revenue analytics is another very interesting area, and I think many organizations are using descriptive analytics for this very purpose, which is very common. Because organizations are always interested to know what the peak sales period is. Throughout the year, let us say from January to December, if I talk about India, there may be many peak periods. For example, one peak period may come around April and May, where you have a wedding season. So, demand may go slightly up. Another peak period may come during Diwali time, when demand can go up. Another peak period may come in December, which may flow up to January, again, it is a wedding season.

Let us say, throughout the year, these three are the important periods for improving your demand. Now, out of these three seasons, which season gives you the highest peak in your sales performance? These kinds of simple analyses will help you understand how we can manage our inventory accordingly. Earlier, the demand for school dresses, or school uniforms, was very high in the last week of June and the first week of July. But now, with data analytics, we are able to understand that this period is shifting to the first week of May or the last week of April. Because the entire season or cycle of school education is changing, the peak period, which used to be June and July, is now April and May.

So, these kinds of analyses are easily possible and very important for all shopkeepers and businesses. Customer segmentation-related descriptive analytics show how many different segments you have, which segment is giving you how many orders, and which is your most priority segment. In fact, this segmentation is done not only by business

houses but also by political parties to identify their target voters and how many votes they can expect from those segments. So, this is quite commonly used by organizations. Inventory management is another very popular case, which I just mentioned with reference to our sales and revenue analytics, that helps you plan your inventory so that whenever your peak period is coming, you should have your inventory ready before that.

Another important thing in inventory is, in year one, year two, and year three, what is your average work in process? What is your average work in process? Average inventory work in process means the average inventory in your pipeline. So, when I use descriptive analytics, I will give these inputs to various improvement exercises, and with those exercises, I will always aim to reduce my WIP year on year. So that the cost involved, and the capital tied up in that work in process, can be minimized, giving me more agility.

It will help me achieve the objectives of lean manufacturing. It will help me improve my efficiency as well. So, that is inventory management kind of thing where descriptive analytics will be used. A lot of use of descriptive analytics will be there, but presently it is slightly out of scope for our discussion in this operations management class, that is financial reporting. Organizations use financial reporting data, and in financial reporting, they also generate balance sheets, income statements, cash flow statements, profit and loss statements, etc.

All these statements are actually the summarization of a huge amount of data points from one particular financial year, one particular quarter of the financial year, etc. So, summarization, actually descriptive analytics, is the summarization of your many data points. You have thousands of data points in a particular period. Now, if those thousands of data points are scattered, you will not be able to make any decision. So, now summarization of those large numbers of data points into a meaningful outcome is descriptive analytics.

So, the entire financial reporting, whether your balance sheet, income statement, profit and loss statement, or cash flow statement, all these are again the summarization of those data points so that you have a meaningful understanding of that entire data. So, these are all simple use cases of descriptive analytics. The list is not exhaustive, but it gives you a

fair amount of idea of how many different ways descriptive analytics can be used in the organization. Now, coming to two important types of characteristics that define your data, as I mentioned at the beginning of this session, one is central tendency, and another is variation. Now, when we talk of central tendency, to know more about these things, you can join or study any basic textbook of statistics.

So, that will give you more clarity. But, for our purpose in this particular course, I am giving you a very brief introduction to the measure of central tendency. Now, when you have, let us say, data of this type. This is demand data for January, February, March, April, and up to December. So, you may have the demand data of 1000, 900, 950, 1100, and so on, up to let us say 1400.

That is how the demand data for a particular product for every month is given to us. Now, this individual data will not give you any significant information. This is important because all our analytics will be based on this raw data only. Now, when we have this raw data with us, the first thing you need to see are these three very basic measures of central tendency. Very basic measures of central tendency: first is the arithmetic mean, which is simply the sum of all these data points Xi. Here, Xi means X January, and i means the period.

So, i will be January, i will be February, and i will also be December. So, you have X January to X December. So, you have a total of i equals 1 to 12, or you can say i equals January to December. Here, i is January, and n is December. You will sum up the demand for all these different periods and divide by 12. That is the arithmetic mean, which is also known in simple language as the average.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

However, the average many times may not be a very good measure of central tendency, as there are certain limitations of the average because it is a very sensitive measure of central tendency which is affected by outliers. For example, here in this data, you see that all the periods are around 1000, 950, increasing slowly and gradually going up to 1400. For example, now consider a particular period, let us say October, for which the demand is just 100. This is known as an outlier, which is completely away from the other data points. Other data points are ranging from 900 to 1400, but there is a period for which the demand is just 100.

Now, this 100 demand of a particular period is going to significantly affect this average calculation, the arithmetic calculation. Similarly, if I consider that this October's demand is not 100, but 10,000, this is also an outlier; other periods' demand is around 1,000 to

1,200, but one period is 10,000. Now, if I take that data of 100 or 10,000, these are going to significantly impact my arithmetic mean calculation. So, therefore, I say that it is a very simple way of knowing the central tendency of your data points, but not very good because it is very significantly or very sensitive to the outliers. Sometimes, if your data is in negative and positive, then in that case, negatives and positives will cancel out.

For example, there is a place, just think hypothetically, where the night temperature of that place is minus, let us say, 10 degrees, and the day temperature is plus 40 degrees peak temperature. So, you will say that the average temperature of this place is plus 40 minus 10 divided by 2. So, this becomes 30 by 2, which is around 15 degrees. So, you will say that this is a very comfortable place; 15 degrees temperature is the average temperature of the city. But, you see that the temperature is going as low as minus 10, which is severe cold, and the day temperature is going as high as 40 degrees, which is severe heat.

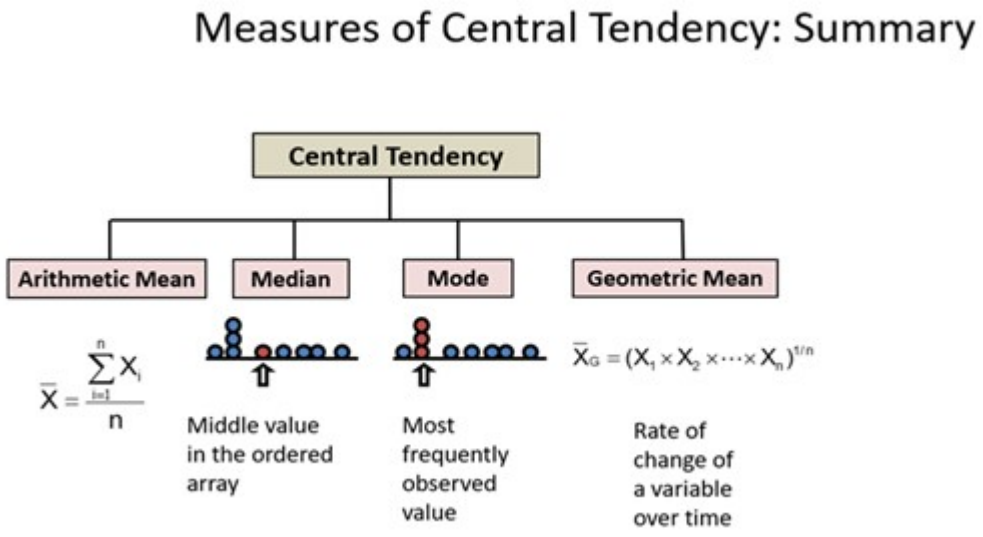
So, it is neither hot nor cold; you are saying the city is very comfortable, with a temperature of 15 degrees centigrade. So, many times, this arithmetic mean is a very misleading kind of simplification of your data points. So, to overcome these issues, we have two more measures of central tendency, which are median and mode. Now, here, with the help of this particular simple diagram, we are trying to explain to you what the median is. The median is exactly the mid value of your data points.

Here, it is important that you arrange your data in a particular order. Whether ascending or descending, it is up to you, but you have to arrange your data in a particular order, and then you will be able to get the midpoint of your data. So, like in this particular case here, these are the four data points, and these are also four data points, and this is the median value, the mid value of the entire data point, the middle point. Mode is another measure, which is the data point that is most frequent in your data set. So, like in the same example where we have 9 data points, out of that, we saw that this particular data point is repeated again and again 3 times, and therefore, this becomes my mode.

So, these are some kind of improvement over the arithmetic mean. And then, one more central tendency measure is possible, which is the geometric mean, and this geometric

mean is where we are taking the geometric mean of all the data points. So, if you have, let us say, three data points. So, it will be X1 multiplied by X2 multiplied by X3, and then you will take the cube root; you will take the cube root of these three data points. So, that is another useful way of knowing; this is generally used when, let us say, we are talking about the growth rate over a particular period, from 2014 to 2024, we have the annual growth rate of our country.

What is the average growth rate for this entire period? The geometric mean will be used for that calculation.



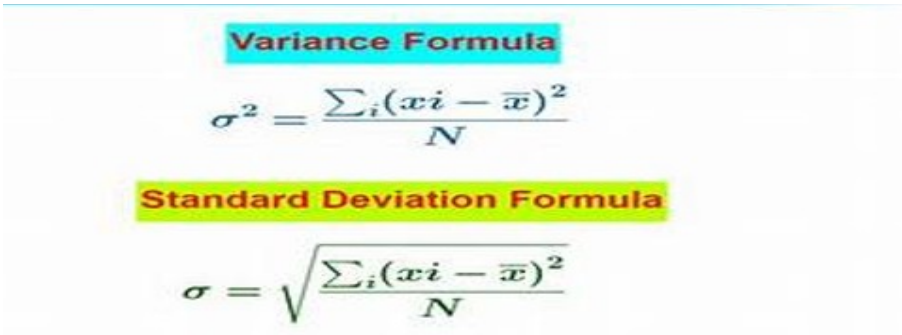
Then, we have measures of variation. So, not only central tendency but variation—what is the spread of your data—that is also very important. There are different types of measures of variation, one of which is range. So, like you have data points from 1000, 1100, 900, 950, 3000. Now, the range will be simply the highest minus the lowest.

Range=Highest-Lowest

So, 3000 is the highest minus 900, that is the range. You have to arrange the data in ascending or descending order. So, the difference between the highest and lowest data points is the range. Another very commonly used measure of variation is variance, where you will use a formula to calculate the average and how the data moves away from that

particular central point—that is the variance. Variance and standard deviation are more or less similar in their way of calculation. The only purpose of variance and standard deviation is to have two separate measures of variation for their use in mathematical calculations.

We cannot add one standard deviation to another standard deviation. So, if it is variance, this is standard deviation. So, actually, variance is the square of standard deviation.

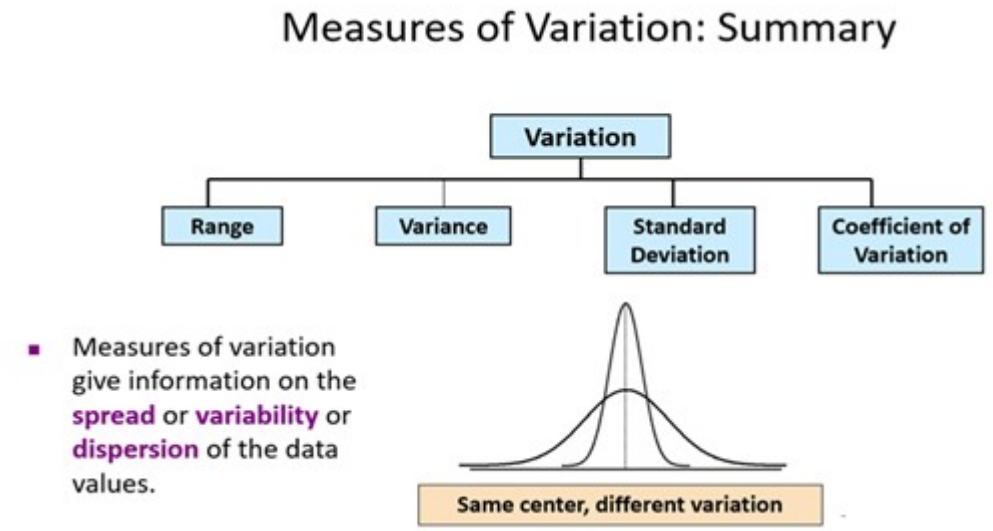


So, if you calculate the variance, the standard deviation is automatically calculated. And then, another thing is the coefficient of variation.

So, when you see that in our previous slide, we talk about the arithmetic mean. So, the relationship between standard deviation and arithmetic mean is the coefficient of variation. Because, in some cases, just by standard deviation, you will not be able to understand the variation in its totality. So, you need to see what the bigger size of the sample is. So, out of that sample, let us say if there is a case where the average value is 500 and another case where the average value is 50.

Now, in the average case of 50 versus 500, for example, I can still bear the variation of 5, but if I say a variation of 5 in this particular case with respect to 50, this variation is very, very high, and this variation of 5 in the case of 500 is not so significant. So, therefore, just by having the standard deviation number, you will not be able to understand the correct picture of variation. So, therefore, the coefficient of variation will help you in

getting a more realistic picture of your variation. Like, in this particular example, you see that different types of cases are there. This is the mean value average.



And based on how the data is distributed, this is one case, and this is another case. You see that the data is distributed according to its spread. So, when you have almost a flat curve, number 1, here you see the variation is high. While in the second case, where you see a peak type of situation, then you can say that the variation is less in that particular case. So, therefore, both these things are required to get a true representation of your data mean as well as your variation.

Excel is a very simple way to get the information about all these calculations. Let me quickly take you to the Excel application, and with that application, we will be able to see how we can do all these kinds of descriptive analytics in a very fast manner. So, if I can go to the Excel sheet. So, here we have some data in front of you which I have already entered. You may use the same data or a different set of data. Like you see, in column A from row number 2 to row number 14, we have kept some random data, arbitrary data.

Now, the simplest kind of analytics which we can do, I am going to put every kind of analytics in column number D. So, first, I am looking to calculate the mean. So, if I write the formula, I hope we all know how to write the formula. So, I have to simply write

average, and I will select the range for all these. So, this is the descriptive analytics related to the average. Let me write this as average.

Then we can calculate the median here. So, without actually telling you that we have to arrange the data in ascending or descending order, all these functions are inbuilt in Excel. So, I need not do that; internally, Excel does all these things, and again, you see that even if I do it manually, I will get the median like this: 2,22,000. So, this is the median. Now, another simple one is the mode.

So, we can use the same formula for mode and let us see what comes. So, this is the mode value, which is 1 lakh. So, all these are the simple descriptive analytics that we are getting. If you can see this data, the mode is 1 lakh, which appears 2 times in this table. Then, let us go for a few more descriptive analytics.

Now, we are going for range; here, we are going to calculate the range. So, the range we will calculate is the highest minus the lowest. Max minus the minimum of these. This is the range of this data, which is 19,62,000; that is the type of range you have. Now, directly, we will go for the standard deviation of this data.

This is the standard deviation for this data. Now, we are doing all these calculations manually for your purpose, so that whatever descriptive analytics you want, you should be able to calculate it. But, there is a possibility that you can use all descriptive analytics in one single go if we use a simple tool in Excel, which is available in the Excel solver. So, let me highlight the use of that. If you go to the data tab at the top of this Excel sheet and then you go to the data solver, and in the solver, when you invoke the solver, data analysis, sorry, in that analysis tool pack, there is an option for descriptive statistics.

And when you click on this descriptive statistics, you need to put the input range. Now, the data which we want to analyze is available from A2 to A14. So, that is my input range by the column. And now, I want to put my answer, let us say, somewhere in G7. So, G7 is where I would like to put my answer, and you will see all the interesting outcomes like the mean.

You can see that this is the same mean which we have just calculated; there may be some slight rounding issues, but it is more or less the same as the median we calculated. So, rather than doing every calculation separately and putting formulas for everything, you can use these things like this standard deviation, which may have some differences, and then the range is there, minimum, maximum, and all these things are available in this simple directly. Now, it is up to you to decide which type of stats out of this entire descriptive statistics is to be used in your case. So, this is a standard template, but you can use this for some kind of more customized, let us say, findings. For example, I just want to count how many house prices are more than 1 lakh.

So, there is a function in Excel that is COUNTIF. So, let us see the COUNTIF function. COUNTIF, and here you need to give the range again, and then you have to set the criteria. Our criteria right now is more than 1 lakh. So, that is the criteria we are putting, and we will see how many house prices are more than 1 lakh.

So, out of 13 house prices, 9 are more than 1 lakh, and only 4 are less than 1 lakh. So, you can play in this way; you can find more and more useful stats which may be required for your purpose. For example, right now, as I said, election time is going on, and results are going to come. Then, a party may be interested—the one which is winning—in knowing: how many are those seats where our winning margin is more than 2 lakh? How many seats where our winning margin is more than 3 lakhs? How many seats which we lost had a margin of loss of, let us say, less than 10,000? Because, then the party may make a strategy that, okay, where the margin of winning for the opposition is less than 10,000? We can work hard, and we may take the seats in the subsequent elections. So, just by using appropriate descriptive analytics, we may find various useful information.

So, with this, we come to the end of this particular session, and I request that you can use more such insights on your own using Excel, and you will see how interesting it is to get different types of insights from the available data.

Thank you very much.