Module 2: Interactive Presentations

Video Transcript

Interactive Presentation: Retail Inventory Management

Let's discuss now, a slightly more complicated or complex example with the retailer. Now, assume that the retailer has another store in Framingham and you also know that the historical demand at this location has a similar pattern to the Boston store, namely, the average demand is 800 and the standard deviation is 250 in its store. And what you would like to do is to use your one warehouse to supply these two stores. Essentially, we need to model the sum of the demand, the cumulative demand across the two stores together. In order to do that, we need to think about the following situation. We have two random variables. Let's just call them x and y, x will be the Boston store and y will be the Framingham store. Each one of them has the same statistics, 800 mean and 250 standard deviation. But what we are interested is not x and y separately; we are interested in what distribution x+y has and specifically, what is the expected total demand of the two stores?

What is the standard deviation of the demand in the two stores together? And what type of distribution that total demand constitutes? So, z will be the sum of x+y. And now, what we want to do is to think about the mean of z and the variance of z. So, we're going to model that or denote that by e and var, respectively. So, the mean of the sum is very easy to figure out. It's just the sum of the means, the mean of x and the mean of y. The variance of z is the sum of the variances of x and y as well. So essentially, the mean of the sum is the sum of the means and the mean of the variance of the sum of two independent random variables is the sum of the variances. If you answer that question and again, you can do that in one command in Python. You plug in the new distribution parameters 1,600 and 354. And the field rate is the same as we had before - 0.95. You get that the result is or the inventory level that you need to have is 2,182.

Let's just compare that to, what I think, the intuition of many of you was that you simply have to scale the answer we received for one store 1,212 by two. And let's just compare the two answers. So, what we see here is that if you just scale the inventory by two, you will think that you need to stock 2,424 units. But that's not the case. In fact, you need to store less, almost 10% less which is only 2,182. And that's a very important physical phenomenon that we call risk pooling or, in this case, inventory risk pooling. What is the intuition of why do you need less inventory? The intuition is that you created flexibility. In that, there is centralized inventory that can be used flexibly to satisfy each one of the stores. The answer of twice times 1,212 would be right, if these two stores and their warehouses were completely separated and we didn't have the flexibility to move things around. In that case, that would be the right answer.

Since you can move things around, you get the advantage of being able to secure the same field rate without scaling the inventory by two. But having saving of about 10% because of this phenomenon, again that we call the inventory risk pulling, I want to draw your attention that even in this case, like in the previous example, when I compare the inventory level to secure field rate of 95% to the mean demand in this case is 1,600. I can actually see that it's again distance from the mean by exactly or approximately 1.64 times the standard deviation, which is in this case, 354. Let's say that we had 10 stores and we wanted to look under total demand distribution. If we apply the same logic like before, we would end up with a normal distributed random variable that has a mean 10 times to 800 in a standard deviation square root 10 of times 250. As we take 10 stores or 10 random variables, the mean is scaled by times 10, but the standard deviation is scaled by the square root of 10, which is 3.2.

If we take 100 stores, the mean would scale by a factor of 100 and the standard deviation again will be scaled by square, will scaled by square root of 100, which is 10. And if in general you're taking end stores, the mean is going to be scaled by n and times and the standard deviation is going to be scaled by square root of n. So, you actually see that the mean and the standard deviations do not scale by the same factor. Before we end up, I would like to take one more minute and talk about the fact and go back to the assumption that we had that x and y, the demand, the random demand, the random variable were assumed to be normally distributed. In practice, what you're going to have to do is to look on your data and make a decision whether this is a reasonable modeling assumption. Again, it's a modeling assumption to model the demand as a normal random variable. Now, how would you do that? What you would like to do is to make sure that your data behaves in a way that doesn't contradict completely this assumption.

So, you can do that by eyeballing it or you can do that by applying some statistical metrics that you can read about in some of the enrichment material that we're going to provide. But what are you looking for? You're looking for roughly symmetric distribution with a single peak, and most importantly, is to have a single peak where you're going to be worried that this assumption is not very, very good. You're going to see in your demand distribution two different peaks that are separated from each other, that's going to make a normal distribution much less attractive to use. And the other thing that you would like to make sure that you don't have too many outliers that most of the values, if not all the values, lie within minus or plus three standard deviations above or below the mean. So, these are two properties that you want to search for. And again, you can use statistical software to make an assessment to what extent your data resemble normal distribution. Again, that reinforces the very important aspect about models. They are not perfect and they are a matter of choice, and modeling choice that you have to make, given the data that you have and given the problem that you're trying to solve.

Interactive Presentation: Managing Patient Appointments

So, the next example is also going to be a relatively simple example, but still representing a very common problem that you have when you have appointment systems and specifically appointments within the healthcare settings. We're going to consider a physician whose ideal workload is 20 patients' appointments per day. Right?

And many of these physicians have different things that they do in their life. For example, in the rest of the day, the doctor might be involved in research, in administrative duties, and so forth. Now most physicians in this country are highly, highly utilized. So, there is a very high demand for doctor's appointments, and the calendar is often full for many weeks in advance and despite of many efforts including automated appointment reminders, calling patients the day before, there is a significant fraction of patients that do not show up to the appointment. That's called the no-show. And in this particular example we're going to assume that on average, 25% of the patients that were scheduled did not show up and that's a very serious problem. Right? Not to mention that it leads to revenue loss. The physician typically received compensation only when they actually had the appointment but more severely, right? More importantly, it actually takes away access from other patients and makes the wait times longer by not using the capacity fully because of this no-show phenomenon.

What should be the strategy of appointment scheduling that the physicians should take in order to secure good utilization on one hand, and on the other hand, making sure that the other risk does not materialize, that you don't invite too many patients and then have to face extra work overtime? X is a function of the number of scheduled patients and then the likelihood that a patient will show up, and that lends us very much to, what we call the binomial distribution. We definitely know from data, what is the probability of success? We know that 75% of the time a patient will arrive, and we're going to assume that that's similar to all patients. And again, this is an assumption. In real life settings, you might figure out that the probability of success is not the same for all patients, but we're going to assume that for the sake of the discussion. Now, let's just think about the other parameter N, the number of trials in the binomial distribution. In our particular case, this is actually our decision. This is a lever that we can actually change right. We decide how many appointments to schedule and then given then no-show or show probability of 75%. No-show 25%, show 75%. We're going to see some random number of patients arriving to the clinic. In this case, the strategy is to schedule 26 appointments per day. When you look on the new distribution of the output, right? This is the new distribution of the output.

What can you see? You see, for example, that there is a substantial chance that you're going to have more than 20 patients showing up. In fact, if I want to know exactly what is the chance that more than 20 patients will show up? What I need to do is to calculate the sum of all direct angles from zero up to 20. Or another way to calculate that is one minus all the rectangles, right to 20. Right? And when I do that, again, you can do that. You don't need to do that yourself. You can use a command in Python to do that. You're going to find out that there is 66% that you're going to have 20 patients or less, and 34% that you're going to have more than 20 patients. So, that means that the physician will have to work overtime 34% of the time. 34% of the days the physician will have to do some overtime, and in some days, it's going to be substantial over time. What we see here is that there's a trade-off here, right? And again, there is no right answer, what is better, 20 or 26? That's for the decision maker to decide. But what we have created here is a model that now allows you to explore how different decisions that you could make, impose different performance metrics and different realities that you can now consider to choose from, and perhaps as a good exercise for you to do after this module, is to figure out what is the number of appointments that the physician should book if she wishes to have the risk of at most 5% chance to have more than 20 patients a day, in which case she will have to spend over time. Right? So now the physician is, as a decision maker, is prescribing a performance target, right? That you would like to ensure. And now you can think about how to use the model that we've created to answer this question.