Transcript for Session 045

Listen to the podcast session, see resources & links:

http://chandoo.org/session45/

Transcript:

Hi and welcome to http://chandoo.org podcast. This is session number 45. http://chandoo.org podcast is dedicated to making you awesome in data analysis, charting, dashboards and VBA using Microsoft Excel.

Today I have an exciting topic for you. This is an Introduction to Monte Carlo simulations in Excel. Before we talk about the topic, I just want to spend a couple of minutes sharing some updates and also a personal message to you.

Let me first start with the personal message - I just want to thank you for tuning into our podcast. I know you have a lot of choices out there; thank you so much for choosing to spend time with me today and listen to what I have to say and help you become awesome in your Excel and work life. Thank you so much.

The second update that I have for you is that finally after one false start, last week on Saturday I completed a 200-km BRM on my cycle. There was an event on September 12 and I was sceptical because the last time that I attempted it - those of you who listened to my podcast in late February or early March would remember this - I rode for about 80-km and then I developed severe knee pain in my left knee and I couldn't ride any further. I just completely broke down and sat beside the road and I couldn't even get up. I sat there for several minutes before I could gather enough strength in my knees to get up and call the organizers and tell them that I am not riding. So, I was very sceptical whether I could do such a 200-km ride barely six months after my knee injury. Well, not so much an injury, but some sort of joint pain. But, I thought of giving it a try because I had been doing a lot of cycling all through August and late July and all through September so far and I thought that I feel good and as fit as ever and that I should probably try it. So, I went on September 12th. There was a lot of forecast for rain and so I was actually praying that it would rain and I would actually drop out but that didn't happen! So, I went there and that's it. I went on the ride and there was some struggle when I was between the 130 and 160 km mark but I guess that was because of the afternoon heat and a bit of drizzle that was there and that caused a bit of annoyance. But, finally, I am glad I completed it. It took me about 12 hours to complete the ride. I started at 4:30 in the morning and I completed it a little before 4:30 in the evening. Out of these 12 hours, the riding time was about 10 hours and 10 minutes

and the balance was break time for breakfast and numerous snack breaks and sometimes I was just frustrated and I would stop and sit and gaze into the sky or scenery. But, the ride was beautiful and although nobody was riding with me - everybody was well ahead of me (just to be modest!) - it was good to have people in company and know that there is somebody ahead and somebody behind - all in all, it was a very good experience. They are doing a 400-km ride next week and I am definitely not going for that because I know what my limits are but I feel really good about being able to complete the 200-km ride and I just wanted to share this with you. Although you are not really cheering me on the road, I know you are listening to all these stories and you are tuning in every week and you are kind of rooting for more riding from me. I can see the emails and all that. So, thank you so much.

The last update before we jump into the podcast that I have for you is about the book. In the last podcast (session 44), I told you that I have co-authored a book on dashboards. It is called Dashboards for Excel and it is being published by Apress and that it would be released on 15th September. But, unfortunately, there has been some delay at the publisher level because they wanted to go in for a new cover design which is taking some more time and so I don't yet know when the book will be released. Today is 16th September and there is no information about the book. As per Amazon and Barnes and Noble, it says that the book will be released on 18th November. I just hope it doesn't take that long but Jordan told me via email that there has been a delay due to the printing of the new covers and it should be out there in the next two weeks. So, I'm waiting for my book with my fingers crossed. Even I haven't read it yet but thank you to all of you who have pre-ordered the book. Thank you so much. It shows a lot of faith that you have in me. And, I am really thankful to you for all the support and love that you give me.

So, let's talk about Monte Carlo simulations. I first heard about Monte Carlo simulations in 2005 and, at that time, I was not using Excel for anything more than simple data entry and maybe adding a bunch of numbers. We had Excel as a subject in our MBA course and in 2005 (or maybe late 2004), one of the second year seniors conducted a class and he told us how to use Excel in various advanced ways like pivot tables, formulas and one of them was Monte Carlo simulations. Of course, he was using VBA to run them but nevertheless. Of course, out of curiosity, I attended the session because I was not using Excel that much in 2005 and I wondered what good this knowledge could be. But, I went for the session and I think that around the same time when he started talking about the VBA code for Monte Carlo simulations, I feel asleep. Just like most of us do in college especially in a B-School where there is a lot of stress and so much to read and do, you usually sleep in classes where you either know things or you are not interested in. So, this was the second and I kind of felt like I would probably never need this knowledge about Monte Carlo simulations.

After about 6-7 years and when I started blogging on http://chandoo.org and I talked about Excel a lot and all that, even then I didn't know much about Monte Carlo simulations but sometime in 2010, Hui who is one of our authors on http://chandoo.org approached me and he said that he has some excellent material on data tables and Monte Carlo simulations. At that time, I didn't even understand or know

much about data tables and Monte Carlo simulations seemed like something that was way over my head. But, I asked him to come on board and share his knowledge because I always wanted to bring these different perspectives to http://chandoo.org. He wrote this mammoth post - more than 3000 words - explaining data tables, one-way, two-way and multi-way data tables, Monte Carlo simulations and some advanced simulations and all sorts of stuff and people loved this. There were comments practically every day to say how awesome the post was. So, a few months later, I sat down and read the entire thing just to make sure I understood the techniques. I was amazed at that time. I wondered why I didn't learn these things earlier in my life. Maybe I have learnt and used them in a different context very slightly but I never really understood the significance of data tables and Monte Carlo simulations.

So, today, I am really happy that I am talking about these because after that point in 2010, I learnt more about them and I have taught these things in my courses - especially in the 50 Ways to Analyze Data program - and more and more people are using these techniques. So, today, I want to introduce the idea of Monte Carlo simulations to you.

Let's first understand the phrase Monte Carlo simulations and what these are. Well, the basic idea is like this - you take something like let's say a coin flip - you know that if it is not a loaded coin there is an equal possibility of both head and tails. If you flip, your probability of getting a head or tail is 50-50. This is general knowledge, right? But, if you want to test this or if you want to see what would happen, for example you could flip the coin 1000 times and record the outcome for each time and then, you can tally the heads versus tails. You might find that this outcome is 49-51 percent or 40-60 percent or something like that. It doesn't exactly match with 50-50. Maybe there is a possibility of that too but when you are doing these kinds of experiments - you are experimenting the coin toss and recording the outcome - you will have to also do the physical task of doing the experiment, i.e. flipping the coin and recording the outcome. For a thousand flips it could maybe take a couple of hours or kill an entire afternoon.

Then comes the process of simulation. This is where you don't flip any coin; you simply simulate what happens if you flip the coin. For example, you could use a random number generator like the RAND() function in Excel and you could simply write the RAND() function in 1000 cells. If the RAND() value is <0.5, we will consider it as a head. If it is >=0.5, we will consider it as a tail. You could come up with some sort of a definition like that and we write 1000 RAND() functions and we simply ask Excel how many heads are there and how many tails are there and what the tally is. This is equivalent to doing the real life coin flips but in the second case you are not doing anything. You write one RAND() function and you write one test function by using the IF formula or something like that and you just drag it down thousand times and you have simulated the coin toss 1000 times.

This is, in a very loose way, what a Monte Carlo simulation is. You are essentially simulating a real life scenario by randomly generating the outcomes of the experiment and you are combining the outcomes

to do some analysis like the percentage of heads versus tails. Now, in the case of a coin toss, it is not so much of a mystery. We all know that there is a 50% probability of a head and 50% probability of a tail. There is really nothing that you can find by simulating it. With simulations you might find that once you simulate very large number of times like a million times or something, your probability of heads versus tails is same as what you would expect which is 50-50. But, if you simulate only two times, you might get 100% heads or 100% tails depending on what your random number generator is thinking at that point of time. So, simulations when you are doing it with computers for simple problems like heads versus tails, there is really not much of a mystery. But, if you were to do these kinds of simulations where the problem is very hard that's where Monte Carlo simulations become very attractive.

This is where let's talk a little bit about the history of Monte Carlo simulations. The original Monte Carlo simulations were developed as part of the Manhattan project where U.S. developed an atomic bomb in the Second World War around 1940's. There was a person called Stanislow Ulam - an American physicist and scientist - and he developed the methodology that we now call Monte Carlo simulations. The idea is that when you are developing an atomic bomb, there are naturally a lot of things that you need to calculate, understand and test before you can actually deploy an atomic bomb. So, instead of trying to figure out a problem like how far a neuron would go (a neuron is one of the particles in an atom) and how much time it takes before it can collide with another thing like another neuron or whatever and the kind of energy it would release depending on that collision etc. - there is really no simple way to answer questions like this if you are trying to solve them mathematically or using the concepts of physics. So, Stanislow Ulam, the American physicist who came up with the Monte Carlo method as a solution, suggested that instead of trying to do that as an experiment or answer the question, why not simulate it. Why don't we simulate a neuron going and hitting something and take the simulations and run them over several times because they have access to fast computers. Now, remember this was the 1940's and so their fast computers would have been like our slowest computers today but still really fast computers for those days. So, why not use the computers and create a mechanism through which we can simulate all of this and come up with calculations. So, it is not a simple random distribution like a head or a tail but there are many variables involved and each variable would have a range of values and one of the values would be randomly picked within that range based on certain distribution of the values etc. And, they came up with a mechanism to understand various specifics using Monte Carlo simulations. Of course, they were calling them simulations and they needed a code name for this because the Manhattan project was a top secret project. So, they came up with a name like Monte Carlo which is a casino where one of the relatives (an uncle) of Stanislow Ulam used to go and play cards. So, they named it Monte Carlo simulations and that's why it is called that. So, this is a bit of history of how this originated. Even before when this happened in 1940, Enriko Fermie (considered as the father of the Nuclear Reactor) in 1930 itself came up with similar calculations even though he never published them. Not for this exact problem but he used the ideas like Monte Carlo simulations to understand certain aspects of nuclear physics etc. So, that's a bit of history of how this originated. Back in those days, there were probably spreadsheets but no Excel or computer software or anything like that.

But, fast forward to 2015, and today we are using Excel and we can pretty much simulate anything using the same concepts. Let me give you one example of a problem which can be simulated using Monte Carlo simulations. Let's take something which is fairly universal, for example, investing in the stock market. This is something that we all can relate to. You are working as an Analyst or Manager and you make some money every month and you naturally care about your future and so you want to put this money somewhere. If you are following any investment journals or magazines or newspaper columns or even blogs, everybody suggests investing in an index fund. For example, if you are investing in an index fund of the New York stock exchange - I don't know how many stocks are traded there - let's say there are 50 stocks traded in that exchange in one of the main indexes - instead of trying to figure out which company to put your money in, you would just buy all the companies in the same proportion as the index does. So, if your New York stock exchange closing value goes up, your investments will also go up. If they go down then your investments will also go down. The many investment professionals, financial planners and finance gurus suggest is that instead of figuring out which company to put your money in, you should simply invest in an index fund and let the market take care of your investments. They also say that over a long period of time, your investments can return 8-9% per annum. We don't really know what is going to happen in the future for sure but we can simulate what that is going to look like using Monte Carlo simulations. For example, we could simply take the accepted wisdom and simply say that if I put in $1000 today, I am going to get 8% return after maybe 5-10 years. But, is that really true? There is a possibility that your investment may not have changed at all or it might have grown at 20% or it might have shrunk by 75%. All of these are possibilities. So, how do you ensure that we come up with a more realistic measure of the growth rather than going by the 8% number that is tossed in every financial publication? Well, one simple method is that we can take a look at the historical market changes, i.e. what is the percentage movement every day in the history. We can take a look at the last 20 years' worth of data - the daily closing price and the percentage change from the previous day. Then, we can randomly simulate a future of 365 days or 700 days or 900 days or whatever, and, for each of those days, we pick one random percentage change from the previous history. So, out of the 20 years' worth of percentage changes, we pick one random value for each of those 365 days and then we take the initial 1000 and we change the 1000 by those random percentages. With this you can really see the kind of growth your investment can have over the next 3, 5, 10 years or a couple of decades. Of course, when you do that, you should be very careful because if you are taking some data - let's say that you do this analysis in early 2000 - markets all over the world had a huge bull run before that. But, in early 2000, there was a dotcom bubble followed by the real estate bubble and then the global financial meltdown. All sorts of things happened for the next 10 years. So, any kind of forecasting that you would have done in 2000, you would not have predicted that your money will go down because everything was rosy until that point. So, you need to take a timeframe that is more realistic than heavily optimistic or pessimistic and use that to simulate the percentage changes when you want to calculate how much your money will be in the future. Of course, this is where the Monte Carlo simulations will simplify the process for you. No matter how much simulation you do, you can never ensure that your money will grow or shrink. That is because markets, and especially stock markets, are quite a bit random and very complex in nature.

Enough commentary about stock markets but I hope you understand the example wherein instead of just relying on a number like 8% or 3%, we come up with a bunch of random percentages by looking at

either historical data or some sort of marketing intelligence report or gut feel and then use that to simulate a random feature and what the value would be at the end of the future. That can be calculated with Monte Carlo simulations.

All of this is fine. The conceptual idea of Monte Carlo simulations sounds really good but how easy is it to do them in Excel? Well, when it comes to implementing Monte Carlo simulations in Excel, there are many ways in which you could do that. You could do the brute force method which is where by using the RAND() and other functions in Excel we will manually run the simulations 6000 or 10000 times just like the coin toss experiment early on in this podcast. That's the easiest method.

The next level is that you could use VBA. Instead of doing this yourself, you could run VBA and using a FOR loop, VBA will loop through your code and generate random possibilities for you. But, these are what I would call early Monte Carlo simulations. If you want to really do something powerful but don't want to spend a lot of time then you must use data tables. Data tables are a powerful feature in Excel and very few people know about this. I say very few people but, obviously, all the advanced users would know about it. Even when I started blogging about Excel 2008, I didn't hear about data tables or use them much until late 2010 as confessed earlier on in this podcast. Data tables are a powerful feature and you won't really use them unless you kind of go into them or somebody introduces them to you. But, once you know about them, and once you feel comfortable around data tables, you would be amazed at the kind of things that can be done with them.

Data tables are essentially like the FOR loop within Excel. So, instead of using VBA or anything, data tables can do this for you. So, we could use data tables to run the simulations for us. Let's say that you want to run 5000 simulations - so the data table will have 5000 rows and it can calculate the outcome for all of those 5000 rows using random combinations and random values for the input variables. Now, all of this might seem abstract to you but this is where I highly encourage you to visit http://chandoo.org/session45 and check out the resources section because I have got a very good example of data tables which Hui has authored on our website - the data tables and Monte Carlo simulations webpage - it is a very comprehensive guide and it explains how you can use data tables and how you can do Monte Carlo simulations with data tables.

Data tables are one of the preferred, best and easiest ways to do Monte Carlo simulations in Excel. Let's take a simple example so that we could actually understand the process behind this. One of the earliest known examples of Monte Carlo simulations is estimating the value of pi (the ratio between the area of a circle and r-squared). Pi is an irrational number and so there is really no way to define it. But, how do you find out what the value of pi is? So, one of the simple things is that you draw a perfect square and you draw a quarter circle within it, i.e. only the 90 degrees portion of the circle and not the entire 360 degrees portion of the circle. So, if you throw some sort of random objects into that square like grains of sand or rice (uniformly sized), the proportion of grains that are inside the circle compared to the entire

square would be pi divided by 4. That will give you an estimate of the value of pi divided by 4. So, you take that estimate and multiply it by 4 and you will get an approximation of pi. Of course, you can draw a square and a quarter circle inside it and then throw a bag of rice on it but I don't think your spouse will appreciate it especially if she or he is anything like my wife (she'd be screaming and asking how I could make such a mess)!! Jokes aside, this is where I think using Monte Carlo simulation is a really fun way to estimate pi. We could generate random X and Y coordinates inside a square. We could say that the square has dimensions of 10 and 10. So, we will generate some random values for X and Y and we ask Excel whether the X and Y coordinate is inside the circle or outside the circle using the simple Euclidean distance calculation for a circle:
$x^2+y^2=r^2$

So, we use that calculation to check whether an X and Y coordinate is inside the circle or not and then we run this test 5000 times or 10000 times for that many random X and Y values. At the end, all you have to do is simply come up with a ratio so that you can find out how many dots are inside the circle versus the total number of dots and multiply that by 4. That will give you an estimate for pi. So, I have made an example workbook for this podcast (pi estimate) using the data table and Monte Carlo simulation idea discussed in this podcast. You can download that example workbook at http://chandoo.org/session45.

While we are talking about pi, there is also another interesting experiment that you can do to estimate pi. Again, this can be done through Monte Carlo simulations where the idea is basically called throwing pins and needles or throwing hot dogs. I don't want to delve too much into that concept because it is a very intricate one and not everybody can understand it. So, I am going to provide a link to an article which basically talks about this along with an Excel workbook that Hui published on our website a while ago. I will link to that on the show notes page at http://chandoo.org/session45. That's a little bit about Monte Carlo simulations and how to do them using data tables and an example of the estimation of pi using Monte Carlo simulations.

Now, let's talk a little bit about things that you should keep in mind when you are doing Monte Carlo simulations. The first and foremost thing that you need to be aware about is that for the input variables for the simulation, for example in our estimation of pi, we are talking about random X and Y coordinates, the kind of distribution that these will follow. In case of random X and Y coordinates, they are what we call uniformly distributed. That means that the value of X from 0 to 10 can be any value; it is uniformly distributed. But, when we talk about things like estimating or simulating stock market movements, let's say that we want to have the market movements daily ups and downs from -10% to +10%, the movement won't be uniform within that range. The likelihood of -10% is not the same as the likelihood of -1%. -1% is more likely than -10% because -10% is like a very rare event. Something like the recent Chinese market collapse has triggered a 10% or maybe close to that change in the global markets. But, other than that, you would usually see the fluctuations around the center, maybe towards the right, especially in a bull market. So, in such cases, we can't go with a uniform distribution. We have to go with maybe a normal distribution like the bell curve or some other type of distribution. So, you need to

understand your input variables and you need to come up with the correct random variables for that. You can't use RAND() or RANDBETWEEN() all the time. You would have to use a proper probability distribution function like NORMINV or something like that and then use RAND() inside that so that you can get variables in that kind of fashion like normally distributed random variables or exponentially distributed random variables or what not. So, understand the distribution of your data and use that in your simulations random number logic. That's the first part.

The second thing is that you need to run enough simulations. If you run only 100 simulations or 20 simulations, you are going to get a very poor estimate of the pi value. But, once you up the number of simulations, like do it 5000 or 10000 or 20000 times, you are then going to get more realistic values. That said, keep in mind that not everything can be estimated. So, that's the third point - simulation versus reality.

Simulations are a good way to understand what would happen and what-if kinds of things but they are not an exact substitute for reality. So, simulation is going to give you some level of confidence or a deeper understanding of your model or reality but it is never a substitute for the actual reality or the actual future that is going to happen. So, keep all of these things in mind when you run your Monte Carlo simulations.

That's about the Monte Carlo simulations. As I said, the resources for you for this podcast would be Hui's excellent article on data tables and Monte Carlo simulations which I will link from our show notes page along with our pi estimate Excel workbook. You can download it and play with it to understand how simulations can work and how you can use them to your advantage. So, that's about it for the Monte Carlo simulations.

I hope you enjoyed this podcast. I just want to thank you once again for tuning into http://chandoo.org podcast this week and becoming awesome. Thank you so much for all your reviews and awesome comments on iTunes. One of the reviews recently dropped on iTunes caught my attention. I just want to share that with you. This is from Mary Van - "I have worked with Excel for a long time in my job but after taking a Power Pivot class recently I realised how much about Excel I did not know. I ran across Chandoo's podcasts and started with session number 1 and am almost caught up now and feeling a bit sad. I actually like the longer episodes. He injects a bit of himself into each and every show and you will learn things in every single one. I especially appreciate the full transcripts. I plan to review those as well as start reading the blog soon. I also like that some episodes are more basic while others are more intense. Loved number 35. Thank you, Chandoo, for all you do to make us awesome." Then, I liked the P.S. part - "P.S. - My sister just named her new kitten Mango!" So, there you go. People listen to my podcast; of course this is a coincidence - I am not sure if his sister is also listening to our podcast but anytime that I hear a name like mango, I just love it!

So, thank you so much once again for tuning into our podcast. If you love it and enjoy the content, please take a minute and write a review on iTunes. This is going to help us because more and more people will discover http://chandoo.org podcast and become awesome in their life. Thank you once again. Visit http://chandoo.org/session45 for all the show notes and resources for this podcast. Stay awesome and I will talk to you next week. Bye.