Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 03

Lecture 12: Seasonality and its Features

Thank you. Hello all. So, welcome to this new lecture in this course on time series forecasting with applications in R. Until last week, what we studied was kind of, we stopped at the ARIMA process. If you remember, in the last session, towards the end, we took up and then in particular we looked at this ARIMA PDQ process that you see on the slides in front of you. So, what we'll do is we'll again pick it up from here.

So, we'll expand more on non-stationary time series per se. Because if you remember in the last lecture, we discussed that almost 90% of the examples are non-stationary, right? I mean, you should expect some trend, or there should be some seasonality, right? Or there should be some pattern within the time series that makes the series non-stationary. So, elaborating and then kind of expanding the non-stationarity aspect of a time series is more important than expanding the stationary time series, okay?

So, if you remember the definition of an ARIMA PDQ process, this is more like a quick revision. In the first place, it is a non-stationary time series process. So, let us say if you have a non-stationary process Yt which follows ARIMA PDQ. So, when do you say that Yt follows ARIMA PDQ? Let us say if you difference this Yt d times, right.

So, if you difference this time series process d times and if the resultant series or if the output series is a stationary ARIMA (p, q) process, then we can actually say that the initial time series Yt happens to be ARIMA (p, d, q). And then again, as discussed in the last session, the first couple of orders, so let us say p and q, stand for the autoregressive and the moving average orders of the process, while the number of times you may want to difference the series is given by d. And then, if you remember, we took up a simple example. So, let us say if you start with a particular process, let us say Yt, right? And then let us say if you apply the differencing operator twice, okay?

So, Nabla operated two times. And let us say that you assume that the resultant series is Wt, which follows an ARMA (2, 1) process. And then obviously, the assumption is that this ARMA (2, 1) process would be stationary, right? Then, we can say that the initial time series Yt happens to follow an ARIMA (2, 2, 1) process, alright. Now, again, the first 2 and the last 1 because they correspond to the AR orders and the MA orders of the stationary ARMA process.

And then, since you have differenced the series 2 times or twice, the middle value of d happens to be 2, okay. Now, just as a quick revision, we will kind of revise the different kinds of differencing that we studied in the last lecture. So, the first one is lag 1 difference. So, lag 1 difference, all of you know that the operator is nabla. So, if you apply this nabla operator in Yt, it kind of gives you Yt minus Yt minus 1.

So, you basically take the current value of the time series and then subtract one past value. So, yt minus yt minus 1. And then, in general, we can actually extend this to multiple or future lags. So, let us say lag 2 difference or lag 3 difference. So, in general, if you have a lag d difference, so differencing at a lag d is nothing but 1 minus b to the power d, right, applied on yt.

So, 1 minus b to the power d applied on yt. So, by the way, if you remember, when is this lag d difference kind of used? It is when there is seasonality. So, if you want to difference directly at a particular lag, so if you look at the formula here, the formula is yt minus yt minus d. So, the lag between the subscripts here is nothing but d. So, we call this a lag d difference. And then the other kind of differencing is the dth difference. So, let us say if you have a strong trend which is not linear but, let us say, quadratic or has some higher power, let us say cubic, right, and so on and so forth.

Then there might be a requirement of differencing it those many times, right? So, in the last lecture, we took an example where if you have a quadratic trend, then the amount of differencing required is twice, right? And then, after you difference twice, we can actually get a stationary model. So, to be able to control the amount of trend we have, we require to difference the series those many times. So, let us say, in general, if you want to apply the dth difference and then formalize nabla raised to the power d, which is nothing but 1 minus b raised to the power d applied on yt.

By the way, in all these formulas, this capital B is nothing but the backshift operator. And then, a couple of remarks or a couple of points here: if the trend MT is a polynomial function of order D, then the Dth difference applied on YT would be stationary. While if the data is seasonal with, let us say, some period D. So, period D means what? So, if you have monthly data, then the period is 12. If you have quarterly data, then the period is 4, and so on and so forth.

So, if the data is seasonal with a certain period D, then a particular lag D difference is applied, and that would kind of remove the seasonality aspect from the series. So, I think this slide sort of sums up all kinds of differences that one can apply, depending on if you have a trend or if you have some seasonality behavior in the series, and so on and so forth. And then, the last slide in this lecture here would be kind of a simulation again. So, what we have in front of us is a simulated sequence of an ARIMA 1, 1, 1 process. Now, by the way, all these things have been done in R. So, all the simulations and all the calculations have been done in R.

So, a particular ARIMA 1, 1, 1 process where the coefficients are nothing but 0.7 and 0.2. So, 0.7 corresponds to the AR coefficient, which is phi 1, and then 0.2 corresponds to the MA coefficient, which is theta 1, right? And then, this is just a simulated series or simulated sequence of that particular ARIMA model. And then, clearly, you can see that the series is not stationary, firstly, right? Because you have like local trends. For example, you have a trend here and then possibly a trend here, right?

But what it could mean is that since both these coefficients are less than 1, right? So, if you difference the series once, right? If you difference the series once, then possibly we can get hold of a stationary ARMA process with orders 1, 1, right? And again, why stationary? Because again, like I said, if you remember the criteria of a particular ARMA process being stationary is that if both these coefficients happen to be between minus 1 and 1, then we can actually say that the ARMA process is stationary.

So, even though this initial ARIMA 1, 1, 1 process is non-stationary, of course, but then differencing this once would produce a stationary ARMA 1, 1 process. So, this is a broad idea about ARIMA processes and then what would happen if you difference such ARIMA processes, and then at the end of the day, if you difference an ARIMA process, we kind of get a stationary ARMA process with the corresponding orders. So now probably in the next slides, or in the next set of slides, we'll kind of expand more on non-stationary aspects, then how do you deal with, let's say, seasonality, right? And then we'll talk about a few tests of stationarity. Can you somehow perform some hypothesis testing to kind of prove or kind of show whether a time series is stationary or not?

And then again, we will kind of expand on different types of seasonality. And then towards the end, we will focus on one particular seasonal model, which is SARIMA. So, SARIMA is a natural extension of ARIMA. So, we will talk about how do you bring in the idea of seasonality along with the trend. All right.

So, we'll now discuss a few aspects of seasonality and its features. All right. So, what exactly do you mean by seasonality? And then, if you remember one prior lecture that we had, we kind of expanded on different kinds of non-stationary behaviors. Right.

And there, we talked about trend. We talked about seasonality. We talked about irregular components. We talked about changing variance. Right.

So, in the next few minutes, we'll specify more about a few aspects of seasonality. So, initially, we will again give a few more examples. So, just to acquaint all of you with the idea of what you mean by seasonality and what are the few examples where seasonal behaviors or seasonal structures can be seen. So, firstly, seasonality is a regular periodic variation where the period of the cycle is generally less than one year. So, again, if you think about temperature data or rainfall data.

So, what one can observe is. So, in the summer months, as we discussed earlier. So, we typically see hotter summer months, right? So, temperatures are at their peaks, right? And the same behavior repeats itself year by year, right? So, the period of repetition, or the period between one peak and the next, is usually less than or equal to one year, right?

And then, such seasonal patterns are usually predictable because, if you want to jot down seasonality using a graph, then seasonality would be something like a repetitive pattern, where the peaks correspond to typical months over the year, whereas the troughs correspond to other months. But then, such patterns are predictable, right? Because When such a series is coming down and the series is at its trough or at the bottom, then one can expect that it should probably rise back again from here. So, such seasonal patterns are usually predictable.

And then, factors behaving similarly at a particular time of the year, month, or week are called seasonal factors or seasonal patterns. So, there should be some repetitive behavior at a particular month of the year, day of the month, or week of the month. So, all these ideas amount to seasonality. And then, in general, seasonality is caused by, let's say, cycles of seasons or different holidays. So, in India, especially, let's say, if you have higher sales during Diwali, Dashara, and Holi, etc.

So, all these kinds of things depend on holidays. Or regular changes in behavior, or let us say biological rhythms and physiology in response to periodic changes in the environment. So, all these are kinds of causes of seasonality. So, if you want to summarize this paragraph, you can write down the causes of seasonality. So, seasonality can be caused due to different reasons.

Let us say cycles of seasons, then different holidays, or let us say regular changes in behavior etc., and so on. Now, I think this slide again gives you several more examples of seasonal behavior. So, let us say animal migration. So, in a particular time of the year, there might be a larger amount of animal migration as compared to, let us say, some other ones, or let us say an increase in sales of coffee or warm clothes during winter, or let us say an increase in sales of fans or ACs during winter. So, let us say, it all depends on when the sales are happening and when the sales are kind of increased or decreased.

By the way, this should be summer. So, let us say increased sales of fans or ACs during summer. So, again, it all boils down to different seasons within a year. Or let's say clothes and firecracker sales during Diwali. So again, due to let's say holidays.

So, this reason could be due to holidays or again related to seasons. Then the next one could be, let's say, increased unemployment in June. So what typically happens in June is that. Since most of the graduates are kind of in the job market, then probably one might see slightly more unemployment compared to some other times where possibly all the graduates have been placed and probably, they've started working in some industry or academia and so on and so forth. But typically, when the college session ends, let us say in April, May, or by mid-June, then starting from June onwards, one can actually observe slightly increased rates of unemployment.

Or the last one could be something like an increase in viral fever cases during the change of seasons. So again, all these are kind of revision examples where one can actually experience some seasonality. Now, why exactly is it important to study seasonality in the first place? So, reasons to study seasonality. Now, of course, there could be many more reasons, but then these are typically some of the important ones.

So the first one could be something like better planning for a temporary increase or decrease in, let us say, labor requirements, inventory, training, periodic maintenance, or let us say supply chain management, etc. So I think the key here is better planning. So better planning in terms of multiple different areas. So let's say inventory management, supply chain management, operations, maintenance, etc. And then once you identify a seasonal pattern or a seasonal repetition, eliminating it from the time series for studying some other components, let's say trend, cyclicality, or the irregular component, becomes all the more important.

So, reasons to study seasonality are to kind of eliminate that. Because, at the end of the day, seasonality amounts to non-stationarity. So, all of us have seen that. So, if a series is non-stationary, one has to kind of identify that and then try to remove it. And then there could be one more reason, which is kind of interesting, in regards to forecasting.

So, let us say forecasting future seasonal patterns, such as, let us say, climate changes. So, there is a term called climate normal. So, what do you mean by climate normal? Climate normal is nothing but the 30-year average of a weather variable for a given time of year. So, climate normal means, on average, what should happen over a 30-year time span. Okay.

So, forecasting future seasonal patterns, let us say climate changes, or let us say future temperatures, future rainfall. So, all these are kind of areas of interest. All right. So, now the next thing we will study is some typical graphs or graphical techniques to kind of detect seasonality. Right.

And I think all these graphs would be kind of easy to digest and understand, and then they will present a much clearer picture as to how you visualize seasonality in multiple examples. Okay. Now, possibly all these could be plotted and then their combinations henceforth. But then, these are typical plots that one can actually draw. So, the first one is, let us say, a simple time series plot or run sequence plot.

So, by the way, this is nothing but the usual scatter plot that we all kind of plot as a very first idea. Or what one can do is have a seasonal plot or a seasonal sub-series plot. Or the next one could be something like multiple box plots. So, corresponding to each month of a year, we can actually plot a single box plot for that month. Or one can actually plot some ACF plots or spectral plots and so on and so forth.

So, we will try to elaborate on a few of these plots, and then we will bring in some examples and show you how to draw these plots and how to visualize them. Alright, so now the first one is a simple time series plot, or it is also called a run sequence plot. So, a run sequence plot is nothing but a scatter plot of the time series data. The only difference is that the x-axis is typically numbered as per the time frame. So, by the way, the example here is average daily gas flow data.

So, the amount of gas that flows from, let us say, the manufacturer to a consumer through pipelines and so on, and then average how much gas flows on a daily basis. And then this is a run sequence plot. So, a run sequence plot means nothing but you are basically plotting the data in a sequence as per the time frame. And then each entry gives you a run, basically. And then this is also called a typical or simple time series plot.

So, on the x-axis, if you observe, we have all the months, right? So, let's say July of 2008, then April of 2009, then January of 2010, then October of 2010, right? Then July of 2011, and so on and so forth. So, I think this data is more like quarterly data, right? So, if you observe a pattern here, the data is more like quarterly data.

So, let's say July, then April of next year, then January, October, July. Obviously, you have a few months which are not being listed on the axis, but then still you have the data for that. But the data is quarterly data. And then here, if you simply observe the plot, you can clearly see a quarterly seasonality. So, you see typical peaks and then typical troughs repeating themselves after every fixed quarter.

So this is a very typical kind of simple time series plot or run sequence plot. So just by observing such a plot, one can actually make out the extent of seasonality. So, one can actually make out the extent of seasonality and the typical behavior or typical time span or the period of seasonality. And then, by the way, this horizontal line that you see here, which is a blue line, gives you the mean. So, let us say if you want to indicate what the average daily gas flows data is and then the average of that average, right, something like that.

So, typically here, the average of the average is close to 500, right. So a few more points on the prior plot, which is the run sequence plot, as to where exactly can you make use of such a plot or the plot helps in identifying what exactly, right? So one can actually identify all these patterns. So seasonality present in the series, of course, right? And one can actually identify any shifts in the location of the seasonal patterns.

So again, if you go back for a second, So for example, this horizontal line is nothing but the location or the mean. So if you observe in any aspects of this graph, if the gas flow value is kind of deviating towards one side of this horizontal line, then one can actually expect some shift in the location, right. But here obviously you cannot make it out because there is no shift in the location.

But then if there was, what would have been observed is that you would have observed some local behaviors where the values are kind of drifting away from that horizontal line, okay. So one can actually observe some shifts in the location or one can actually observe some shifts in the variation. So shifts in the location and shifts in the variation also. So how exactly? So shifts in the variation could mean something like, so again I will try to replicate the same plot.

So let us say this is the horizontal blue line which indicates the mean. And let us say initially the average gas flow is kind of behaving like that. But then towards the end the average gas flow is kind of behaving like that. So, even though you have a seasonal behavior clearly, the variation is not equal, right? So, you have a fanning out kind of pattern.

So, initially the average amount of gas flow is kind of restricted in a tighter range, but then as you go down the quarters or down the years, if the average gas flow is kind of increasing in that manner, then one can actually expect that the variance also is kind of being affected. So, such a plot is kind of really important in identifying all these ideas. And the last one is one can actually identify the presence of any outliers. All right. So now the next plot is again a simpler one.

It is called a seasonal plot. So, what exactly do you mean by a seasonal plot, and how do you draw a seasonal plot? So, the time series is plotted against the individual seasons in which the data was observed. And then, one advantage of the seasonal plot is that the underlying seasonal pattern can be visualized more clearly. And then, the third advantage is that it sort of identifies years when the pattern changes.

So, let us say if you have temperature data and if you want to kind of identify in which year the temperatures were slightly higher than all the other years or in which years did you observe harsher winters or harsher summers. So, all these reasoning-based questions can be answered using a seasonal plot in a visual manner. So, we will see an example of a seasonal plot. So, again, we are using the same data as before, which is the average daily gas flows data, and then this is a very typical seasonal plot, alright. Now, what exactly is happening here?

So, on the x-axis, if you observe, we have all the months in a particular year. So, let us say January, February, March, all the way up to December, okay. So, this would be our typical year, okay. And then, in the graph itself, you see different colors. So, black, orange, gray, and then blue.

So, each color kind of corresponds to one season or one particular year. Alright, so let us say what happened in, let us say, 2020 might be given by the black line. Then what happened in, let us say, 2021 might be given by a different color, let us say, orange. Then what happened in 2022 would be given by, let us say, a silver color, and then what happened possibly in 2023 might be given by the blue color. Alright.

So, on the x-axis, you have within-year, so each month being plotted, and then the graph itself kind of gives you the behavior across years. So, one particular curve or one particular color kind of signifies one particular year. So, in this case, one can actually compare between years, right? So, let us say if you talk about August, right? So, during the August month in, let us say, 2022, which is the blue line okay, or 2023, which is the blue line, as compared to, let us say, 2020, which is the black line.

So, can you see a difference here? So, let us say the black line is way above the blue line at this point, at the August point, right? So, what do you mean by that? So, you had, on average, the daily gas flow was much larger in the year 2020 as compared to 2023 or 2022, and so on, okay.

So, one can actually compare in a particular month what changes one can observe across years of the same data. So, this is a typical seasonal plot. So, what we do is we have a particular year plotted on the x-axis and then values corresponding to different years in the graph itself. Now, we will kind of shift to slightly different data. So, the next data is the airline passengers data, and you probably must have seen this data even earlier.

So, what we have here is that, firstly, this data is inbuilt in R, all right, and then this kind of gives you the number of airline passengers who are traveling in thousands over all these years. So, from 1949 all the way up to 1960, and then, of course, you have all the months. So, this data is monthly data, firstly, right? And again, without even plotting anything, just by looking at the structure of the data and the name of the data, one can actually assume that there has to be some seasonality involved. Because the number of people who travel by airlines might be slightly more in summers than in winters.

Because in summers, one has a few more holidays as compared to winters, and so on and so forth. So, what we have here is we have monthly data from January, February, all the way up to December, and then over all the years. So, let's say all these are the individual values. So, 112, 115, 145, 171, etc. So, using this data, we will create a few more plots.

So, the next one is called a seasonal plot of the data set. So again, I think this is the exact same plot that we had earlier. And probably this would be slightly clearer than the first one, than the earlier one. Because what we have here is, we are kind of again on the x-axis, we have all the months. So, January, February up to December.

And then we are essentially plotting the number of airline passengers within that year on a yearly basis. So, each color kind of represents one year. So, 1949, 1950, 1951, 1952 all the way up to 1960, all right? And then for each year, you have a different curve, right? So, again, the same story. So, let us say if you go back to the month of, let us say, July or August, right?

So, one can actually see a drastic difference between the top curve and the bottom curve, right? So, whichever years they are. So, I think the bottom curve corresponds to 1960, the top curve corresponds to 1949. Or exactly the opposite, exactly the opposite, my bad. So, the top curve corresponds to 1960 while the bottom curve corresponds to 1949.

And one can see huge differences in the number of passengers who traveled in airlines between 1949 and 1960 in the same month. So let us say July or August or, for that matter, any other month. So, in that sense, a seasonal plot kind of gives you a lot of information, not just for a particular season and a particular year, but across all the years as a comparison. Now, the next one is called a seasonal sub-series plot. So, what do you mean by that?

So, such a graph kind of makes the evolution over time clearer. So, what do you mean by that? So, we have a graph, and then inside the graph, we have some blue lines. So, those blue lines kind of represent the means of the individual groups or the averages of the individual groups. So, one can actually see the averages clearly.

So, one can actually figure out where exactly the average is aligned also. However, this plot has a disadvantage: between-season analysis is difficult. For example, one cannot compare the number of passengers between two months. So, let us see if you want to compare between July and August or between July and December. So, it is slightly hard to compare such a thing.

But we will take up one example of this. So, this is a very typical seasonal sub-series plot. So, now what is happening here is again on the x-axis you have all the months. And then, typically, let us say if you take any particular month, right? So, let us say January.

So, this is more like the extent of the number of airline passengers in January over all the years. So, 1949 to 1960. So, the minimum is this much. This is the maximum. And then the blue line sort of represents the average.

Okay. So, in January, over all the years from 1949 to 1960, what is the behavior in the airline passengers? Right. So, if you see, for example, July or August. Right.

So again, the same thing. So, in July, over all the years from 1949 to 1960, the extent of the number of passengers is roughly from 100. So, let's say this is 100. Right. To something like 600.

Right. So, one can actually get a range of the particular observation one is trying to estimate in a particular month over all the years, along with the average that is plotted. Okay. So, probably in the next session, what we will cover is we will try to extend this idea about seasonality, and then, like we discussed earlier, we will try to merge all these ideas into how you can define a particular model of seasonality, which is called SARIMA. Okay.

Thank you. Thank you.