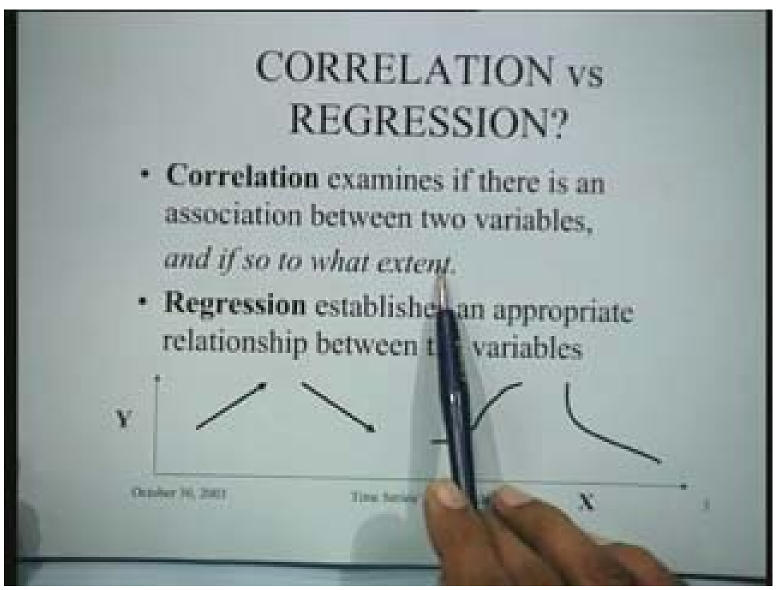
Project and Production Management
Prof. Arun Kanda
Department of Mechanical Engineering
Indian Institute of Technology, Delhi

Lecture - 35
The Analysis of Time Series

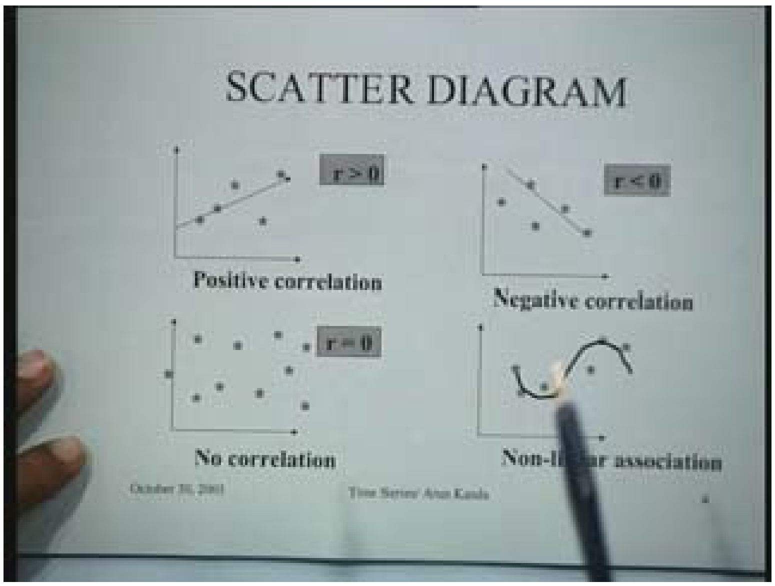
In the last lecture we had looked at the basic principles of forecasting. We have identified some of the major methods that are used for developing the demand forecast. In today's lecture we are going to be talking about the analysis of time series. The development of forecast is based on the time series. Specifically we are going to be talking about methods of correlation, regression and time series, decomposition as they are applied to forecasting value of the demand. To begin with let us try to recall what we mean by this common term, generally known as correlation. We are all familiar with this term and we know that correlation to some extent measures the degree of association between two variables or two series. This is a central concept which is used in forecasting and therefore we must clearly understand the distinction between correlation and regression. Correlation examines if there is an association between two variables and if so, to what extent.

(Refer Slide Time: 02:19)



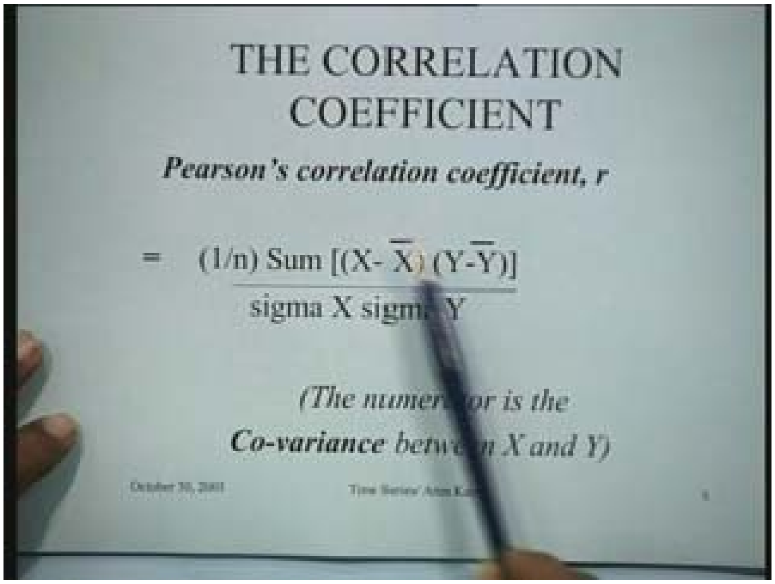
The correlation coefficient measures the degree of association, between two variables and regression on the contrary establishes a unique relationship between the variables. If there are two variables y and x and you find that they behave in a certain fashion, discovering the relationship between the variables is actually the business of regression. This distinction that should be kept in mind and more importantly correlation and regression are both used in the context of forecasting in various ways. So we will see in a short while how they are used. For instance if the relationship between two variables x and y is plotted on a diagram known as the scatter diagram, you would find for instance in the first graph here. Generally as the one variable increases the other variable also tends to increase, so this is said to be a situation where the correlation is positive. The correlation coefficient between the two variables would be positive when it shows a trend of this nature. On the other hand when the relationship of this nature, i.e., if one particular variable increases and the other variable decreases then the correlation coefficient is set to be negative. Negative correlation therefore really means that as one variable increases, the other decreases. That is the physical significance of the correlation coefficient. These variables are sort of randomly distributed the way we have in this third diagram

(Refer Slide Time: 04:03)



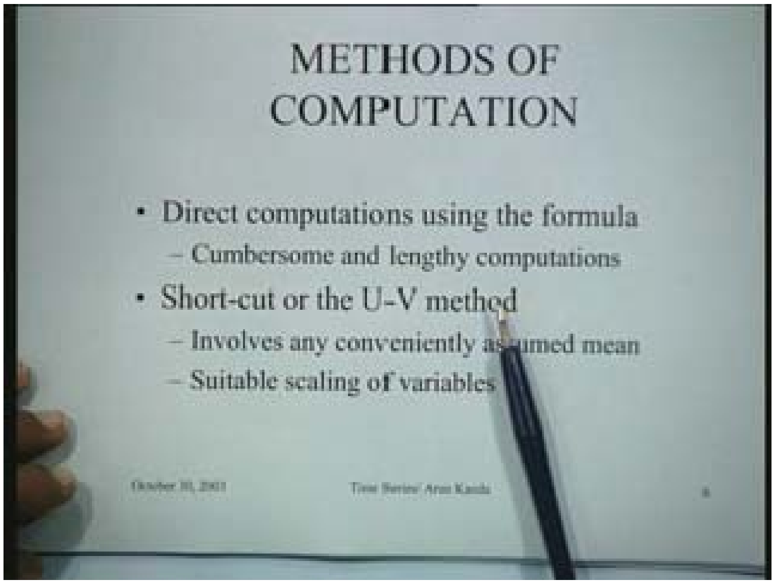
The implication is that there is no correlation between them, they are totally uncorrelated and this would mean that the correlation coefficient is zero. What can also happen is that the variables would establish some kind of non linear association which would essentially mean that the different points, the correlation coefficient would vary from positive to negative depending up on how the situation is. The most commonly used correlation coefficient is the Pearson's correlation coefficient r which is measured by this particular formulae which says if there are n data points and you have the series x and y defined on this end points, then one up on n sum of x – x bar into y – y bar the summation of this expression. Over all the end points divided by sigma x and sigma y is actually defined as the correlation coefficient.

(Refer Slide Time: 05:18)



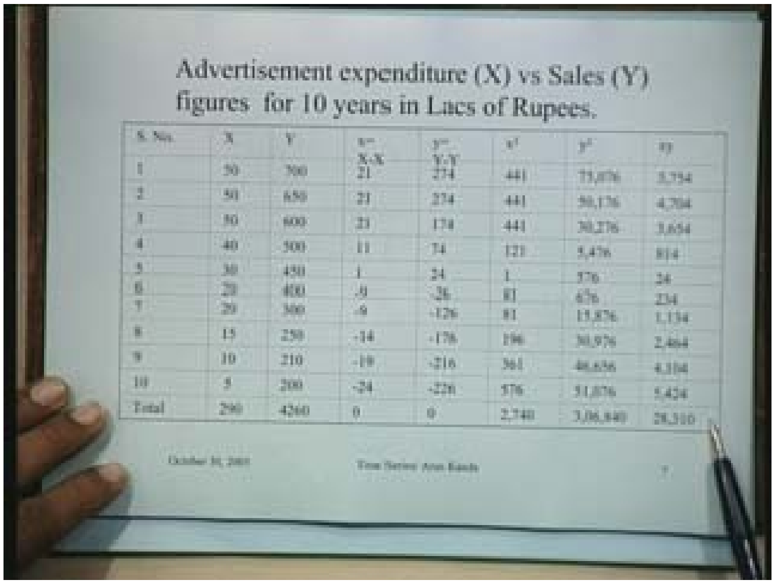
This is needed in a number of applications where we are expected to calculate the correlation coefficient. The numerator which is this expression 1/n sum of x – x bar into y – y bar is generally defined as the co variance between x and y. So covariance divided by the sigma x and sigma y will actually give you the correlation coefficient and this is how it is. There are various ways of computing the correlation coefficient between the two series, one of the methods is the direct computation using the formula we had just defined. This is generally a cumbersome and lengthy method for the purposes of computations.

(Refer Slide Time: 06:12)



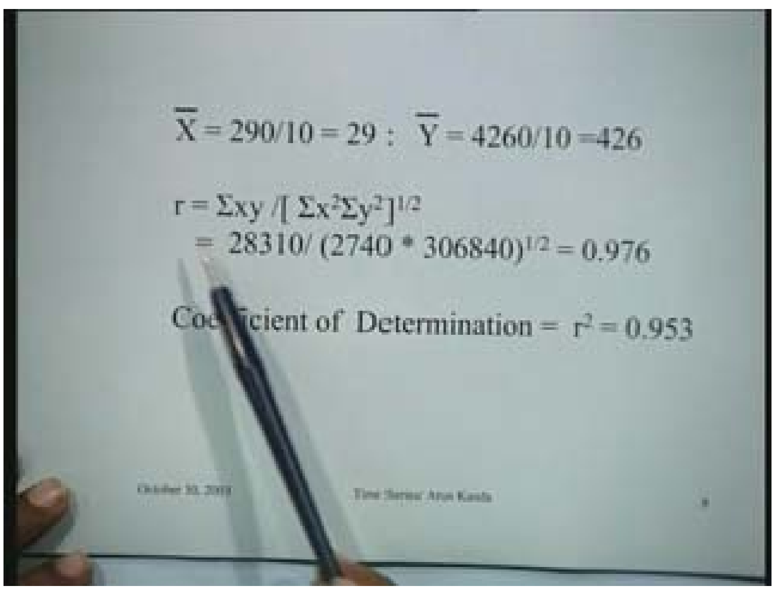
One can also use the short cut or the UV method which is basically like shifting the origin. It involves any conveniently assumed mean and then suitable scaling of variables. This is generally pretty easy for purposes of computation. For instance if you have these two series, let us say x is the advertisement expenditure in 10 years that a company port sale and the sales figure over 10 years is in lakhs of rupees which are shown here. We are trying to find out whether there is any relationship between the advertisement expenditure and the sales. We want to find out the correlation coefficient between these two variables. What you can easily do is if you want to use the simpler procedure you can take for instance any value corresponding to the x – x bar y – y bar and then get these particular values for instance what we have done here is if you assume that this is 1, then these values will be –9. This value would be eleven in terms of x – x bar these particular values and similar values would be there for y – y bar having computed this small x and small y you can calculate the series x square y square in the form of a table.

(Refer Slide Time: 07:47)



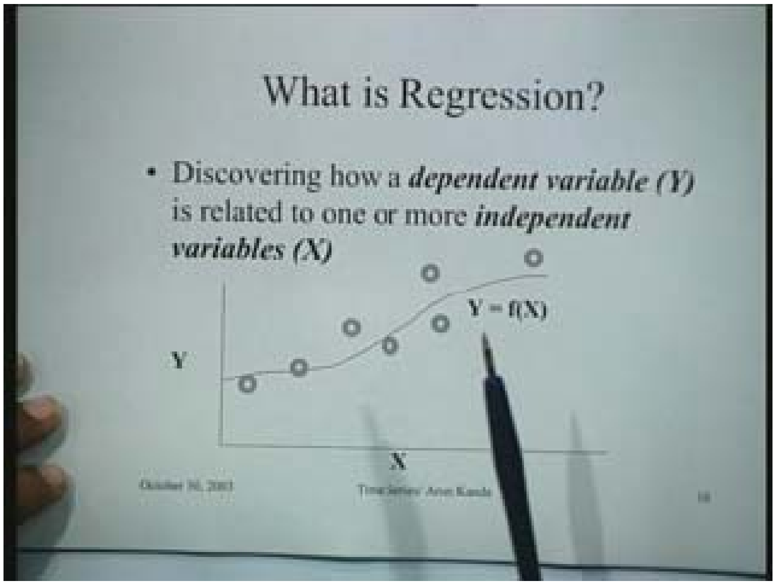
Compute the value of x, y and get the totals of the bottom of these particular things and once you do that you know for instance that x bar is simply 290 by 10 which comes from the values that we had computed at the bottom similarly value for y bar.

(Refer Slide Time: 08:06)



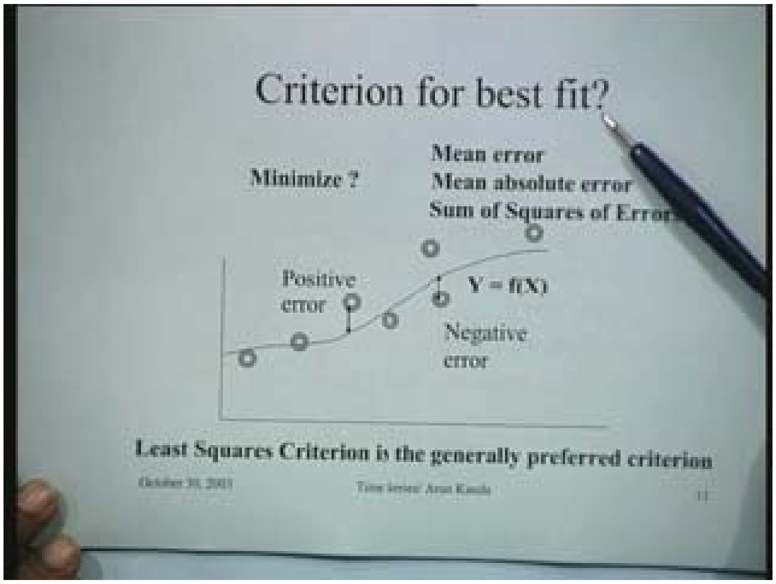
This particular coefficient which is the correlation coefficient is defined as summation x y divided by summation x square into summation y square under the root. These values are now available at the bottom of the table very conveniently, you can get 0.976. Therefore the coefficient is a determination which is the square of the correlation coefficient will be 0.953. This was just to illustrate that the computation for the correlation coefficient and the coefficient determination can in fact be done in a very simple way. After this let us come to regression. Regression is the methodology that is quiet often used for instance in trend extrapolation and casual models for defining a function which can be used for purposes of making forecast of demand. So essentially if we try to ask ourselves this question as to what is regression basically the purpose of regression is discovering how a dependent variable y is related to one or more independent variables x or x_1, x_2, x_3 and so on up to x_n

(Refer Slide Time: 09:37)



Essentially from the demand history, demand is the function of time that might be available with us or with y as the function of x in general. We are trying to determine a relationship of the dependent variable y as the function of the independent variable x. I think it is important to know how you define your dependent and independent variables. The dependent variable is generally the variable that you are interested in, that is the demand and the independent variables are the inputs that you have available at based up on whatever data is available, generally that is how it is. In fitting a function the question that arises very naturally is what is the criterion for best fit?

(Refer Slide Time: 10:31)

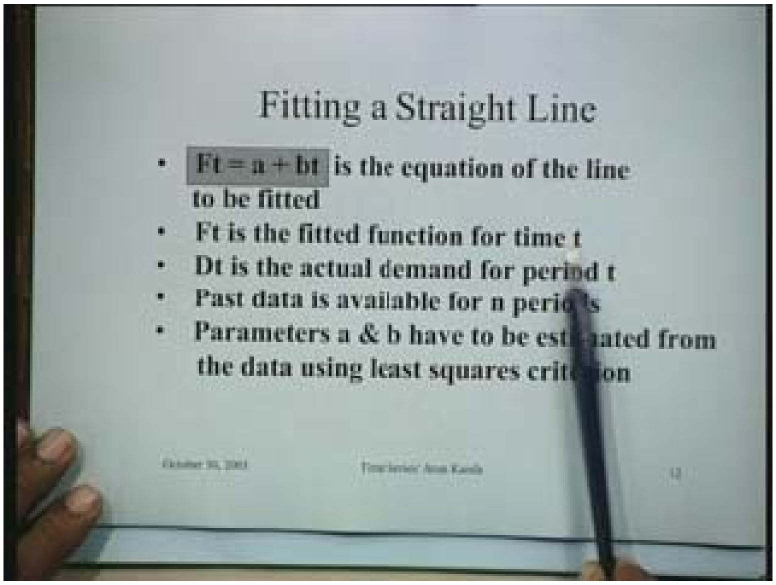


In fact a number of criteria have been suggested for fitting this particular function. You see whenever you fit a function y is equal to f of x what would happen is this is the plot of the function, some points will be above the functions, some points will be below the function in variable. This is the positive error; this is the negative error in the sense, so you would like to ensure that there are no errors. In fact if you could develop a function which passes through all these points without any errors that would in fact be error less regression function. What happens is that normally since that is not possible, we try to develop a criterion which is based on these error functions.

One of the criterion that has been used is the mean error that means you calculate the error here, some positive error, some negative error and ultimately we calculate the mean and you are interested in minimizing the mean error. Clearly this is not a very good criterion because if there are two points on opposite sides, no matter whether the error is 20 +/ – 20 or whether the error is +/ – 100, both of them could be rated equally by this criterion. Therefore you would like to have a criterion which measures the degree of fit to a better extent. Another criterion is the mean absolute error or the mean absolute deviation sometimes knows as MAD, the mean absolute deviation. Your objective function is take the modulus of this, modulus of this, the modulus of this, so every where you get a positive contribution from each point, and then you take the sum so that is the mean absolute error which is the sum of the modulus values of all the errors and you treat that as a criterion. The major defect with this criterion is what the major defect with this criterion is.

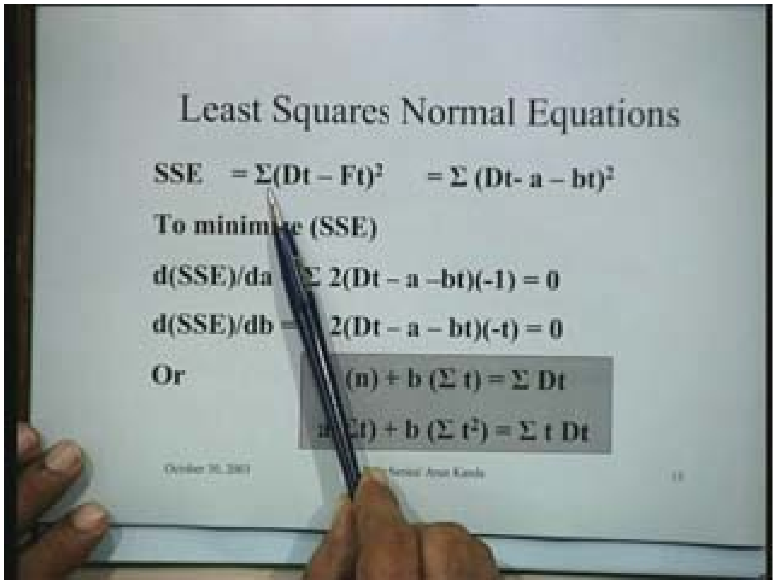
What do you think would be the major defect with this criterion? Obviously the major defect with this criterion is that first you are dealing with the modulus function, now the modulus function is not a mathematically well behaved function. You can integrate it. It has the cusp at that particular point and therefore you cannot perform all the mathematical operations of taking the derivative and equating the derivative to zero or things at that type therefore this is not popular. The third criterion which is most commonly used in fact is to try to minimize the sum of the squares of the errors. So if it is negative it becomes positive. The sum of the squares is the well defined mathematical function which can be utilized for you can use the methods of classical calculus to determine the maximum or the minimum in such situations therefore this is done. Moreover squaring the error is like magnifying the error to some extent, if the error is two, by squaring it you have four. So you are focusing on the error and your objective function is now the sum of the squares of the error and it is this sum of the squares that you try to minimize. It is for this reason that this is generally the most commonly used criterion for determining the correlation coefficient. As we said the least squares criterion that is minimizing the sum of the squares of errors is the generally preferred criterion for reasons that we have just discussed.

(Refer Slide Time: 14:51)



Let us start with the simplest possible situation. Suppose you are interested in fitting a straight line, that is the forecast at time t = a + bt is the equation of the line to be fitted,
Ft is the fitted function of the forecast for time t, Dt is the actual demand. For period t which is available, past data is available let us say for n periods and parameters a and b have to be estimated from the data using least squares criterion. That is in fact the problem, so what can be done in such a situation simply is to develop our objective function, which is the sum of the squares of the errors. The sum of the squares of the errors is nothing but the actual demand Dt – forecast Ft whole square and sum of over all the data points.

(Refer Slide Time: 15:26)



Since our forecasting function Ft is a + bt when you substitute it here, you will get this value. Now the sum of the squares of the error is our objective function which we want to minimize and the two unknown in this particular equation or the two unknown parameters a and b. So you take the partial derivatives of the sum of the squares of the errors with respect to a with respect to b and you get in fact two into Dt – a – bt into – 1= 0 and 2 into Dt – a into bt into – t which is the coefficient of bt is equal to 0. What do you find if we rearrange this equations, such that the coefficients of a and b are here. A coefficient is simply in this particular situation it will be simply n, because this is summed of over n and two cancels out and the coefficient of b is simply going to be summation t and summation Dt would come up on this particular side here that we are trying to talk about and similarly a into summation t + b into summation t square is equal to summation t into summation Dt. Now the big this is an important equation that is why it is boxed.

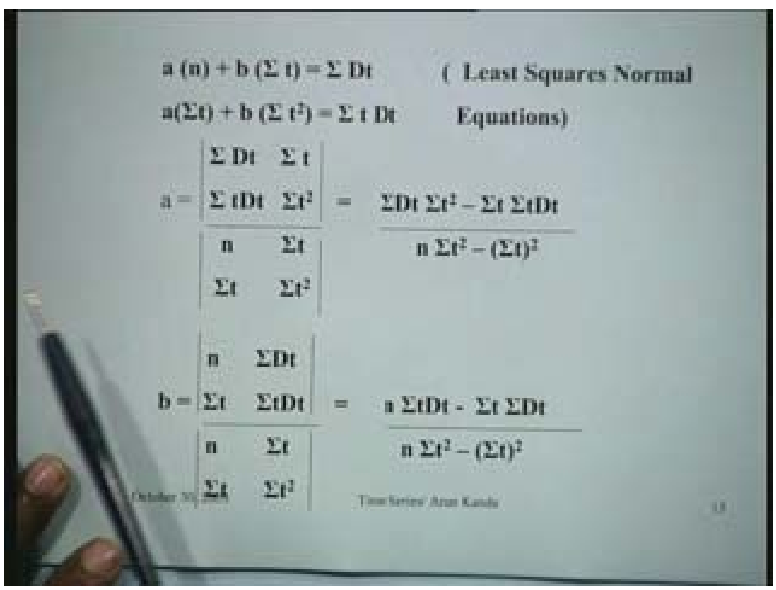
What do you notice in this equation? There are two unknown parameters a and b, so it gives the relationship between these two unknown parameters, everything else is available from the data. We have got two linear equations in two unknowns. The two unknowns being a and b, so you can solve them by Cramer’s rule or any other method for solving linear equations. Sum of squares of errors taking derivates, put it equal to zero, these equations are generally called the least squares normal equations and the point to notice is that the least squares normal equations are all linear equations here.

In fact when we use regression as a method for forecasting, the type of regression is generally characterized by the type of least squares normal equations. If the least squares normal equations are linear, the way they are here the method is called linear regression. Even if I had a function like a + bt + ct square which will be a quadratic equation, when you develop the least squares normal equations, we would again get three equations between a b and c and those equations will still be linear.

Of course these terms summation tq etc will appear but it would still be linear therefore even fitting a function of the type a + bt + ct square is also a problem of linear regression although the function is non linear this in fact is something that beginners find difficult to digest, because they say that if you are dealing with the non linear function, how can it be linear regression? It is linear regression primarily because of the linearity of the least squares normal equations and this point must be understood. That is your choice, you have to make a choice judiciously and making the choice judiciously simply means that you can look at the data and see which would be the kind of function will appear you can try different guesses. Once you made different guesses we will have an objective function or the sum of squares of errors which will tell us which fit is better. That is the normal way, so a characteristic about the least squares normal equations, they are two linear simultaneous equations in two unknown parameters a and b which can be solved by any of the well known methods such as Cramer's rule equations are called least squares normal equations. If you look at the structure of this equation, you will find that there is a very easy way of determining the solution for practical purposes.

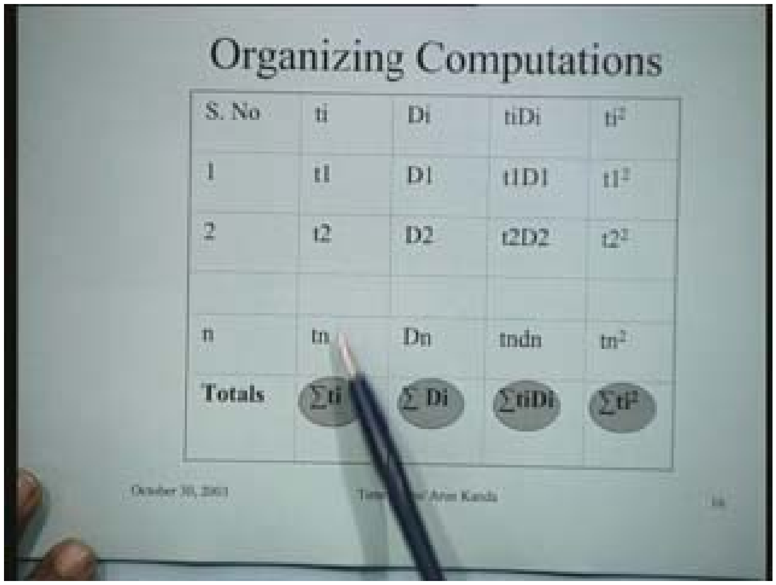
These are the equations, these are the coefficient, these are a and b which are variables, the unknown parameters and we had these coefficients and so on. What happens is if we use Cramer's rule for instance, we would have nothing but in the denominator we will have this determinant and summation t, summation t, summation t square. This would appear in the values of both a and b and as for as the numerator is concerned, you will simply substitute this column in this equation.

(Refer Slide Time: 20:45)



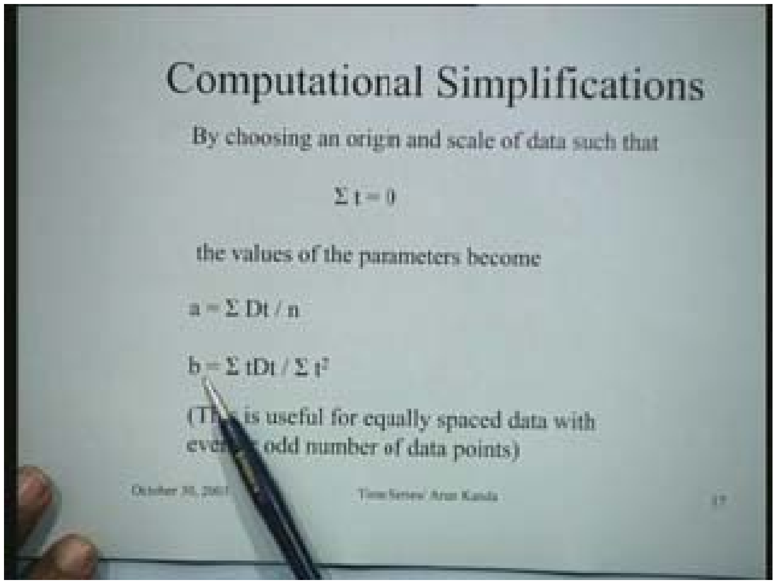
You substitute for this summation Dt summation t Dt and similarly for b you simply substitute this value in the second column here and you get the values of a and b after a bit of simplification. If you want to fit a straight line in general what you notice from this expression is that you would require the values of this particular thing, that is you would require summation Dt, you would require n, you would require summation t, you would require summation t square and you would require summation tDt and then by plugging in these values you can easily get the values of a and b.

(Refer Slide Time: 21:43)



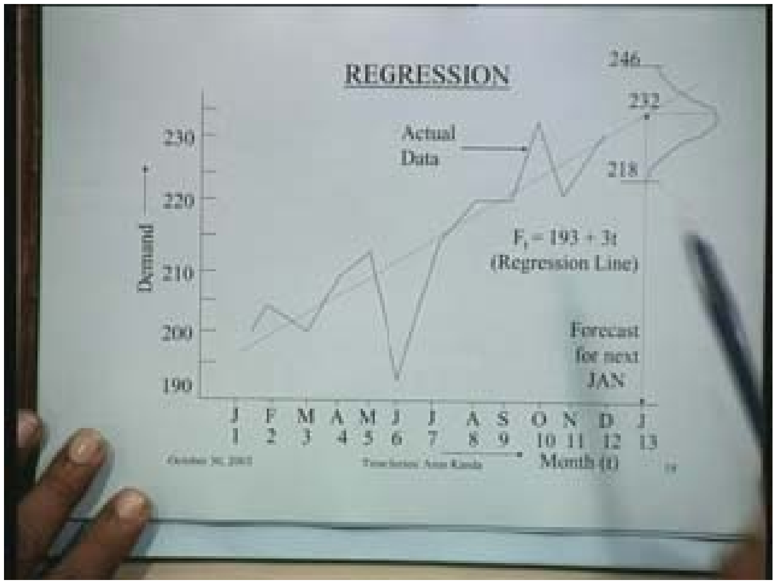
What you can easily do is based up on this requirements you can for instance organize your computations and the organizing of the computations means make a table, which means this is ti, which is your time. This is the demand which is the data available du in general, so this is just multiply these two columns you will get tiDi and then square this column you will get ti square. Then at the end you can sum them up so you get these four totals which you are interested in. This is summation ti, summation Di, summation tiDi and summation ti square. Once you have these computation available you can directly obtain the values of a and b by using the formulas which we just derived.

(Refer Slide Time: 22:39)



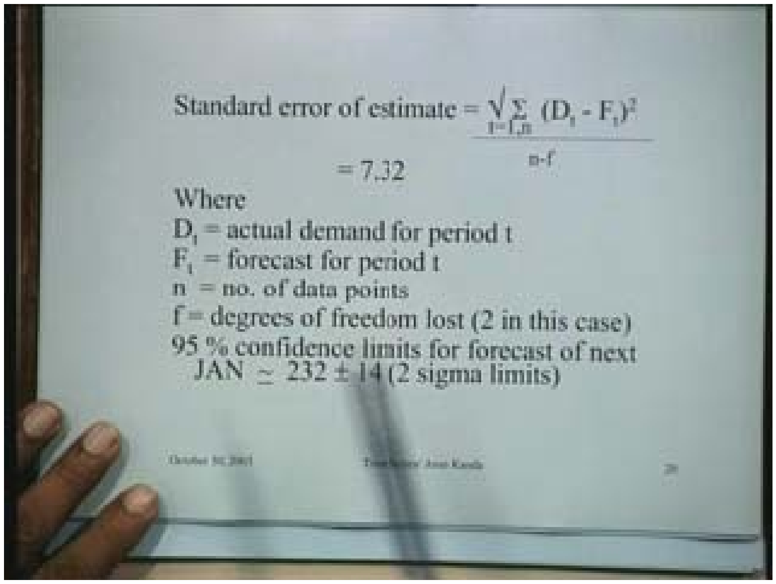
There are some computational simplifications in these calculations, one, by choosing an origin and a scale of data such that summation t is zero. The values of the parameter become very simple a is equal to summation Dt/n which is the average demand and b, the other parameter a summation tDt divided by summation t square. So this is useful for equally spaced data with even or odd number of data points because you can always choose an origin, if you have for instance and even number of points, what would you do? You can take one point as – 1, the next one can be + 1. Then since the difference of 2, subsequent points can be – 3 – 5 – 7, this condition is satisfied. On the other hand if you have an odd number of points the middle value would be taken as zero and + – 1, so one is the scale factor and then this ensures this condition and therefore is very easy to find out. Let us take an example. Suppose we have the demand of a company from January to December and this is the actual demand in units sold for the company during January to December, and we want to fit a regression equation. Let us say, to begin with a straight line on this particular function, by using the methods that we have just indicated you would find that the regression equation is a straight line whose equation in this case is Ft is = 193 + 3t and the actual demand data is actually fluctuating above and below this particular line has shown here.

(Refer Slide Time: 24:07)



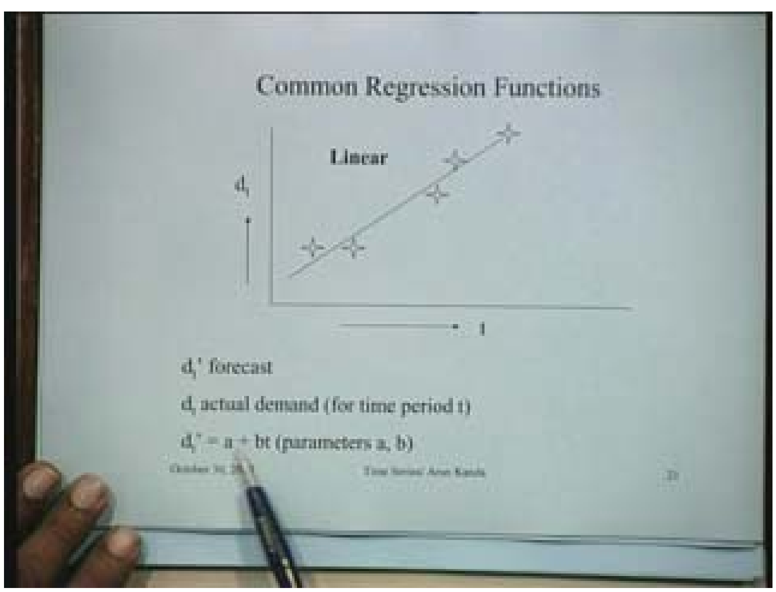
Then you can use this particular equation to forecast, let us say the demand, for next January. How do you forecast the demand for next January? You will take t is equal to 13 because this is the thirteenth period, plug it into this formula and the value that you get is 232. You make a forecast for next January as 232. What we can do is from the errors we can also estimate the standard error of estimate and find out the range in which the forecast will lie. The standard error of estimate for this example demand minus forecast is the error. So sum of the squares of the error divided by n – f where f is the number of degrees of freedom lost which is equal to 2 because two parameters a and b is estimated.

(Refer Slide Time: 24:59)



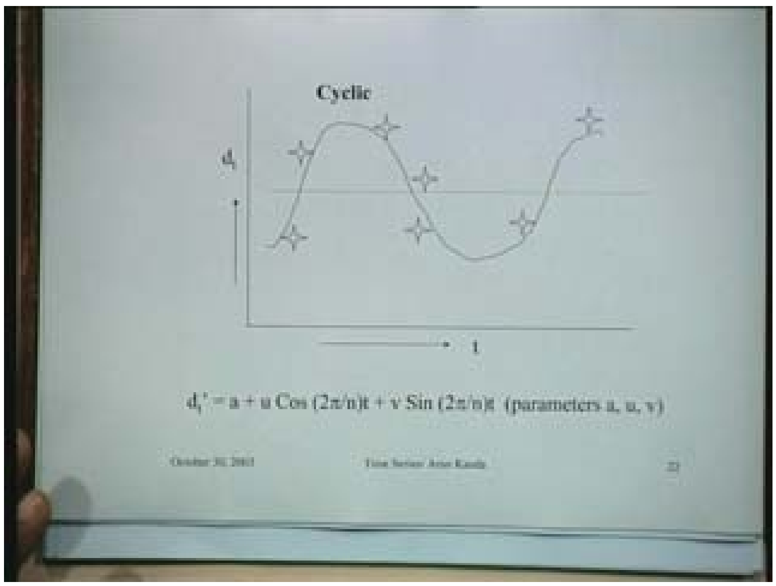
This comes out to be approximately 7, so if you want the 95 percent confidence limits of the forecast for the next January, it will be 232 + (– 14) which is two sigma limits and therefore the interesting thing is that you can get from regression not only the expected demand for next January, but also make a statement that there are 95 percent chances. The demand for next January will lie between 218 and 246 and the demand is 232 that is the interpretation.

(Refer Slide Time: 26:04)



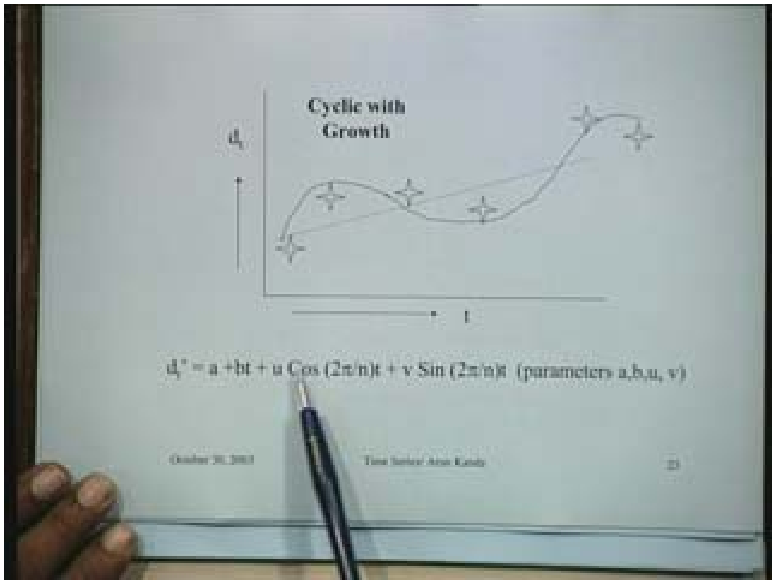
Regression could be applied to a variety of function and we have just seen the application of regression to a straight line situation. For a straight line let us say if d dash t is the forecast and Dt is the actual demand, basically our equation d dash t is equal to a + bt and there were two parameters a and b which we estimated the way we did using least square regression and you will be able to get a straight line. We might on the other hand look at cyclic demand or seasonal demand. The demand for woolen garments, the demand for air conditioner, and the demand for all seasonal products would be something of a sinusoidal way like this. If the periodicity of this particular cycle is capital N or small n as we have shown here, the equation of this line is d dash t is = a + u Cos 2 pi/n into t + v sin 2 pi/n into t. If we use this equation in the same manner and compute the least squares normal equation, we can estimate the three parameters a, u and v. This will be determined again by linear equations.

(Refer Slide Time: 27:10)



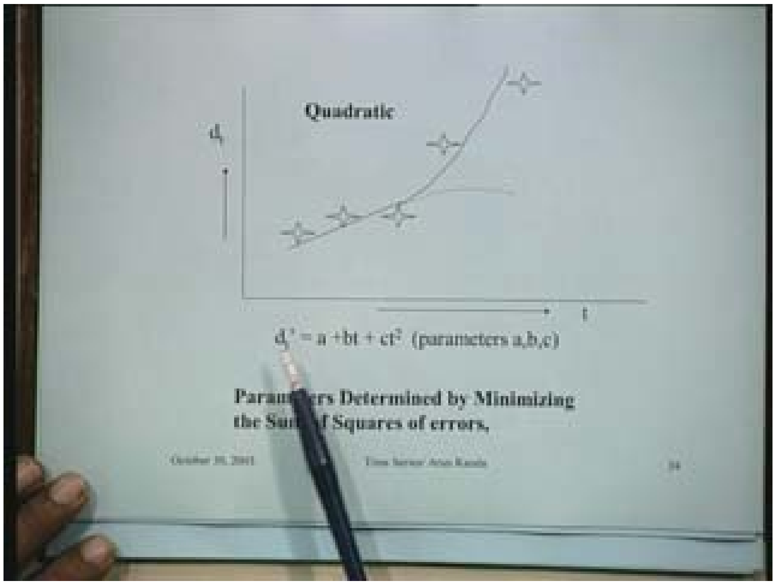
This is also linear regression although the function that we are estimating is cyclic. You might have a situation where the demand is cyclic with growth that means it has a cycle but the peak in January 2002 is lower than the peak in January 2003. So there is a pattern of growth. What we do is a + bt is the component of growth and this is the periodicity, that is, this is the cyclic component, the sinusoidal component.

(Refer Slide Time: 27:53)



So u Cos 2 pi/n into t + v sin 2pi/n into t, if you were going to use this, you would need to estimate a, b, u and v, these are the four parameters. In this case you would have four equations which will determine the parameters. The least squares normal equation for this particular place and the procedure would be similar.

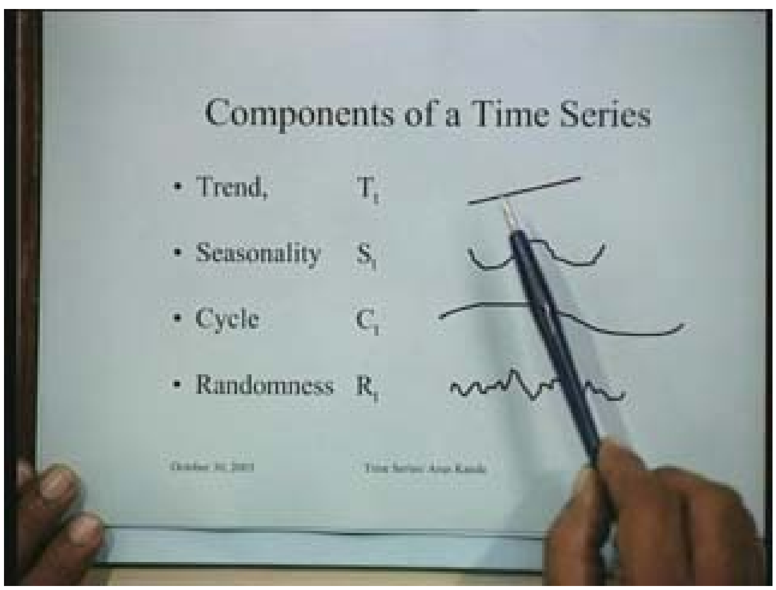
(Refer Slide Time: 28:26)



We might have a situation where the demand is to be represented as a quadratic function that is d dash t is equal to a + bt + ct square. What would happen is that if c is positive then the demand will take this form if c is negative, the demand will take this form physically. Therefore depending upon the data available, you can determine the parameters by minimizing the sum of the squares of errors in the usual fashion. These are some instances of commonly used regression functions which you can now utilize for purposes of determining the appropriate forecasting function. Now every common method of forecasting demand is time series analysis using decomposition, so we are next going to be talking about time series analysis using decomposition and see how this particular method of forecasting is actually applied in practice. The basic idea behind this particular method is that you know, what is a time series? A time series is a function which varies with time. When you talking about the demand history in a certain period, the sales are 20, next is 25 then it could be 18 and then could be 50. The time series are something that changes with time. It is a general term. What is actually assumed in a time series decomposition method is that every time, series that you encounter has these components. It has a trend.

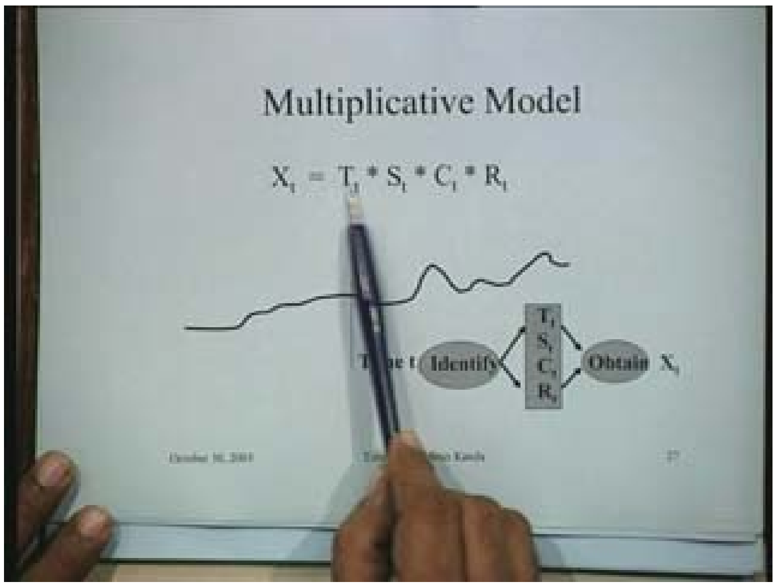
What is the trend generally? It is trying to go up or go down, so that is the trend component. Secondly it might have seasonality. That means the values might tend to go down and then go up and then go down and up. So that is the seasonality component of the demand. Seasonality is generally in over a year, over the seasons. You might have a cyclic component. A cyclic component means that there would be a general period of growth and then a depression and so on. So basically here we are talking about large business cycles. There might be a business cycle and therefore this could be imposed on the demand that is another cyclic component.

(Refer Slide Time: 31:13)



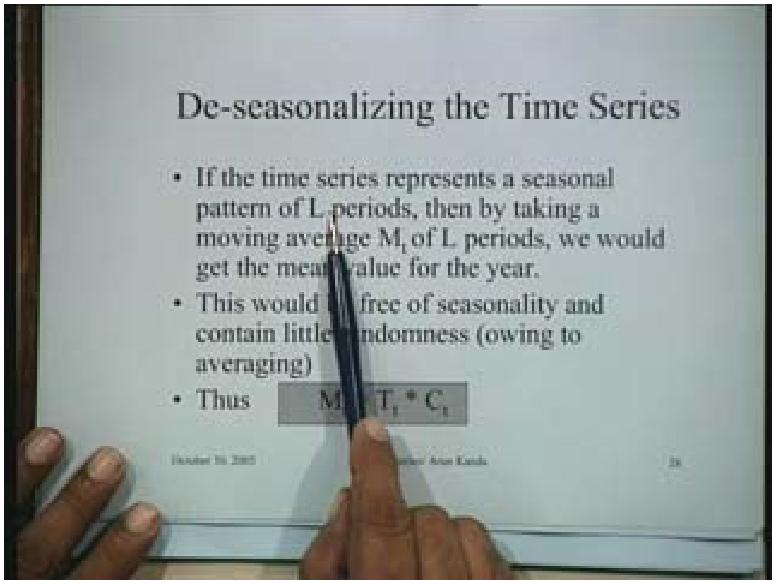
Apart from these there is always some randomness, which means something goes up and down randomly along these cases. So the basic philosophy in time series decomposition is that we try to doubt these components. Once we know the components we then try to utilize these components to forecast the demand that is what it is. The time series is decomposed into these various components and then these components are put together and you can forecast the trend separately. You can forecast these. Then from that forecast you can try to build the demand for the time series, that is the basic idea. The most commonly used method for determining these components is the use of a multiplicative model, although the additive model is also used in some situations but this is the most popularly used model in time series decomposition. What it says is that the forecast for time period t will be the trend for time period t multiplied with the seasonality component, for time period t multiplied with the cyclic component, time period t multiplied with the random component for time period t.

(Refer Slide Time: 32:49)



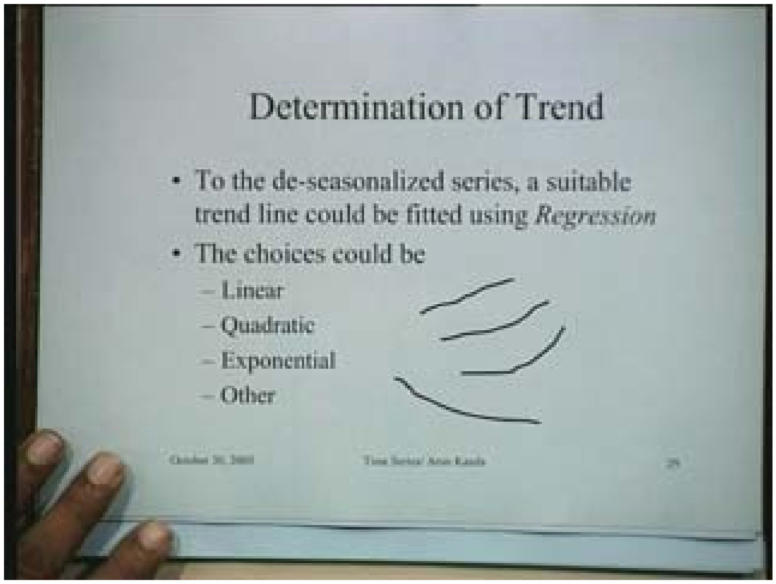
You have to actually get these components and these components will also be the functions of time and you can put them together. The basic philosophy is that if this is the time series which is generally going at a particular time t, if you want to find out, you have identify the trend, the seasonality, the cycle and the randomness. Having obtained these components, you can put them together and use this (Refer Slide Time: 33:18) formula and obtain Xt, that is what it is in the multiplicative model. A very important concept in the development of the time series decomposition is the notion of what they called de-seasonalizing the time series. The basic concept is that if the time series represents a seasonal pattern of L periods, then by taking moving average M_t of L periods, we would get the mean value for the year. I think this is obvious, for instance if there are four seasons, each season is of three months duration. If we take that demand average for three successive periods that would in fact be the average demand over these various periods.

(Refer Slide Time: 34:38)



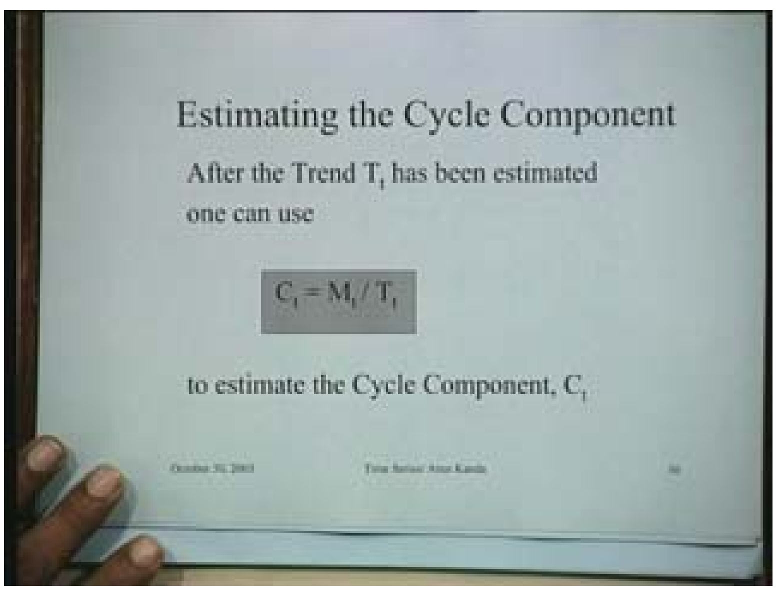
What it means in general is if you recant that the periodicity or the moving of the seasonal pattern is of L seasonality pattern, then you take a moving average of L periods and you would get the mean value of the year. This would be free of seasonality and contain little randomness. Why? because of averaging. Averaging always trends to reduce the randomness and therefore the moving average at time t will be the product of the trend and the cyclic component, because we have taken care of the others factors then de-seasonalizing the demand means getting rid of the seasonal effect. How do you get rid of the seasonal effect? By taking a moving average of the recusant number of periods, so you get rid of that. If you want to now determine the trend, how will you determine the trend? You take the deseasonalized series which we have just determined. A suitable trend line could be fitted using regression. So you use regression again and the choices could be very simple. It would be linear, quadratic, exponential or other function which ever function that you want to use it. Once you have a deseasonalized series which is like the average, how is it behaving?

(Refer Slide Time: 36:09)



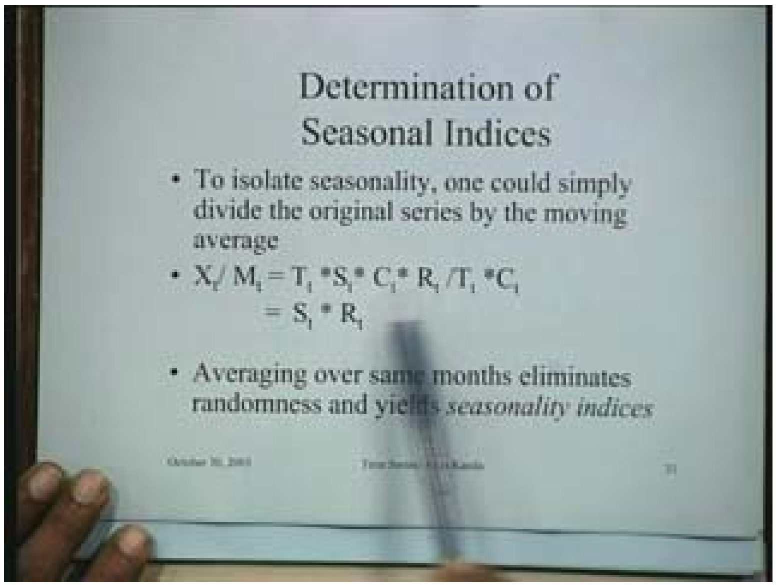
You try to find out what would be the best function to fit this. This can be determined by regression or it can be determined by inspection depending up on the degree of accuracy you want. That function will be the trend line Dt that we are talking about. Then how do you determine the cycle component after the trend T_t has been estimated? One can use $C_t = M_t / T_t$.

(Refer Slide Time: 36:30)



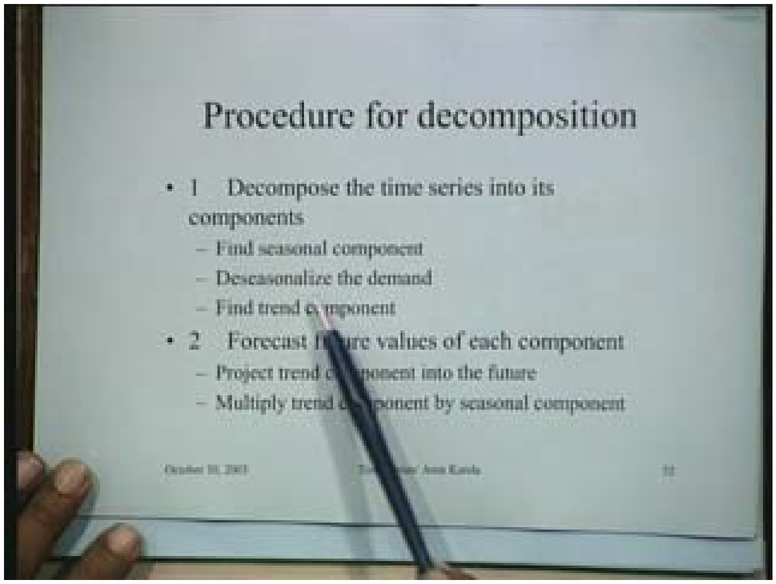
This is the moving average, the moving average was the product of the trend and the cyclic component and therefore to determine the cyclic component you can simply divide the moving average by the trend and this will give you the cyclic component, this is the idea. What remains now is we talked about the three major elements here. What remains is how exactly do you determine the so called seasonality factor or the seasonality indices in this particular situation? To isolate seasonality, one could simply divide the original series by the moving average, because the moving average is an average and everything will be above or below it. This ratio X_t / M_t will be nothing but this is X_t, the whole thing T_t S_t C_t R_t divided by M_t which is T_t C_t and basically what you find here is that by dividing this original value by the moving average, you get the seasonality multiplied with the random factor.

(Refer Slide Time: 37:42)



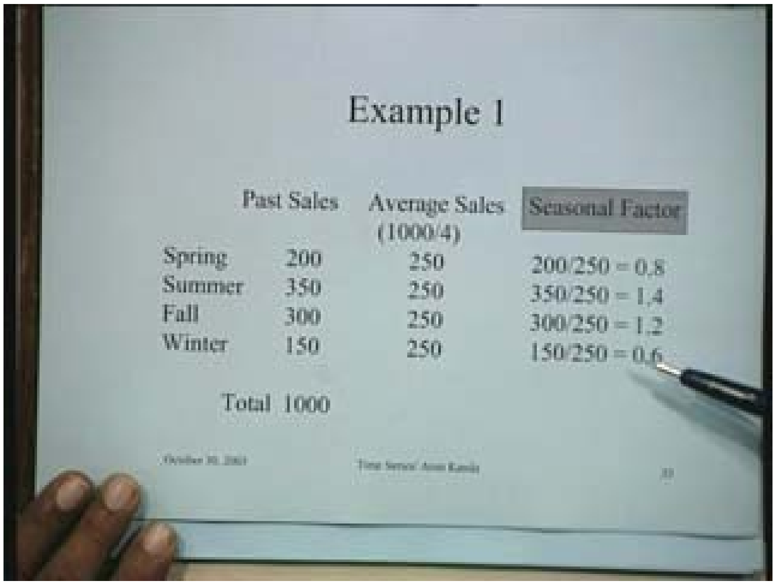
Averaging over same months eliminates randomness to some extent. This way we get the seasonality factor only that means dividing the series by the moving average. We in fact get the seasonality factors and these are generally known as seasonality indices. So we had basically seen the procedure that will be applied to determine seasonality and then this can be done through dividing the original series by the moving average or the moving average of L periods and then identifying the underlined trend which can be through regression and the seasonality indices. Now that we have got these components we can put them together and get the entire forecast. The basic procedure for time series decomposition can we summarize as these two steps.

(Refer Slide Time: 39:14)



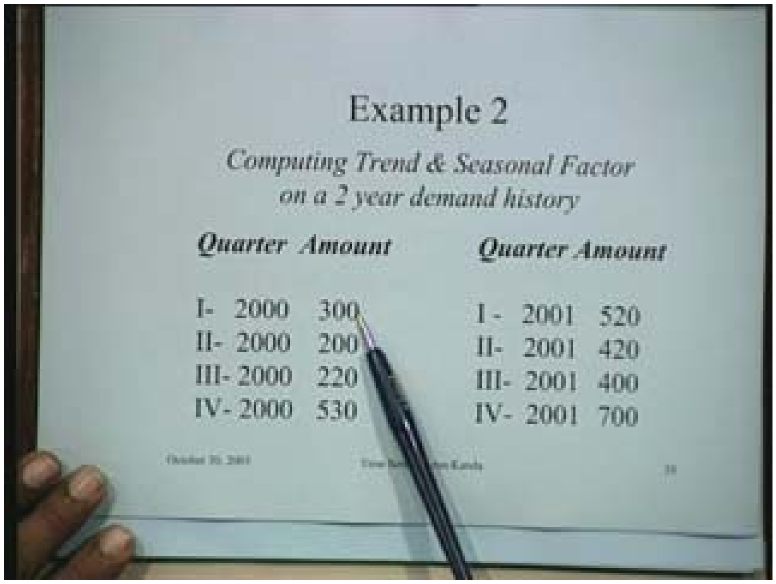
One is that you decompose the time series into the components i.e., first find seasonal component, then deseasonalized the demand and then find the trend because once you de-seasonalize, it is the series for which you find out the trend, then utilizing this information you forecast the future values of each component which are these components and then you project the trend component into the future, multiply the trend component by the seasonality component and this will basically give you the forecast for the next period. This is essentially time series decomposition which holds in this particular fashion. This was basically an over view of the procedure. What we will do now is to take three progressive examples to illustrate how the procedure will be applied. These examples will start from a relatively simple example to a more complicated example to a relatively real life example of how the procedure for decomposition is going to be applied so that you get an appreciation of this particular method. What we can do is let us take the first example, the first example is suppose demand is available to us for the last year and the past sales for different quarters are, during the spring quarter it was 200 during the summer quarter, it was 350 during the fall quarter, it was 300 and during the winter it was 155. Suppose this demand is available, this is data is available and from this data, let us see how we can compute the seasonality factor in this particular situation. We can see that the total demand over the year is the sum of all these values which is 1000. So now the average demand for each of these quarters divided by four, 1000 divided by 4 because there are 4 quarters. Average demand for each of these four quarters is 250 each. The basic idea is that the average demand was 250 in each quarter. Yet in each quarter demand was actually fluctuating. So what you can do is actual demand divided by the average 200 divided by 250 is actually 0.8. This becomes the seasonality factor for spring. What it shows is that on the average, the demand during spring was 80 percent of the average demand, that is what it shows and nothing else. Similarly during summer the seasonality factor was 350 divided by 250 which is 1.4 and this is 1.2 and this is 0.6. These are now our seasonality factors which show that in different seasons like in spring it is 80 percent of the average during summer the sales are 140 percent of the average during fall 120 percent of the average and during winter just 60percent of the average

(Refer Slide Time: 42:31)



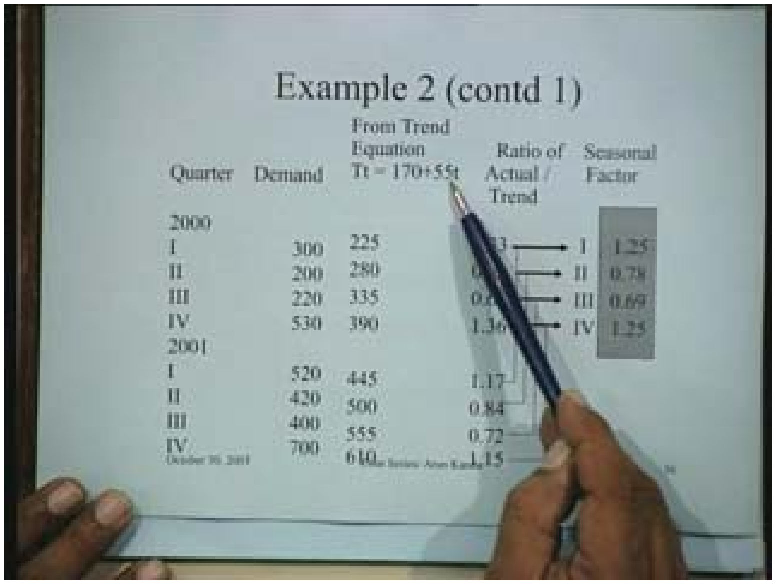
These are the factors and you have now identified the seasonality factor for this particular example. Let us say we want to make a forecast using this seasonal factor. How do we do it? What we can say is that last year our demand was 1000 we expect that next year it will go up by 10 percent. So let us say that the expected next year demand is the total of the 1100. This could actually be developed by regression or any other technique but we are just trying to identify how exactly we will put the components together. So we say that the total demand for next year is likely to be 1100 which means that that the average sales during the period for the next year is likely to be 1100 divided by 4, so 275. This is the average for next year. What we now do is we assume that the seasonality factor that you have computed in the last year will continue to operate in the next year. For the spring season it was 0.8, 1.4, 1.2 and 6. We utilize the seasonal factors there, multiply this and you get the next year's forecast. This is the intension. The intension was to obtain a forecast for next year. So we know that the demand for spring is likely to be 220. The demand for summer is likely to be 385. This is likely to be 330 and this is likely to be 165. This was a very simple example to basically give you an understanding of the use of seasonal factors. The average demand which is the basic trend is that we assume rows by 10 percent. Let us take a slightly more complicated example. Let us say that we want to compute the trend and the seasonal factor on a two year demand history.

(Refer Slide Time: 44:47)



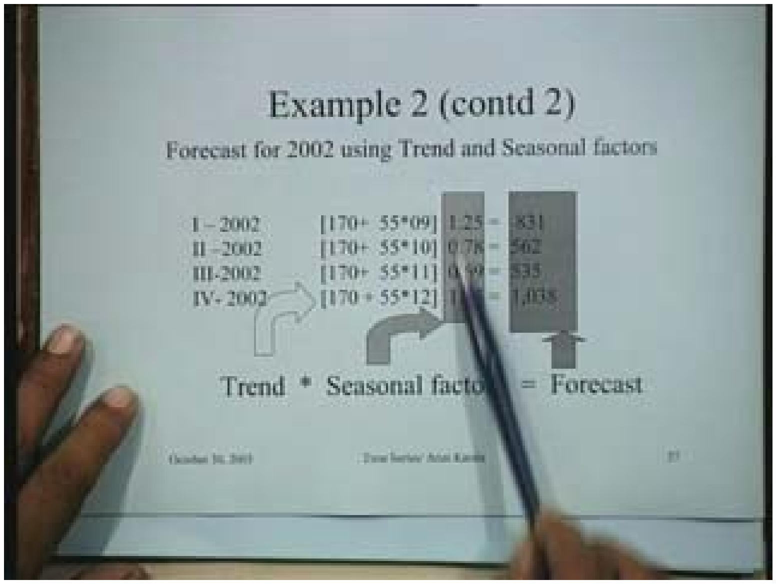
This means we have data for two years. This is the first quarter of the year, 2000 so demand is 300, 200, 220, 530 and then 520, 420, 400 and 700. This is actually the demand data available to us for the past two years. So how do we get the factors from this particular situation? What we do is we have these four data points for 2000. These four data points of these four seasons for 2001 and the demand is 300, 200, 220, and 530 and so on. To this data we fit a regression line. To these 8 points we fit a regression line and the fitted regression line is trend of these points which is 170 + 55t.

(Refer Slide Time: 45:46)



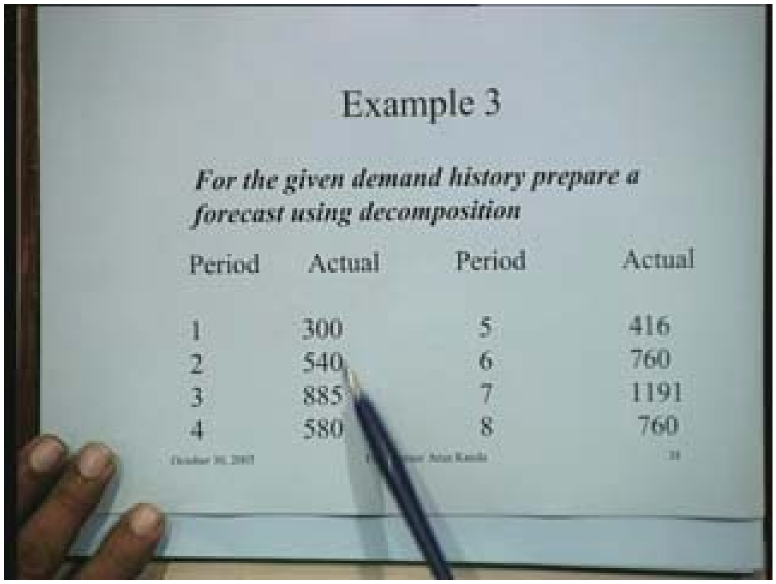
We establish the trend equation to these 2 points and we get this and from this trend equation, we can actually calculate what the demand would be. For instance if you take t =1, this would be 225 t is = 2, it will be 280 etc. These are the values that you get from the underlined trend from the regression line. Then what we can do is we are interested in finding out the seasonality factors. What you do is you take the actual to the trend. This is the actual that you have, which is 300 divided by 225. It is 1.33 and 0.71, .661, 0.36 that is for the first year. When you do this analysis for the second year, this will be 520 by 445 which is 1.17, 0.84, 0.72, and 1.15. Now these are different. What you do is to calculate the seasonality factor for the two year arisen; we take the average of this that means 1.33 and this particular value 1.17. The average of this is 1.25 and then this is 0.78, this is 0.69, this is 1.25. What we have ultimately established is the seasonality factors for each of the quarters, so the first quarter, second quarter, third quarter, fourth quarter and also the supplies in each year. Now we can use this information to develop a forecast. Forecast for 2002 i.e., the next year using trend and seasonality factors would be we use the trend line 170 + 55 into the 9^{th} period, the 10^{th} period, 11^{th} period, and 12^{th} period. Whatever it is multiplied with the seasonality factors which would just computed.

(Refer Slide Time: 47:58)



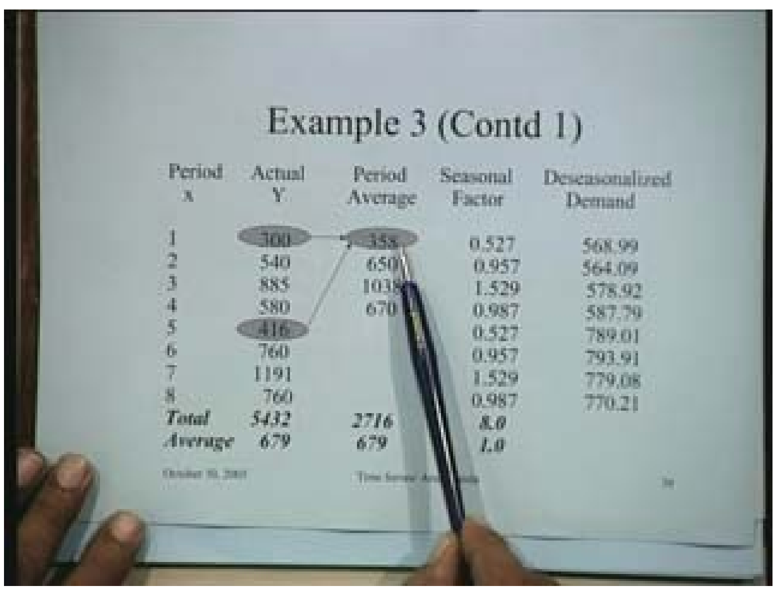
1.25, 0.78, 0.69, 1.25 and this becomes therefore your forecast for the next four quarters of 2002 which takes into consideration trend and seasonality. This is the method that you would adopt for I mean understanding both. We talk in the first example mainly about seasonality here; we talked about both trend and seasonality. Let us take a third example in which let us say we have a given demand history, we want to prepare a forecast using decomposition.

(Refer Slide Time: 48:31)



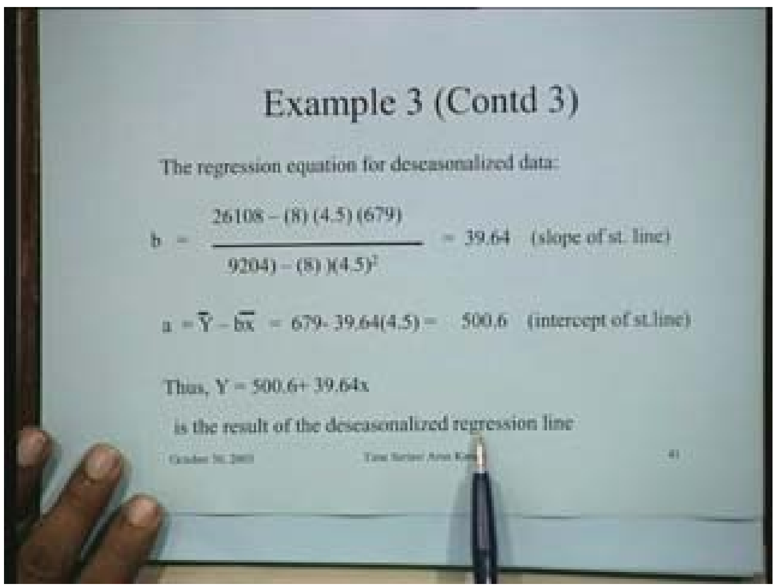
How do we do it, in general these are the eight periods, 1, 2, 3, 4 like first year and second year this is the actual demand which is available to us. What we do in this particular situation is we again have the actual demand over eight periods. So the total demand is this, so 679 is the average demand here.

(Refer Slide Time: 48:51)



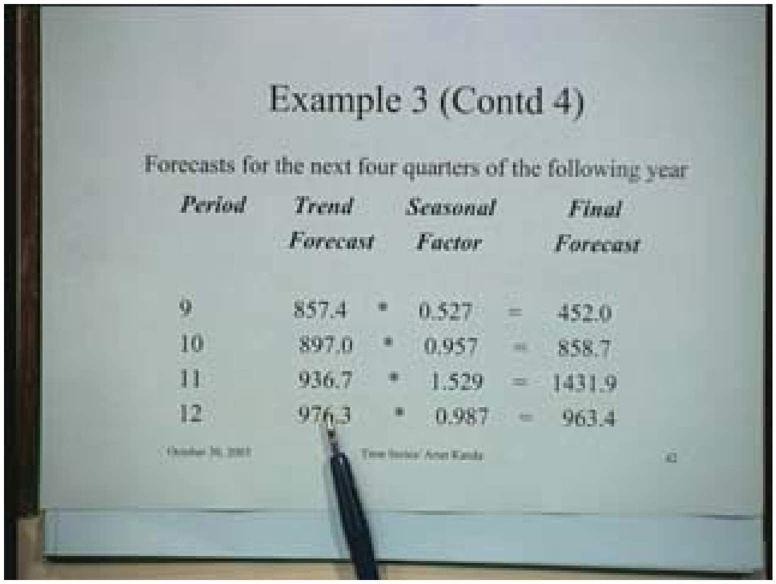
Then what we do is we can calculate the period average. What is the period average? The first period in the first year and the corresponding first period in the second year, so take the average of 300 and 416, this becomes the period average which is 358, 650, 1038 and so on and this is just for four periods, so the total will be half, i.e., 2716 and the overall average for this is 679, we have constructed this. Then you can find out exactly the seasonality factors. What would be the seasonality factors? The seasonality factor would in fact in this particular situation be, this particular value divided by the average. So 358 divided by 679, 650 divided by 679 is 0.957. This value divided by 679 is more than 1 which is 1.529 and then 670 by this is 0.987. That way we have defined it because you calculated this period averages. These factors will repeat themselves for the next year and therefore having got these seasonality factor, we can get the deseasonalized demand. How do you get the deseasonalized demand? You have to actually demand, take out the actual divided by the seasonality factor that you have. So 300 divided by 0.527, you get this particular value and similarly you get these values which are now deseasonalized demands. Then we from the deseasonalized demand have to identify the trend so to the deseasonalized demand, you fit the regression equation. These are the values, this is the deseasonalized demand, sums average is y, and this is y, x square xy and so on.

(Refer Slide Time: 51:11)



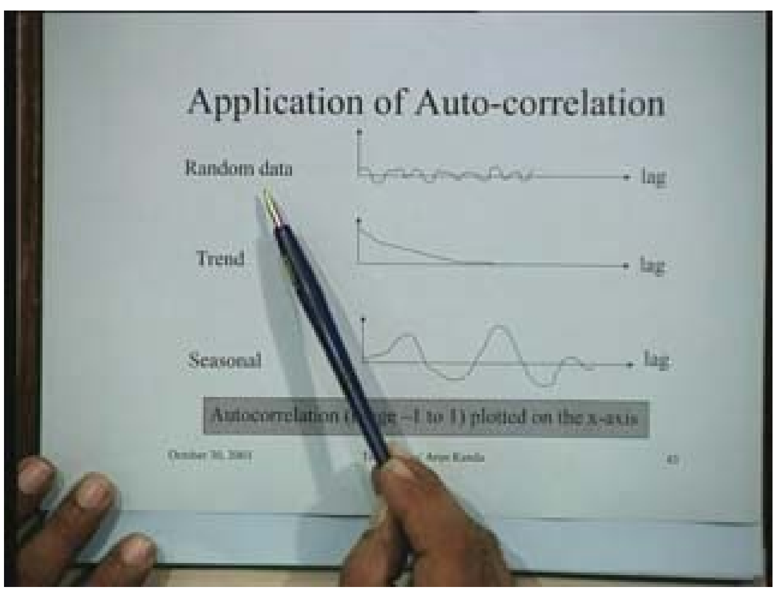
You follow the routine procedure of fitting a straight line and ultimately for this straight line you get the two coefficients a and b and the values of a and b are b is 39.64 which is the slope of the straight line, a is 500.6. This is the equation of the deseasonalized regression line 500.6 + 39.64 that is what we will do. Using this information now you can build a forecast. How will you build a forecast? Suppose you are interested in the forecast for the next four quarters of the following year that is period number 9, 10, 11 and 12.

(Refer Slide Time: 51:36)



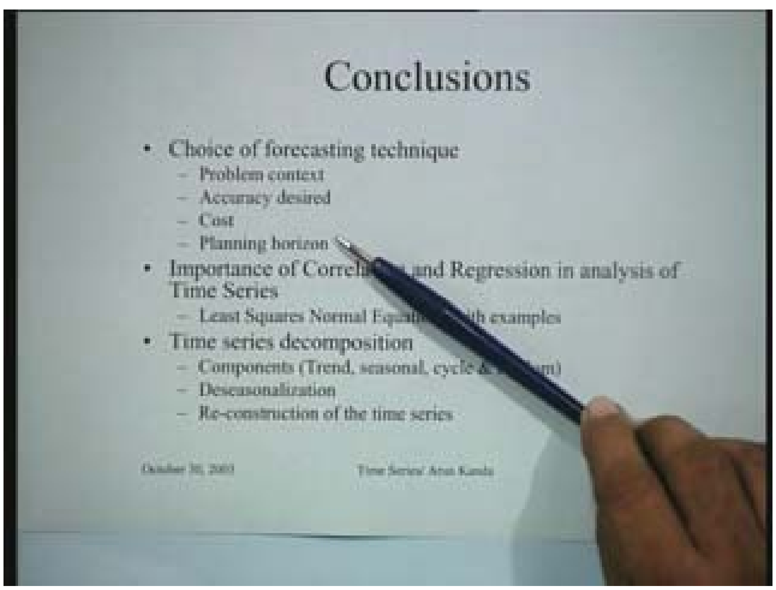
This is the trend forecast which you get from the regression equation which you have determined 857, 897 etc. The seasonality factors are the ones that you computed earlier so you multiply this and this becomes your final forecast. You get the forecast for the 9, 10, 11, 12th period. Finally the important thing to notice that to identify random data or to identify a trend or to identify a seasonal data, especially if you are using box Jenkins models, you need procedures to identify what kind of a trend it is.

(Refer Slide Time: 52:31)



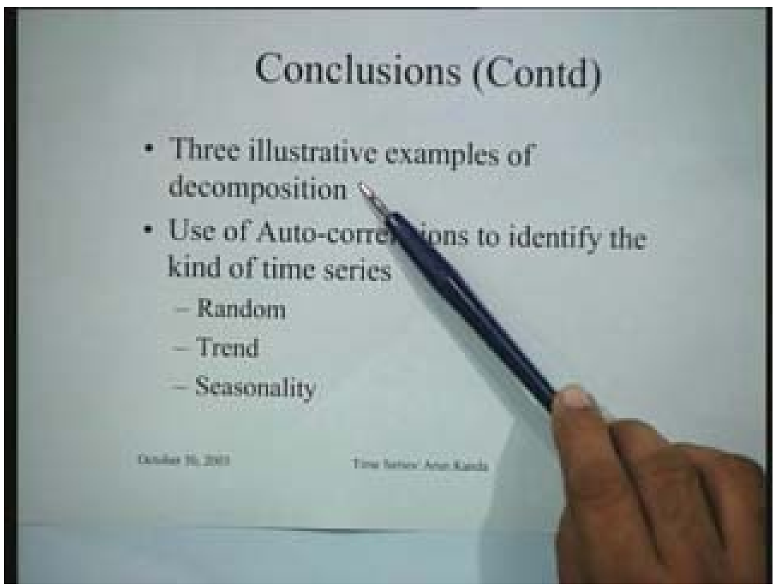
What can be done is that we make use what is known as auto correlation that means you take the correlation of that particular series. If you displace the series by one period and you find out the correlation between the two series, you have auto correlation of one lakh, auto correlation of two lakh and so on. So this is the auto correlation, which is nothing but the correlation coefficient computed by displacing the series by 1 or 2 or 3, so if it is random data what is shown is that the correlation coefficient generally trends to be small and going randomly up and down in the line. If there is a trend between operating within the forecast within the time series then it is generally found that the auto correlation coefficient trends to have this lakh. If it is seasonal the auto correlations also determines some kind of seasonality. The auto correlation are in the range of – 121 which are plotted on the x axis, so this kind of a plot is generally very useful to identify the kind of relationship that exists within the time series as to whether it is random or trend or seasonal. It is more like trying to say that you examine a patient's blood and identify different tests on it. Different test will actually show you the presence of different kinds of diseases. In the same way if you take that example if you look at a time series and calculate the auto correlation, the nature of the correlation will tell you whether it is this or this or this. This becomes the systematic approach to basically handle some of these essential questions. So let us try to see some of the major conclusion that we can derive, from what we have talked about today, the first thing that must be kept in mind is that we talked about a number of different forecasting procedures. We talked about regression; we have talked about time series, decomposition and so on. So the question that is Paramount is how do you choose the forecasting technique? The forecasting technique is dependent to a large extent on the problem context on the accuracy, you desire on the Cost that you are willing to inquire and on the planning horizon which exist freely.

(Refer Slide Time: 55:19)



Regression would be an appropriate technique if you are assured that the kind of chance caused which operated in the past will continue to operate in the future. Under that system whatever happened, could be extended into the future, so it would be a good technique but if this is not happening and the time series is behaving in some erotic fashion then time series decomposition for an arbitrary time series would be a good way because you determine the seasonality coefficient. You determine the underline trend; you determine the underline cycle and utilize this information for making a forecast. The importance of correlation and regression in the analysis of time series was emphasized the least squares normal equations with examples where illustrated so that you know how to estimate the parameters which is the basic operation required in regression. Then we talked about time series decomposition where the various components of the time series like the trend, the seasonality, the cycle and the randomness where identified and considered separately. Then from this you could build the whole thing. This building was done by first deseasonalizing demand that means getting the ripping the demand of any seasonality and it was to that deseasonalized demand that you actually fit in some regression data or some straight line so that you can construct the time series and then you went about to the process of reconstructing the time series by putting the coefficient or the components together. We also saw three illustrated examples of time series decomposition.

(Refer Slide Time: 57:11)



They were designed basically to add progressive stages of difficulty so that you could get an appreciation of the concept of seasonality, trend cycle and so on. Finally we say that auto correlation could be used to identify the kind of time series and depending upon the behavior of the auto correlation, you could have a random or a trend or a seasonality factor into which we are looking at. So with this we conclude our discussion on forecasting. We have looked at forecasting from the point of view of its importance in planning decision, and then we have seen various methods of forecasting subjective and objective and in today's lecture we have looked at some of the procedures like regression, time series, decomposition which are essentially used which are objective procedures to develop a time series forecast. Now based upon this analysis you will try to use forecasting for the various planning decisions in a company. When we start the next lecture we shall be talking about aggregate production planning where the forecast of demand is available and we see how to plan the operations of a company over the next six months or one year as the casement.

Thank you!