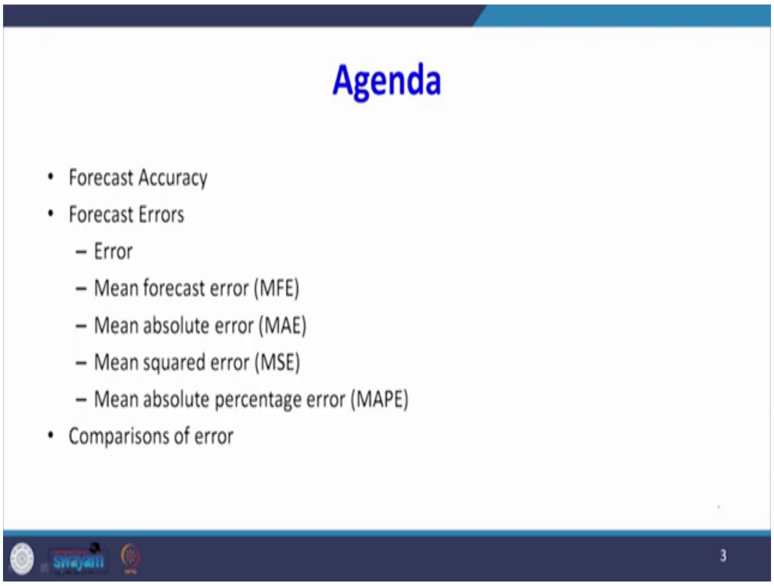
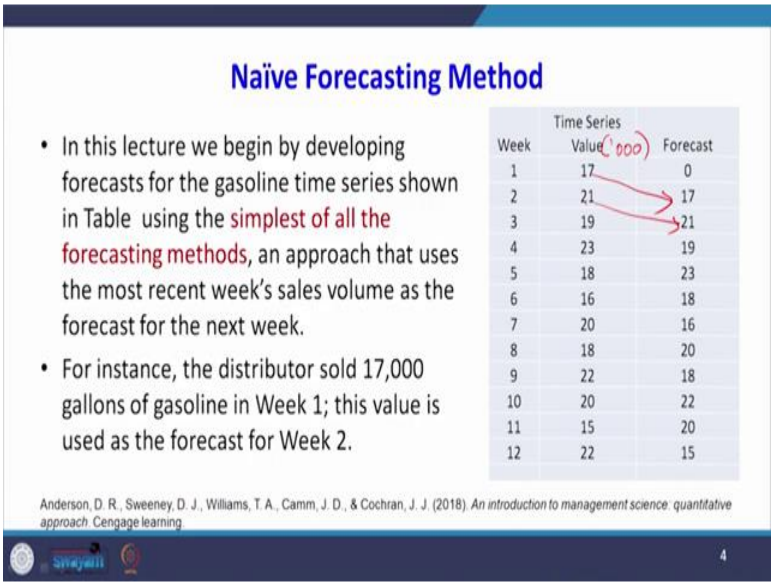
Decision Making with Spreadsheet
Prof. Ramesh Anbanandam
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 57
Time Series Analysis and Forecasting – II

So, dear students, we are discussing time series forecasting techniques; in the previous lecture, I explained various patterns in the time series data. Those patterns are used to select appropriate forecasting techniques. In this lecture, I am going to discuss measures of forecasting accuracy. These measures are also used to select appropriate forecasting techniques.

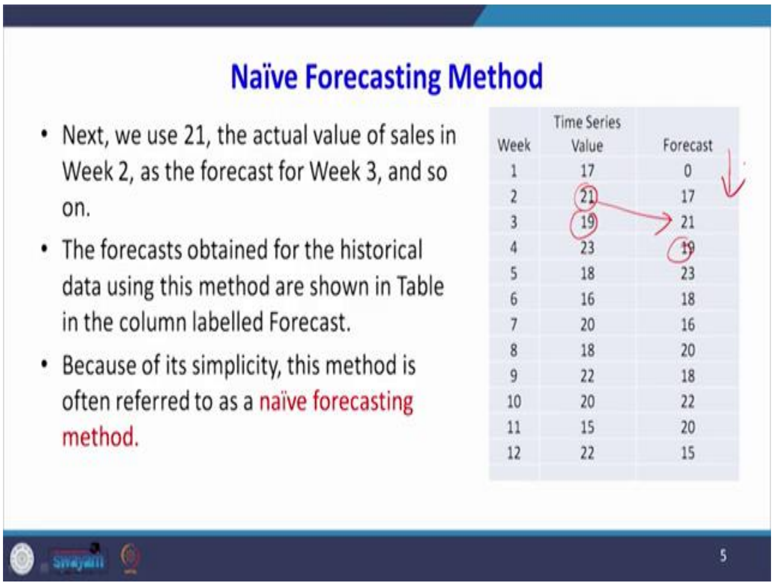


So, the agenda for this lecture is. First, we will explain what forecasting accuracy is. Then, I am going to discuss various forecasting errors. So, first, I will talk about what an error is. Then I will talk about mean forecast error, mean absolute error, mean squared error, and mean absolute percentage error. Later, I will compare 2 forecasting techniques; based on their accuracy level, we can recommend which forecasting technique is more accurate.



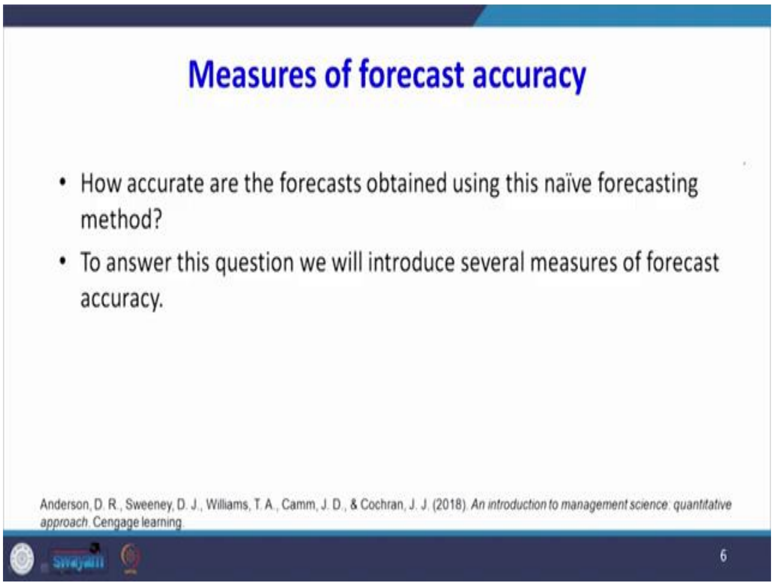
The first forecasting technique that I am going to explain is called the Naïve forecasting method. So, in this lecture, we begin by developing a forecast for the gasoline time series shown in a table using the simplest of all the forecasting methods. That approach uses the most recent week’s sales volume as the forecast for the next week. Suppose you look at the table, and there are 12 weeks left.

Week 1 sales are given 17000 gallons. These values are in terms of thousands. See, in the first year, we cannot forecast it; the second year's forecast is nothing but this value. The previous year's actual value, the third year is the second year's actual value, and the 4th year is the third year's actual value. So, this type of forecasting technique is called the Naïve forecasting method.

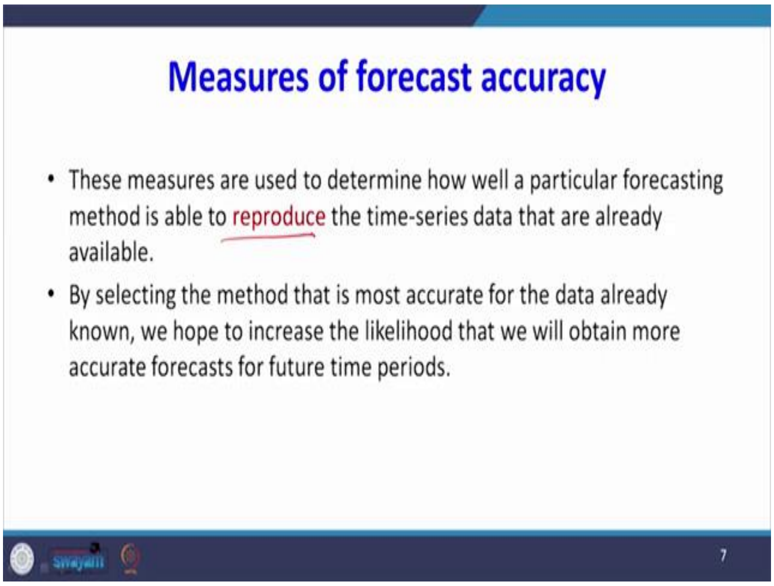


Next, we use 21; the actual value of sales in week 2 is 21 as the forecast for week 3, and so on. So, what is happening? The actual value is 21, so this will be the forecasted value for week 3. So, the actual value of week 3 will be the forecasted value for week 4. The forecasts obtained for historical data using this method are shown in the table in the column labeled forecast.

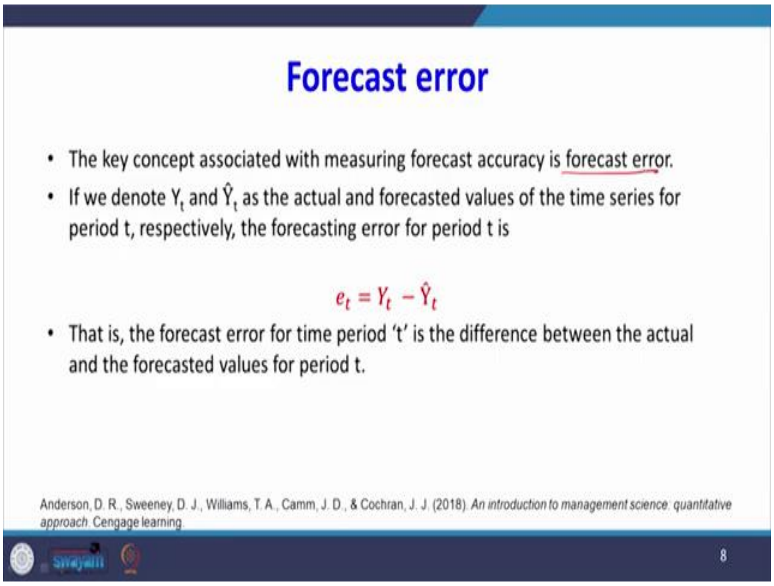
So, here I have shown everything. Because of its simplicity, this method is often called a Naïve forecasting method.



Now, I am going to discuss measures of forecast accuracy. We have forecasted by using the naïve forecasting method. How accurate are the forecasts obtained using this naïve forecasting method? To answer this question, we will introduce several measures of forecast accuracy.



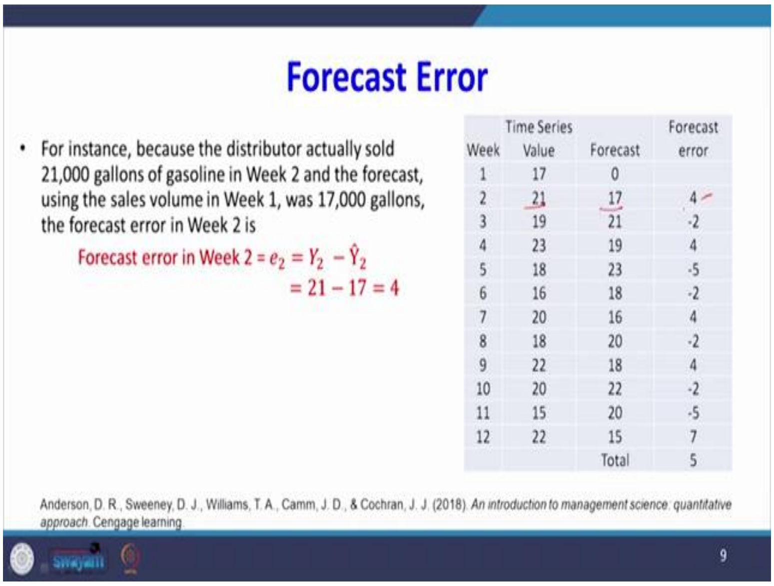
These measures are used to determine how well a particular forecasting method is able to reproduce the time series data that are already available. By selecting the method that is most accurate for the data already known, we hope to increase the likelihood that we will obtain a more accurate forecast for our future time period.



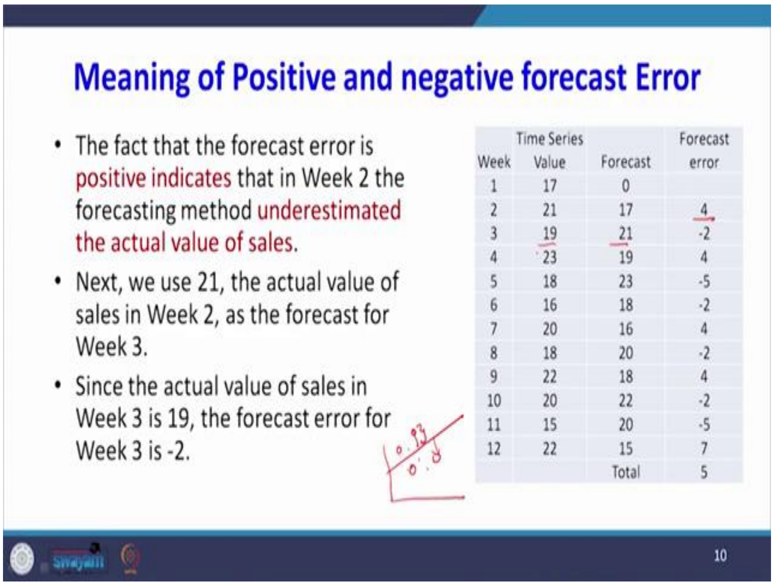
The first measure is forecast error. The key concept associated with measuring forecast accuracy is forecast error. If we denote Y_t and $\hat{Y}_t$ as the actual and forecasted values of the time series for period t, respectively, the forecasting error for period t is

$$e_t = Y_t - \hat{Y}_t$$

That is the forecast error for time period t is the difference between actual and the forecasted values for period t.

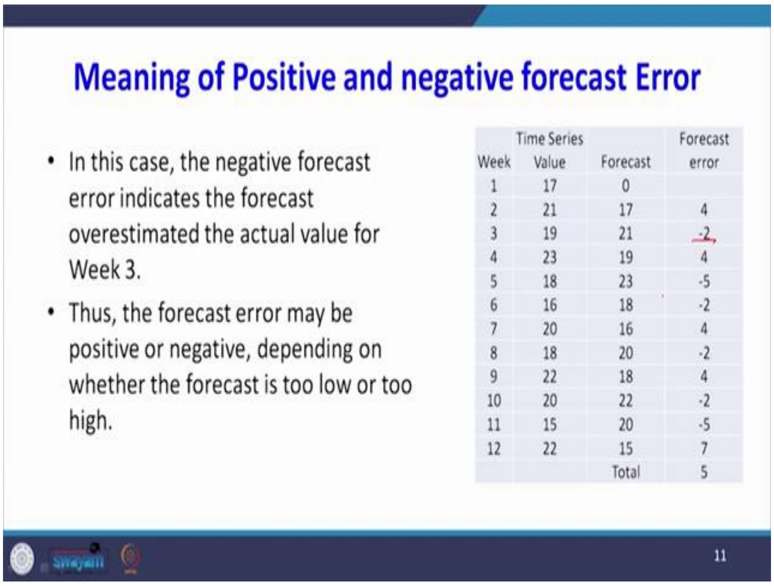


Now, in this table, I am going to explain what is the forecast error. Look at the value 4; how did we get the 4? Actual is 21 for week 2. The forecasted value is 17. So, 21 – 17 is 4. 4 is the forecast error; see this: 21 – 17 = 4.

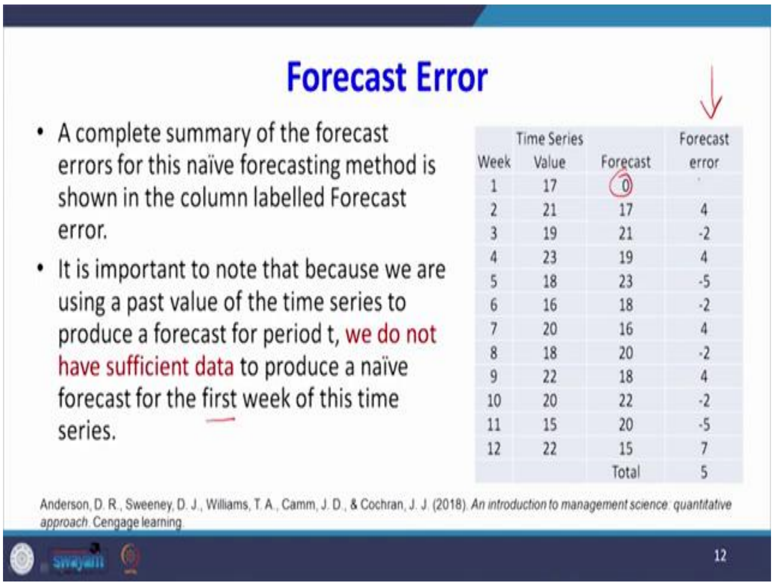


Here, what is the meaning of positive and negative forecast error? Assume that I have data like this. I suppose I have data like this data like this suppose. I have forecasted this. So, here, this error is a positive error. So, what has happened? If the error term is positive, the forecasting method you used underestimated the actual value. If the forecasting error is negative, we have overestimated the actual value.

That is the meaning of positive and negative forecast errors, which I will explain with the help of this table. The fact that the forecast error is positive indicates in week 2, you see that in week 2, the forecast method underestimated the actual values of the sales. Actually, it is a 21, but we forecasted 17, so the difference is forecast error. Next, we see the 19 and 21 actual is 19, but we have forecasted as a 21, so we have overestimated. That is why we are getting negative errors.

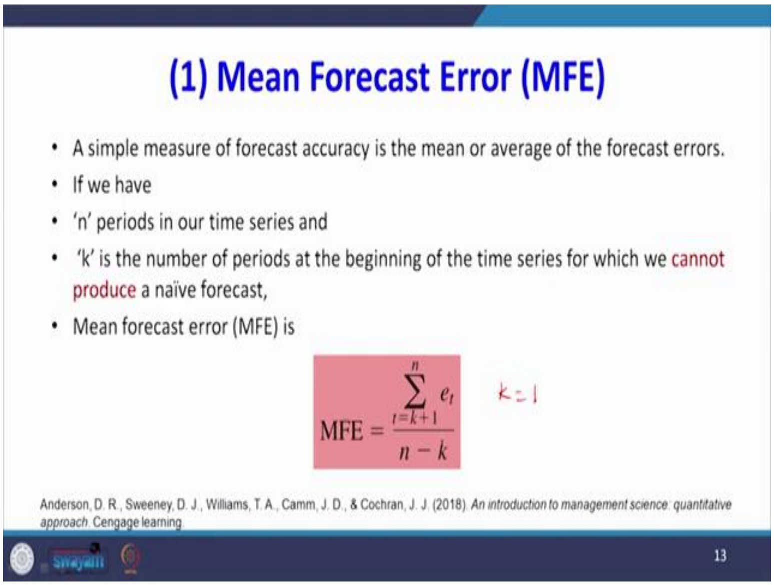


In this case, the negative forecast here indicates that the forecast overestimated the actual value for week 3; see, we have overestimated. Thus, the forecast error may be positive or negative, depending on whether the forecast is too low or too high.



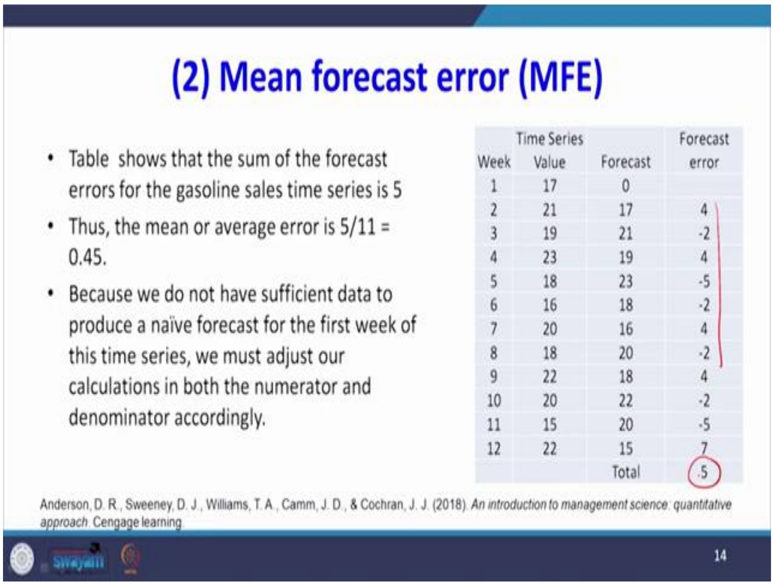
A complete summary of the forecast error for this naïve forecasting method is shown in the column labeled forecast error. Here, I have shown all the forecast errors. It is important to note that because we are using the past value of the time series to produce a forecast for period t, we do not have sufficient data to produce a naïve forecast for the first week of this time series.

See that here we are not able to forecast because we do not have the actual data there. So, there won't be any error, not error. So, we cannot find out forecast errors for week 1.



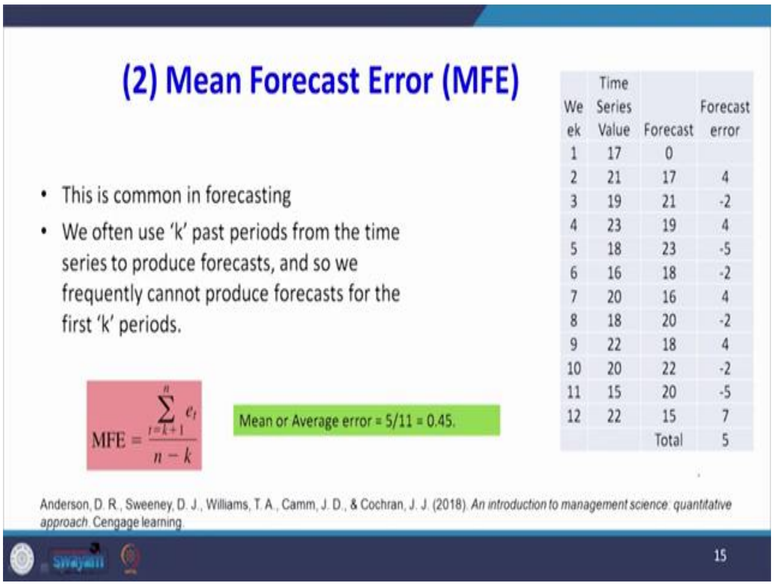
Now, the first measure of the forecast is mean forecast error. A simple measure of forecast accuracy is the mean or average of the forecast error. If we have n periods in our time series and k is the number of periods at the beginning of the time series for which we cannot produce a naïve forecast that is a k. In our problem, k = 1, because for the first data set, we cannot forecast for the first week's data we are not able to forecast.

So, the mean forecast error MFE is the sigma of t = (k + 1). The value of k = 1 because for the first data one data set, we do not have the value. So, we must start from the second week to the 12th week. We have to add all the errors divided by (n – k). So, we must divide it into 11. So, the answer will be a mean forecast error.

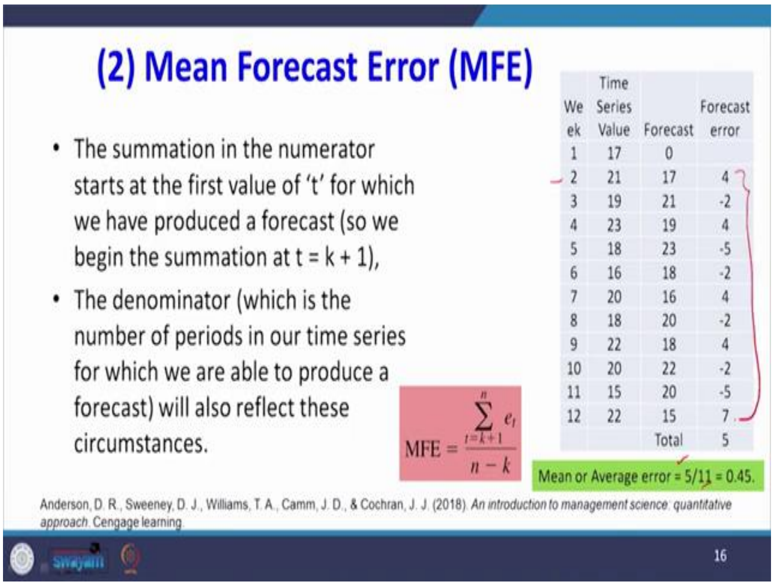


The table shows that the sum of the forecast error for the gasoline sales time series is 5; see, when you add the forecast error, we are getting a sum equal to 5. So, as per our formula, some are divided by a number of data; the number of data is 11 because we have forecasted only for the 11 weeks; for the first week, we do not have. So, when you divide it, we get the value of 0.45.

Because we do not have sufficient data to produce a naïve forecast for the first week of this time series, we must adjust our calculations in both the numerator and the denominator accordingly.

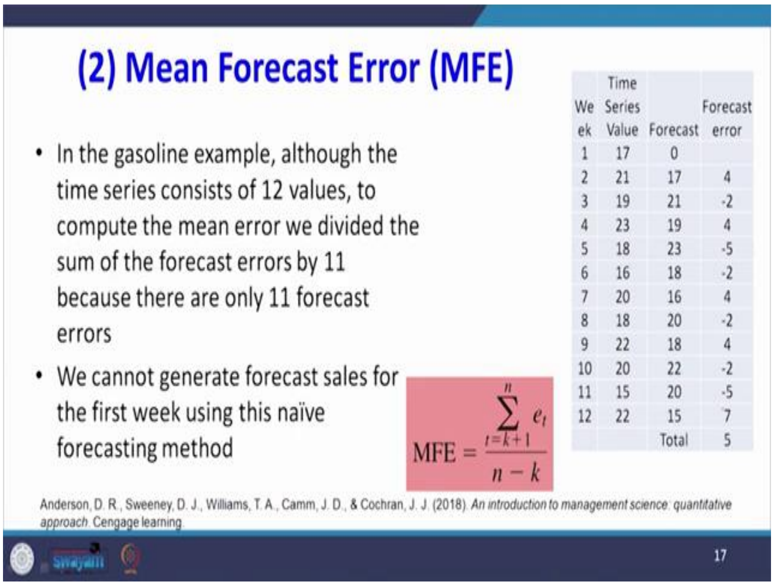


The second type of measure of forecast is mean forecast error. This is common in forecasting. We often use k past periods from the time series to produce forecasts, and so we frequently cannot produce forecasts for the first k period. So, the formula is the sum of the error divided by (n – k). So, the answer is 0.45.

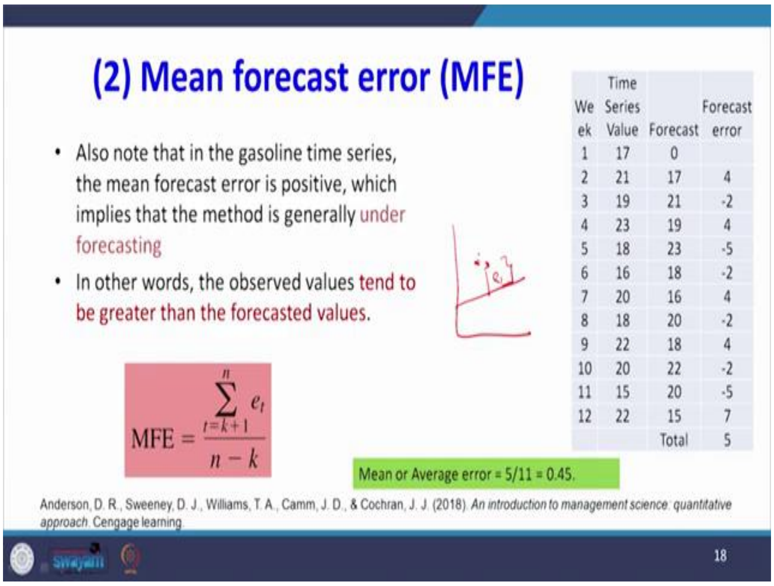


The summation in the numerator starts at the first value of t for which we have produced a forecast. So, we begin the summation t = (k + 1). So, we have to start from the second week onwards. So because the value of k = 1, the denominator is the number of periods in our time series for which we are able to produce a forecast. So, out of 12, we are able to produce for 11 weeks. So, it is (12 – 1). This will be 11.

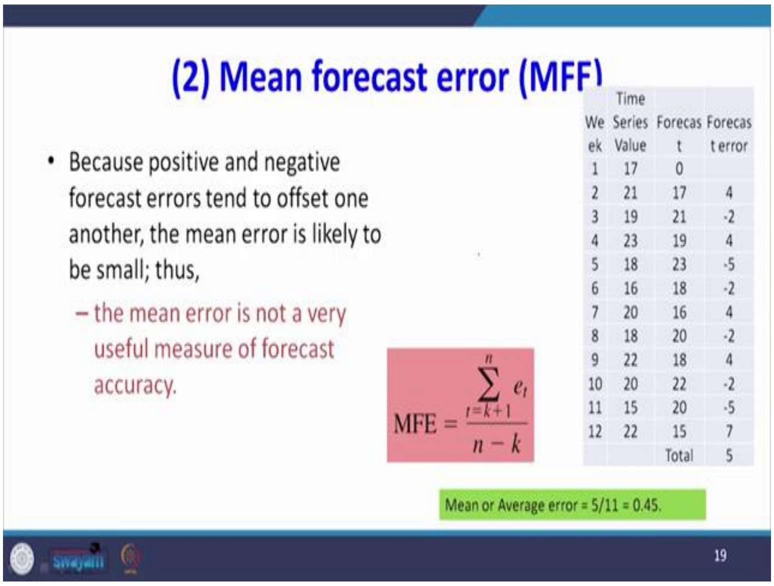
That is why you divided it by 11; here, 5 divided it by 11. And we have added only five from the second week onwards.



In the gasoline example, although the time series consists of 12 values to compute the mean error, we divided the sum of the forecast error by 11 because there are only 11 forecast errors. We cannot generate forecast sales for the first week using this naïve forecasting method.

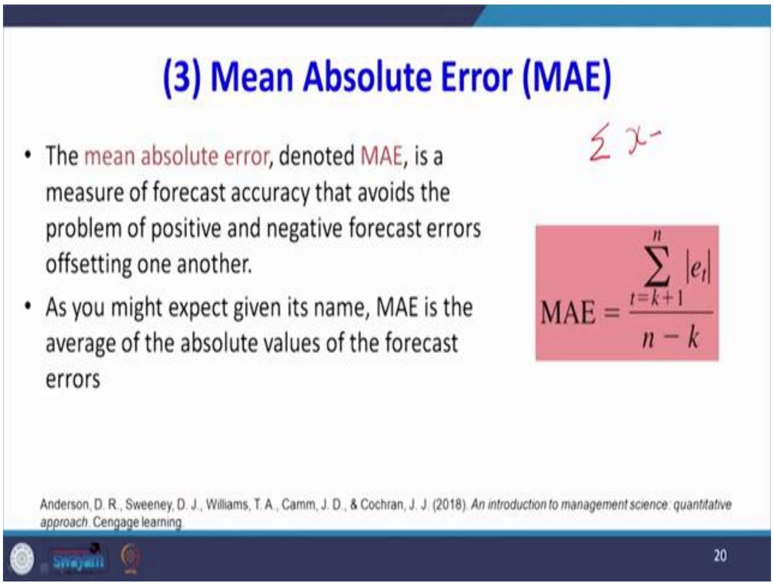


Also note that in the gasoline time series, the data which is given on the right-hand side of the mean forecast error is positive 0.45, which implies that the method is generally under forecasting. As I told you, when we can get the positive error, suppose values like this, so this is your positive error. So, what have you done? In other words, the observed value tends to be greater. So, this point is above then, the forecasted values.



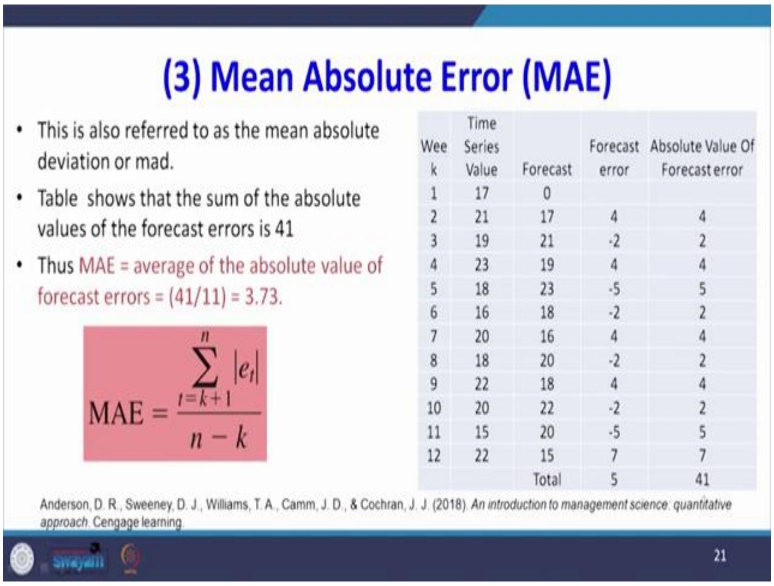
Week	Time Series Value	Forecast	Forecast error
1	17	0	
2	21	17	4
3	19	21	-2
4	23	19	4
5	18	23	-5
6	16	18	-2
7	20	16	4
8	18	20	-2
9	22	18	4
10	20	22	-2
11	15	20	-5
12	22	15	7
		Total	5

Because positive and negative forecast errors tend to offset one another, the mean error is likely to be small. Thus, the mean error is not a very useful measure of forecast accuracy. So, this method is, is not a very useful measure of forecast accuracy.

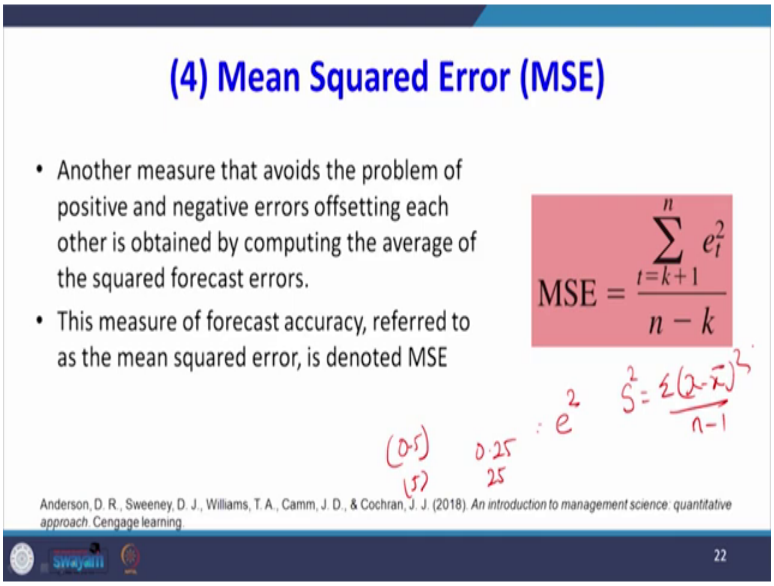


So, what we are going to do, we are going to use another technique called mean absolute error. The mean absolute error denoted by MAE is a measure of forecast accuracy that avoids the problem of positive and negative forecast errors offsetting one another. As you might expect, given its name mean absolute error is the average of absolute values of the forecast error. So, what have you done?

We have taken the absolute value of the error, and we are only adding that for that. Otherwise, when you add the error, there is a possibility it may also become 0.



This is also referred to as the mean absolute deviation MAD. The table shows some of the absolute values of the forecast error is 41. You see that, so first, I found the forecast error. So, wherever negative I have converted into positive. Then, when you add it, you are getting 41. So, when you divide 41 by 11, you get 3.73.

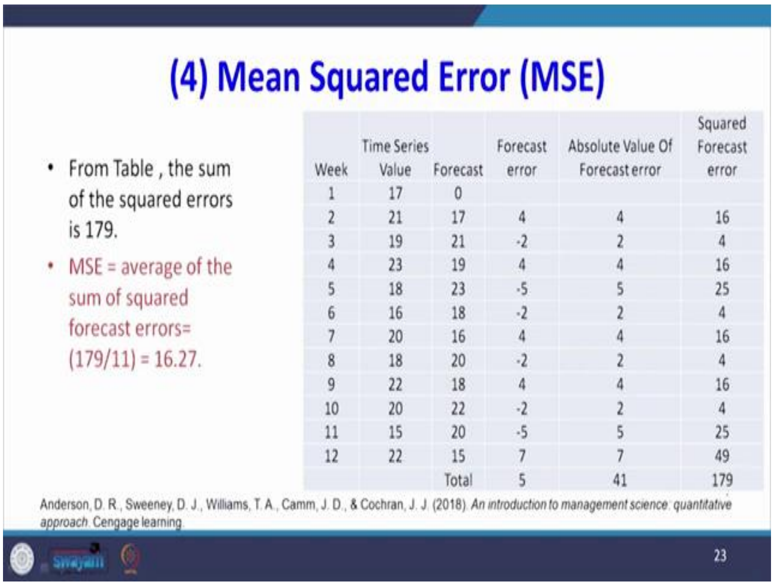


The next measure is a mean squared error. Another measure that avoids the problem of positive and negative errors offsetting each other is obtained by computing the average of squared forecast error. Previously we did not square the error. Simply, we have added the error; then, we have divided the number of periods. Here, it is 11 because only for 11 periods we have forecasting. So, this measure of forecast accuracy is referred to as mean squared error.

There is a reason why we are squaring this error. Suppose the deviation is, say, 0.5. When you square it, it will become 0.25. If the error is 5, when you square it, it is 25. So, what is the advantage of squaring? Is it amplified if there are more deviations? If there is less deviation, the amplification is much less. That is why the variance formula also you might recollect. The variance formula is the sigma of x – x bar that is nothing but an error, square divided by n – 1.

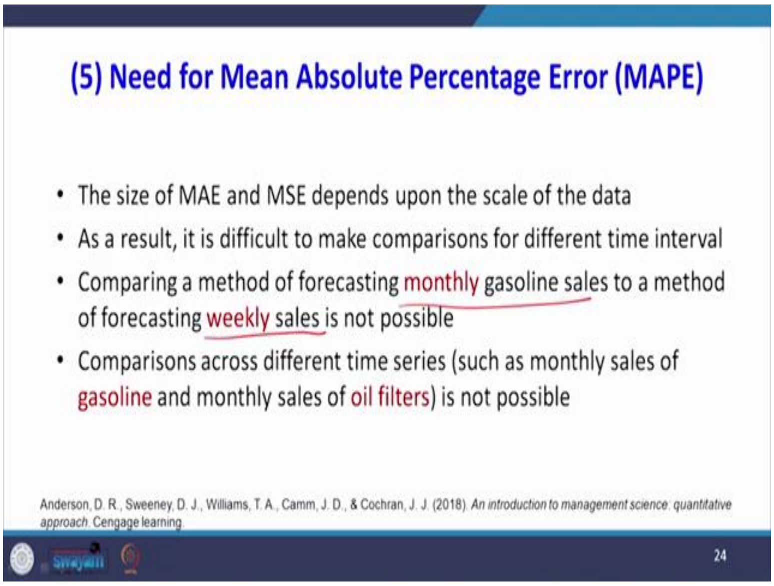
$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

The purpose of squaring is to give a higher penalty for a higher error for a larger deviation and a lesser penalty for a smaller error or smaller deviation.



Week	Time Series Value	Forecast	Forecast error	Absolute Value Of Forecast error	Squared Forecast error
1	17	0			
2	21	17	4	4	16
3	19	21	-2	2	4
4	23	19	4	4	16
5	18	23	-5	5	25
6	16	18	-2	2	4
7	20	16	4	4	16
8	18	20	-2	2	4
9	22	18	4	4	16
10	20	22	-2	2	4
11	15	20	-5	5	25
12	22	15	7	7	49
		Total	5	41	179

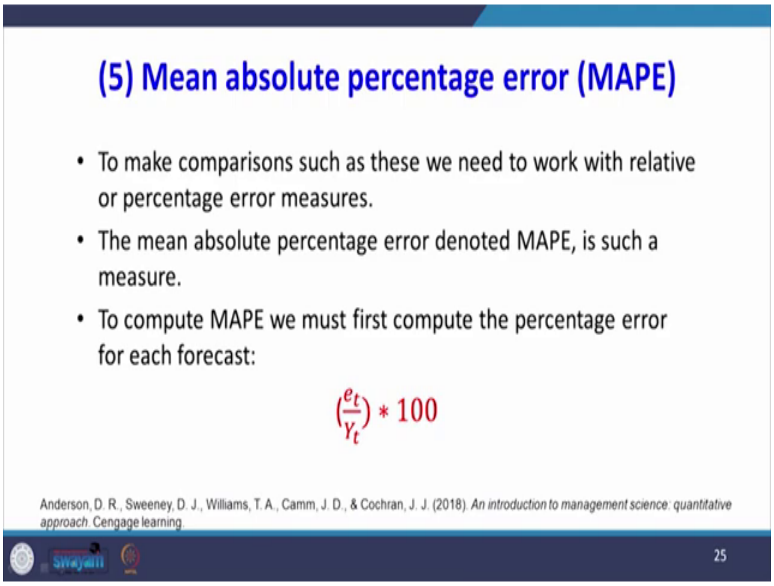
So, here I have explained how to find out the mean squared error. So, what I have done first is forecast errors. Then, I squared the error. Then, I summed it up to 179. So, there are 11 data sets, so it is 16.27. So, the mean squared error is 16.27.



The size of the mean absolute error and mean squared error depends upon the scale of the data. Now, I am going to introduce another measure of forecasting accuracy. That is called mean absolute percentage error. Before that, I will explain why these measures of this mean absolute percentage error measures are required. The reason is that previously, we have used mean absolute error and mean squared error.

These depend upon the scale of the data. As a result, it is difficult to make comparisons for different time intervals. For example, comparing a method of forecasting monthly gasoline sales to a method of forecasting weekly sales. That is not possible. So, this comparison is not possible by using your mean absolute error and mean squared error. Another example is a comparison across different time series, such as monthly sales of gasoline and monthly sales of oil filters, because these are two different products.

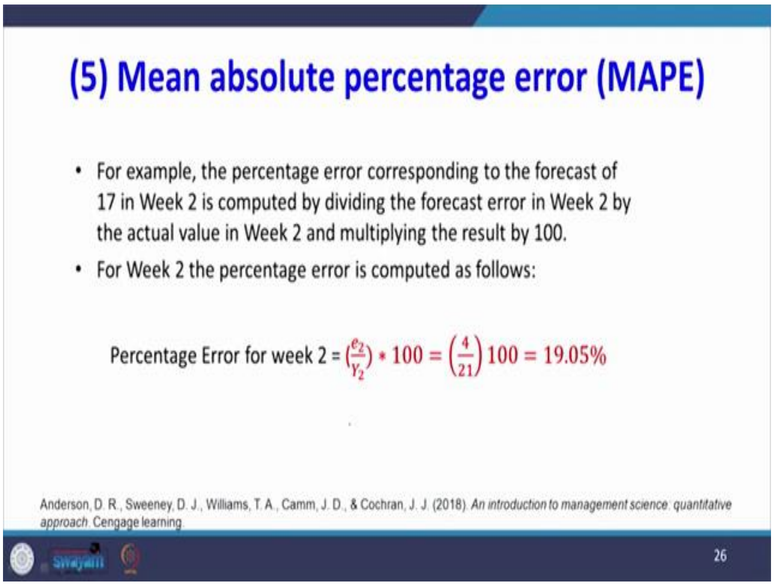
This comparison is also not possible to overcome this drawback. We are going to introduce another measure of forecasting accuracy. That is called mean absolute percentage error.



To make comparisons such as these, we need to work with relative or percentage error measures. The mean absolute percentage error denoted by MAPE is such a measure. To compute the mean absolute percentage error, we must compute the percentage error of each forecast. That is an error for time t divided by actual data, then multiplied by 100, which is the percentage error for each forecast.

To compute MAPE we must first compute the percentage error for each forecast:

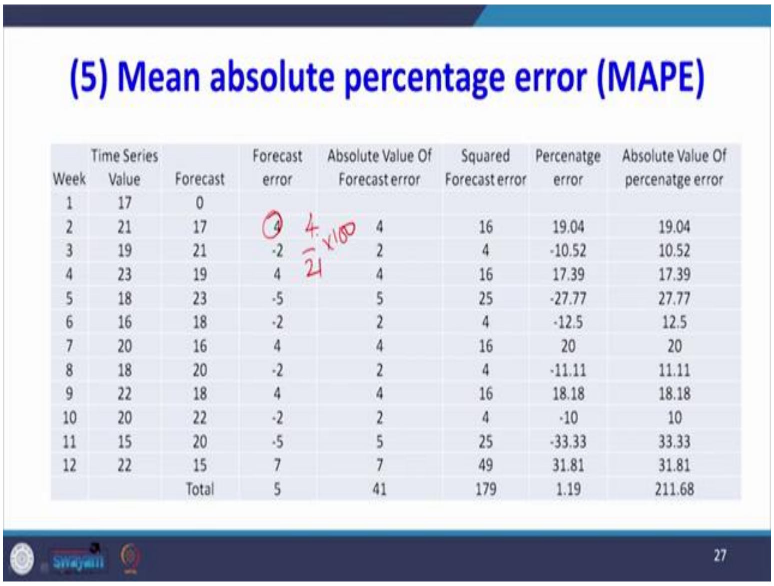
$$(\frac{e_t}{Y_t}) * 100$$



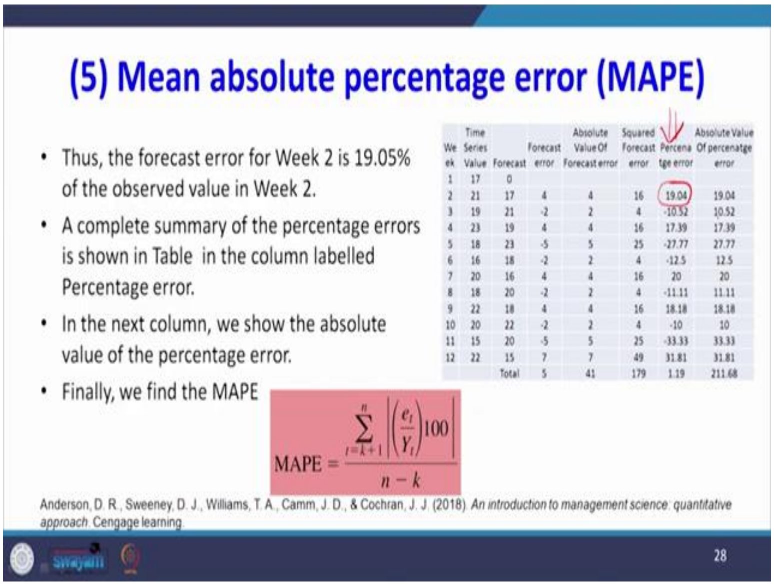
For example, the percentage error corresponding to the forecast of 17 in week 2 is computed by dividing the forecast error in week 2 by the actual value in week 2 and multiplying by a hundred. For week 2, the percentage error is computed as follows. So, the percentage error for week 2 is equal to error two upon the actual value for week 2; the error is 4, and the actual value is 21 multiplied by hundred, which will be 19.05%.

Percentage Error for week 2;

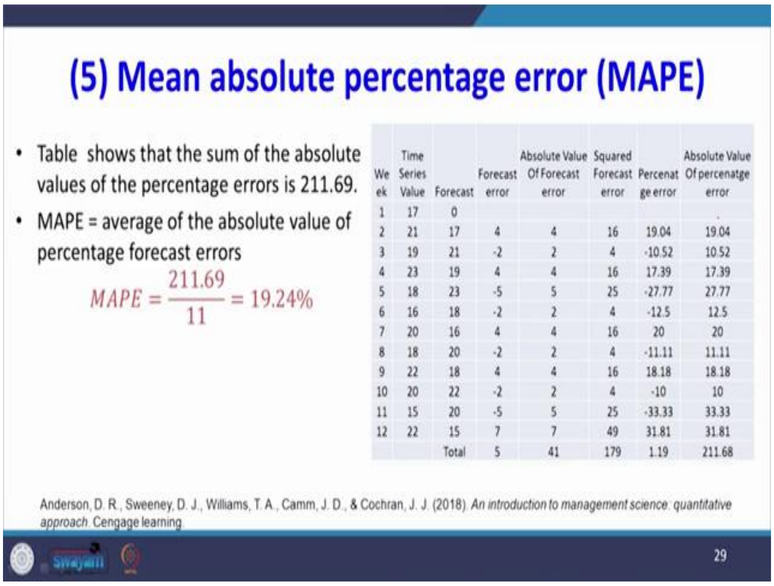
$$\left(\frac{e_2}{Y_2}\right) * 100 = \left(\frac{4}{21}\right) 100 = 19.05\%$$



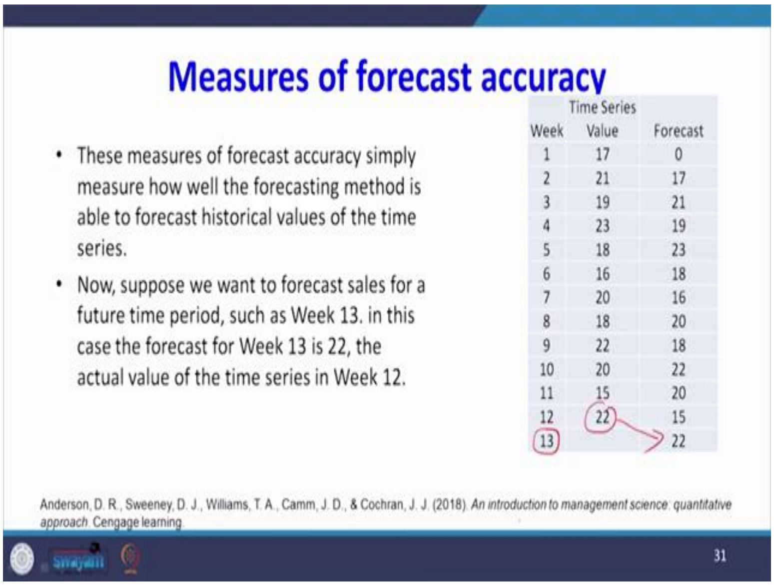
I will show it in the next table; see this. The error is 4; the actual value is 21. So, 4 upon 21 multiplied by 100 will be getting 19.04 % like this. We can do it for all 12 weeks.



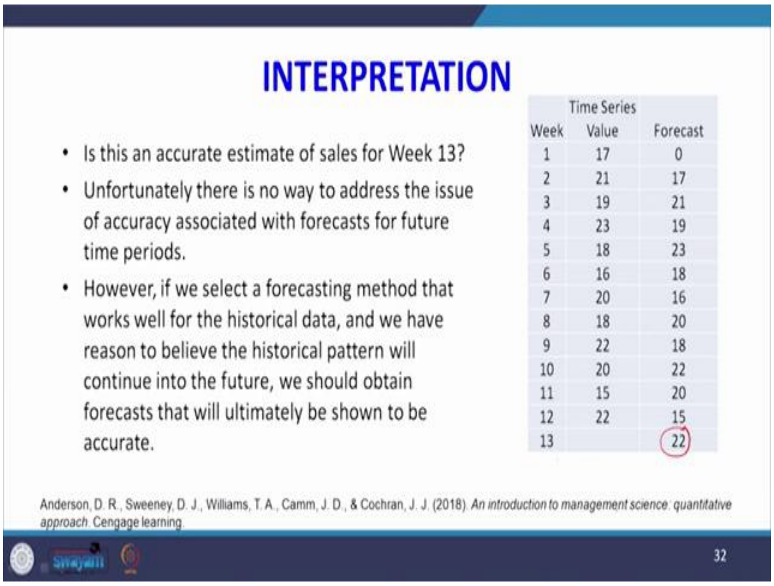
Thus, the forecast error for week 2 is 19.05% this one. Similarly, for the next one, minus 2 upon 19 will be getting –10.52%. So, a complete summary of the percentage errors is shown in the table in the column labeled percentage error, which I have provided. In the next column, we show the absolute value of the percentage error. So, here, minus 10.52, I have taken only a positive value. Finally, we divide by the number of forecast errors. That is your 11. So, then you will get the mean absolute percentage error.



This table shows that the sum of absolute values of percentage error is 211.6 here; when you divide by 11, you are getting 19.24 %.

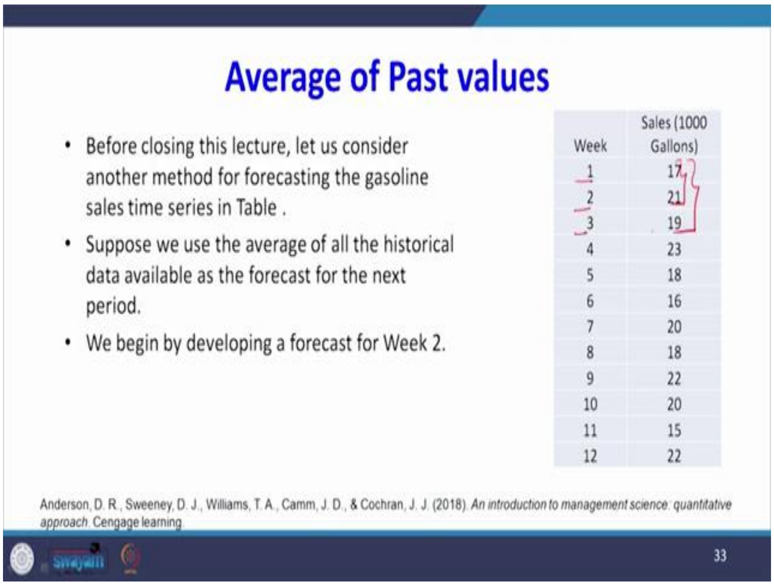


In summary, using the naïve most recent observation forecasting method, we obtained the following measures of forecast accuracy: mean absolute error, is 3.73, mean squared error is 16.27, and mean absolute percentage error is 19.24%. These measures of forecast accuracy simply measure how well the forecasting method is able to forecast historical values of the time series. Now, suppose you want to forecast sales for a future time period, such as week 13, for example, here, week 13. In this case, the forecast for week 13 is 22 because we are using the naïve method, 22. So, the actual value of the time series in week 12 is 22.

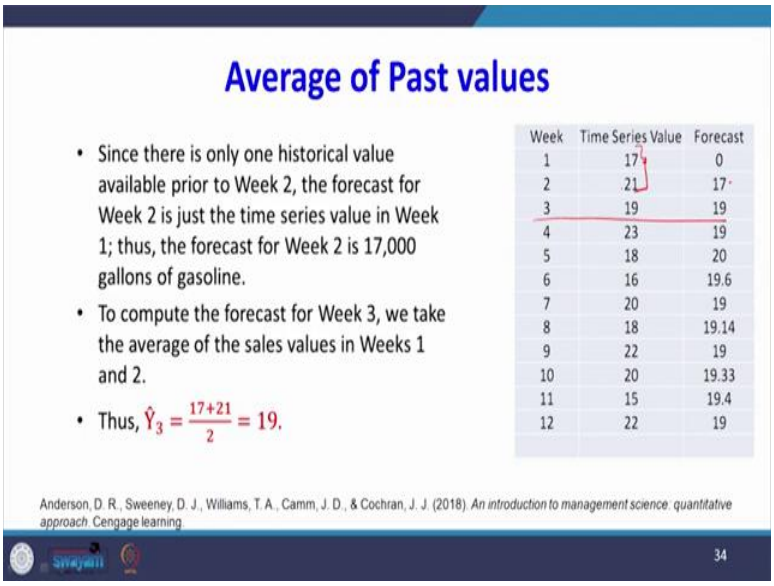


Is this an accurate estimate of sales for week 13? Is it correct? Is it accurate? Unfortunately, there is no way to address the issue of accuracy associated with forecasts of future time periods. However, if we select a forecasting method that works well for the historical data we have a reason to believe the historical pattern will continue into the future. We should obtain a forecast that will ultimately be shown to be accurate.

So, we have forecasted for week 13. The forecasted value is 22. There is no guarantee. This pattern will be repeated in the future also. If you believe that the same pattern of sales for the first 12 weeks will also be replicated in the future, then we can use this naïve forecasting method.

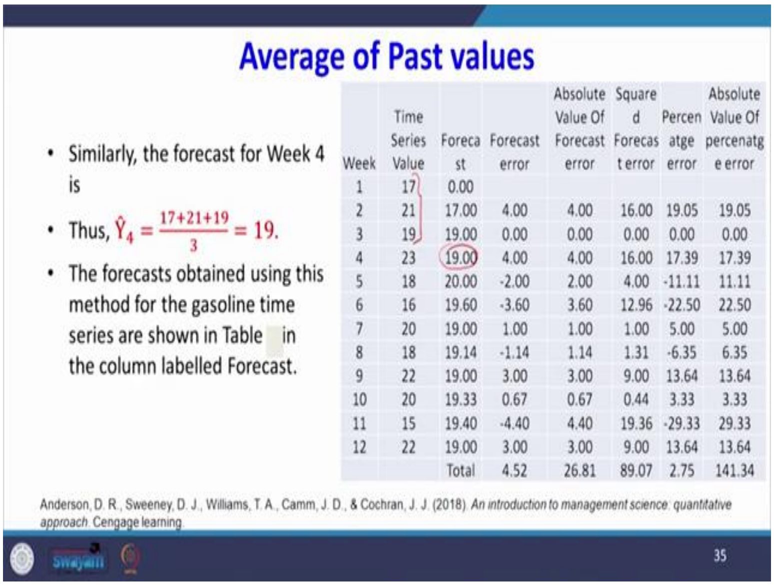


Now, I am going to take another type of forecasting method that is average of past values. So, we begin by developing a forecast for week 2.



Since there is only one historical value available prior to week 2, here is only one historical data. The forecast for week 2 is 17,000 gallons of gasoline. To compute the forecast for week 3, we take the average of the sales for weeks 1 and 2. So that is (17 + 21) upon 2, it is 19. So, the forecasted value for week 3 is 19.

$$\hat{Y}_3 = \frac{17+21}{2} = 19$$



Similarly, the forecast for week 4 so, the week 4 we have to find out the forecast value. So, what we have to do. We have to average these 3, (17 + 21 + 19) upon 3, we are getting 19. So, this 19 will be the forecasted value for week 4. The forecast obtained using this method for the gasoline time series is shown in this table the column labeled forecast.

$$\hat{Y}_4 = \frac{17+21+19}{3} = 19$$

Average of Past values

n	11
MAE	2.43
MSE	8.09
MAPE	12.84

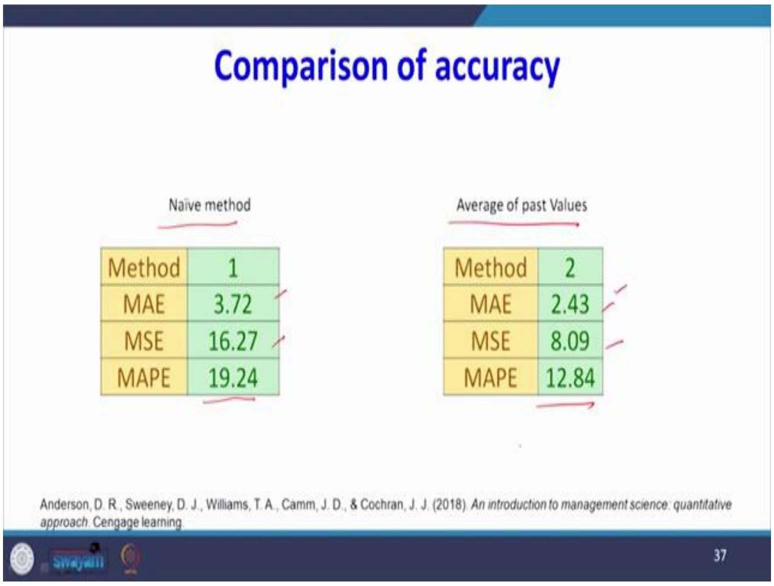
Week	Time Series Value	Forecast	Forecast error	Absolute Value Of Forecast error	Squared Forecast error	Percenatge error	Absolute Value Of percenatge error
1	17	0.00					
2	21	17.00	4.00	4.00	16.00	19.05	19.05
3	19	19.00	0.00	0.00	0.00	0.00	0.00
4	23	19.00	4.00	4.00	16.00	17.39	17.39
5	18	20.00	-2.00	2.00	4.00	-11.11	11.11
6	16	19.60	-3.60	3.60	12.96	-22.50	22.50
7	20	19.00	1.00	1.00	1.00	5.00	5.00
8	18	19.14	-1.14	1.14	1.31	-6.35	6.35
9	22	19.00	3.00	3.00	9.00	13.64	13.64
10	20	19.33	0.67	0.67	0.44	3.33	3.33
11	15	19.40	-4.40	4.40	19.36	-29.33	29.33
12	22	19.00	3.00	3.00	9.00	13.64	13.64
		Total	4.52	26.81	89.07	2.75	141.34

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2018). *An introduction to management science: quantitative approach*. Cengage learning.

36

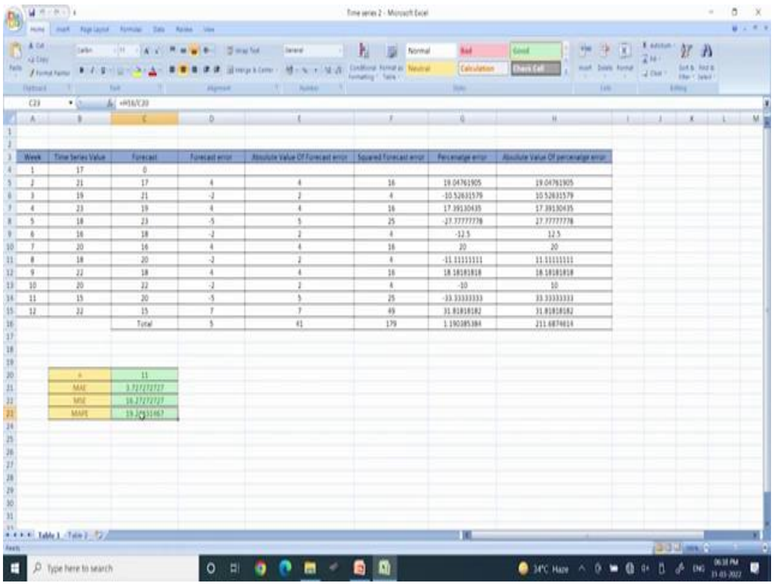
By using the forecasting method called *average of past values*. I have found a mean absolute error. Look at the table that I have solved in Excel. At the end of this lecture, I will show you the Excel spreadsheet. So, the time series is their time series value. I have found the forecast by having an average of the previous values. I found the forecasting error. Then I found the absolute value of forecast error. Then I found the squared forecast error.

Then I found the percentage error. Percentage error is an error upon the actual value. Then I found the mean absolute percentage error. So, by using a similar, I found the total also. So, the mean absolute error is 26.81 upon 11. You will get 2.43. The mean squared error is 89.07. Upon 11 will get 8.07. So, the mean absolute percentage error is 141.34 upon 11. You will get 12.84. So, these forecasting measures are for different types of forecasting techniques, that is, the average of past values.



Now, we have to compare which method is most accurate. The first method we used naïve, and the second forecasting method, in which we used an average of past values, looked at the different measures of forecast accuracy. When you compare the mean absolute error, this is 3.72. This is only 2.43, so this is better. The mean squared error is 16.27, and this is 8.09. Again, the average of past values in this method is better.

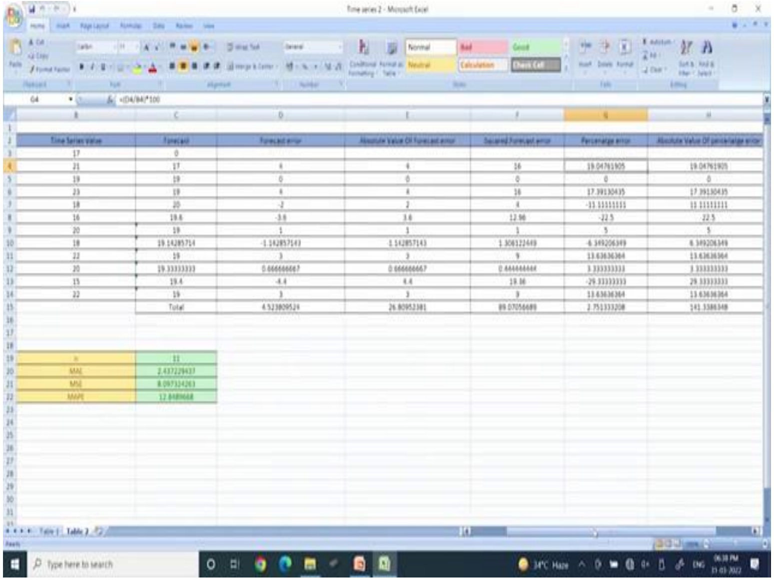
For the mean absolute percentage error, this is 19.24. Here, it is only 12.84. So, when we compare, even though the data sets are different. The second method, which is the average of past values method, provides an accurate forecast than the naïve method.



Dear students, now I am going to explain how to use a spreadsheet for forecasting using the naïve method and using the average of past value method. I have 12 weeks of sales; the time series values are given as 17, 21, 29, 23, and so on. So, if you use the naïve method, it is 17 because the week 2 forecast is for week 1 actual sales. Then, 21 and 21 are week 2, and the actual sales are week 3 forecasted values. Then I found an error.

We know what the error is. Actual value minus forecasted value, so we got 21 – 17. So, I have dragged it. 21– 17 is 4. Then 19 – 21 is –2 and so on. Then I used Excel; there are options for absolute abs that will provide only the absolute value. That is the positive value. Then I have squared that. Then I found the percentage of the error, the error upon actual value. For example, d5 is divided by b5, followed by a percentage error.

Then I found the using absolute function. I found the absolute value for percentage error. Here, I found the total because I wanted to know the average. So, we know n is 11, so the mean absolute error is the total value divided by 11, then the mean squared error for the mean square. It is 179 divided by 11, so getting 16.27. The mean absolute percentage error is 211.68; upon 11, you will get this value.



Time Series Value	Forecast	Forecast error	Absolute Value Of Forecast error	Squared Forecast error	Percentage error	Absolute Value Of percentage error
17	0					
21	17	4	4	16	19.04761905	19.04761905
19	19	0	0	0	0	0
23	19	4	4	16	17.39130435	17.39130435
18	20	-2	2	4	-11.11111111	11.11111111
16	19.6	-3.6	3.6	12.96	-22.5	22.5
20	19	1	1	1	5	5
18	19.14285714	-1.142857143	1.142857143	1.306122449	-6.349206349	6.349206349
22	19	3	3	9	13.63636364	13.63636364
20	19.33333333	0.666666667	0.666666667	0.444444444	3.333333333	3.333333333
15	19.4	-4.4	4.4	19.36	-29.33333333	29.33333333
22	19	3	3	9	13.63636364	13.63636364
	Total	4.523809524	26.80952381	89.07056689	2.751333208	141.3386348

n	11
MAE	2.437229437
MSE	8.097324263
MAPE	12.8489668

Similarly, for the second forecasting method, the average of the previous values is used. For the first week, there is no forecast for the second week. You see that there is only one value: 17 upon 1 17. For the third week, I have added (17 + 21)/2, that is 19. For the 4th week, I have found the average of the previous week's actual sales, which are 17, 21, and 19. So, I got 19. Then, as usual, I found the error actual minus forecasted value.

Then I found the absolute error, then the squared error, then the percentage error. Then, I found the absolute value of the percentage error. Then I have done the sum by using that sum, and the number of data sets is 11 here also. So, I found mean absolute error, mean squared error, and mean absolute percentage error. Then, I compared which method was the most accurate method.

So, we are concluding that averaging the past values for forecasting the next period is the best way of forecasting. There is a more accurate method when compared to the naïve method. Dear students, in this lecture, I have discussed measures of forecasting accuracy. What are the measures I have discussed? First, I have talked about the error, then mean forecast error, mean absolute error, mean squared error, and finally, mean absolute percentage error. These forecasting errors will help to select appropriate forecasting techniques. Thank you.