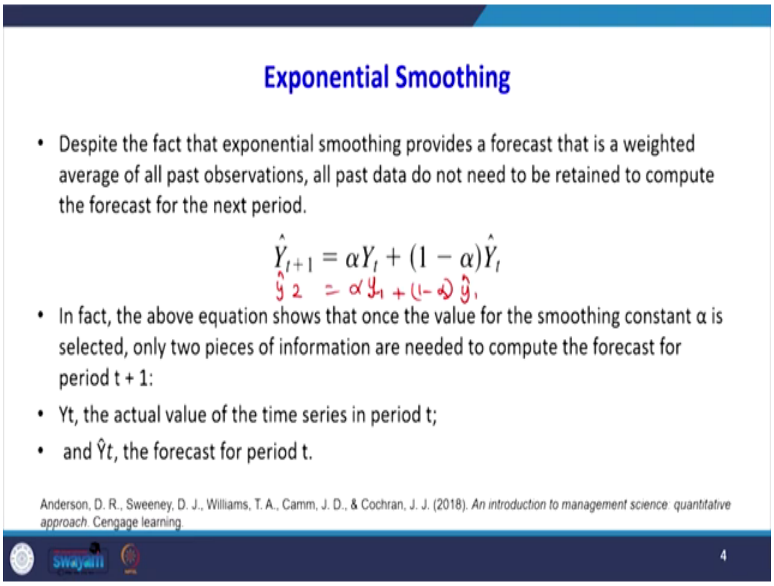
Decision Making with Spreadsheet
Prof. Ramesh Anbanandam
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 59
Time Series Analysis and Forecasting – IV

So, dear students, in the previous lecture, I gave an introduction to exponential smoothing. In this lecture, I am going to explain the workings of the exponential smoothing model with a sample problem. I am also going to explain how to use a spreadsheet for exponential smoothing. After that, I am going to discuss an interesting and powerful forecasting technique called the regression model.

In this lecture, I will explain how to build a regression model in the next lecture. I will explain how to use this model for forecasting. So, the agenda for this lecture is exponential smoothing problems and linear trend forecasting using a regression model.

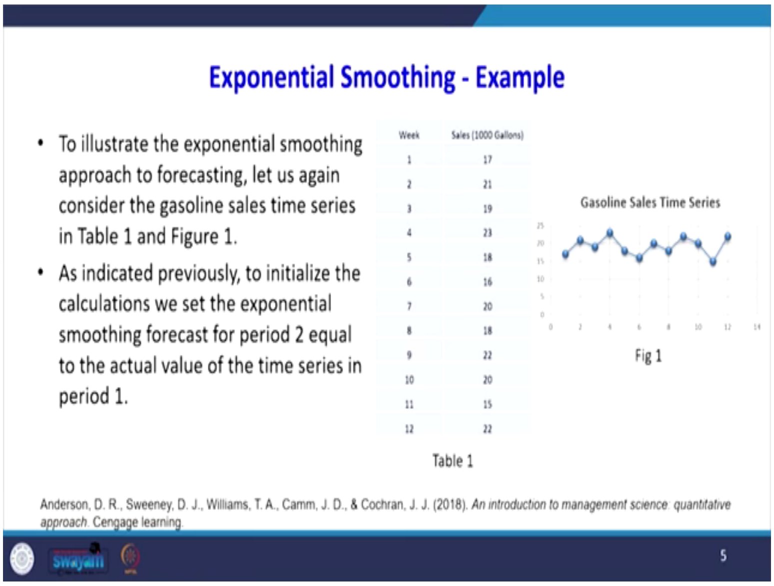


First, we will start with exponential smoothing, despite the fact that exponential smoothing provides a forecast that is a weighted average of all past observations. All past data do not need to be retained to compute the forecast for the next period, for example, here. Look at the formula;

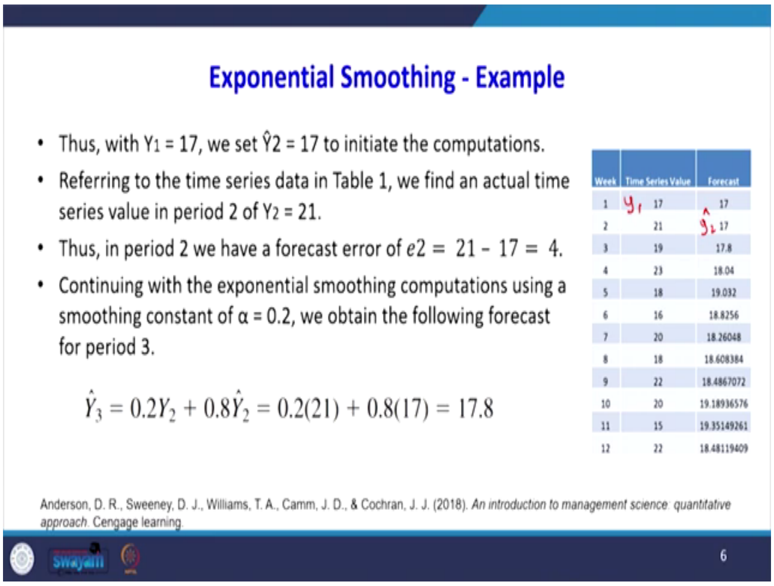
$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t$$

The above equation shows that once the value for the smoothing constant α is selected, only two pieces of information are needed to compute the forecast for the period (t + 1): Yt, the actual value of the time series in period t, and $\hat{Y}t$, the forecast for period t.

That means for forecasting the next period; I need the previous period's actual value with alpha weightage plus the previous period's forecasted value with (1 – α) weightage. The above equation shows that once the value of the smoothing constant alpha is selected, only two pieces of information are needed to compute the forecast for period 2 + 1. What are the two pieces of information? Y_t is the actual value of the time series in period t, and Y_t is the forecast for period t.



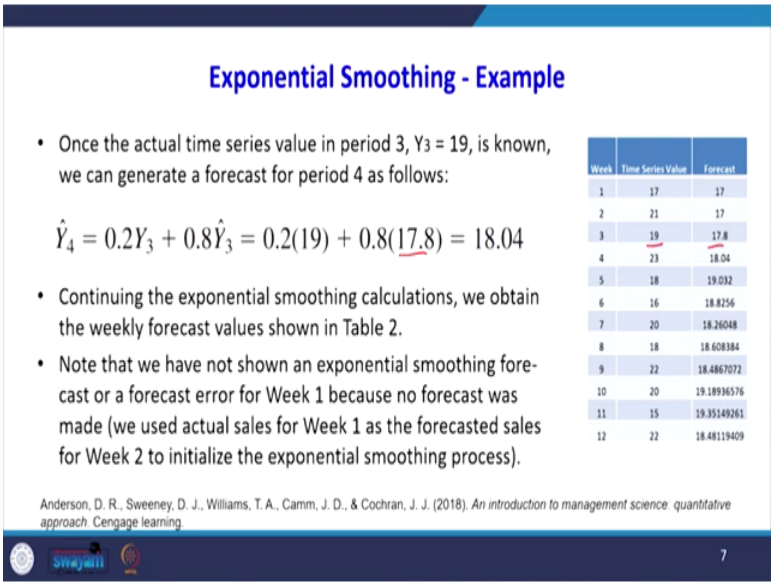
To illustrate the exponential smoothing approach for forecasting. Let us again consider the gasoline sales time series in Table 1 and Figure 2. The reference for this problem is from Anderson et al. As indicated previously, to initialize the calculations, we set an exponential smoothing forecast for period 2 equal to the actual value of the time series in period 1. This was the given data. For this data set, we are going to find the exponential smoothing.



Thus, with Y1 = 17, we set Ŷ2 = 17 to initiate the computations. Referring to the time series data in Table 1, we find an actual time series value in period 2 of Y2 = 21. Thus, in period 2 we have a forecast error of Continuing with the exponential smoothing computations using a smoothing constant of α = 0.2, we obtain the following forecast for period 3.

$$\hat{Y}_3 = 0.2Y_2 + 0.8\hat{Y}_2 = 0.2(21) + 0.8(17) = 17.8$$

So, the forecasted value for period 3 is this value 17.8.

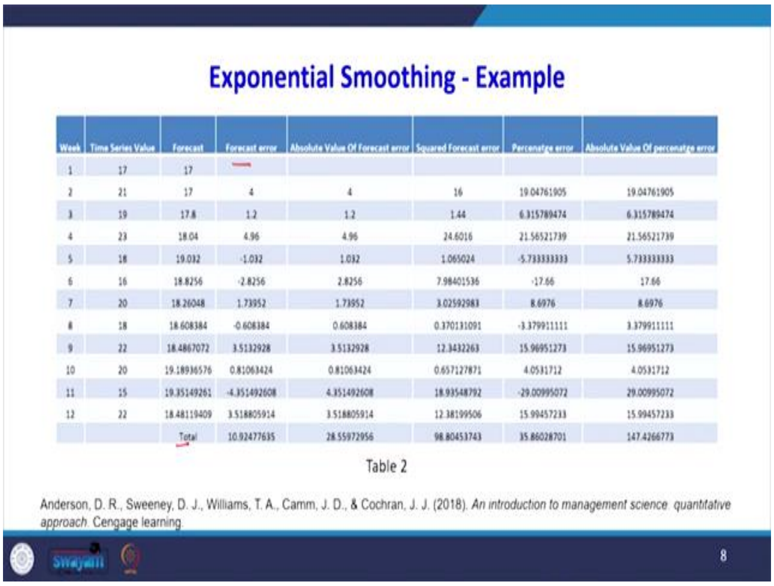


Once the actual time series value in period 3 is 19, is known, we can generate a forecast for period 4. Once the actual time series value in period 3, Y3 = 19, is known, we can generate a forecast for period 4 as follows:

$$\hat{Y}_4 = 0.2Y_3 + 0.8\hat{Y}_3 = 0.2(19) + 0.8(17.8) = 18.04$$

So, we have got the forecasted value for the 4th period. That is 18.04.

Continuing the exponential smoothing calculations, we obtain the weekly forecast value, as shown in the table. Note that we have not shown an exponential smoothing forecast or a forecast error for week 1; continuing with the exponential smoothing calculations, we obtain the weekly forecast value shown in a table that is in the next slide I will show you.

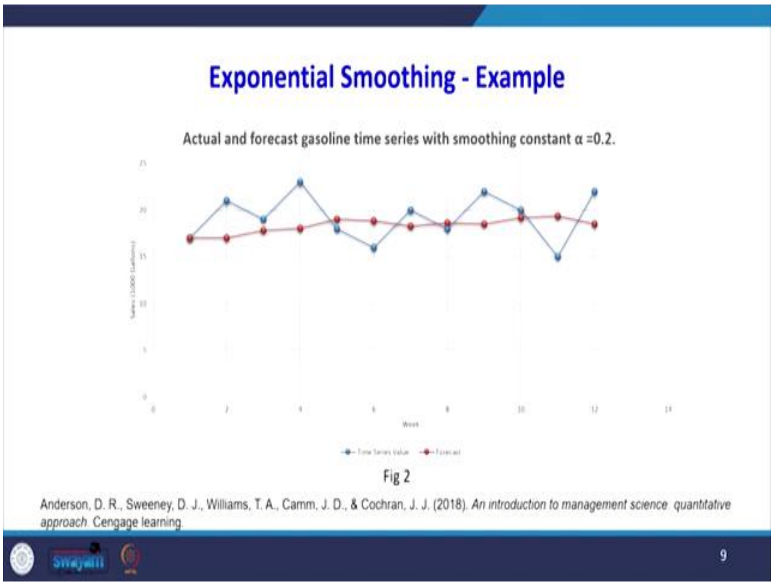


Week	Time Series Value	Forecast	Forecast error	Absolute Value Of Forecast error	Squared Forecast error	Percenatge error	Absolute Value Of percenatge error
1	17	17					
2	21	17	4	4	16	19.04761905	19.04761905
3	19	17.8	1.2	1.2	1.44	6.315789474	6.315789474
4	23	18.04	4.96	4.96	24.6016	21.56521739	21.56521739
5	18	19.032	-1.032	1.032	1.065024	-5.733333333	5.733333333
6	16	18.8256	-2.8256	2.8256	7.98401536	-17.66	17.66
7	20	18.26048	1.73952	1.73952	3.02592983	8.6976	8.6976
8	18	18.608384	-0.608384	0.608384	0.370131091	-3.379911111	3.379911111
9	22	18.4867072	3.5132928	3.5132928	12.3432263	15.96951273	15.96951273
10	20	19.18936576	0.81063424	0.81063424	0.657127871	4.0531712	4.0531712
11	15	19.35149261	-4.351492608	4.351492608	18.93548792	-29.00995072	29.00995072
12	22	18.48119409	3.518805914	3.518805914	12.38199506	15.99457233	15.99457233
		Total	10.92477635	28.55972956	98.80453743	35.86028701	147.4266773

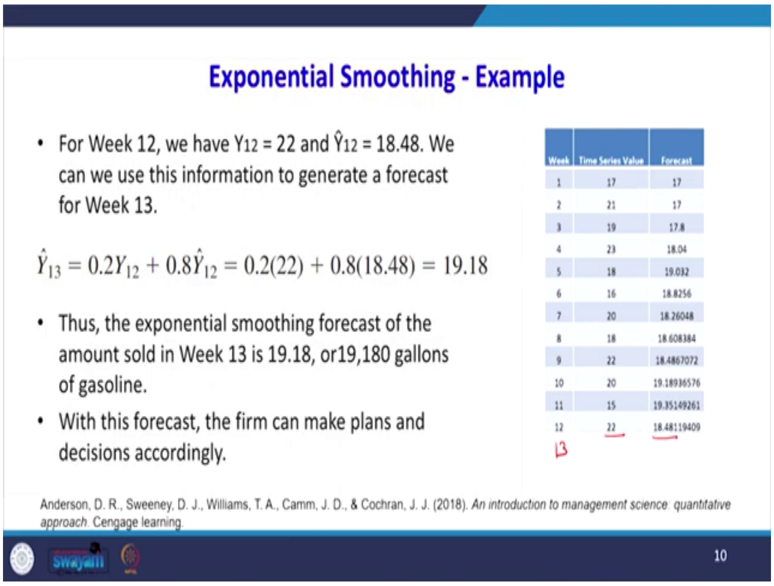
Table 2

This is the complete exponential smoothing where α = 0.2. We have not shown an exponential smoothing forecast or a forecast error for week 1 because no forecast was made. We used actual sales for week 1 as the forecasted sales for week 2 to initialize the exponential smoothing process. So, this is the point that we do not make any errors here because we have taken the actual values of period 1 as the forecasted value.

So, what have we done here? We found the forecast error. Then, the absolute value of forecast error. Then we found the squared error, then the percentage error, then the absolute value of the percentage error. At the bottom of this table, we have found the sum of this forecast error that will be used to calculate some of the accuracy measures for the forecasting.



The actual and forecast gasoline time series with smoothing constant α = 0.2. So, the actual value is blue in color. The forecasted value is the red in color. You see that it has been leveled.

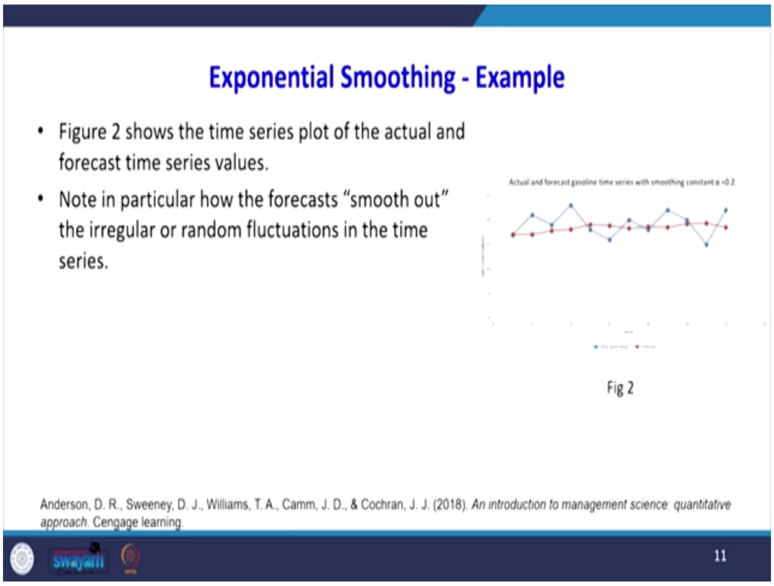


For Week 12, we have Y12 = 22 and Y12 = 18.48. We can use this information to generate a forecast for Week 13.

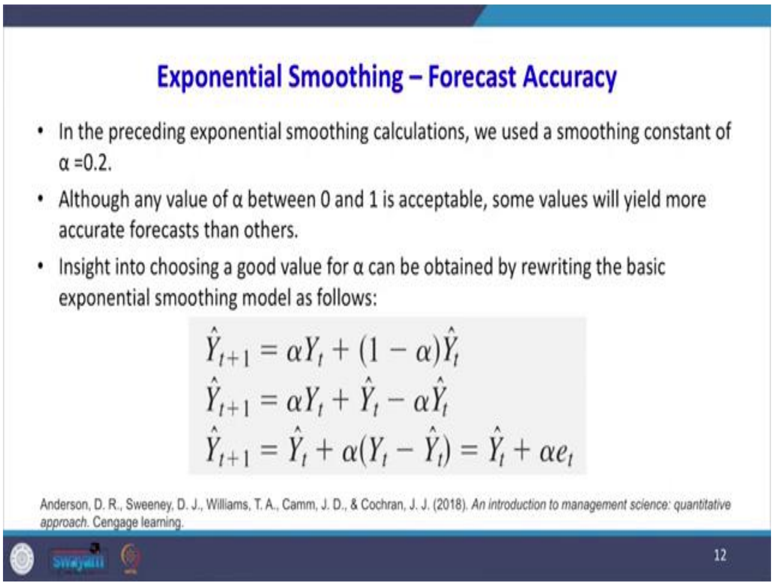
$$\hat{Y}_{13} = 0.2Y_{12} + 0.8\hat{Y}_{12} = 0.2(22) + 0.8(18.48) = 19.18$$

So, in week 12, the actual value is 22, and the forecasted value is 18.48. When you substitute there, we are getting 19.18.

Thus, the exponential smoothing forecast of the amount sold in week 13 is 19.18. Or in actual units, it is 19180 gallons of gasoline. With this forecast, the firm can make plans and decisions accordingly.

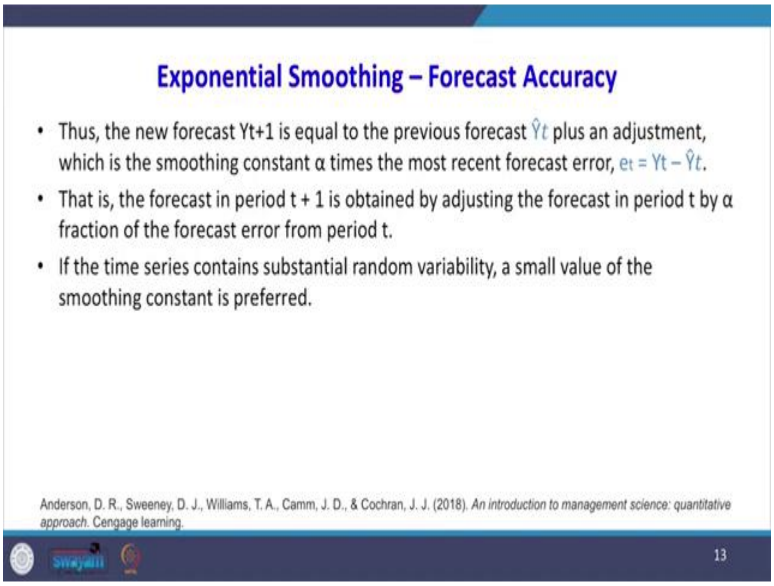


The figure shows the time series plot of actual and forecast time series values, noting, in particular, how the forecast smooths out the irregular or random fluctuation in the time series. Initially, the blue color says that there is some irregular pattern. By using exponential smoothing the irregulars are removed.



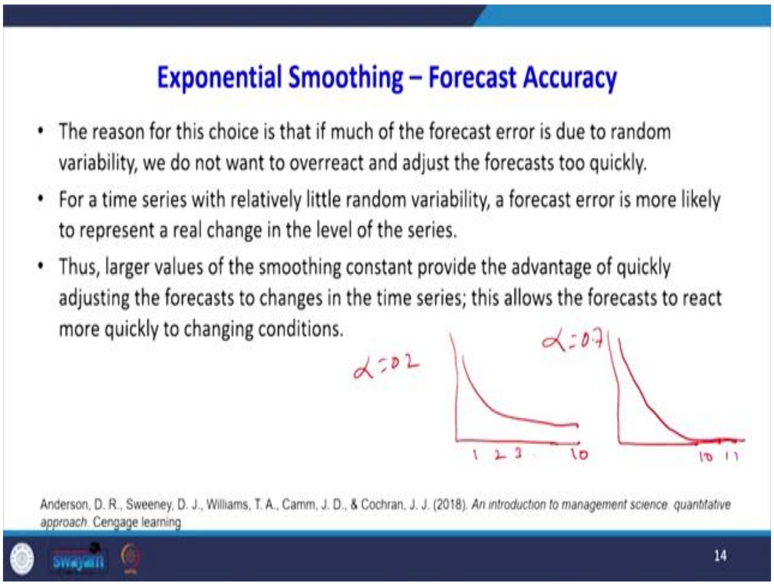
Now, we talk about the accuracy of this exponential smoothing. In the preceding exponential smoothing calculations, we used a smoothing constant of α = 0.2, although any value of alpha between 0 to 1 is acceptable. Some values will yield more accurate forecasts than others; Insight into choosing a good value for an alpha can be obtained by rewriting the basic exponential smoothing model as follows.

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t$$
$$\hat{Y}_{t+1} = \alpha Y_t + \hat{Y}_t - \alpha \hat{Y}_t$$
$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) = \hat{Y}_t + \alpha e_t$$



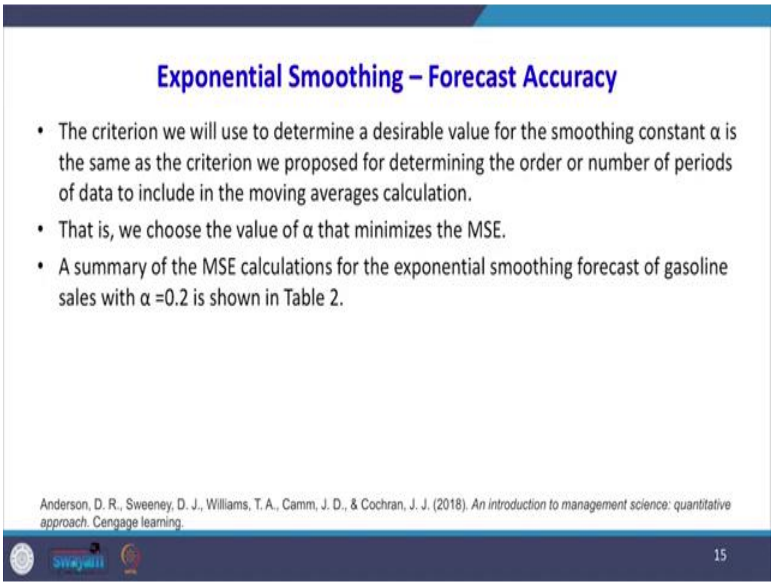
Thus, the new forecast Yt+1 is equal to the previous forecast $\hat{Y}t$ plus an adjustment, which is the smoothing constant times the most recent forecast error, et = Yt – $\hat{Y}t$.

That is, the forecast in period (t + 1) is obtained by adjusting the forecast in period t by an alpha fraction of the forecast error from period t. If the time series contains substantial random variability, a small value of the smoothing constant is preferred.

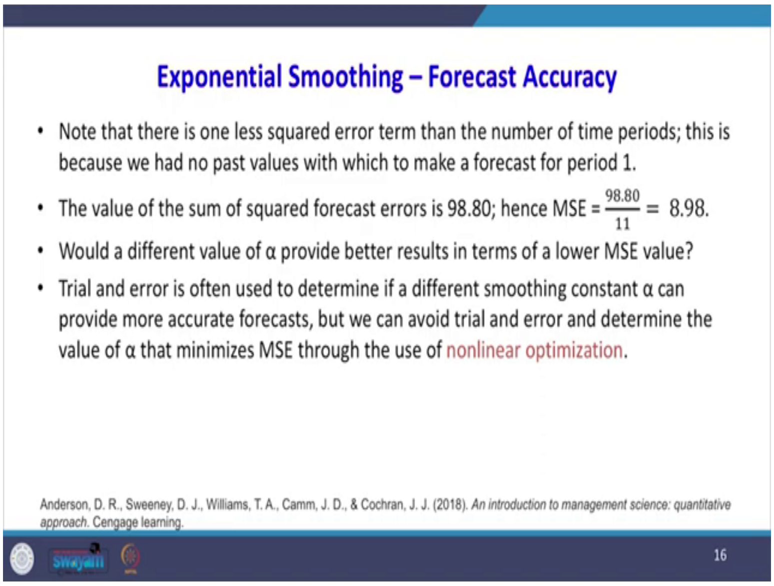


The reason for this is that if much of the forecast error is due to random variability, we do not want to overreact and adjust the forecast too quickly. For a time, series with relatively little random variability, a forecast error is more likely to represent a real change in the level of the series. Thus, the larger values of the smoothing constant provide an advantage of quickly adjusting the forecast to changes in the time series.

So this allows the forecast to react more quickly to changing conditions. In the previous slide, I also explained how the value of alpha will affect the result. For example, when α = 0.2, our value was like this, even saying this is here 1, 2, 3, and even the 10th year data also got some values. Suppose the alpha = 0.7. So, what will happen? That initial value is getting some values; the value for year 10 and year 11 almost is becoming 0.



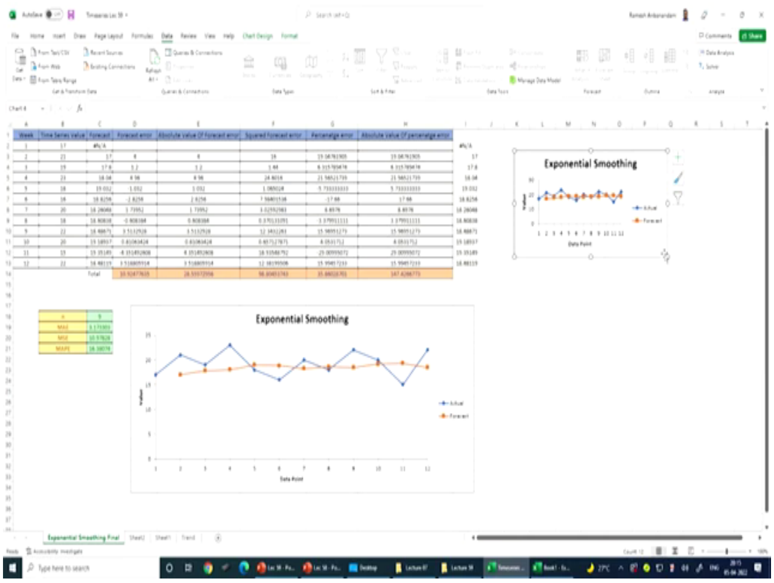
The criterion we will use to determine a desirable value for the smoothing constant alpha is the same as the criterion we proposed for determining the order or the number of periods of data to include in the moving average calculations. The order is k. That is, we choose the value of alpha. That minimizes the mean squared error. A summary of the mean square error calculation for the exponential smoothing forecast of gasoline sales with α = 0.2 is shown in the table that I have already explained.



Note that there is one less squared error term than the number of periods. This is because we had no past value with which to make the forecast error for period 1. So, we have 12 years of data, but we are divided by 11. Because for the first year, we do not have any errors. The value of the sum of squared forecast error is 98.80. So, when you divide this into 98.80 by 11, we are getting 8.98.

Would a different value of alpha provide a better result in terms of a lower mean square error value? Trial and error are often used to determine if a different smoothing constant alpha can provide a more accurate forecast. However, we can avoid trial and error and determine the value of alpha that minimizes mean squared error through the use of non-linear optimization. This non-linear optimization is already explained in my previous lectures.

So that chapter is under non-linear optimization problems that we can refer to, how to get the right value of alpha. That will minimize the squared error, so here, minimizing squared error is a non-linear objective function. There, what value of alpha will square will minimize the mean squared error. So alpha is the most appropriate.



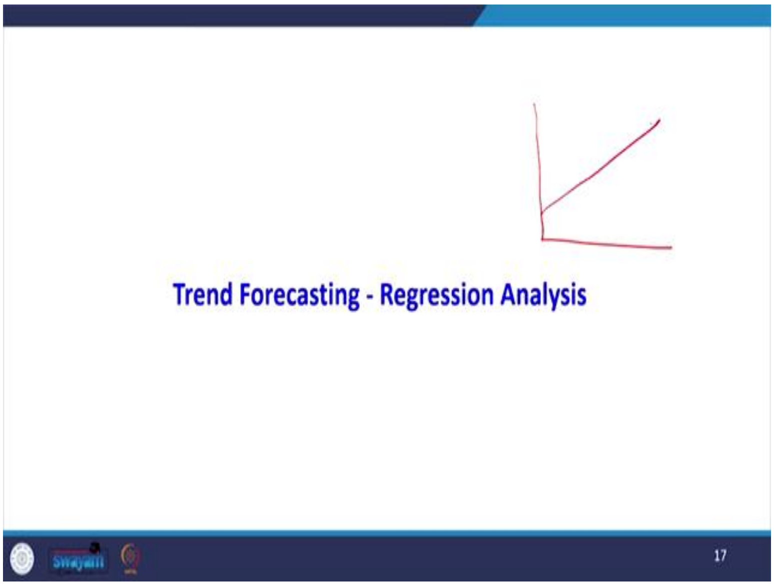
Now, I am going to explain how to use a spreadsheet for doing exponential smoothing forecasting. So, I have taken 12 weeks of data, and I have the sales value. So, you see that C3 looks at the 17.8 that is C3 that is our year C3 looks at the formula of their α into B3 that is 21 + (1 – α) into C3. So, I need to know that actually in Excel. So, in the C column, I used alpha into Yt + (1 – α) into Yt.

So, I have got an exponential smoothing value for the C column. Then, I found the forecast error in the D column. Then, I found the absolute value of forecast error. So, what equations have I used? Sorry, what functions are used in Excel? There is an absolute function value. Then I found the squared error actual minus predicted value. Then, I found the percentage of error. The percentage of error is the error divided by the actual value.

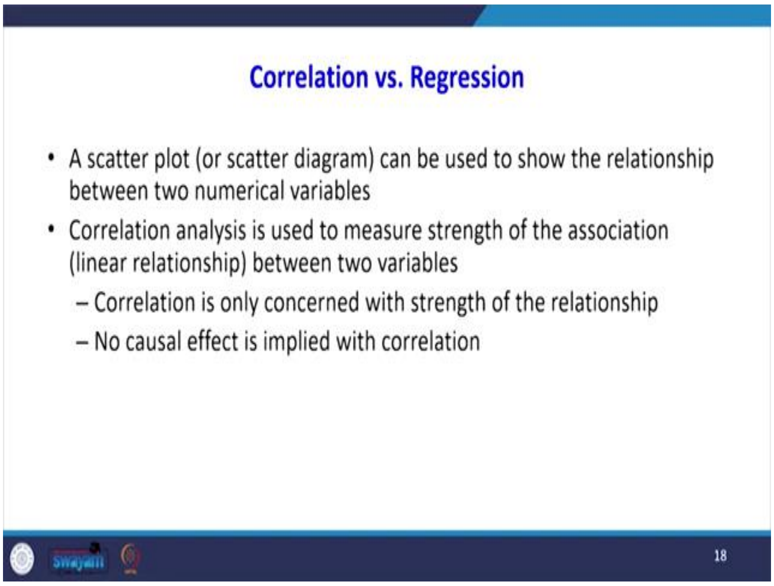
Then, I found the absolute value of the percentage error. In the end, I have summed up all the totals. So, from that, I have found the mean absolute error, then the mean squared error, and the mean absolute percentage error. This is one way to do Excel. But in Excel, there is a built-in function that is there. Now, I am going to explain how to use a spreadsheet for exponential smoothing. So, go to data analysis and choose exponential smoothing. Here, we have to select the input range.

Then, here, for the damping factor, you have to take the (1 – α), so our alpha value is 0.2, and the damping factor is 0.8. There is a label, so I have to select the output range. You see that I have selected the output range not for week 1, week 2 onwards, and week 2 to 12 weeks. So, when I press ok. Look at the values in C and values in I. Both are the same, so we can use

Excel for exponential smoothing, and we have the corresponding plot also. So, the orange color shows the forecasted value, and the blue line shows the actual value.

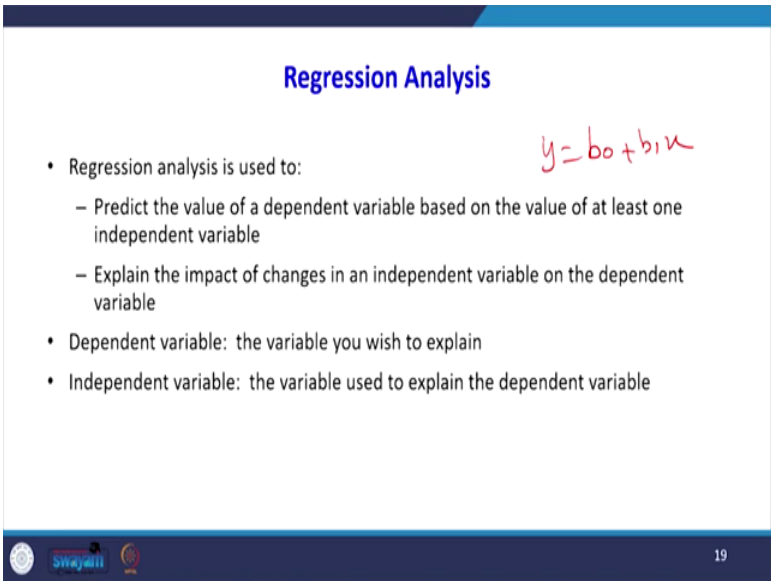


Dear students. Next, we are going to discuss trend forecasting. With trend forecasting, we know what a trend is. What is the trend? The trend will be like this. So, what we are going to do for this is we are going to do a regression analysis.



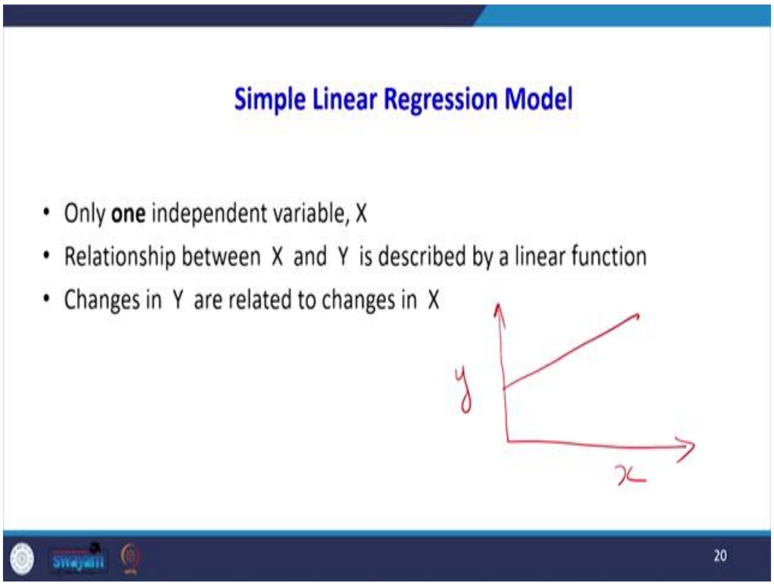
First of all, what is the difference between in terms of application correlation and regression? A scatter plot or scatter diagram can be used to show the relationship between two numerical values. Correlation analysis is used to measure the strength of association, which is a linear relationship between two variables. Correlation is only concerned with the strength of the relationship. No causal effect is implied with the correlation.

However, in regression, we can say which is independent and which is a dependent variable. But in correlation, we will not say which is the dependent variable or which is the independent variable. That is a major difference between correlation and regression analysis.

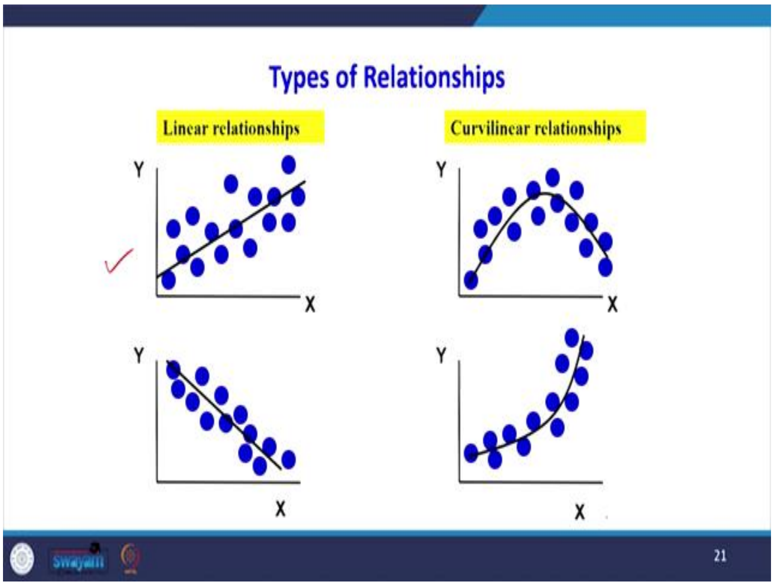


Now, we will discuss the regression analysis used to predict the value of the dependent variable based on the value of at least one independent variable. It explains the impact of changes in the independent variable on the dependent variable. Suppose I write this way: $Y = b_0 + b_1x$. This is an example of regression analysis. What is it saying here? Y is the dependent variable, and x is the independent variable.

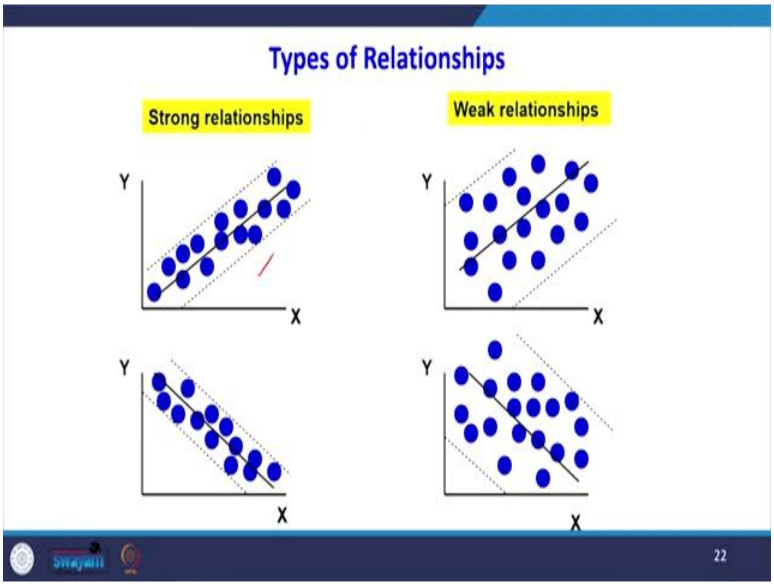
What is the dependent variable? The variable you wish to explain. What is the independent variable? The variable is used to explain the dependent variable.



First, we will discuss a simple linear regression model where we are going to use only one independent variable. So, in simple regression analysis, only one independent variable will be used. The relationship between X and Y is described by a linear function. So, changes in Y are related to changes in X. So, if I used only one independent variable this way, this is an example of a simple linear regression model.

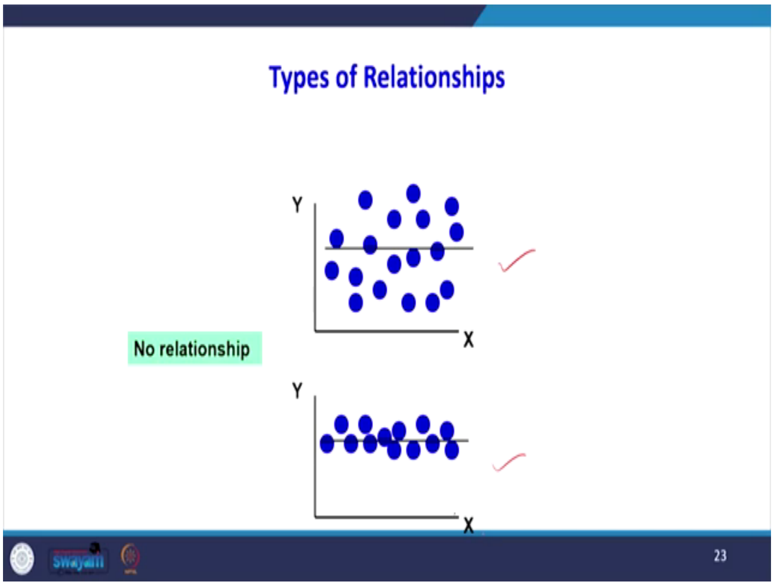


We may have different types of relationships. So, this one follows a linear relationship. In a positive linear relationship, X increases, and Y also increases. The bottom one if X increases, Y decreases. On the right-hand side, we can see curvilinear relationships. However, in this regression, we are going to consider only the linear relationship models.

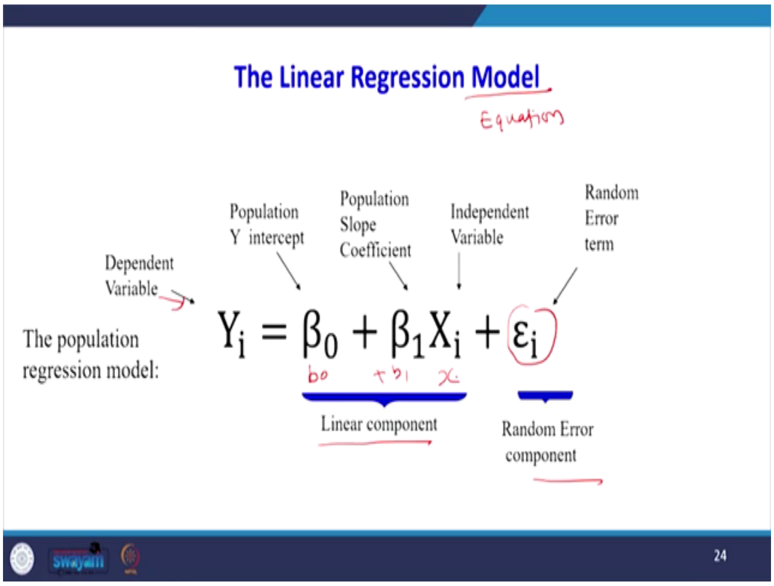


And types of relationships, whether it is a strong relationship or weak relationship. Look at this picture on the left-hand side. Here are all the actual points that the blue dots represent actual points. These are packed together. So, it has a strong relationship; the dispersion is very low. But look at the right-hand side. The variation is more So, which means it is the weak relationships.

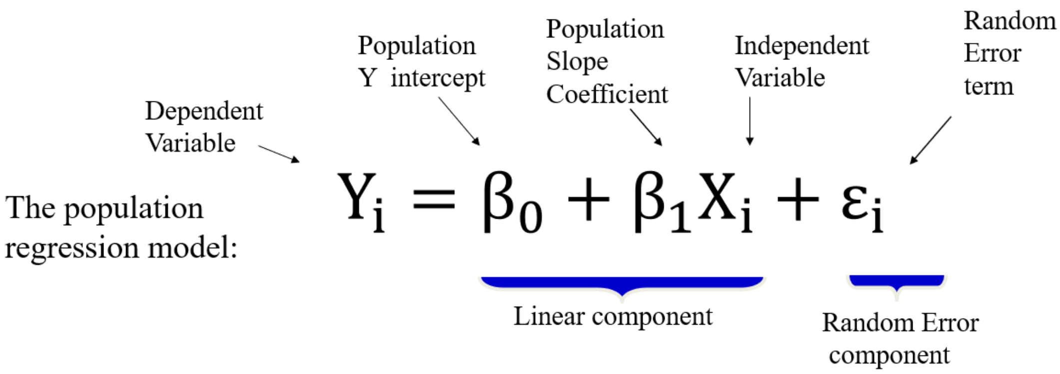
The top one has a weak relationship with a positive slope, and the bottom one has a weak relationship with a negative slope.



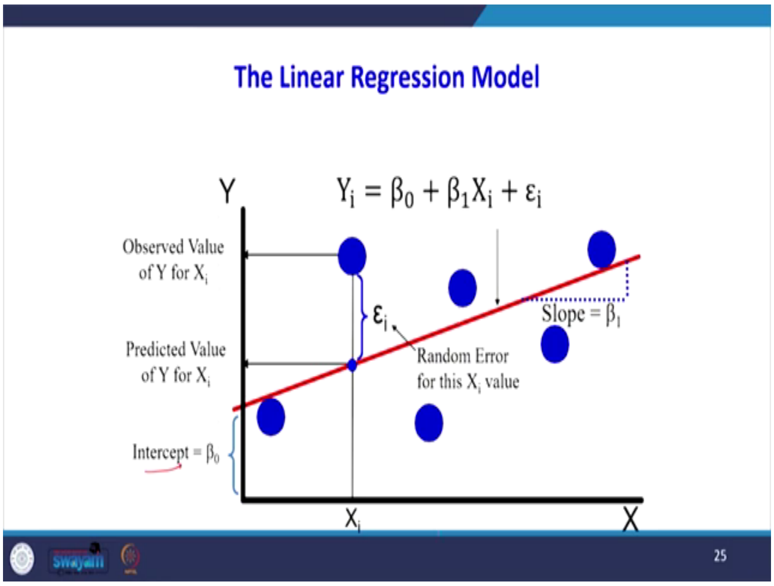
Sometimes, there may not be a relationship at all; see the top one, X and Y, there is no connection. It is the level type of data that is whatever may be the value of X; the value of Y is somewhat similar to a constant. So, here we can say, for example, here X and Y are independent. That means what? The value of Y is not affected by the value of X.



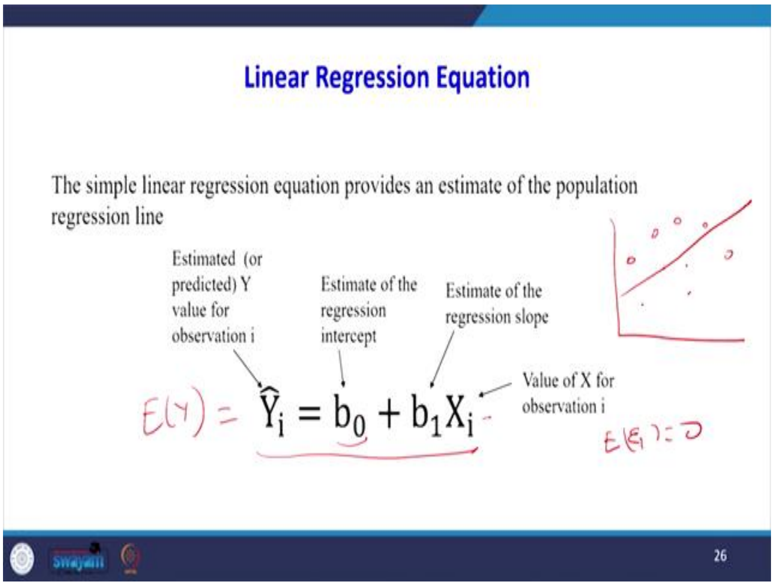
Now, regarding the linear regression model, you should be very careful. Sometimes, I say linear regression model, and sometimes, I say linear regression equation. If we say the model that represents the population, Yi dependent variable is equal to beta 0 population Y intercept plus beta 1 population slope coefficient XI independent variable plus a random error term. So, this component is a linear component.



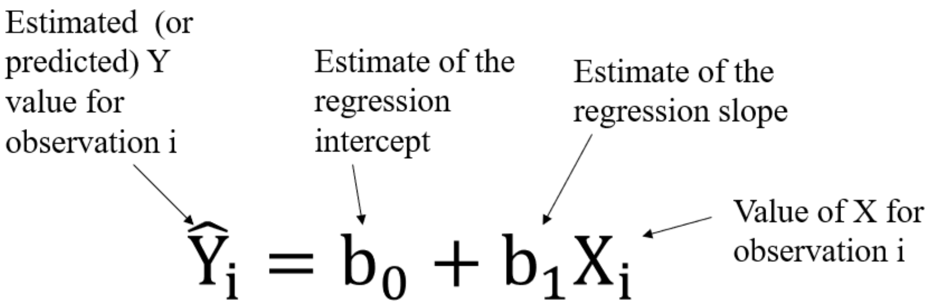
This is an error component. This equation expresses the regression model for the population. Because we have included an error term here. Not only that, all capital letters represent the regression model for the populations, beta 0 beta 1 in case I write b0 + b1, this represents the regression equations for the sample. That is the difference. There won't be any error term, and I will explain why.



Now, look at the linear regression model. The beta 0 represents the Y-intercept. This epsilon I represents the error term. That is, what is the error term? Difference between actual value and predicted value. This red line shows the regression model, and the slope is beta 1.

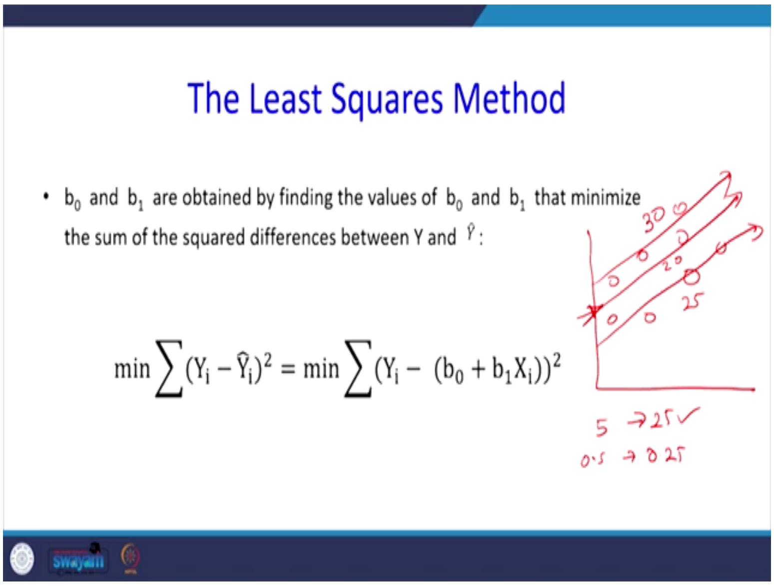


Now, you see, I am saying linear regression equations, and when I say equations, look at the equation here that all are in small letters, beta 0, sorry, $b_0 + b_1x$, and see this there is no error term here. If there is no error term, we call it a regression equation for the sample. Why is there no error term? That is how we got the values of b_0 and b_1. It takes care that error is minimized.



So, what will happen? When you say the expected value, this Y is nothing but the expected value of our Y equal to what we can write the expected value of y. That is the mean of the Y value, which is very important. The estimated or predicted Y value of observation is not the actual value. So, what I mean here, not actual value, is suppose the data is like this. The equation which I am drawing is not the actual value, it is the mean value or the expected value.

So, what will happen to any error term? The expected value of the error term is equal to 0. What is the meaning? There will be positive errors and a negative error. If you add that the expected value of error will become 0 that is why there is no error term here.

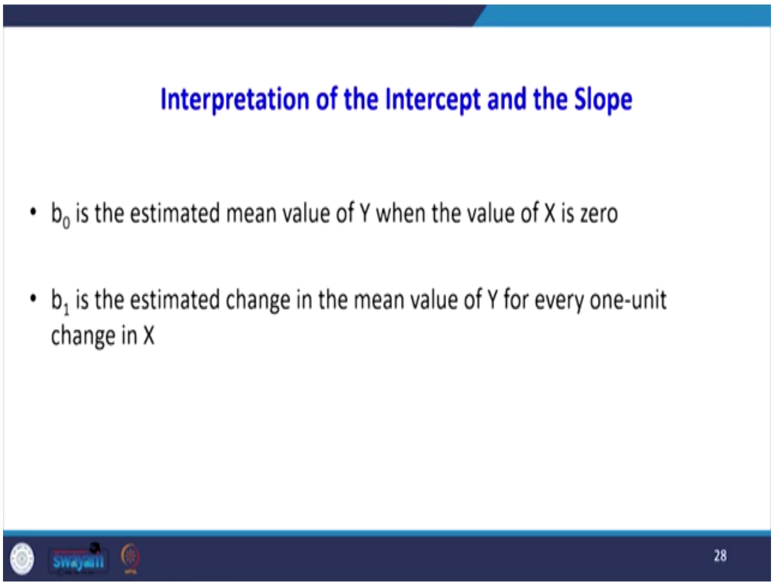


Here, b_0 and b_1 are obtained by finding the values of b_0 and b_1. That minimizes the sum of the squared differences between Y and Ŷ. The way the logic behind how we are finding this in slope and the y-intercept is that the sum of the square of the error has to be minimized. For example, if I have data like this, there are different ways I can draw a regression equation like this.

I can draw a regression equation like this, or I can also draw a regression location like this. But how do you know the best regression equation? That means for each regression type. We have to find out the error. We have to square the error, and then we have to sum the error. For example, in the first one, the sum of the squared error is 25. For the second one, the sum of the squared error is, say, 20. For the third one, the sum of the squared error is, say, 30.

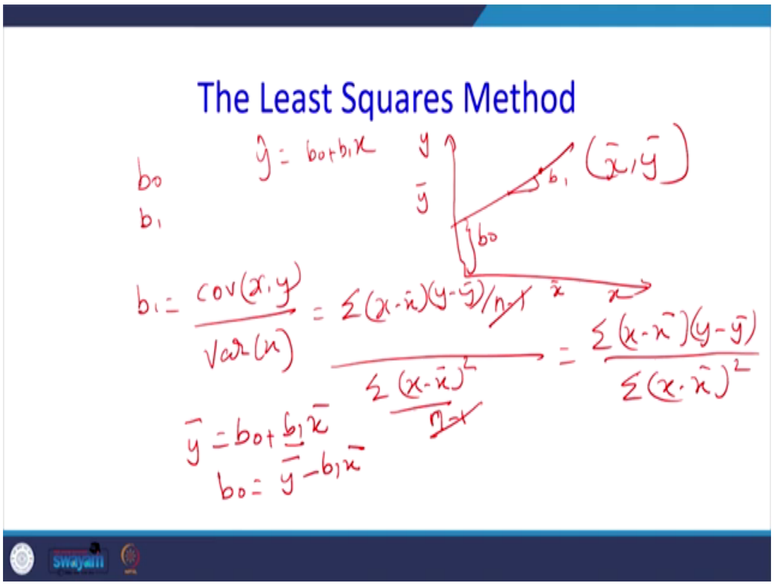
So, we can say the second one, the middle one line is the best equation because that has minimized some of the square of the error. That is the concept of your least square method. (Yi – Ŷ) whole square. Because I am squaring the error. So that it will become positive become positive. There are more deviations when you square it. It will become higher values.

For example, if the deviation is 5 when you square it, it will become 25. If the deviation is 0.5 when you square it, it will become 0.25. That is the reason we are squaring the error. If you are not squaring, the sum of the error will become 0. The purpose of squaring is to provide a higher penalty for higher deviations and a lesser penalty for lesser deviations. That is the beauty of this squared transformation.



b_0 is the estimated mean value of Y when the value of X is 0.

b_1 is the estimated change in the mean value of Y for every 1-unit change in X.



Now, to construct a regression equation, we need two things: one is b0, and the other one is b_1. Suppose I write $\hat{Y} = b_0 + b_1x$; suppose the regression equation is like this. So, this length is b_0. This slope is b_1. There is a detail in the calculus on how we get b_1 and b_0. But I will tell you the logic behind this b_1. So, the b_1 is nothing but covariance between x and y divided by the variance of x.

What is the covariance?

Step 1: **Covariance Formula**

$$\text{Cov}(X,Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1}$$

Step 2: **Variance of X**

$$\text{Var}(X) = \frac{\sum (X - \bar{X})^2}{n - 1}$$

Step 3: **Slope Calculation (Cancellation of $n - 1$)**

$$m = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

$$m = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

Step 4: **Finding the Y-Intercept (b)**

The equation of a straight line is:

$$Y = mX + b$$

Solving for b:

$$b = \bar{Y} - m\bar{X}$$

Final Regression Equation:

$$Y = mX + b$$

Where:

- m is the slope,
- b is the Y-intercept,
- $\bar{X}$ and $\bar{Y}$ are the means of X and Y.

The other logic is if you want to draw the best regression line, it has to pass through this is the x-axis; this is the y-axis.

$$\hat{Y} = b_0 + b_1 X$$

Since the regression line must pass through the mean point $(\bar{X}, \bar{Y})$, substituting $X = \bar{X}$ and $Y = \bar{Y}$:

$$\bar{Y} = b_0 + b_1 \bar{X}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

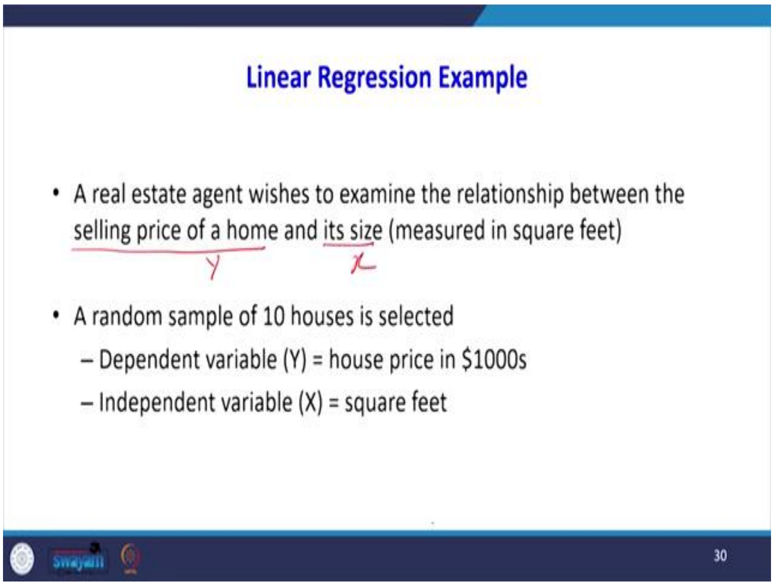
From regression, we already know:

$$b_1 = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}$$

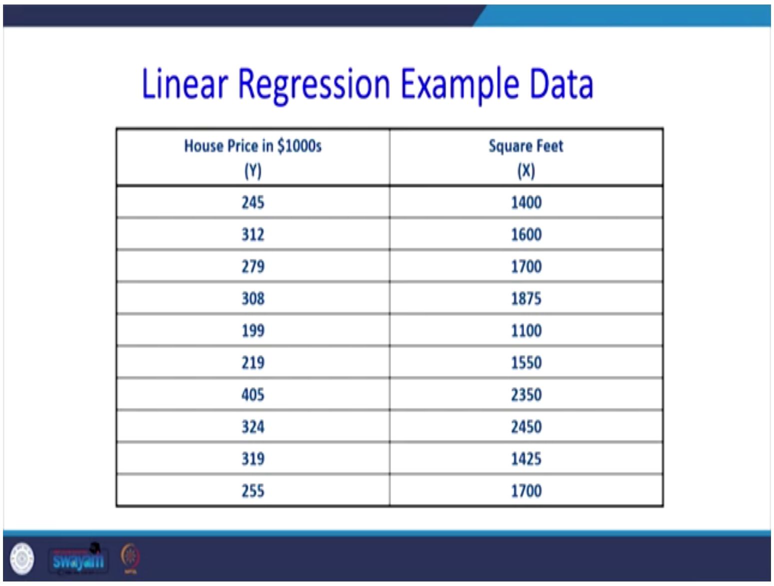
Thus, the final equation for the regression line becomes:

$$\hat{Y} = (\bar{Y} - b_1\bar{X}) + b_1 X$$

In our class, we are going to use the spreadsheet directly, but you should know what is the logic behind this slope and y-intercept.



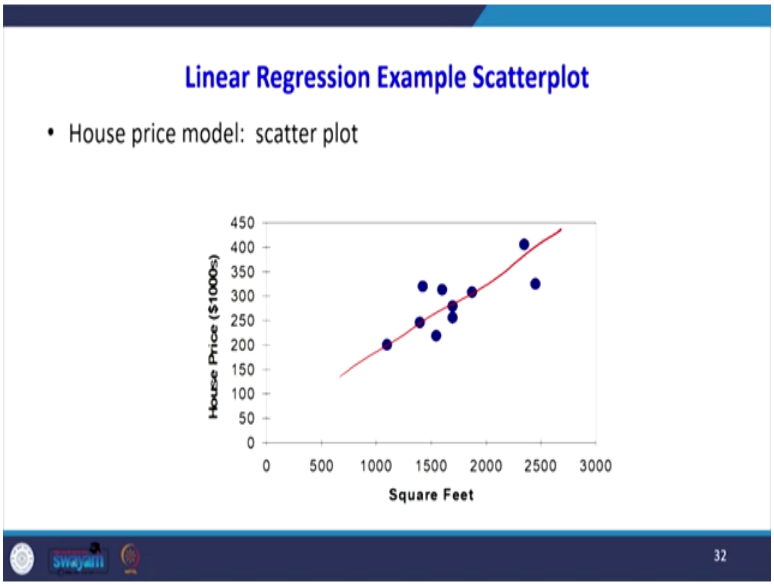
So, I have taken one sample problem. A real estate agent wishes to examine the relationship between the selling price of a home and its size. That is measured in square feet here. The selling price of the home is taken as your dependent variable, and its size is taken as your independent variable. A random sample of 10 houses is selected. So, the dependent variable is house price, and the independent variable x is square feet.



Linear Regression Example Data

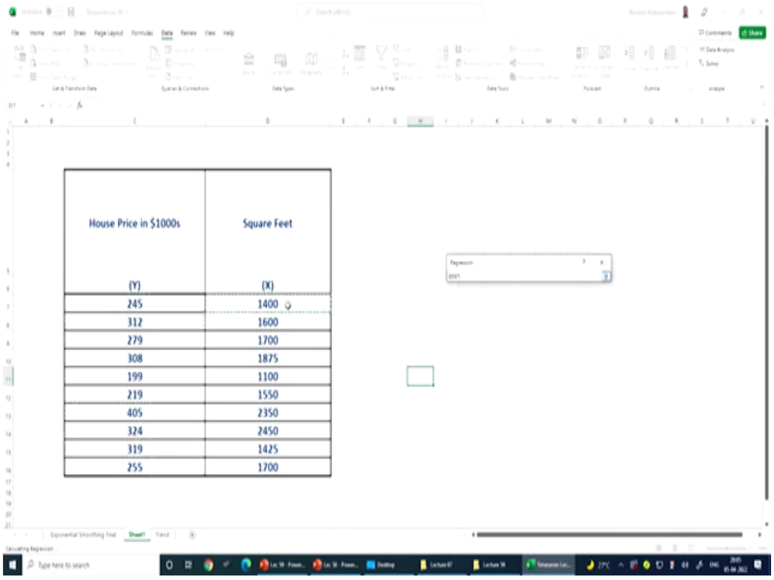
House Price in $1000s (Y)	Square Feet (X)
245	1400
312	1600
279	1700
308	1875
199	1100
219	1550
405	2350
324	2450
319	1425
255	1700

This is an X value. This is a Y value. I am going to use a spreadsheet for this data. How do I do the regression analysis?



Before doing the regression analysis the first step is to plot the scatter plot. The purpose of a scatter plot is to get a rough idea of the relationship between x and y. You see that you can see a kind of trend. That means we can plot a regression model. Suppose this scatter plot shows that there is no relationship at all. Then, there is no point in doing the regression analysis.

So, the first step for drawing a regression equation is to draw a scatter plot to confirm that there is a relation between x and y.

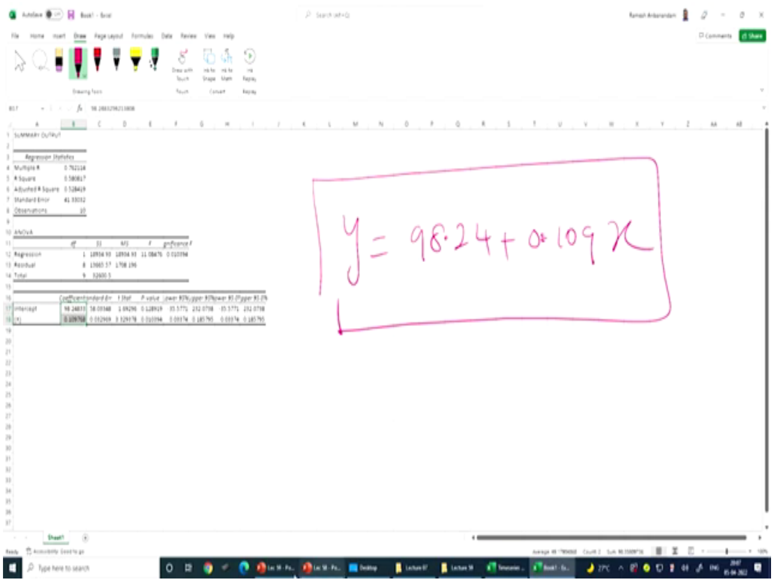


Dear students, now I am going to explain how to do the regression analysis with the spreadsheet. So, you have to click data. I have the data here, so, the data I have a data analysis tab for some students. they may not have this data analysis tab. For those people, what you have to do is an add-in in Excel, so you have to go to file. Then you have to go to options. Then you have to click on add-in.

So, you see that here the analysis tool pack is activated. If it is not there, it will be under inactive application add-ins. So, you have to select that you have to press go. You will get you have to press go. You see that already. I checked it because in this system I already have installed it is already checked. When you press it on your computer, also it will be enabled. Then, you can press the data tab. You will also get the data analysis tab.

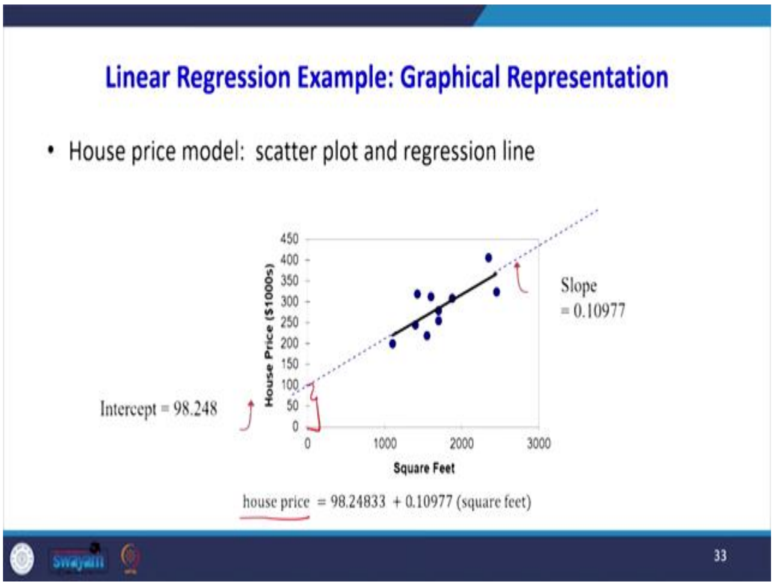
So, go to data analysis, and then go to regression. So, click ok. You see that first, you have to select the Y dependent variable. So, select the dependent variable, including labels. Next, you select the input range here. We have only one input, so I am selecting only one. If there is more than one input, you have to select all the inputs at a time. Then click the labels; there are other components there, such as residuals, standard residuals, residual plots, and normal probability plots.

I will explain this in the next lecture, but at present, we are going to predict we are going to construct only a regression model. So, press ok. Now, we have the y-intercept as 98.24, and the slope is 0.109.



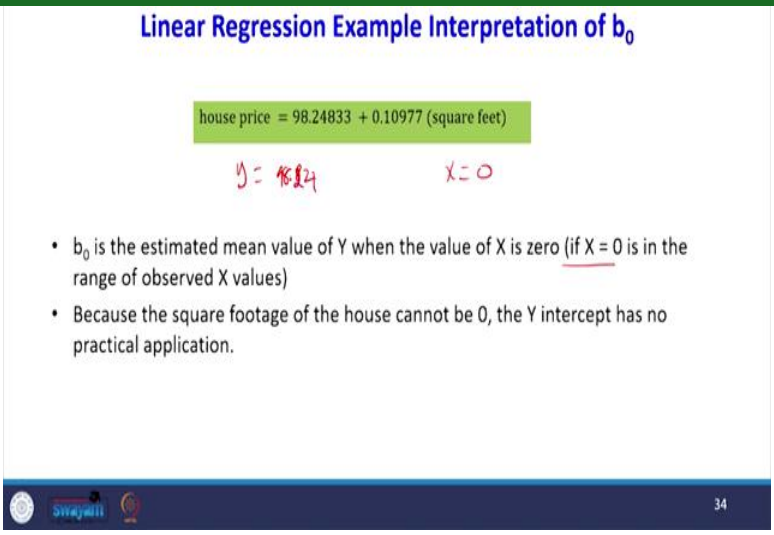
You see, the y-intercept is 98.24, and the slope is 0.104. There are other values there like r, r square, adjusted r square, and standard error, and we have got the ANOVA table also, and there are other ways to interpret. For example, t values there, p values there, a lower 95 % limit is there, and an upper percentage limit is what I will explain in the next class. But at present we are looking at only the intercept 98.24 and the slope 0.109.

So, how can we write the regression equation? For example, we can write the regression equation Y = 98.24 + 0.109 x, so, this is our regression equation. Now, we will go back to the presentation. I will explain pictorially what is this 98.24?



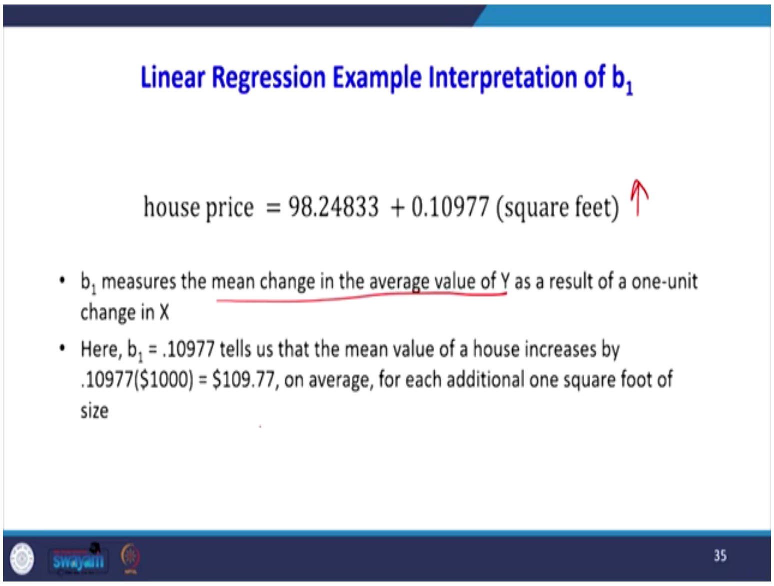
First, you have done the scatter plot. Look at the intercept. So, this value is your Y-intercept, and the slope is 0.10977. So, what is the dependent variable? So, is the house price equal to 98.24 + 0.109 square feet? There is a small rounding error there.

$$\text{house price} = 98.24833 + 0.10977\ (\text{square feet})$$



So, how to interpret the Y-intercept, b_0 is the estimated mean value of Y when the value of X = 0; if there is no independent variable, the b_0 is an approximate value of our Y value. That is, if X = 0 is in the range of the observed X value, here b_0 is the mean of all Y values because the square footage of the house cannot be 0. Then, the Y-intercept has no practical applications. So, the Y-intercept sometimes cannot be interpreted because of this context.

So, when you substitute X = 0, that means there are no square feet. We are getting the price of 98.24. But how is it possible? If there are no square feet at all, how can 98.24? How can a house price of 98.24 so there is no meaning to interpret the Y-intercept?



But we are going to interpret the coefficient of square feet as X. So, b 1 measures the mean change in the average value of y. Remember, the house price is an average value. It is not the exact value. So, the average value of Y as a result of 1 unit change in X. So, if the square feet are increased by 1 unit, what is the corresponding change in the house price? That is the meaning of slope, so, here, b_1 = 0.10977, which tells us that the mean value of a house is increased by 10977.

In actual terms, it is 109.77 dollars on average for each additional 1 square foot of size. Dear student, in this lecture, I have explained how to use a spreadsheet for exponential smoothing models for the sample problem. In the next lecture, I will also explain the fundamental concepts of building a regression model for forecasting. I will discuss how to use the regression model for prediction. Thank you.