

## Recall:

- **Example:** Relationship between advertising expenditures and sales (weeks 2 and 4)

A small-business owner wants to assess the effects of advertising on sales levels.

- Paired observation data were collected.
- Each pair consisted of monthly advertising expenditure and monthly sales levels (in thousand dollars).

(Values are in thousand dollars)

Advert	Sales
$x$	$y$
1	30
3	40
5	40
4	50
2	35
5	50
3	35
2	25

**Estimated model:**

$$\hat{y} = 23.03 + 4.83x$$

Intercept
Slope

### Slide 2:

In this video, I will show you how you could estimate a regression model using Microsoft Excel.

Let us go back to the example on relationship between advertising expenditure and sales.

- In the last video, we estimated the regression model and got an intercept equal to 23.03 and the slope equal to 4.83. These are point estimates of the intercept and slope coefficients.
- The good thing with statistical softwares, such as Microsoft excel, is that it can estimate the regression model in few simple steps.
  - The regression output reports a number of other statistics which are very helpful in terms of statistical analysis.
  - In this video, we will focus on generating the excel output and show that we can get the estimated equation through excel. We will also discuss the interval estimates of the two parameters reported in the excel output. The next video will discuss how we could use the regression output to assess fitness of the model.

# Regression in Microsoft Excel

**Step 1:** Enter the data on x and y in an Excel spreadsheet.

**Step 2:** Go to the 'Data' tab  
(This will display 'Data Analysis' top right).

**Step 3:** Click on 'Data Analysis'  
(This will pop up the Data Analysis box).

**Step 4:** Select "Regression" from the given options,  
Click Ok.

**Step 5:** Provide details of where the data is located and select  
output options.  
Click OK. This will generate the required regression output.

## Slide 3:

This slide summarizes the five steps procedure for estimating a simple linear regression model. Let us go through these steps one by one.

## Regression analysis in Excel

### Step 1:

Enter the data on X and Y in an Excel spreadsheet.

The screenshot shows an Excel spreadsheet with the following data:

	Advert	Sales
	<i>x</i>	<i>y</i>
1	1	30
2	3	40
3	5	40
4	4	50
5	2	35
6	5	50
7	3	35
8	2	25

Blue callout boxes indicate the data locations:

- Data on X is located in cells A1 to A9
- Data on Y is located in cells B1 to B9

### Slide 4:

The first step is to open an excel file and enter the data.

- You should be already familiar with this as you frequently used excel in this subject.
- You should note the location of the data in the spreadsheet as you would need this information later in the final step to generate the regression output.
  - Notice that in our case, the X data is located in column A, in cells A1 to A9, and the Y data is located in column B, in cells B1 to B9.
  - Also note that the first row of each column lists the name of the data series.

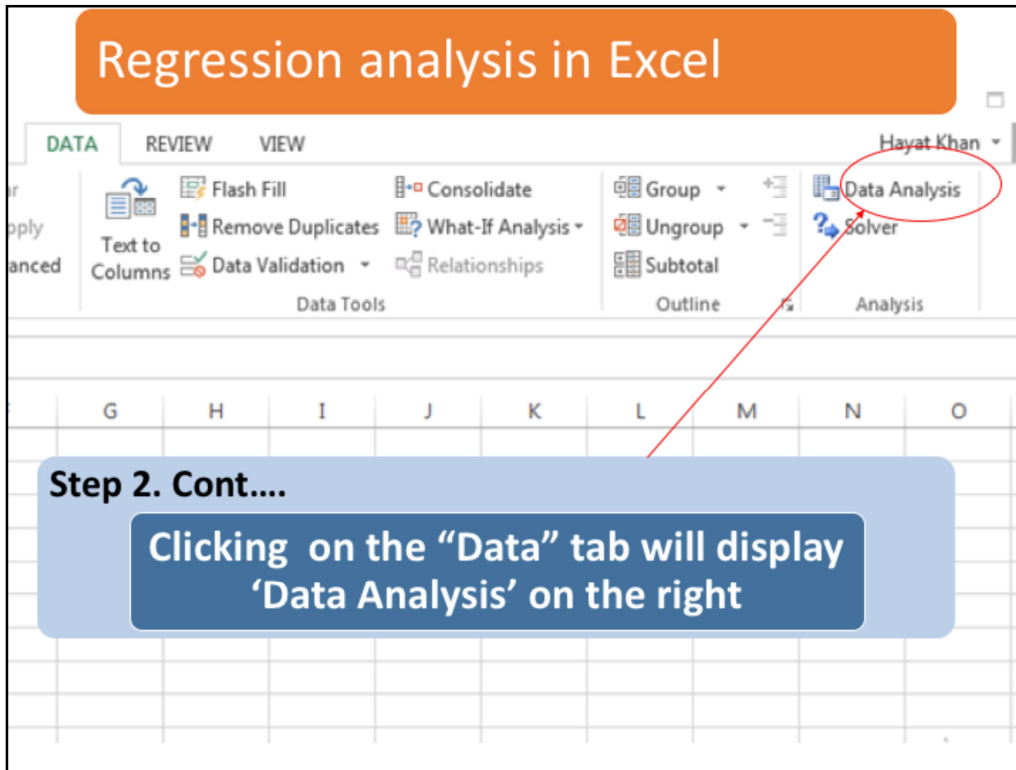
## Regression analysis in Excel

**Step 2:**  
Go to the 'Data' tab

	A	B	C	D	E	F	G	H	I
1	x	y							
2	1	30							
3	3	40							
4	5	40							
5	4	50							
6	2	35							
7	5	50							
8	3	35							
9	2	25							

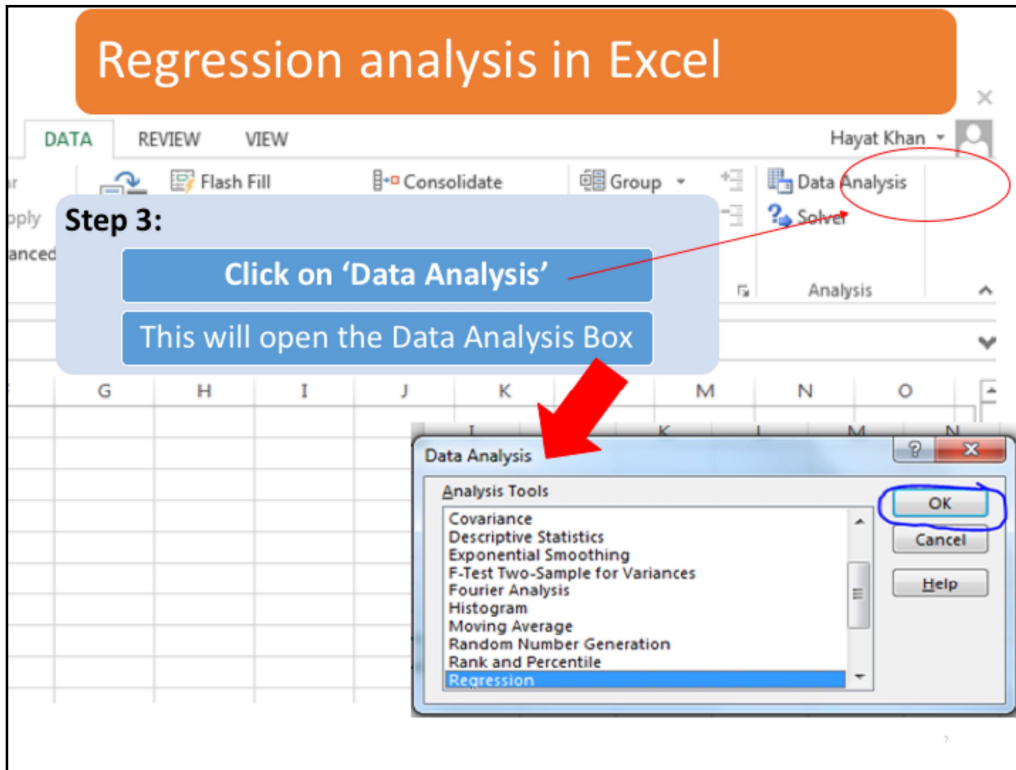
### Slide 5:

On top of the excel sheet you will notice different tabs. The second step is to go to the (Data) tab. You can do so by simply clicking on the tab.



**Slide 6:**

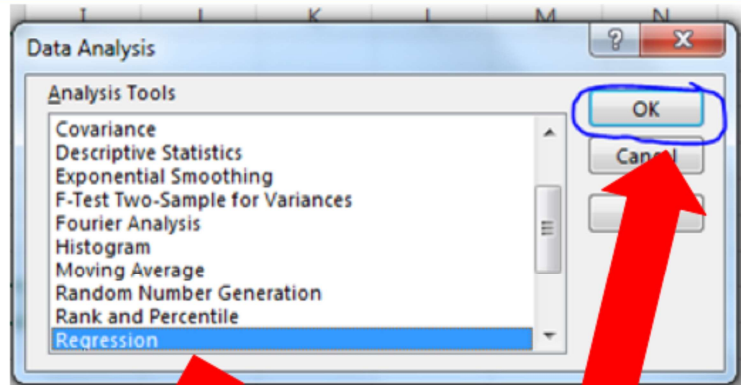
This will display Data Analysis option on the top right, as you can see here.



**Slide 7:**

The third step requires you to click on the Data Analysis tab. This will open a box, named Data Analysis. The box lists a number of analysis tools.

## Regression analysis in Excel



### Step 4:

Select "Regression"  
from the Analysis  
Tools options

Click Okay (This will  
open a regression  
Window, see next)

### Slide 8:

One of the items on the list will be regression. The fourth step is to select the regression option from the analysis tools and click OK.

**Step 5:**

	A	B
	x	Y
2	1	30
3	3	40
4	5	40
5	4	50
6	2	35
7	5	50
8	3	35
9	2	25

**Step 5:**

- Write B1:B9 in Input Y Range (as that's where Y is located)
- Write A1:A9 in Input X Range (as that's where X is located)
- Select labels and confidence level
- Select Line Fit plots
- Click OK

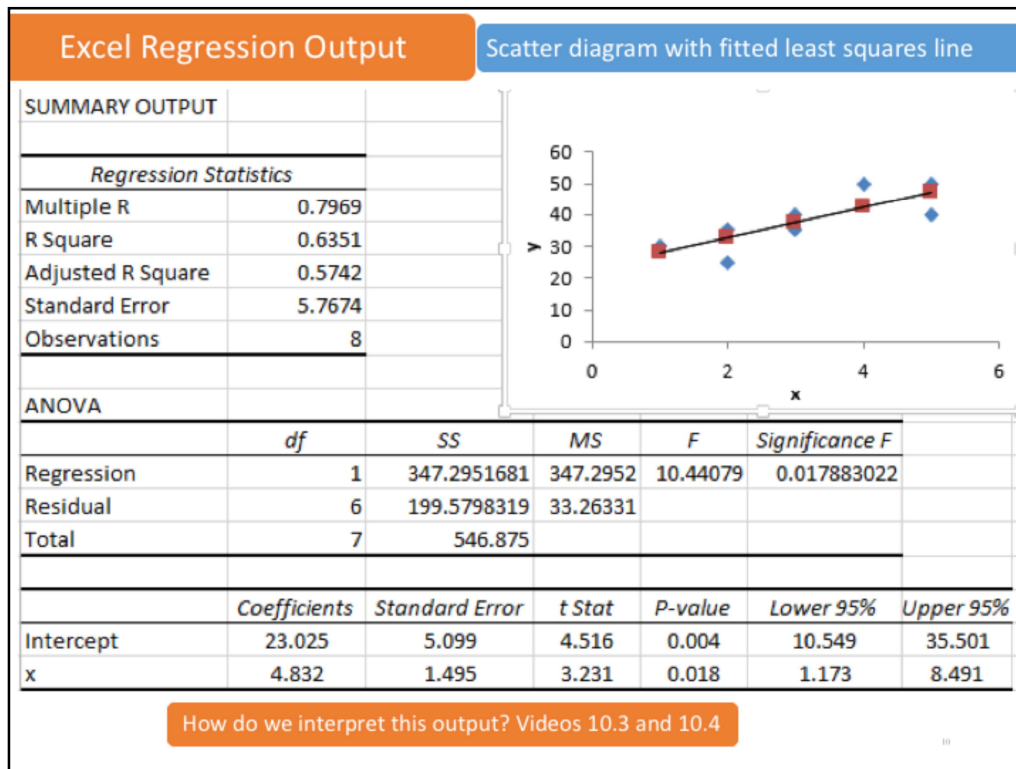
### Slide 9:

This will open a new box, labelled as Regression.

In this box, we need to tell Excel where the data is located, along with some more options.

- Insert the cell reference for data on variable Y in (the input Y range reference cells), and data on variable X in (the input X range reference cells). It is important to remember that Excel treats variable Y as the dependent variable and variable X as the independent variable.
  - Since the data on Y is located in cells B1 to B9, we write B1:B9, which means B1 to B9.
  - Similarly, we write A1:A9 in the input X range, which tells Excel that the data on variable X is located in cells A1 to A9.
- After identifying the location of the data, you should now select labels. This tells Excel that the first row contains names of the variables. The default confidence level is 95%, which will report 95% confidence intervals of the betas.
- You may also select the Line fit plots option which will generate a scatter plot of X and Y and X and Y hat.

Finally, when you click okay, excel will generate its final output.

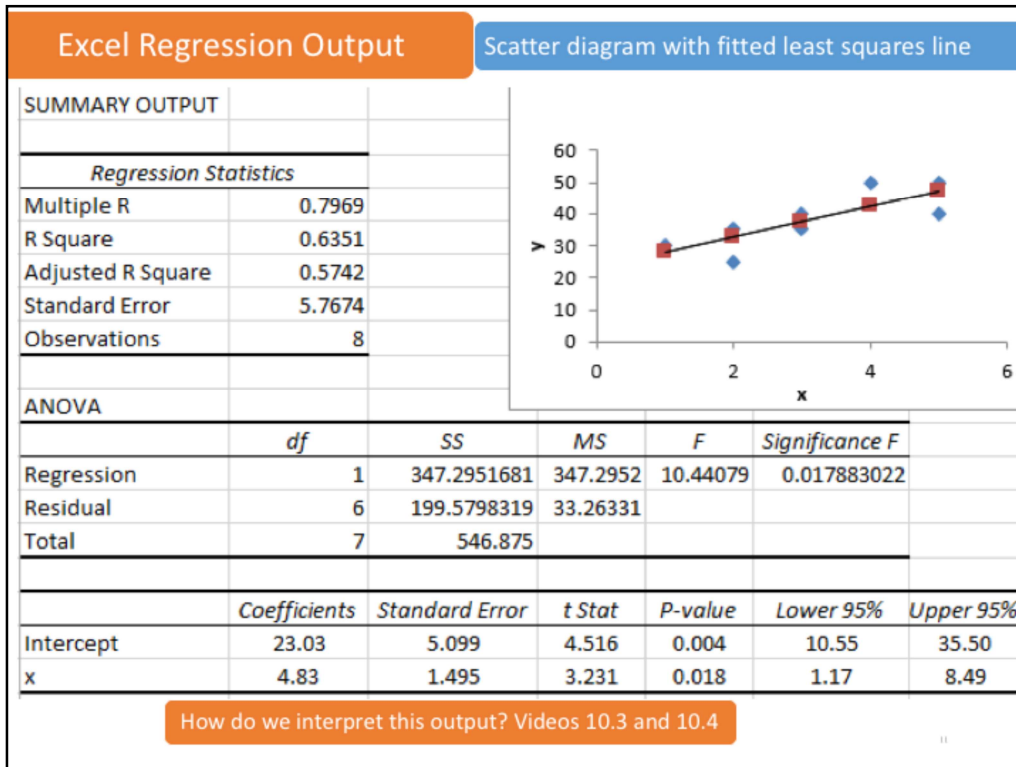


**Slide 10:**

The excel output is reported here. Here,

- we will discuss point estimators of the slope and intercept coefficients, which are reported in the bottom left of the regression output table, and
- interval estimators of these coefficient which are reported at the bottom right.

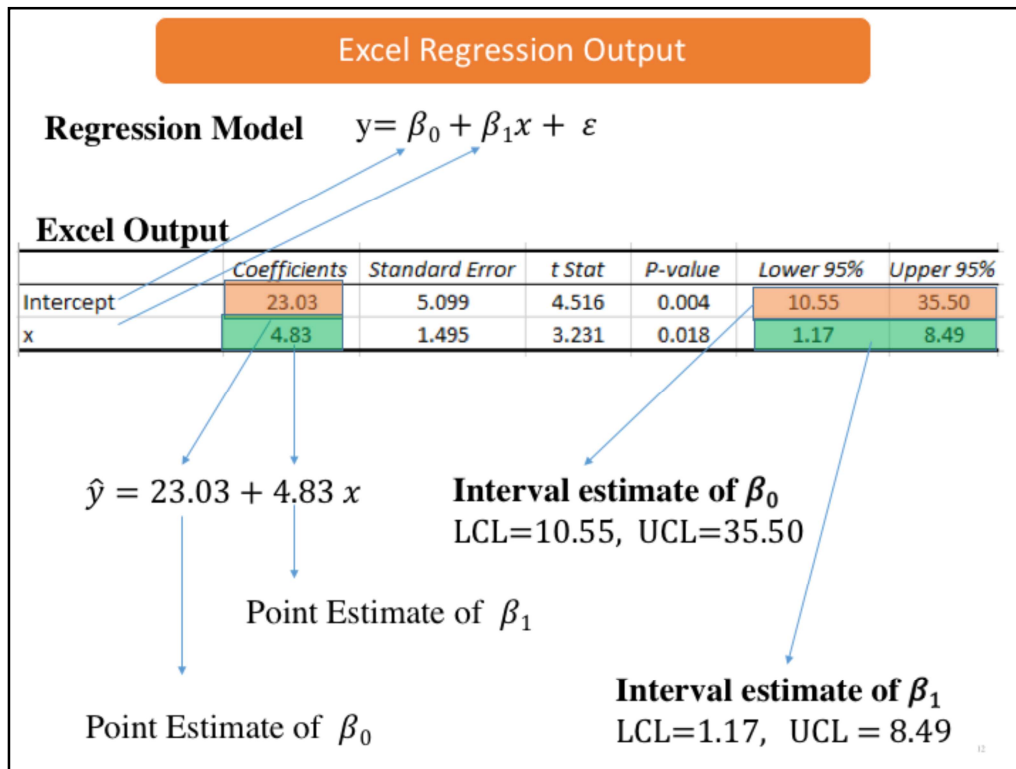
We will discuss the rest of the output in the next video.



**Slide 11:**

Recall that the regression output is for the regression model,  $Y$  equal to  $\beta_0$ , plus  $\beta_1$  times  $x$ , plus the error term.

- Recall that  $\beta_0$  is the intercept and  $\beta_1$  is the slope coefficient.
- The Excel output reports point estimates of these coefficient here.



**Slide 12:**

- The value of the intercept is 23.03 and that of the slope is 4.83.
- The equation of our estimated regression model is therefore  $\hat{y} = 23.03 + 4.83x$ . Notice that this is the same equation we derived in the previous video using the formulas for the intercept and slope.

The lower confidence limit for 95% confidence interval of the slope is 1.17, whereas the upper confidence limit is 8.49. Similarly, the 95% of lower and upper confidence limits for the intercept are 10.55 and 35.50 respectively.

Let us derive these limits for the slope coefficient.

## 95% confidence intervals of estimates

**Regression Model**  $y = \beta_0 + \beta_1 x + \varepsilon$

### Excel Output

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	23.03	5.099	4.516	0.004	10.55	35.50
x	4.83	1.495	3.231	0.018	1.17	8.49

Confidence interval of  $\beta_1$ :

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, n-2} S_{\hat{\beta}_1}$$

95% Confidence interval of  $\beta_1$ :

$$\hat{\beta}_1 \pm t_{0.025, 6} S_{\hat{\beta}_1}$$

From t-table:  $t_{0.025, 6} = 2.447$

$$\begin{aligned} LCL &= \hat{\beta}_1 - t_{0.025, 6} S_{\hat{\beta}_1} \\ &= 4.83 - 2.447 * 1.494 \\ &= 1.17 \end{aligned}$$

$$\begin{aligned} UCL &= \hat{\beta}_1 + t_{0.025, 6} S_{\hat{\beta}_1} \\ &= 4.83 + 2.447 * 1.494 \\ &= 8.49 \end{aligned}$$

### Slide 13:

Since the population variance of the estimate are not known, the confidence intervals are based on the t-distribution.

- The general formula for the confidence interval of beta-1 with a tail probability of alpha is  $\beta_1\text{-hat} \pm t_{\alpha/2, n-2} S_{\hat{\beta}_1}$ .
- For alpha equal to 0.05, which corresponds to 95% confidence interval, and n equal to 8, the interval is  $\beta_1\text{-hat} \pm t_{0.025, 6} S_{\hat{\beta}_1}$ , with 6 degrees of freedom, times the standard error of  $\beta_1\text{-hat}$ .
- Both,  $\beta_1\text{-hat}$  and its standard errors are reported in the regression output. From the t table, we find that  $t_{0.025, 6}$  equal 2.447.
  - Substituting these values in the lower and upper limits formula gives a lower limit of 1.17 and an upper limit of 8.49.
    - The 95% confidence interval of the slope coefficient is that it lies between 1.17 and 8.49.

## Interpretation:

- **Point Estimate: 4.83.**

- When advertising expenditures increases by \$1,000, on average, sales increases by 4.83 dollars

- **95% confidence Interval : LCL=1.17, UCL= 8.49**

- We estimate that each additional \$1000 expenditures on advertising results in an increase in sales between \$1170 and \$8490 using a 95% confidence interval.

14

### Slide 14:

With this in mind, let us now interpret these value.

- The point estimator of the slope coefficient is 4.83.
  - This means that when advertising expenditure increases by \$1000, on average, sales increase by \$4830.
- Given the 95% lower confidence limit of the slope is 1.17 and the upper confidence limit is 8.49, We estimate that each additional \$1000 expenditures on advertising results in an increase in sales between \$1170 and \$8490.

This concludes our analysis for this video. You may now proceed to the next video.