

Regression in excel

- ▶ Yash Khurana TY1G2233662
- ▶ Heeta Parmar TY1G2233674
- ▶ Gauravi shinkar TY1G2233692
- ▶ Viraj koparkar TY1G2233663
- ▶ Jay kunjir TY1G2233705

- ▶ Guided By- Sanket B
- ▶ Github ID : bsanketm

Regression Model

- A multiple regression model is:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

Such that:

- ✓ y is dependent variable
- ✓ x_2 and x_3 are independent variables
- ✓ β_1 is constant
- ✓ β_2 and β_3 are regression coefficients
- ✓ It is assumed that the error u is independent with constant variance.

- We wish to estimate the regression line:

$$y = b_1 + b_2 x_2 + b_3 x_3$$

Regression Analysis in Excel

- We do this using the **Data analysis Add-in** and **Regression**.
- Example:

	A	B	C
1	CARS	HH SIZE	CUBED HH SIZE
2		1	1
3		2	8
4		3	27
5		4	64
6		5	125
7			

Regression Analysis in Excel

The screenshot shows a Microsoft Excel window titled "carsdata.xls [Compatibility Mode] - Microsoft Excel...". The ribbon is visible at the top with tabs Home, Insert, Page Layout, Formulas, Data, Review, and View. The Data tab is selected. On the far right of the ribbon, the "Data Analysis" button is highlighted.

The main area of the screen displays a data table with three columns: CARS, HH SIZE, and CUBED HH SIZE. The data points are:

	A CARS	B HH SIZE	C CUBED HH SIZE
1	1	1	1
2	2	2	8
3	2	3	27
4	2	4	64
5	3	5	125
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

The cell A8 is currently selected. A regression dialog box is open in the foreground, titled "Regression". The "Input" section contains the following settings:

- Input Y Range: \$A\$1:\$A\$6
- Input X Range: \$B\$1:\$C\$6
- Labels
- Constant is Zero
- Confidence Level: 95 %

The "Output options" section contains the following settings:

- Output Range: \$A\$8
- New Worksheet Ply: (empty)
- New Workbook

The "Residuals" section contains the following settings:

- Residuals
- Standardized Residuals
- Residual Plots
- Line Fit Plots

The "Normal Probability" section contains the following setting:

- Normal Probability Plots

Buttons for OK, Cancel, and Help are visible on the right side of the dialog box.

Regression Analysis in Excel

- The regression output has three components:
 - Regression statistics table
 - ANOVA table
 - Regression coefficients table.

SUMMARY OUTPUT						
Regression Statistics						
	df	SS	MS	F	Significance F	
Multiple R		0.895828018				
R Square		0.802507837				
Adjusted R Square		0.605015674				
Standard Error		0.444400903				
Observations		5				
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	1.605015674	0.802507837	4.063492063	0.197492163	
Residual	2	0.394984326	0.197492163			
Total	4	2				
Coefficients						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.896551724	0.764398064	1.172885918	0.361624318	-2.39238769	4.185491139
HH SIZE	0.33646813	0.422703764	0.79599038	0.509506953	-1.482279374	2.155215634
CUBED HH SIZE	0.002089864	0.013113794	0.159363815	0.888021498	-0.054334235	0.058513964

Interpreting Regression Statistics Table

Regression Statistics

- The standard error here refers to the estimated standard deviation of the error term u .
- It is sometimes called the standard error of the regression. It equals $\text{sqrt}(SSE/(n-k))$.
- It is not to be confused with the standard error of y itself (from descriptive statistics) or with the standard errors of the regression coefficients given below.
- $R^2 = 0.8025$ means that **80.25%** of the variation of y_i around its mean is explained by the regressors x_{2i} and x_{3i} .

		Explanation
Multiple R	0.895828	$R = \text{square root of } R^2$
R Square	0.802508	R^2
Adjusted R Square	0.605016	Adjusted R^2 used if more than one x variable
Standard Error	0.444401	This is the sample estimate of the standard deviation of the error u
Observations	5	Number of observations used in the regression (n)

Interpreting Regression Statistics Table

Regression coefficients table

- The regression output of most interest is the following table of coefficients and associated output:

	Coefficient	St. error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.89655	0.76440	1.1729	0.3616	-2.3924	4.1855
HH SIZE	0.33647	0.42270	0.7960	0.5095	-1.4823	2.1552
CUBED HH SIZE	0.00209	0.01311	0.1594	0.8880	-0.0543	0.0585

Interpreting Regression Statistics Table

Regression coefficients table

- A simple summary of the previous output is that the fitted line is:

$$y = 0.8966 + 0.3365x + 0.0021z$$

Exercise

Y	X1	X2
39	15	110
44	9	65
50	10	90
64	12	100
65	3	160
55	13	130
66	15	105
12	2	20
92	20	240
81	17	95

Regression analysis in excel using scatter plot charts with Trendline

You can use Microsoft Excel scatter charts when you want to do a quick and brief regression analysis. This method also uses the least squares method. In addition to simple linear regression, Trendline gives you the option to fit your data in to other regression models such as, exponential; logarithmic; polynomial; power and moving average.

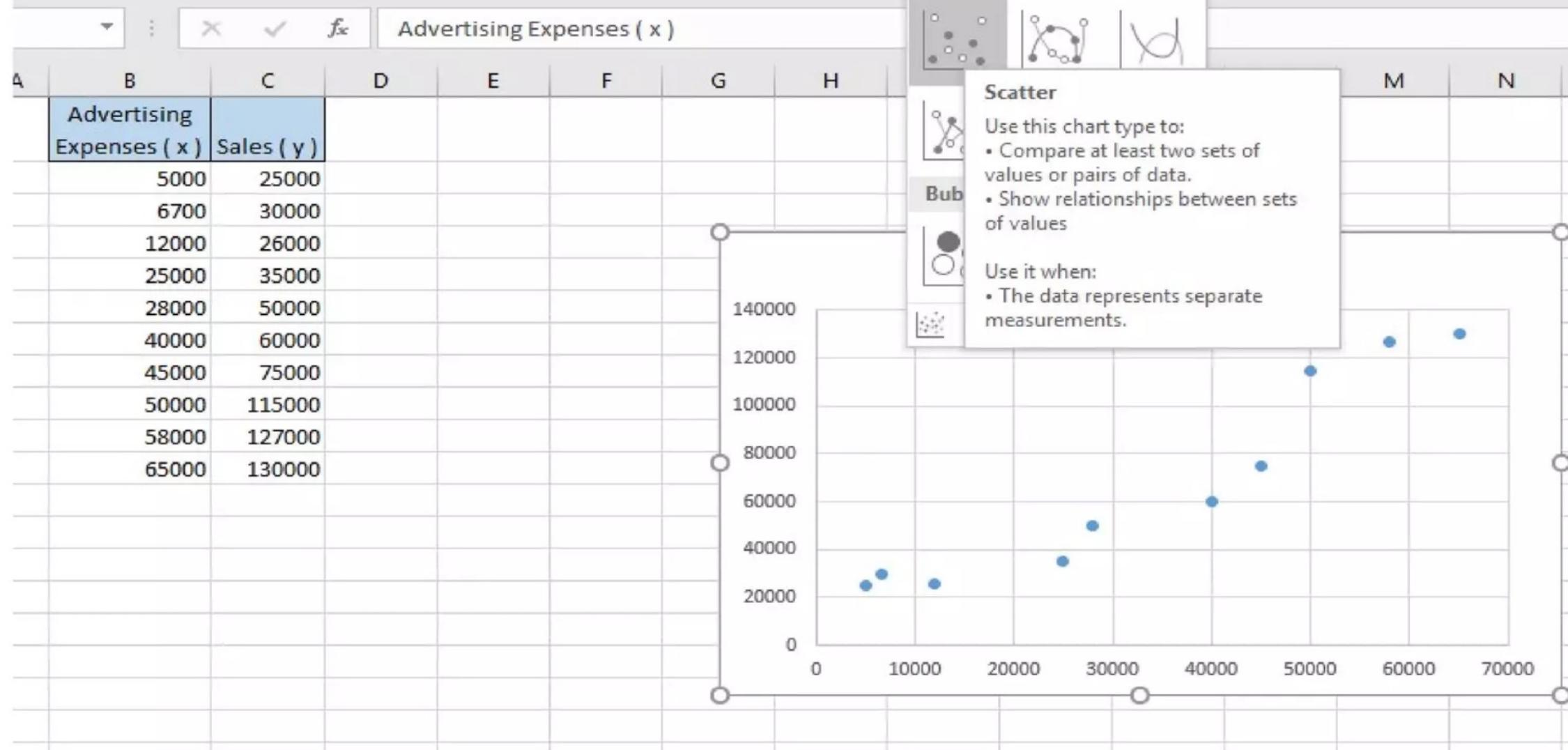
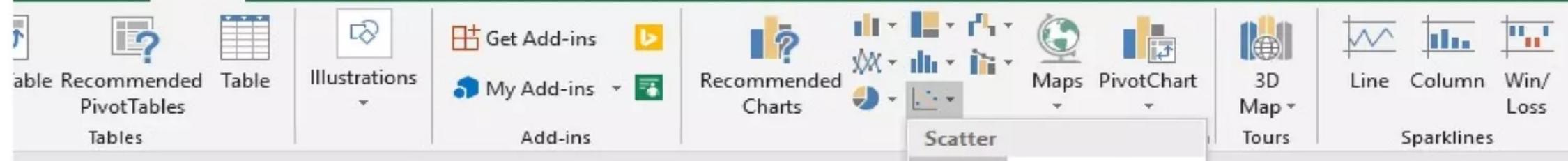
Regression analysis procedure in excel using trendline option

Step 01: Prepare your data in two adjacent columns. Make sure that your independent variable, x is in first column and the dependent variable, y is in next column.

Step 02 : Select both columns having X and Y values. You have the option to select with or without column headers.

Step 03: Go to → “Insert” Tab → “Charts” group → click “Insert scatter (X,Y) or Bubble chart” button.

Select any of the Scatter Chart type provided in the drop menu. I prefer the first chart type having only points.



Now you will see that the “Format Trendline” pane appears right side of the Microsoft Excel window.
Step 09: Configure the trendline options as follows.

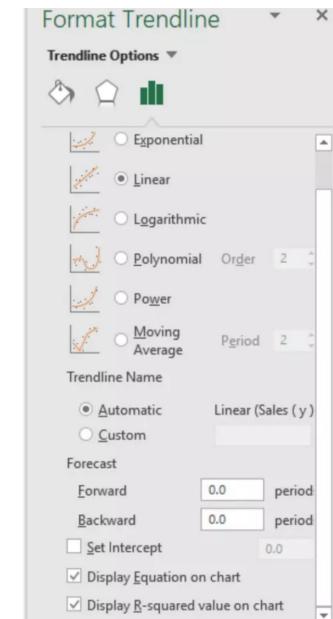
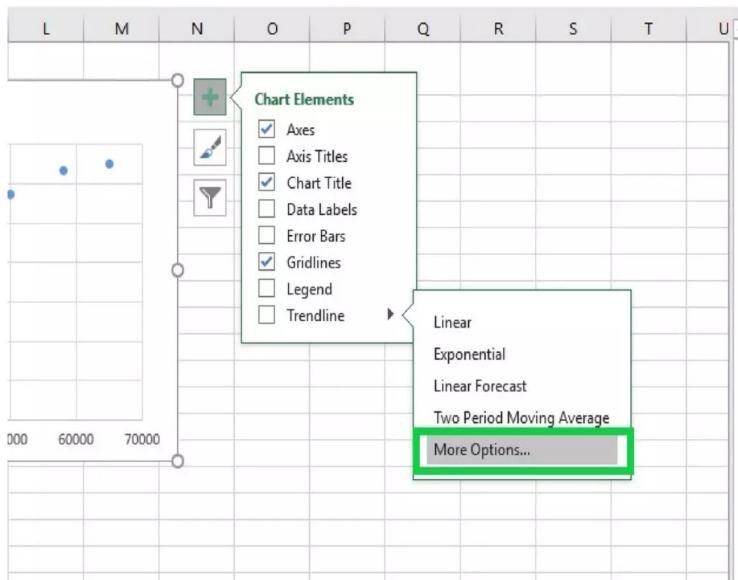
Step 04 : Click anywhere on the scatter chart. Three buttons will appear top right corner just outside the chart area.

Step 05 : Click the “Chart Elements” button which looks like thick ‘+’ symbol.

Step 06 : Move the mouse pointer on to the “Trendline” item of the appeared drop down menu.

Step 07 : Click the small black right-arrow head which appears in “Trendline” menu item.

Step 08 : Click “More options” menu item.



- 1) Select radio button for “Linear”.
 - 2) Select the checkbox for “Display Equation on chart”.
 - 3) Select the checkbox for “Display R – squared value on chart”.
- Now, you can see the regression equation and R² value above the trendline.

Regression analysis in Excel using formulas or worksheet functions

There are times that you only need to find regression coefficients. In that case you can simply use Excel worksheet functions or formulas. SLOPE(), INTERCEPT() and RSQ() are the main worksheet function you will need to find linear regression coefficients. Prepare your independent (X) and dependent (Y) variable values as in previous cases.

	B	C	D	E	F
1	Advertising Expenses (x)	Sales (y)			
2	5000	25000	Slope (m)	=SLOPE(C2:C11,B2:B11)	
3	6700	30000	Intercept (b)	=INTERCEPT(C2:C11,B2:B11)	
4	12000	26000	R-squared	=RSQ(C2:C11,B2:B11)	
5	25000	35000			
6	28000	50000			
7	40000	60000			
8	45000	75000			
9	50000	115000			
10	58000	127000			
11	65000	130000			

Slope (m) 1.869325316
Intercept (b) 4733.681688
R-squared 0.896682361



Calculate the slope of the regression line

Step 01 : Insert “= SLOPE ()” formula within a desired cell.

Step 02 : For the first parameter, select the Excel cell range that you have entered the Y-values which is the dependent variable. And for the second parameter select the cell range that you have entered the X-values which is the independent variable.

Step 03 : Press “Enter”.

Now we got the value for the slope of the regression line.

Calculate the intercept of the regression line

Step 01 : Insert “= INTERCEPT ()” formula within a desired cell.

Introduction

- If your data points are non-linear, clearly will not fit a linear regression (a straight line through all data points), it might be ideal for polynomial regression.
- Polynomial regression, like linear regression, uses the relationship between the variables x and y to find the best way to draw a line through the data points.

Limitation of Linear Model

- Generally, the data which are having very low variance with linear characteristics will suit the simple linear model.
- If the data points which shows different variation with non-linear characteristics then linear model will not fit best and may result high variance and low R² value for the model.
- In this constraint, it is advisable to take the non-linear regression models rather than optimizing the linear models.

POLYNOMIAL REGRESSION

Polynomial regression is one of the machine learning algorithms used for making predictions. for ex. It is widely applied to predict the spread rate of COVID-19 and other infectious diseases.

The polynomial regression model may contain one, two or more than two predictor variables. each predictor variable may be present in various powers. this polynomial model is called a second order with one predictor variable because the single predictor variable is expressed in the model to the first and second powers

Polynomial Regression

- The polynomial regression is an extension of linear model with higher order.
- We can derive different polynomial models depending upon different order and number of variables.

Second order polynomial with one variable.

$$Y = aX^2 + bX + c + \epsilon$$

Second order polynomial with two variable.

$$Y = aX_1 + bX_2 + c + \epsilon$$



**THANK
YOU**

