

## LEARNING STATISTICS THROUGH SIMULATION – A CASE STUDY OF REGRESSION ANALYSIS

**Dr. Anita Chaware**

(Ph.D) is an Associate Professor and  
Head, Dept. of Computer Science,  
SNDT Women's University,  
Juhu Campus , Mumbai 49

---

### *Abstract*

*An essential part of introductory statistics is linear regression analysis. All areas of real life are full of data and this data needs to be studied. Statistic education is thus the need of every field of higher education. An essential part of introductory statistics is linear regression analysis. Many students find it difficult to understand the basics of linear regression analysis. In this paper, we explain how to perform regression analysis using simulation using Excel Data Tables and standard methods. This paper demonstrates how to compute various regression formulas and draw the line of regression. Students will be able to understand the liner Regression and the line of regression better using this method. Our early analysis demonstrates that using this strategy in the classroom improves student learning.*

---

**Keywords :** Regression analysis, Collinearity, Scatter graphs,

---

### **Introduction**

Our daily lives offer several opportunities to use statistics. Numerous fields, including business, engineering, the natural and social sciences, can benefit from statistical information. Therefore, almost all degree programmes require students to take at least one statistics course. Many instructors are pursuing research on how to best teach and comprehend statistical topics like sample distributions, confidence intervals, hypothesis testing, and p-values because many students find them difficult to understand [1], [2], [3]. Computers have significantly changed how statistics are taught at all levels. They have significantly changed how statistics are taught at all levels. The advancement of technical resources has enhanced statistics education and learning. Many students studying at the

beginning level do not have a solid knowledge of the topics due to the traditional teaching method, which consists of books, lectures, and mathematical derivations. Computer simulation techniques (CSM) have been used in the classroom by statisticians to teach challenging ideas. Students can experiment with random samples using computer simulation approaches to help them understand statistical concepts that are difficult and abstract.[4][5].

Students' understanding is enhanced when computer simulations are used to illustrate fundamental ideas and guide them toward the understanding of key principles.

The benefits of the computer simulation is more when students drew multiple samples from a variety of populations in order to observe the sampling distribution of the mean. Through the use of simulations, instructors put students in settings where they can apply what they have learned in class. Since the 18th century, simulation training has been a feature of healthcare curricula; as technology has advanced, it has continued to change. Comparing simulation tools to conventional teaching techniques reveals various advantages. They are able to provide users with feedback that enables users to enhance their knowledge and skills. They also enable interaction between users and simulated equipment.

The two primary characteristics that make up the definition of computer simulation are as follows:

A actual or hypothetical system can be represented in the form of a computer model, that can be utilized to access information on the system's behaviour.

It is possible to conduct experiments, in which case modifying the input data will result in different output values.

Author in [7] was one of the first authors to notice the potential of spreadsheets in a simulation framework for the purpose of teaching difficult ideas of statistics. He did so in the context of teaching complicated concepts. He utilised spreadsheet simulation as a tool for strategic planning in order to calculate the power of the tests. The results of the evaluation revealed that the students who made use of their applications performed noticeably better than their peers. Mills has conducted a thorough analysis of the published research on computer simulation approaches that may be utilised in any and all areas of statistics to assist students in gaining a better understanding of complex ideas.

In addition to this there are web-based apps created in computer languages such as Java and JavaScript, amongst other languages. In general, students who are just beginning their studies in

statistics do not possess the essential abilities to generate programmes that would mimic the information that can be accessed on these online sites. In addition, in order to use web-based apps in a classroom setting or anyplace else, one must first have access to the internet.

The paper has no requirement of being remain connected to the internet in order to use an Excel demonstration. Students are able to execute these simulations and visualise the ideas that are being taught as long as they have access to a computer that is equipped with Excel.

In this paper, we describe how to use Excel standard data sheets and formulas to perform simulation.

This paper is organized as follows. The next section gives a brief overview of linear regression with an example. The next section introduces the simulation of linear regression using Excel Data Tables. The later sections provide drawing the line of regression with the graph facility of excel and a comparison of two teaching methods, and concluding remarks.

## OVERVIEW OF REGRESSION ANALYSIS

In the field of any research, regression analysis is one of the methods that is utilised the majority of the time. Researchers in the research field are able to assess associations between one independent variable and one dependent variable by using a technique called regression analysis. The dependent variable in any research is often the outcome that we are interested in (for example, sales in marketing research ), whereas the independent variables are the tools at our disposal with which we may accomplish those outcomes (e.g., pricing or advertising).

With the regression analysis,

It is possible to determine whether or not independent variables have a significant association with a dependent variable.

It also indicates the relative strength of different independent variables' impact on a dependent variable. Make predictions and estimations.

### For example:

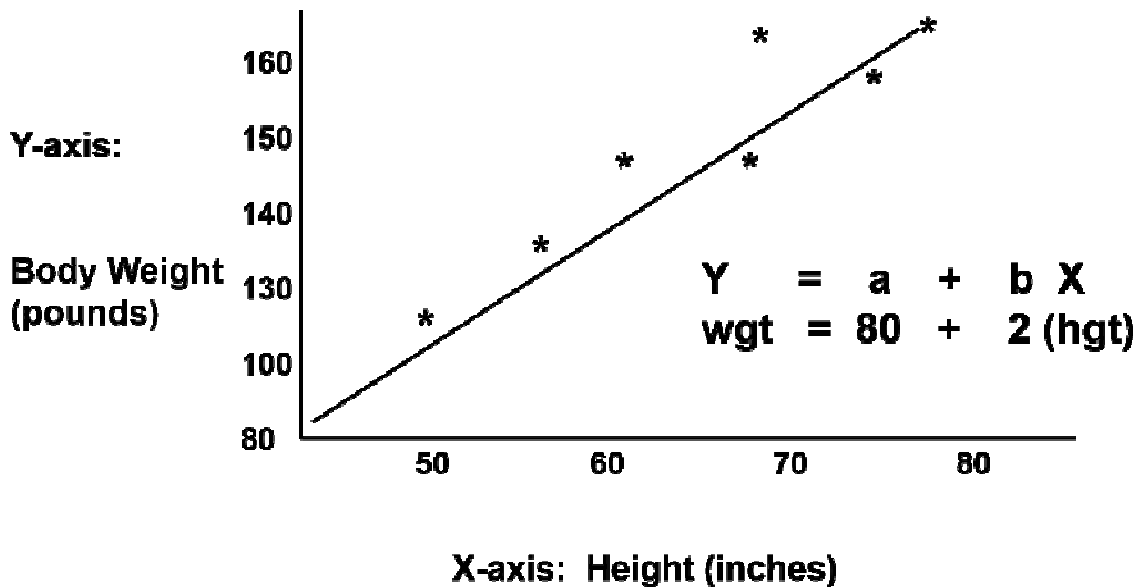
The process of identifying relationships via the use of mathematical models is known as regression analysis. Imagine, for the purpose of discussion, that height was the sole factor that determined body weight. If we were to plot height (the independent or 'predictor' variable) as a function of body weight (the dependent or 'outcome' variable), we might see a very linear relationship, as shown in the diagram below. Height is the independent variable, while body weight

is the dependent or 'outcome' variable.

The equation for a line,  $Y = a + b(x)$ , might also be used to represent this connection. In this equation, 'a' refers to the Y-intercept, while 'b' refers to the slope of the line. If we knew an individual's height, we could use the equation to make a prediction about their weight. In this particular illustration, if an individual had a height of 50 inches, we would estimate that his weight would be:

$$\text{Weight} = 60 + 2 \times (50) = 160 \text{ lbs}$$

In this simple linear regression, we are examining the impact of one independent variable on the dependent variable under the study.



1.1 scatter plot depicting the line of regression equation. [Figure from Lisa Sullivan, <https://sphweb.bumc.bu.edu/>]

### III. SIMULATION USING EXCEL DATA SHEETS

The approach to solve any statistical problem in excel is using the built-in Excel functions. You may design a straightforward formula that will allow you to add, subtract, multiply, or divide the data in your worksheet. Always begin a simple formula with the equal sign (=), then list any mathematical equations having numeric values or cell address, and then list any arithmetic operators like the plus (+), minus (-), asterisk (\*), or forward slash (/) signs.

The common example in this paper is to study the impact of number of hours(Y) studied by the

student for a subject on the marks(X) got by him in the exam. The independent variable (“X” column: the “independent ”variable) is time in weeks, and the outcome of interest (“Y” column, the dependent variable) is total marks obtained in the subject exam.

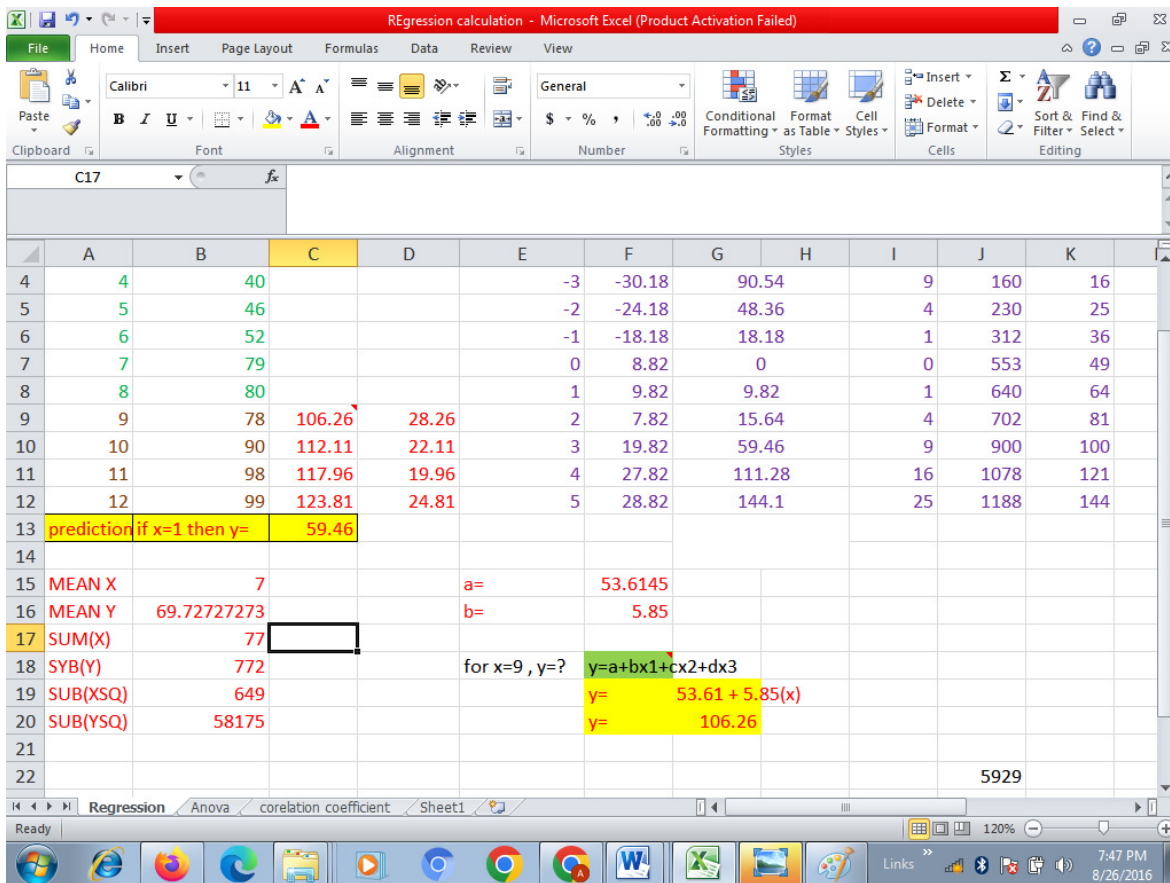
To start with excel- enter the your data in columns as seen below.

study hrs (X)	Marks (Y)	Predicted	Difference / Error	x-xmean	y-ymean	*(x-xmean) (y-ymean)	sq(x-xmean)	SUM (XY)	X <sup>2</sup>	Y <sup>2</sup>
2	50			-5	-20.18	100.9	25	100	4	2500
3	55			-4	-15.18	60.72	16	165	9	3025
4	40			-3	-30.18	90.54	9	160	16	1600
5	46			-2	-24.18	48.36	4	230	25	2116
6	52			-1	-18.18	18.18	1	312	36	2704
7	79			0	8.82	0	0	553	49	6241
8	80			1	9.82	9.82	1	640	64	6400
9	78	106.26	28.26	2	7.82	15.64	4	702	81	6084
10	90	112.11	22.11	3	19.82	59.46	9	900	100	8100
11	98	117.96	19.96	4	27.82	111.28	16	1078	121	9604
12	99	123.81	24.81	5	28.82	144.1	25	1188	144	9801
Find a	53.6145	Regressi on Equation	Y=ax+b	So if x=1 then y=?						
Find b	5.85	Y=53.61(1)+5.85		=	59.46					

Where a and b formula are given in the images.

$$a = \bar{y} - a\bar{x}$$

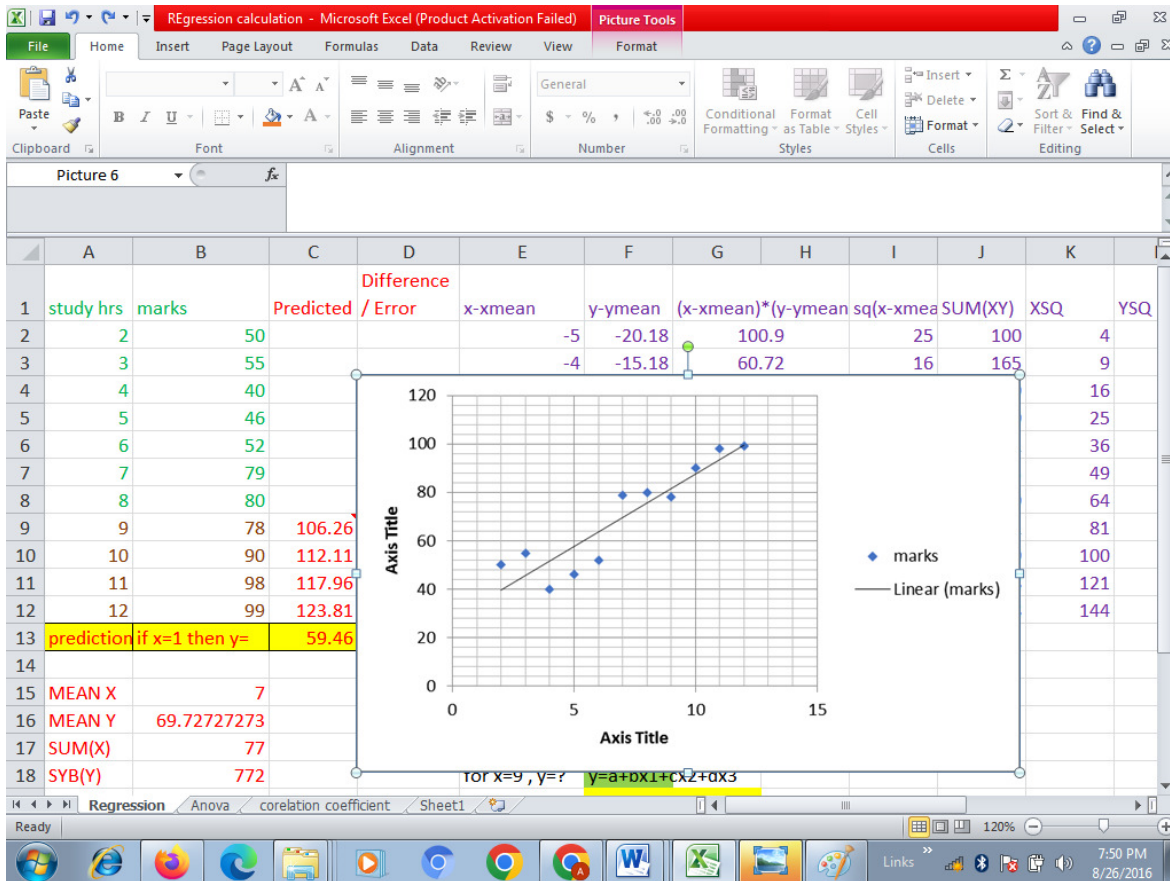
$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$



### Creating Graphs in Excel

Create your regression curve by making a scatter plot. Add the regression line by choosing the “Layout” tab in the “Chart Tools” menu. Then select “Trendline” and choose the “Linear Trendline” option, and the line will appear as shown above. To add the line equation and the R2 value to your figure, under the “Trendline” menu select “More Trendline Options” to see the “Format Trendline” window shown below. Select the boxes next to “Display equation on chart” and

“Display R-squared value on chart” and you are all set.



### Conclusion

Up until very recently, the majority of statisticians who teach the subject have relied on the mathematical derivation of test statistics whose sampling distributions may be determined by analogy. These mathematical derivations need a level of expertise that is beyond that of the majority of students enrolled in basic statistics schools. The computer simulation technique, on the other hand, which is detailed in this work, does not need students to have any prior mathematical knowledge and helps them get a deeper comprehension of the ideas. As a result of advances in technology, we are now in a position where we may employ computer-intensive approaches as an alternative to the conventional methods of educating students in statistics. We have shown how simulation may be used to teach students about topics related to simple linear regression analysis. Excel is a user-friendly spreadsheet programme, and a large number of pupils have access to

it. Based on our first research, it appears that using a computer simulation as an instructional method will help students better comprehend the material. Utilizing Excel's built-in functions and data templates is another way to tackle advanced concepts like regression, which may also be done.

### References :

1. H. Callaert, 2002. "Understanding statistical misconceptions." Proc. Of the 6 th Int. Conference on the Teaching of Statistics. Cape Town, South Africa, July 2002.
2. C. Keeler and K. Steinhorst, "A new approach to learning probability in the first statistics course." Journal of Statistics Education. Vol. 9, no. 3, 2001.
3. D. Tempelaar, "Modeling students learning of introductory statistics." Proc. of the 6 th Int. Conference on the Teaching of Statistics. Cape Town, South Africa, July 2002.
4. P. Cobb, "Where is the mind? Constructivist and sociocultural perspectives on mathematical development." Educational Researcher, vol. 23, pp. 13-20, 1994.
5. S. Tishkovskaya and G. Lancaster, "Statistical Education in the 21<sup>st</sup> century : a review of challenges, teaching innovations and strategies for reform," Journal of Statistical Education, vol. 20, no. 2, 2012,
6. R. C. delMas, J. Garfield, and B. L. Chance, " A model of classroom research in actions: Developing simulation activities to improve students' statistical reasoning." Journal of Statistics Education, vol.7, no.3, 1999,
7. P. Ecklund, "Introduction to excel 2007 data tables and data table exercises," 2009