

Teaching Linear Regression with Python and Excel: Methods, Tools, and Practical Tasks

Neli Kalcheva
Faculty of Computer Sciences and
Automation
Technical University – Varna
Varna, Bulgaria
n_kalcheva@tu-varna.bg

Maya Todorova
Faculty of Computer Sciences and
Automation
Technical University – Varna
Varna, Bulgaria
maya.todorova@tu-varna.bg

Ginka Marinova
Faculty of Computer Sciences and
Automation
Technical University – Varna
Varna, Bulgaria
gmarinova@tu-varna.bg

Firgan Feradov
Faculty of Computer Sciences and
Automation
Technical University – Varna
Varna, Bulgaria
firgan.feradov@tu-varna.bg

Abstract— This article analyses various approaches to teaching linear regression using Python and Excel. The inductive and deductive methods, as well as problem-based learning, are examined. Examples of using the scikit-learn library or Excel functionalities for building linear regression models are described. Practical tasks are presented to support the understanding of regression analysis concepts.

Keywords— Excel, linear regression, Python, teaching methods.

I. INTRODUCTION

Linear regression is a fundamental statistical method widely applied in various fields such as economics, engineering, social sciences, medicine, and data analysis. Understanding this method is essential for students studying data analysis and machine learning.

Various pedagogical methods can be used in teaching linear regression, encouraging analytical thinking and the application of mathematical models. The main approaches include:

a) Deductive method – concepts are presented theoretically and then applied to examples [1]. Deduction starts with general principles and concludes with specifics [2], working in a “top-down” direction, from theory to data that may support or contradict the theory [3].

b) Inductive method – students derive relationships through the analysis of specific examples [1]. Induction involves moving from specifics to general conclusions [2],

following a “bottom-up” direction by using participants' insights to develop broader themes [3].

Researchers suggest combining both methods, as successful learning requires integrating inductive and deductive approaches for better comprehension [4], [5].

c) Problem-based learning (PBL) – students work on real-world tasks, explore data, and formulate hypotheses [6]. The goal of this type of learning is long-term knowledge retention and the development of practical skills for applying knowledge in real situations [7].

Python and Excel are widely used tools for data analysis. Python offers libraries such as NumPy and SciPy for data processing, while Excel provides built-in functions for statistical analysis. Scikit-learn enables the construction of regression models, whereas Excel uses the "Data Analysis ToolPak" for regression analysis.

This article aims to analyse the effectiveness of different methods for teaching linear regression and to propose approaches for integrating them into the learning process. By combining the analytical power of Python with the structured approach of Excel, teaching becomes more interactive and accessible to students with varying levels of preparation.

II. MATERIALS AND METHODS

The deductive method is particularly useful in the early stages of learning, where students need to acquire fundamental knowledge, while the inductive method is

Online ISSN 2256-070X

<https://doi.org/10.17770/etr2025vol3.8556>

© 2025 The Author(s). Published by RTU PRESS.

This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

successfully applied to develop critical thinking when working with real data.

A. Deductive Method

The deductive method starts with presenting the theoretical foundations of linear regression, followed by practical application. The main stages include:

- Explaining the linear regression equation (1);

$$\hat{y} = a + bx \tag{1}$$
- Defining the least squares method;
- Using Python and Excel to apply the model to real data.

This method is beneficial for students who prefer a systematic approach to mathematical concepts.

B. Inductive Method

The inductive method works by analysing real-world examples from which students independently identify patterns. In the context of linear regression, this approach can be implemented through:

- Providing a dataset on sales and advertising;
- Visualizing the data using scatter plots;
- Discussing potential relationships between variables.

The goal of this method is for students to conclude that a linear relationship exists between variables, which can be modelled using a regression equation.

C. Problem-Based Learning

Problem-based learning places students in a situation where they must independently analyse data and build a regression model. For example:

- Providing data on shopping mall attendance based on temperature;
- Students constructing a regression model and analysing the obtained results;
- Discussing additional factors that may influence attendance.

This method fosters analytical skills and encourages independent thinking.

III. RESULTS AND DISCUSSION

A. Building a Linear Regression Model in Microsoft Excel

Teaching linear regression begins with a deductive approach, where students first learn theoretical concepts and then apply them in practical tasks using Excel. This method ensures structured presentation but carries the risk of passive knowledge absorption. To increase student engagement, practical examples are introduced after the theoretical material. These examples are made available in a shared online file, facilitating independent task execution and encouraging the application of acquired knowledge.

Task 1: The table below presents data on the results of a Basic Programming 1 test and a Basic Programming 2 test for six students. Determine whether a relationship exists between the test scores. Create the following table in Excel. Construct a linear regression model. Generate a graph of

the data. Make a prediction for the Basic Programming 2 test result if a student scores 70 on the Basic Programming 1 test.

TABLE I TEST SCORES COMPARISON

Student	Programming Test Scores	
	Basic Programming 1 Test (X)	Basic Programming 2 Test (Y)
A	75	78
B	100	95
C	66	64
D	50	40
E	88	80
F	20	40
G	55	50
H	95	90
I	73	76
J	82	85

To successfully complete this task, students should have basic knowledge of Excel.

The steps for completing the task involve:

- Creating a data table;
- Calculating the mean for each column;
- Computing the coefficients a and b ;
- Formulating the regression equation;
- Predicting the Basic Programming 2 test score for $X = 70$ by substituting it into the regression equation;
- Creating a scatter plot via Insert → Scatter;
- Adding a trendline by selecting Add Trendline;
- Enabling Display Equation on Chart and Display R-squared value on Chart.
- Expected result: learners should obtain the following graph (Fig. 1):

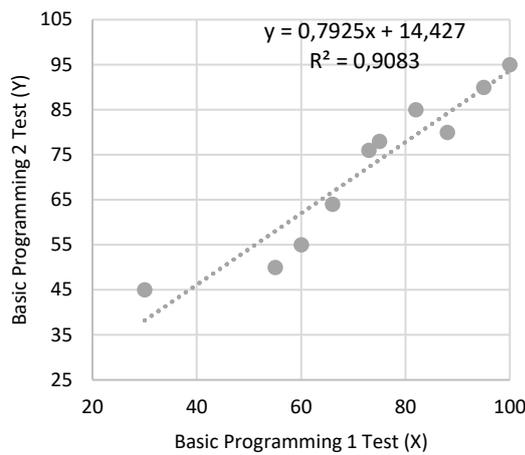


Fig. 1. Linear regression model of the test results.

After constructing the regression equation, students compare the manually calculated values of coefficients a and b with the automatically generated equation in Excel. Any differences (if present) are analysed, with a particular focus on predicting the Basic Programming 2 test score for X = 70.

Task 2: Table II presents data on the average temperature and precipitation levels for each month of the year. The objective is to determine how temperature influences precipitation and to construct a regression model using Excel's Regression tool.

TABLE II MONTHLY AVERAGE TEMPERATURE AND PRECIPITATION

Month	Climatic Indicators	
	Average Temperature (°C)	Precipitation (mm)
January	-2.5	42.2
February	-3	25.4
March	5.5	52.2
April	10.7	58.4
May	19.5	65.2
June	24	68.3
July	28.8	70.2
August	26.5	65.7
September	18.2	60.8
October	9.8	61.2
November	2.3	45.9
December	1.2	46.4

Steps for Completing the Task:

- Create a data table containing temperature and precipitation values.
- Open the "Data Analysis" tool from the "Data" menu in Excel.
- Select "Regression" from the available analysis tools.
- Set Input Y Range (dependent variable) and Input X Range (independent variable).
- Enable "Line Fit Plots" to generate a graph of the linear relationship.

Once the task is completed, the regression analysis results should be examined. An inductive approach is applied in this case, where students first work on the problem and then discuss the expected results, with theoretical concepts introduced as needed to aid in the analysis.

After processing the data, the generated output tables are analysed. The Summary Output provides an overview of the regression results, while the ANOVA table presents key indicators such as degrees of freedom, regression sum of squares, residual sum of squares, null hypothesis (H₀), alternative hypothesis (H₁), t-statistic, confidence intervals, intercept, and regression coefficients. The Residual Output contains residual values, showing the differences between predicted and actual values.

Finally, the generated graph illustrating the linear dependence between the dependent and independent variable is examined.

B. Building a Linear Regression Model in Python

In a follow-up lesson, problem-based learning (PBL) can be applied, encouraging students to discover solutions through analysis. A shared file containing predefined tasks is provided to students in an online learning environment.

Task 3: Table III presents real estate price data in euros based on property area in square meters.

TABLE III REAL ESTATE PRICES BASED ON PROPERTY AREA

Property	Real Estate Characteristics	
	Area (m ²)	Price (€)
Property 1	70	129,000
Property 2	45	80,000
Property 3	90	165,000
Property 4	100	190,000
Property 5	55	105,000
Property 6	85	150,000
Property 7	60	110,000
Property 8	120	225,000
Property 9	95	175,000

Property	Real Estate Characteristics	
	Area (m ²)	Price (€)
Property 10	50	95,000
Property 11	110	205,000
Property 12	130	250,000
Property 13	75	135,000
Property 14	65	122,000
Property 15	80	140,000

Based on the data in Table III, construct a linear regression model in Python. Calculate the coefficient of determination (R^2). Determine the regression parameters—slope and intercept. Plot the regression line and predict the price of a property with an area of 75 square meters.

Task 4: Table IV provides data on the number of visitors to a shopping centre based on air temperature over several days. The aim is to examine the relationship between temperature and customer visits. Based on the data, build a linear regression model and use it to predict the number of visitors when the temperature is 14 °C.

TABLE IV CUSTOMER VISITS BASED ON TEMPERATURE

Day of the Month	Table Column Head	
	Temperature (°C) (X)	Number of Customers (Y)
1	6	52
2	8	54
3	10	66
4	8	60
5	22	98
6	18	80
7	15	72
8	12	65
9	25	110
10	5	48

Task 5: Table V presents data on the time customers spend in a store and the amount spent. Build a linear regression model in Python. Calculate the coefficient of determination. Extract the regression parameters—slope and intercept. Plot the regression line. Predict the spending for a customer who spends 12 minutes in the store.

TABLE V TIME SPENT IN STORE VS. AMOUNT SPENT

Record	Customer Purchase Data	
	Time Spent in Store (minutes)	Amount Spent (€)
1	5	8
2	7	10
3	10	13
4	15	18
5	20	22
6	25	25
7	30	30
8	35	35
9	40	42
10	45	48

To successfully complete these tasks, students should have basic knowledge of Python programming. They receive instructions for independent exploration of the scikit-learn library [8] and its key linear regression classes, while the instructor facilitates the process by guiding discussions and asking targeted questions.

In this lesson, the instructor acts as a facilitator rather than a direct source of knowledge, aiming to develop students' lifelong learning skills [6].

The steps for completing the task require:

- Importing the scikit-learn library
- `from sklearn.linear_model import LinearRegression`
- Creating a model of type LinearRegression
`model = LinearRegression()`
- Training the model using the fit() method, where X contains the independent variables and y is the dependent variable:
- `model.fit(X, y)`
Predicting new values with the predict() method:
`predictions = model.predict(X_new)`
- Evaluating the model's quality using the score() method
`score = model.score(X, y)`

After completing each task, students analyse the obtained coefficients, the regression line, and the coefficient of determination values. At the end of the lesson, a discussion is held on the applications of linear regression in the real world, emphasizing the importance of this method for forecasting and data analysis.

IV. CONCLUSION

The use of Python and Excel in teaching linear regression provides an effective approach for introducing students to regression analysis. Excel is a convenient tool for quick analysis and visualization, while Python offers

greater flexibility and automation. By combining the inductive and deductive methods, along with problem-based learning, the teaching process becomes more effective and engaging. Future lessons may expand to include topics such as multiple, polynomial, and logistic regression, with Python and Excel continuing to support deeper understanding and real-world data analysis.

ACKNOWLEDGMENTS

This work was supported by the Technical University of Varna.

REFERENCES

- [1] L. K. Soiferman, "Compare and contrast inductive and deductive research approaches," Online Submission, 2010.
- [2] W. M. K. Trochim, *The Research Methods Knowledge Base*, 2nd ed. Atomic Dog Publishing, 2006.
- [3] J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications, 2007.
- [4] A. Onwuegbuzie and N. Leech, "Taking the 'Q' out of research: Teaching research methodology courses without the divide between quantitative and qualitative paradigms," *Quantity and Quality*, vol. 39, pp. 267-296, 2005.
- [5] M. Tiittanen, "The deductive and inductive instructional approach controversy in Eswatini context," *Global Journal of Foreign Language Teaching*, vol. 10, no. 2, pp. 139-158, 2020.
- [6] C. E. Hmelo-Silver, "Problem-based learning: What and how do students learn?" *Educational Psychology Review*, vol. 16, pp. 235-266, 2004.
- [7] F. Dochy, M. Segers, P. Van den Bossche, and D. Gijbels, "Effects of problem-based learning: A meta-analysis," *Learning and Instruction*, vol. 13, no. 5, pp. 533-568, 2003.
- [8] Scikit-learn Developers, "Scikit-learn Documentation", [Online]. Available: <https://scikit-learn.org>. [Accessed: Feb. 24, 2025].