

# Linear regression methods in mathematical statistics

Asaad Naser Hussein Mzedawee

[asaad.nasir@qu.edu.iq](mailto:asaad.nasir@qu.edu.iq)

University Of AL-Qadisiya-College of Administration and Economics

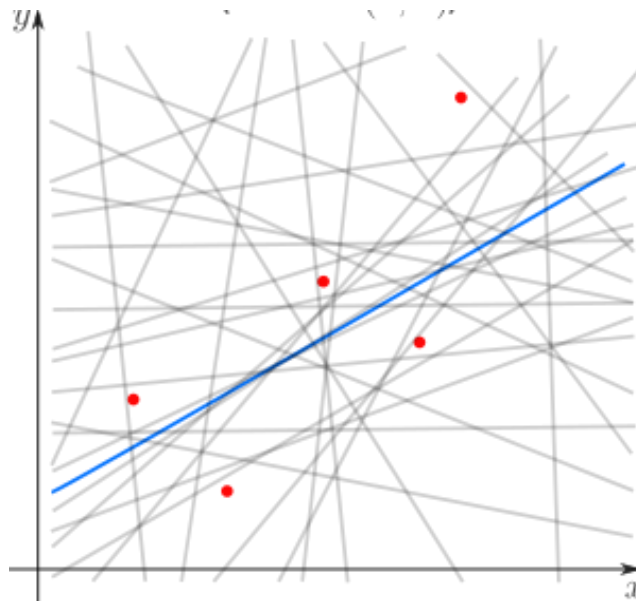
**Abstract:** This article discusses the main methods of linear regression in mathematical statistics. Linear regression is one of the fundamental tools of data analysis and is widely used in various fields ranging from economics and finance to medicine and sociology. The article describes in detail the basic concepts of linear regression, including the model, parameter estimates, hypothesis testing, model selection and interpretation of the results. Practical examples of using linear regression methods on real data are also presented.

**Keywords:** linear regression, mathematical statistics, model, parameters, hypotheses, model selection, interpretation of results.

## Introduction

Regression is a way to select from a family of functions the one that minimizes the loss function. The latter characterizes how much the test function deviates from the values at the specified points. If the points are obtained in an experiment, they inevitably contain measurement error, noise, so it is more reasonable to require that the function conveys the general trend, rather than accurately passing through all points. In a sense, regression is an "interpolating approximation": we want to draw the curve as close to the points as possible and still keep it as simple as possible in order to capture the overall trend. For the balance between these conflicting desires, the loss function (in English literature, "loss function" or "cost function") is responsible.

Linear regression is a powerful data analysis technique that allows you to predict the values of unknown data based on already existing related known data. The essence of this method is to mathematically model the dependent variable that we want to predict, based on known or independent variables, using a linear equation. An inherently simple linear regression method attempts to plot a line graph between two data variables,  $x$  and  $y$ . As an independent variable,  $x$  is plotted along the horizontal axis. Independent variables are also called independent variables or predictor variables. The dependent variable  $y$  is plotted on the vertical axis.  $Y$  values can also be called response variables or predicted variables.



**Figure1. Simple Linear Regression with the image  $(1, x)$**

One of the main tasks of linear regression is to find the best fit of a linear model to the available data. The least squares method is the most common approach to solving this problem. It provides an optimal fit to the model by minimizing the sum of squared distances between predicted values and actual data. Thus, linear regression allows us to create a linear model that best describes the relationship between known and unknown data.

To understand how linear regression works, it is important to understand the concept of a linear equation. A linear equation is a mathematical expression that describes the linear relationship between variables. In the case of univariate linear regression, we have two variables: the dependent variable (the parameter that we are trying to predict) and one independent variable (the known value that we use to predict). The linear equation in this case has the following form:  $Y = a + bX$ , where  $Y$  is the dependent variable,

X is the independent variable, and a and b are the coefficients that we need to determine in order to achieve the best fit of the model. Determining the optimal coefficients a and b is the goal of the linear regression method itself. To do this, a special procedure called model parameter estimation is used. The procedure for estimating parameters is based on analyzing the available data and minimizing discrepancies between predicted values and actual data. From this we can conclude that linear regression is also a statistical method based on probability analysis.[9]

In addition to predicting unknown values, linear regression provides other useful information. For example, it allows you to evaluate the relationship between variables. The Pearson correlation coefficient shows the degree of linear relationship between variables and can be useful for studying the relationship between different factors. Also, linear regression allows you to evaluate the significance of each parameter of the model and determine which variables have the greatest impact on the result. It is important to note that linear regression has its limitations and assumes several prerequisites that must be met in order to obtain reliable results. One of the main prerequisites is the linearity of the relationship between variables. If the relationship between variables is not linear, then linear regression may produce inaccurate or incomplete results. It is also important to check for outliers, heteroscedasticity, and autocorrelation of data, as these factors may distort the analysis results. Thus, linear regression is a powerful data analysis technique that allows you to predict the values of unknown data based on already existing related data. It is based on constructing a linear equation and determining the optimal coefficients, and also allows you to evaluate the relationship between variables. However, it is important to consider the limitations and prerequisites of this method in order to obtain reliable analysis results.[10]

The relevance of the study of linear regression methods in mathematical statistics lies in their wide application in various fields, ranging from economics and finance to medicine and sociology. Linear regression is one of the main tools for analyzing the relationship between a dependent variable and explanatory variables. The main goal of the study is to develop and improve linear regression methods to make them more accurate, efficient, and reliable. The research is aimed at studying various aspects and problems related to the use of linear regression, such as multicollinearity, outliers, heteroskedasticity, etc. The aim of the research is also to determine the optimal methods for estimating linear regression parameters, such as the least squares method, maximum likelihood method, and others. In addition, the research aims to find optimal methods for selecting variables to improve the predictive power of linear regression models.[3]

The results of the research on linear regression methods in mathematical statistics will help improve statistical data analysis, forecasting, and decision-making based on statistical models. They will be useful for researchers, statisticians, economists, and other professionals who work with data and analyze the dependence of variables.[15]

Linear regression models are relatively simple and provide an easily interpreted mathematical formula for making predictions. Linear regression is a recognized statistical technique that is easily applied to software and computing. Companies use it to reliably and predictably transform raw data into business intelligence and useful analytics. Scientists in many fields, including biology and behavioral, environmental, and social sciences, use linear regression to perform preliminary data analysis and predict future trends. Many data science methods, such as machine learning and artificial intelligence, use linear regression to solve complex problems.[12]

Linear regression methods in mathematical statistics are one of the fundamental tools for analyzing and predicting data. They are used to evaluate the relationship between dependent and independent variables, as well as to build a mathematical model that can be used to predict future values. First of all, linear regression assumes that the relationship between variables can be described by a linear function. A linear regression model can be expressed as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Where Y is the dependent variable,  $X_i$  is the independent variables,  $\beta_i$  is the regression coefficients, and  $\varepsilon$  is the residuals. [6]

The main task of linear regression methods is to estimate the regression coefficients  $\beta_i$ . For this purpose, the least squares method (OLS) is used. The OLS minimizes the sum of squares of the difference between the real and predicted values of the model.[4]

Example of solving a linear regression problem: Let's say we have data on the income (Y) and the number of years of education (X) of 100 people. To estimate the relationship between these variables, we use linear regression. After applying the least squares method, we get the following model:

$$Y = 2000 + 500X$$

This means that each year of education corresponds to an increase in income by 500 units. Thus, linear regression methods in mathematical statistics allow us to analyze and predict the relationship between variables, which has a wide range of applications in various fields, including economics, sociology, medicine, etc. [11]

Linear regression methods are an important tool in mathematical statistics and are widely used in various spheres of life. Linear regression is a statistical model that allows you to investigate the relationship between a dependent variable and one or more

independent variables. It is based on the assumption of a linear relationship between variables and is an equation of a straight line that best describes the observed relationship.[2]

Initially, linear regression methods were used in statistics to study economic and social phenomena. However, over time, they have found application in other areas as well. In biology, for example, linear regression is used to study the relationship between two biological variables, such as an organism's weight and height. In medicine, it is used to analyze the relationship between physiological indicators and diseases.[9]

One of the main advantages of linear regression is the ability to predict the values of the dependent variable based on the known values of the independent variables. This allows you to make informed decisions and predict future results. In addition, linear regression allows you to determine the degree of influence of each independent variable on the dependent variable, which makes it indispensable in data analysis and decision-making. However, keep in mind that linear regression is based on the assumption of linear dependence of variables, which may not be applicable in some cases. In addition, it is sensitive to the presence of outliers in the data, which can distort the results of the analysis. Therefore, before applying linear regression methods, it is necessary to carefully check the compliance of the model with the collected data and, if necessary, use additional analysis methods. [15]

In general, linear regression methods are one of the main tools of mathematical statistics and are essential in life for analyzing data and making informed decisions in various fields.

Linear regression methods are one of the most common tools in mathematical statistics for analyzing the relationship between a dependent variable and one or more independent variables. They belong to the class of supervised learning methods, which means that there are pre-known pairs of training data that include the values of the dependent and independent variables. Before applying linear regression methods, you need to make sure that several assumptions are met. First, a linear relationship between the dependent and independent variables is required. Second, the regression residuals (the difference between observed and predicted values) are expected to have a normal distribution. Finally, multicollinearity problems are eliminated when independent variables strongly correlate with each other.[1]

Examples of calculations using linear regression methods in real life are widely used. One popular example is real estate price forecasting. The dependent variable is the price, and independent variables can be the area of the house, the number of rooms, the year of construction, the distance to the city center, and other factors. Using linear regression methods, you can estimate the impact of each of these factors on the price and build a model that will predict the price of real estate based on the available data. Another example is market trend analysis and sales forecasting. The dependent variable is sales volume, and independent variables can be the product price, advertising budget, seasonal factors, and other factors. Linear regression methods allow you to estimate the impact of each of these factors on sales and create a model for predicting future sales.[5]

In both cases, linear regression methods allow you to analyze the influence of various factors on the dependent variable and predict its values based on the available data. This allows you to make informed decisions and plan further actions in real life.[7]

## Reference

1. Gafarov O. S. Methods of linear regression in mathematical statistics. - M.: Publishing House of the Moscow Institute of Economics, 2012.
2. Demidovich B. P., Maron I. A. Numerical methods for solving systems of linear algebraic equations. - M.: Nauka, 2006.
3. Zeldner G. Linear regression in Economics. - M.: Higher School of Economics, 2006.
4. Kolmogorov A. N., Chistyakov V. P. Handbook of Mathematical Statistics. - M.: Nauka Publishing House, 1974.
5. Makarov I. S., Golubtsova E. A. Linear regression. - M.: Yurait, 2016.
6. Marchenko T. V., Parfenov D. A. Linear regression: systemic and structural approaches. - M.: KNORUS, 2013.
7. Nagornova E. Yu., Soloshenko T. I. Linear models in data analysis. - M.: Finance and Statistics, 2018.
8. Pilipenko T. A. Linear models in economics. - M.: Publishing house "Dashkov and K", 2010.
9. Seleznev S. V., Vaganova E. I. Lectures on mathematical statistics. - M.: Moscow University Press, 2002.
10. Stewart J. Linear regression and modeling. - M.: DMK Press, 2003.
11. Tyrtshnikov E. E. Solving linear problems by QR decomposition methods. - M.: FIZMATLIT, 2005.
12. Fu Shu-Hui. Introduction to statistical regression. - M.: Higher School of Economics, 2012.
13. Chechurin A. Yu. Linear regression: theory and applications. - M.: Lan, 2008.
14. Shirvani A. H., Moser D. S. Linear and nonlinear regression. - M.: Williams, 2005.
15. Yatsenko V. S. Linear regression. - M.: Finance and Statistics, 2002.