Multidiszciplináris Tudományok, 14. kötet, 3. szám (2024) pp. 152-159. **https://doi.org/10.35925/j.multi.2024.3.13**

APPLICATIONS OF LINEAR REGRESSION IN DIFFERENT DISCIPLINES

József Túri
associate professor, Institute of Mathematics, University of Miskolc
3515 Miskolc, Miskolc-Egyetemváros, e-mail: jozsef.turi@uni-miskolc.hu

Abstract
In this paper we examine the linear regression and its application in different disciplines. Linear regression is used in economics, engineering sciences, natural sciences and even in medicine. The applicability of linear regression lies in the fact that many phenomena in the a forementioned fields of science can be described with the help of linear regression. Although there are some phenomena that cannot be described by linear regression, most non-linear regressions can be reduced to linear regression by a simple transformation. Another advantage of linear regression is that it is very easy to apply and perform calculations with the help of a computer.

Keywords**:** *simple linear regression, multivariate linear regression, applications of linear regression, computer implementation of linear regression.*

1. Introduction

The linear regression is a widely used method in several sciences (for example in economics, engineering sciences, natural sciences, and even in medicine) The reason for this is that the method is simple to apply, and its computer implementation is also easy to solve. Furthermore, the use of linear regression is facilitated by the fact that there is not only a univariate case, but also a multivariate one. Furthermore, even if linearity is not met, in many cases we can trace the problem back to linear regression. The basis of linear regression is that in practice very often the plotted points fall approximately on a straight line. This is exactly the essence and basis of linear regression, that the points lie approximately on a straight line. However, we emphasize that it is usually only approximate. And we look for a straight line and fit it to the points that best fits them. However, we can still solve the problem relatively easily.

Linear regression has an extensive literature. Zou, Tuncali and Silverman gave a good summary of the correlation and simple linear regression (Zou et al., 2003). Poole and O'Farrell show the assumptions of the linear regression model (Poole et al., 1971). Naseem, Togneri and Bennamoun made the linear regression for face recognition (Naseem et al., 2010). Raposo shows the evaluation of analytical calibration based on least-squares linear regression for instrumental techniques (Raposo, 2016). Ludbroock examines the issues of regression selection (Ludbroock, 2010). Groeneboom and Hendrickx examines the place and position of regression within statistics (Groeneboom et al., 2018). Baždarić, Šverko, Salarić, Martinović and Lucijanić provide a detailed overview of linear regression (Baždarić et al., 2021). Aalen show A linear regression model for the analysis of life times (Aalen, 1989). Nie, Chu, Liu, Cole, Vexler and Schisterman give a method for linear regression with an independent variable subject to a detection limit (Nie et al., 2010). Vovk, Nouretdinov and Gammerman examine on line

predictive linear regression (Vovk et al., 2009). Yang, Liu, Tsoka and Papageorgiou give the mathematical programming for piecewise linear regression (Yang et al., 2016). Austin and Steyerberg examined the number of subjects required for linear regression per variable (Austin et al., 2015). Zhang, Khalili and Asgharian give a post-model-selection inference in linear regression models which an integrated review (Zhang et al., 2022). Ferraro, Coppi, González and Colubi show a linear regression model for imprecise response (Ferraro et al., 2010). Groß's comprehensive monograph is a great help for those interested in linear regression and its users (Groß, 2003). Yuan and Yang discuss when and where linear regression can be used (Yuan et al., 2005). Mumtaz and Petrillo present the application of linear regression to evaluate the impact of green supply chain management on industrial organizational performance (Mumtaz et al., 2018). Krämer and Sonnberger: Krämer test the linear regression (Krämer et al., 1986). Filzmoser and Nordhausen give an overviewt the robust linear regression for high-dimensional data (Filzmoser et al., 2021). The article also deals with multivariate regression. Multivariate linear regression also has an extensive literature. Tranmer, Murphy, Elliot and Pampaka give a description for Multiple Linear Regression (Tranmer et al., 2020). Nimon and Oswald gives a specific approach, which is also reflected in the title of the article: Understanding the Results of Multiple Linear Regression: Beyond Standardized Regression Coefficients (Nimon et al., 2013). Sinnakaudan, Ghani, Ahmad and Zakaria give a very well-written works: showing the multiple linear regression model for total bed material load prediction (Sinnakaudan et al., 2006). Wang, Liying, Wu, and Guan examine the multiple linear regression modelling for compositional data (Wang et al., 2013). Mahmoud examine the phase I analysis of multiple linear regression profiles (Mahmoud, 2008). Etemadi and Khashei also investigate the multiple linear regression with remarkable results (Etemadi et al., 2021).

2. The univariate linear regression

The univariate linear regression is very common and can be used in many different sciences. The disadvantage is that in some cases the method is not complex enough. Nevertheless, it can be used many times.

The basic scheme is as follows. We take the basic points: X_1, X_2,…,$X_{(k-1)}$, X_k and Y_1, Y_2,…,$Y_{(k-1)}$,Y_k (these measurement points can be said to be random variables). It is worth including the obtained data in a table (see below, Table 1.).

Table 1.

Y=aX+b	1.	2.	…	(k-1).	k.
X	X_1	X_2	…	$X_{(k-1)}$	X_k
Y	Y_1	Y_2	…	$Y_{(k-1)}$	Y_k

In Figure 1. below illustrates when we can apply linear regression. Linear regression can be used if the points in question fall approximately on a straight line, which can be seen in the first figure.

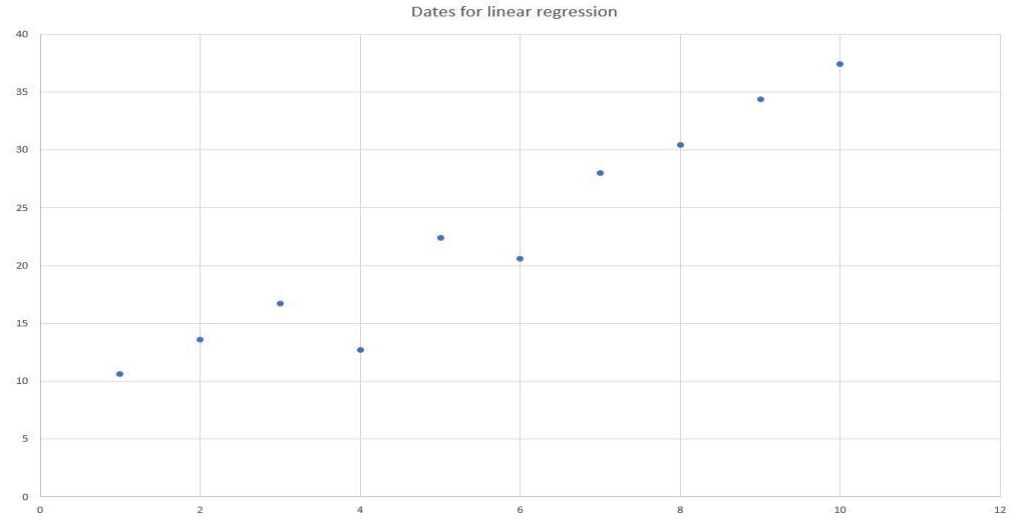

Figure 1.

If the points lie approximately on a straight line, we can apply linear regression (see Figure 2). When regression is applied, the equation of the straight line is analytically obtained. This means that for the equation of the line Y=aX+b, at the end of the regression, we get the constants a and b. This is important because this way we can also make predictions for those values a, b that are not in the set at the base point. So we get the equation of the line from the information extracted from the base points and the points belonging to them.

Formally writing down the equations from which we obtain the parameters a and b:

$$Y_1=aX_1+b,$$

$$Y_2=aX_2+b,$$

$$Y_3=aX_3+b,$$

$$\ldots\ldots.$$

$$\ldots\ldots..$$

$$Y_{(k-1)}=aX_{(k-1)}+b,$$

$$Y_k=aX_k+b.$$

By plotting the equation of the resulting line, we can see that the points fit approximately well on its parallel (see Figure 2.).

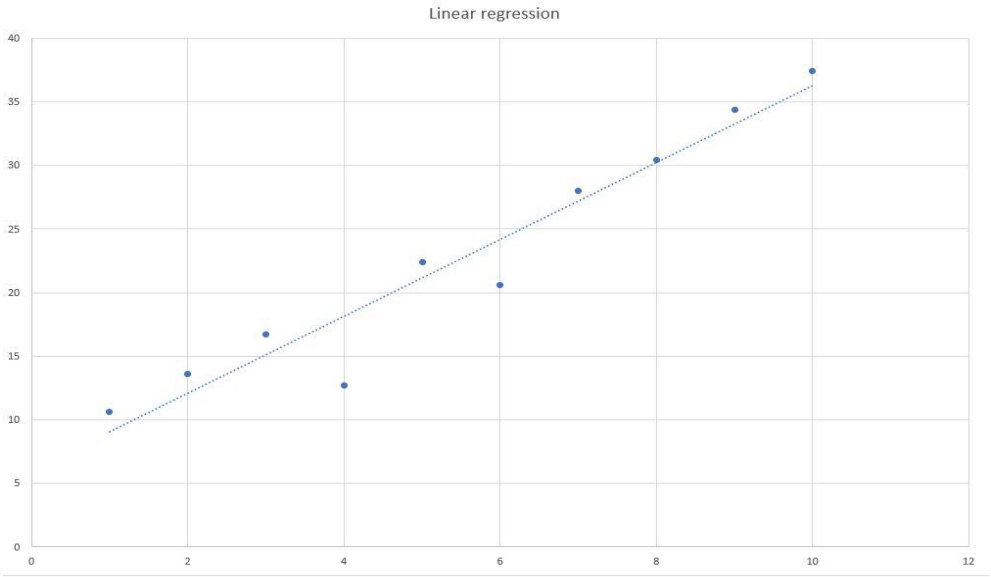

Figure 2.

The concrete implementation of linear regression is, of course, done with the help of a computer. Excellent software is available to implement different statistical methods. Examples include SPSS, SAS, R, etc. Figure 3 shows that, for example, linear regression can be performed very simply with the help of SPSS.

For example, linear regression can be used in countless ways in economics. For example, when we examine the revenues of different companies depending on the years, the obtained pairs of points (year, revenue) in many cases fall approximately on a straight line (of course, we get a straight line with different parameters). If the GDP of a country increases (decreases), then consumption in that country also increases (decreases) and usually the relationship between the two things is linear. Thus, linear regression can also be used here: the (GDP, consumption) coordinates are located approximately on a straight line.

Of course, linear regression can be used in many cases in the technical sciences as well. For example, in the chemical sciences, there is an approximately linear relationship between pressure and temperature, so linear regression can be used. But of course the same can be said about mechanical engineering, electrical engineering, logistics and other technical disciplines as well, i.e. linear regression can be applied to several phenomena.

Of course, medicine also uses the method of linear regression. For example, the use of dosages of many drugs in healing shows an approximately linear relationship (at least within certain limits).

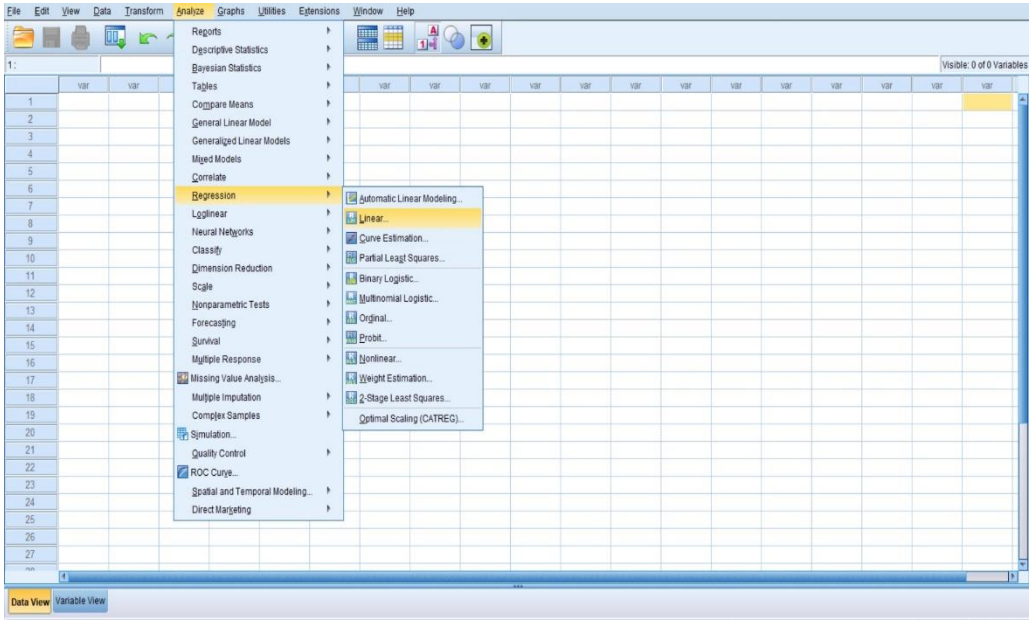

Figure 3.

Of course, linear regression can be used in most scientific fields.

3. The multivariate linear regression

Multivariate linear regression differs from univariate linear regression in that it can take several properties into account. Considering the technical sciences, for example, we can take several factors into account here, and if they fall approximately on a hyperplane, then multivariate linear regression can be used. Of course, univariate linear regression is a special case of multivariate linear regression. Using multivariate linear regression differs from using univariate only in that several independent variables must be substituted into the formula. However, the mathematical background of multivariate linear regression is shockingly more complex than that of univariate linear regression.

The general form of multivariate linear regression is formalized as follows:

$$Y_1=a_rX_{1,1}+a_{r-1}X_{2,1}+\ldots+a_1X_{r,1}+a_{0,}$$

$$Y_2=a_rX_{1,2}+a_{r-1}X_{2,2}+\ldots+a_1X_{r,2}+a_{0,}$$

$$Y_3=a_rX_{1,3}+a_{r-1}X_{r-2,3}+\ldots+a_1X_{r,3}+a_{0,}$$

.....

.....

.....

$$Y_{k-1}=a_rX_{1,k-1}+a_{r-1}X_{2,k-1}+\ldots+a_1X_{r,k-1}+a_{0,}$$

$$Y_k=a_rX_{1,k}+a_{r-1}X_{2,k}+\ldots+a_1X_{r,k}+a_{0.}$$

The formulas can also be written in the following form:

$$Y=AX+b,$$

where $Y=(Y_1,Y_2,\ldots,Y_k)^T$ and $A=(a_0,a_1,a_3,\ldots,a_r)^T$ and X is the corresponding matrix with $X_{i,j}$ (i=1,2,…,k; j=1.2,…,k) elements. If the determinant of X is not zero, then the above equation can always be solved.

Multivariate linear regression can still be plotted in the case of two variables (see Figure 4.). The figure shows that the points shown in the figure are located approximately on a plane. Of course, in the case of three variables, the illustration is no longer possible.

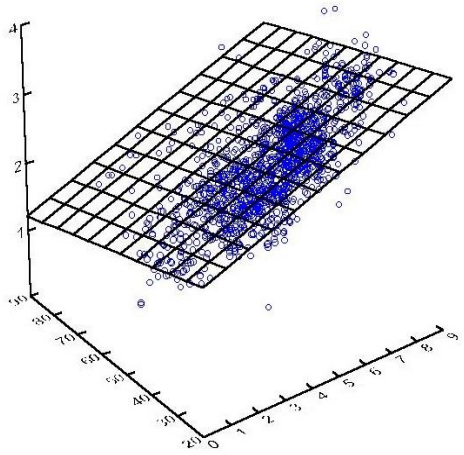

Figure 4.

The multivariate linear is even more widely used than its univariate version.

Consider an economic example: if, for example, we take a sample, i.e. we ask several people, what is their income (X_1), how many square meters do they live in (X_2), how much do they spend on food (X_3), how much do they spend on clothing (X_4), how much do they spend on culture (X_5) and what is your income (Y). Here X_1, X_2, X_3, X_4, X_5 are the independent variables, while Y is the dependent variable. From the data, we can determine $A=(a_{i,j})$, so for any X_1, X_2,…,X_r we can determine Y, i.e. the individual's income.

Of course, multivariate linear regression is also widely used in engineering: if, for example, we consider the concentration of a metal alloy (X), the temperature of the metal (X) and look for the tensile strength (Y), linear regression can usually be used.

Of course, we can give countless more examples of scientific fields where multivariate linear regression can be applied

4. Summary

In this paper, the linear regression and its application is presented, including the computer implementation. In the first part of the article, we deal with linear regression, illustrating the method with examples. Then we turn to multivariate linear regression, where examples (economics and engineering) also make the method more understandable.

References

[1] Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). *Multiple Linear Regression* (2nd ed.); Cathie Marsh Institute Working Paper 2020-01. https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/2020-1-multiple-linear-regression.pdf

[2] Zou, H. K., Tuncali, K., Silverman, S. G. (2003). Correlation and simple linear regression. *Radiology*, 227(3). **https://doi.org/10.1148/radiol.2273011499**

[3] Nimon, K. F., Oswald, F. L. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods, 16*(4), 650–674. **https://doi.org/10.1177/1094428113493929**

[4] Poole, M. A., & O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, 52, 145–158. **https://doi.org/10.2307/621706**

[5] Naseem, I., Togneri, R., & Bennamoun, M. (2010). Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(11), 2106–2112. **https://doi.org/10.1109/TPAMI.2010.128**

[6] Raposo, F. (2016). Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review. *TrAC Trends in Analytical Chemistry*, *77*, 167–185. **https://doi.org/10.1016/j.trac.2015.12.006**

[7] Ludbrook, J. (2010). Linear regression analysis for comparing two measurers or methods of measurement: But which regression? *Clinical and Experimental Pharmacology and Physiology*, *37*, 692–699. **https://doi.org/10.1111/j.1440-1681.2010.05376.x**

[8] Groeneboom, P., & Hendrickx, K. (2018). Current Status Linear Regression. *The Annals of Statistics*, *46*(4), 1415–1444. https://www.jstor.org/stable/26542832, **https://doi.org/10.1214/17-AOS1589**

[9] Baždarić, K., Šverko, D., Salarić, I., Martinović, A., & Lucijanić, M. (2021). The ABC of linear regression analysis: What every author and editor should know. *European Science Editing, 47*. **https://doi.org/10.3897/ese.2021.e63780**

[10] Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Med., 8*, 907–925. **https://doi.org/10.1002/sim.4780080803**

[11] Nie, L., Chu, H., Liu, C., Cole, S. R., Vexler, A., & Schisterman, E. F. (2010). Linear regression with an independent variable subject to a detection limit. *Epidemiology, 21*(4), S17–S24. **https://doi.org/10.1097/EDE.0b013e3181ce97d8**

[12] Vovk, V., Nouretdinov, I., & Gammerman, A. (2009). On-line predictive linear regression. *The Annals of Statistics*, *37*(3), 1566–1590. http://www.jstor.org/stable/30243678, **https://doi.org/10.1214/08-AOS622**

[13] Yang, L., Liu, S., Tsoka, S., Papageorgiou, L. G. (2016). Mathematical programming for piecewise linear regression analysis. *Expert Systems with Applications*, *44*, 156–167. **https://doi.org/10.1016/j.eswa.2015.08.034**

[14] Sinnakaudan S. K., Ghani A. Ab., Ahmad M. S., Zakaria N. A. (2006). Multiple linear regression model for total bed material load prediction. *Journal of Hydraulic Engineering*, *132*(5), **https://doi.org/10.1061/(ASCE)0733-9429(2006)132:5(521)**

[15] Austin, P. C. Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology*, *68*(6), 627–636. **https://doi.org/10.1016/j.jclinepi.2014.12.014**

[16] Zhang, D., Khalili, A., Asgharian, M. (2022). Post-model-selection inference in linear regression models: An integrated review. *Statist. Surv*., *16*, 86–136. **https://doi.org/10.1214/22-SS135**

[17] Ferraro, M. B., Coppi, R., González Rodríguez, G., Colubi, A. (2010). A linear regression model for imprecise response. *International Journal of Approximate Reasoning*, *51*(7), 759–770. **https://doi.org/10.1016/j.ijar.2010.04.003**

[18] Groß, J. (2003). *Linear Regression*. Springer, Lecture Notes in Statistics (LNS), 175. **https://doi.org/10.1007/978-3-642-55864-1_2**

[19] Yuan, Z., & Yang, Y. (2005). Combining linear regression models: When and How? *Journal of the American Statistical Association*, *100*(472), 1202–1214. **https://doi.org/10.1198/016214505000000088**

[20] Berndt, E. R., & Savin, N. E. (1977). Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica*, *45*(5), 1263–1277. **https://doi.org/10.2307/1914072**

[21] Wang, H., Shangguan, L., Wu, J., Guan, R. (2013). Multiple linear regression modeling for compositional data. *Neurocomputing*, 122, 490–500. **https://doi.org/10.1016/j.neucom.2013.05.025**

[22] Mahmoud, M. A. (2008). Phase I Analysis of multiple linear regression profiles. *Communications in Statistics - Simulation and Computation*, *37*(10), 2106–2130. **https://doi.org/10.1080/03610910802305017**

[23] Mumtaz, U., Ali, Y., Petrillo, A. (2018). A linear regression approach to evaluate the green supply chain management impact on industrial organizational performance. *Science of The Total Environment, 624*, 162–169. **https://doi.org/10.1016/j.scitotenv.2017.12.089**

[24] Krämer, W, Sonnberger, H. (1986). *The Linear Regression Model Under Test*. Physica-Verlag Heidelberg, Heidelberg. **https://doi.org/10.1007/978-3-642-95876-2**

[25] Etemadi, S., Khashei, M. (2021). Etemadi multiple linear regression. *Measurement*, 186. **https://doi.org/10.1016/j.measurement.2021.110080**

[26] Filzmoser P, Nordhausen K. (2021). Robust linear regression for high-dimensional data: An overview. *WIREs Comput Stat*., 13:e 1524. **https://doi.org/10.1002/wics.1524**