International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 13, Issue 5 (Sep.-Oct. 2024), PP. 23-27

Modelling of prime numbers with regression analysis

Şenol Çelik*
[1]*Biometry Genetics Unit, Department of Animal Science, Agricultural Faculty, Bingöl University, Bingöl, Türkiye*
**** Correspondence:*** *senolcelik@bingol.edu.tr*

Abstract
In this study, it is aimed to analysis prime numbers, which have an important place in mathematics, with regression methods. The 169 prime numbers from 2 to 1009 are considered as dependent variables, and the numbers from 1 to 169 corresponding to them are considered as independent variables. Four regression methods, namely simple linear, quadratic, cubic and exponential regression methods, were analyzed comparatively. The coefficients of determination (R^2) of simple linear, quadratic, cubic and exponential regression models were 0.9961, 0.99963, 0.99976 and 0.7719, respectively. The parameter coefficients of all models were statistically significant ($p<0.001$). The cubic regression method with the highest R^2 value was found to be the most appropriate regression method for modelling prime numbers.
Keywords: *Prime number, regression, determination coefficient, model.*

Date of Submission: 23-08-2024 Date of acceptance: 03-09-2024

I. Introduction

The multiplicative building blocks of the number system are known as prime numbers. When a number is prime, its product cannot be obtained by multiplying it with any lower natural number. For example, the prime number 11 cannot be divided into smaller parts; only 1 X 11 equals 11. Conversely, a composite number can be written as the product of two or more prime elements. Two times two times three is the composite number 12. Any whole integer greater than one is either the product of a certain set of primes or a prime. The fundamental theorem of arithmetic refers to this truth, which the ancient Greeks knew, since it is so essential to the system of natural numbers (Pomerance, 1982).

For a very long time, prime numbers and number theory were regarded as abstract, pure mathematics subjects with no real-world applications outside of their intrinsic intricacy and beauty. However, prime numbers were determined to be the building block for hash tables, pseudo-random number generators, and public key cryptography techniques in the 1970s. Furthermore, a full cycle of potential rotor positions was created by developing rotor machines with prime or co-prime numbers of pins on each rotor, enabling more secure communication. Prime numbers have been employed in computer science and encryption as well as in musical creation (Zaman, 2024).

Even though studying prime numbers could not have any immediate applications in real-world situations, it can still draw in a broad range of fans that value mathematics' intricacy and beauty. Mathematicians may learn about fundamental structures and ideas that underpin the broad and intriguing field of mathematics by exploring the characteristics and interactions of prime numbers. An intriguing development in mathematics is the finding of a novel solution to the prime number theorem. Mathematicians have always been fascinated by the prime number sequence, and discovering new, precise solutions to the sum and the infinite sum has greatly advanced our knowledge of it (Goldstein, 1973; De Vas Gunasekara et al., 2015; Moree et al., 2018).

An integer number that has precisely two divisors, 1 and itself, is called a prime number. For the query "prime number," Semantic Scholar [1] provides more than 8.9 million articles; of these, slightly more than 1.6 million are from the fields of computer science and mathematics; the remaining entries are from other fields like biology or geology, illustrating the significance of prime numbers to other disciplines.

There are interesting studies on prime numbers conducted by some researchers in various fields (Kumchev and Nedeva, 1998; Rosu, 2003; Ghidarcea, and Popescu, 2024).

The aim of this study is to define prime numbers and determine the most appropriate regression model by performing regression analysis of prime numbers.

II. Materials and Methods

Prime numbers

The sequence of prime numbers, which begins

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, \dots,$$

has held untold fascination for mathematicians, both professionals and amateurs alike. The prime number theorem, the fundamental theorem that will be covered, makes it possible to forecast the distribution of prime numbers, at least roughly (Goldstein, 1973).

Let x be positive real number, and let π(x)=the number of primes ≤ x. Then the prime number theorem asserts that

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log(x)} = 1$$

where log x denotes the natural log of x.

$$\pi(x) = \frac{x}{\log x} + o\left(\frac{x}{\log x}\right), \qquad (x \to \infty),$$

where $o\left(\frac{x}{\log x}\right)$ represents a function f(x) with the property

$$\lim_{x\to\infty} \frac{f(x)}{x/\log(x)} = 0.$$

Regression analysis

Simple linear regression is a very straightforward approach for predicting a quantitative response Y on the basis of a single predictor variable X. It assumes that there is approximately a linear relationship between X and Y . Mathematically, it can be written this linear relationship as (James et al., 2023).

$$Y = \beta_0 + \beta_1 X$$

Quadratic regression model was given following (Liu et al., 2005).

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2$$

Cubic regression model effects of the independent variable was considered as shown in the following model (Heinrichs et al., 1992):

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

The distribution of exponents finds extensive applications in the sciences, particularly in the field of science biology. However, its use is not restricted to this domain; it is also employed in science pathology and other related domains. Regression exponent modeling can be used to depict data that has been gathered in this manner. It makes it easier for the researcher to see and investigate any changes that take place. One method of nonlinear regression is exponential growth modeling. The following is the exponential growth formula (Rohimet al., 2020):

$$Y = \beta_0 e^{\beta_1 X}$$

In order to compare the prediction performance of the regression models, the following goodness of fit criteria were determined (Willmott and Matsuura, 2005):

Coefficient of Determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Adjusted Coefficient of Determination

$$Adj.R^2 = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Mean-square error (MSE) given by the following formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

III. Results

The graphs of prime numbers between 2 and 101 are shown in Figure 1, those between 2-503 are shown in Figure 2 and those between 2-1009 are shown in Figure 3.

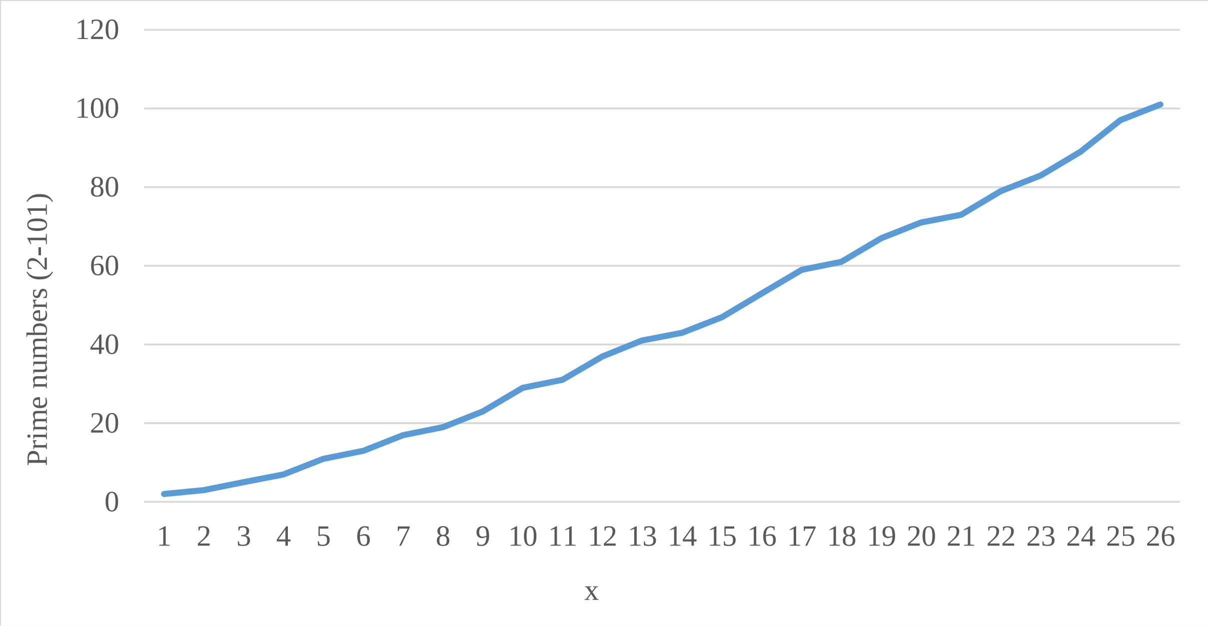

Figure 1. Prime numbers between 2-101

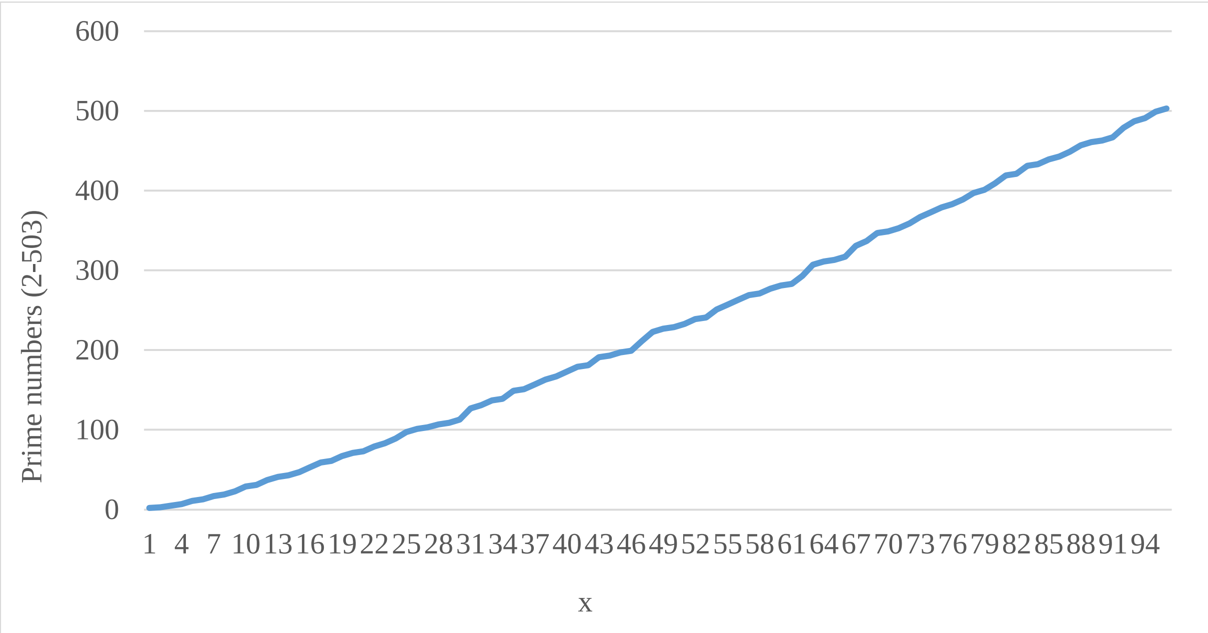

Figure 2. Prime numbers between 2-503

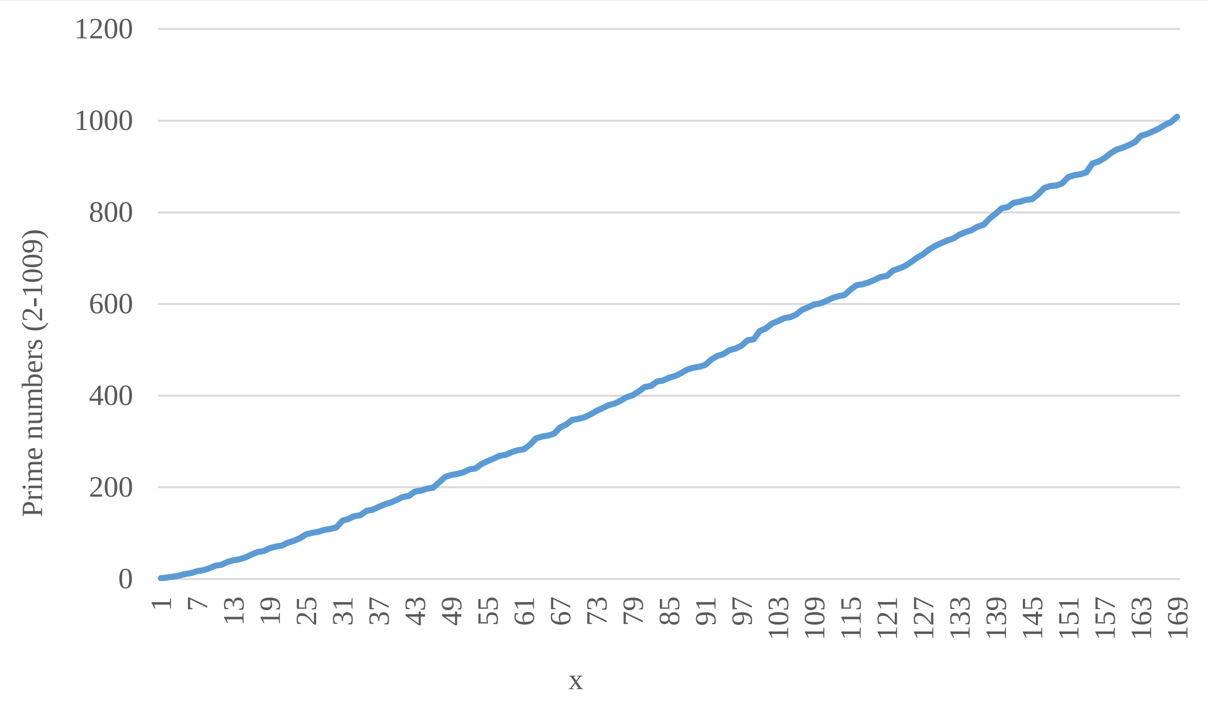

Figure 3. Prime numbers between 2-1009

When all the figures were analyzed, prime numbers were found in the same trend. Regression analyses of prime numbers up to 1009 were performed and the results obtained are presented in Table 1. Linear, quadratic, cubic and exponential regression models were applied. The parameter coefficients of the applied regression models are presented in Table 1.

Table 1. Parameter coefficient of regression analysis

Model	Constant (β_0)	β_1	β_2	β_3
Linear	-64.329	6.127 ***		
Quadratic	-23.717	4.702 ***	0.008 ***	
Cubic	-14.141	4.035 ***	0.018 ***	-0.000038 ***
Exponential	47.176	0.022 ***		

*** ($p<0.001$).

As shown in Table 1, the parameter estimates of linear, quadratic, cubic and exponential regression models were statistically significant ($p<0.001$). All regression models analysed can be evaluated. According to the obtained parameter coefficients, 4 different regression models were written as follows. The overall significance of the regression models was also tested and all models were found to be generally significant ($p<0,001$). Goodness of fit tests were applied to determine the most appropriate regression model (Table 2).

Linear regression model,

$$Y = -64.329 + 6.127\,X$$

Quadratic regression model,

$$Y = -23.717 + 4.702\,X + 0.008\,X^2$$

Cubic regression model,

$$Y = -14.141 + 4.035\,X + 0.018\,X^2 - 0.000038\,X^3$$

Exponential regression model,

$$Y = 47.176\,e^{0.022\,X}$$

Table 2. Goodness of fit of the regression model

Model	R^2	Adj. R^2	MSE
Linear	0.9961	0.996	356.249
Quadratic	0.99963	0.99962	34.28
Cubic	0.99976	0.99976	21.989
Exponential	0.7719	0.7706	0.335

As seen in Table 2, R^2, AdjR^2 and MSE statistics were calculated. R^2 values of linear, quadratic, cubic and exponential regression models were calculated as 0.9961, 0.99963, 0.99976 and 0.7719, respectively. Adj. R^2 values of the same models were 0.996, 0.99962, 0.99976 and 0.7706, respectively. The cubic regression model with the largest R^2 and Adj. R^2. The cubic regression model with the largest R^2 and Adj. R^2 values was selected as the most appropriate model.

IV. Conclusion

The graphs of prime numbers in different intervals (2-101, 2-503 and 2-1009) were analyzed and it was seen that they have similar tendencies to each other. Regression analyses of prime numbers up to 100, 500 and 1000 were performed. In all cases, the cubic regression model was obtained as the most appropriate. It is concluded that regression analysis can be applied on prime numbers and can create meaningful models.

References

[1]. Bickel, Peter J.; Doksum, Kjell A. 2015. Mathematical Statistics: Basic Ideas and Selected Topics. Vol. I (Second ed.). p. 20.

[2]. De Vas Gunasekara ARC, Jayathilake AACA, Perera AAI. 2015. Survey on prime numbers. Elixir Appl. Math, 88:36296-36301.

[3]. Ghidarcea, M., Popescu, D. 2024. Prime Number Sieving—A Systematic Review with Performance Analysis. Algorithms, 17(4), 157.

[4]. Goldstein, L.J. 1973. A History of the Prime Number Theorem. The American Mathematical Monthly, 80(6), 599–615. https://doi.org/10.1080/00029890.1973.11993338

[5]. Heinrichs, A.J., Rogers, G.W., Cooper, J.B. 1992. Predicting body weight and wither height in Holstein heifers using body measurements. J. Dairy Sci., 75: 3576–3581

[6]. James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J. 2023. Linear regression. In An introduction to statistical learning: With applications in python (pp. 69-134). Cham: Springer International Publishing.

[7]. Kumchev, A., Nedeva, T. 1998. On an equation with prime numbers. Acta Arithmetica, 83(2), 117-126.

[8]. Liu, H., Tarima, S., Borders, A.S., Getchell, T.V., Getchell, M. L., Stromberg, A.J. 2005. Quadratic regression analysis for gene discovery and pattern recognition for non-cyclic short time-course microarray experiments. BMC bioinformatics, 6, 1-17.

[9]. Moree P, Petrykiewicz I, Sedunova A. 2018. A compu tational history of prime numbers and riemann zeros. arXiv preprint arXiv:1810.05244.

[10]. Pomerance, C. 1982. The search for prime numbers. Scientific American, 247(6), 136-147.

[11]. Rohim, R.A.A., Ahmad, W.M.A.W., Ismail, N.H., Ghazali, F.M.M., Alam, M.K. 2020. Modeling the growth of bacteria streptococcus sobrinus using exponential regression. Pesquisa Brasileira em Odontopediatria e Clínica Integrada, 20, e5380.

[12]. Rosu, H.C. 2003. Quantum hamiltonians and prime numbers. Modern Physics Letters A, 18(18), 1205-1213.

[13]. Semantic Scholar. 2023. Available online: https://www.semanticscholar.org/ (accessed on 26 February 2023).

[14]. Willmott C., Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30:79-82.

[15]. Zaman, B.U. 2024. New prime number theory. Annals of Mathematics and Physics, 7(1), 158-161.