NOLEGEIN Journal of
Operations Research & Management

ISSN: 2581-5458
Volume 4, Issue 2, 2021
DOI (Journal): 10.37591/NJORM

http://mbajournals.in/index.php/JoORM/index

Research NJORM

Validation of n-variable Linear Regression Model

Krishna Kumar Chaudhary[1], Anjay Kumar Mishra[2,*]

Abstract

Purpose: *The researcher generalizes regression of three variables to n-variable by induction method with the help of cofactor of a matrix so that it assist to the predictor and estimator to estimate and predict dependent variable with the help of independent variables as applied in the relationship between dependent variable GDP and independent variables agriculture, industry, and service sector by using the n-variable regression model at initial condition. This seems a highly applied tool. So, its verification for validity is most and it is attemped here.* ***Design/methodology/approach:*** *To verify the n-variable linear model, both the methods are applied for the same data and estimated the best fit of the line and checked after applying the correlation coefficient matrix method and multiple regression equation using deviations from mean, zero order correlation coefficients, and standard deviation, do the fluctuation existed in the line or not, is estimated.* ***Findings/result:*** *The estimation of dependent variable with the help of independents variables can be estimated through the regression model found to be valid.* ***Originality/value:*** *It is pure research and a useful tool for estimation for multiple variables has been developed and validated.*

Keywords: Application, estimation, matrix, n variable, validation

INTRODUCTION

The estimation of likely value of one variable from the known value of the other variable is regression analysis. The cause and effect relationship is clearly indicated through regression analysis. Linear regression model is just an association between a dependent variable and influenced variables. In our daily life, a single dependent variable is impacted by lot of independent variables so to overcome with the problem Chaudhary and Mishra (2021) [1] has developed a development of n-variable regression model which is highly applicable technique to show the predictor and estimator can predict and estimate the impact of dependent variable on independent variables.

*__Author for Correspondence__
Anjay Kumar Mishra
E-mail: anjaymishra2000@gmail.com

[1]Assistant Professor, School of Engineering, Madan Bhandari Memorial Academy, Morang, Nepal
[2]Research Professor, Srinivas University, India, and Research Director, Madan Bhandari Memorial Academy, Morang, Nepal

Received Date: January 07, 2022
Accepted Date: January 11, 2022
Published Date: January 14, 2022

Citation: Krishna Kumar Chaudhary, Anjay Kumar Mishra Validation of n-variable Linear Regression Model. NOLEGEIN Journal of Operations Research & Management. 2021; 4(2): 26–29p.

Recently the model is applied in analysis of GDP using the n-variable regression model by Chaudhary and Mishra (2021) [2] where the association is shown up between GDP and influenced factors industry, agriculture, and services. Even in a recent study by Chaudhary and Mishra (2021) [3] on impact of agriculture on economic development of Nepal using statistical model, the least square method using the Aldrich et al. [4] and [5] concept which shows that agriculture has a significant positive impact on economic development of Nepal and agriculture is backbone for an economy without which the economy neither can function nor can be survived.

OBJECTIVES

The main aim of the article is to develop an association between dependent variable and independent variables and verify whether the multiple regression equation using deviations from mean, zero order correlation coefficients, and standard deviation method and n-variable regression model gives the same model. If model will have fluctuation what is the fluctuated value.

RESEARCH METHODOLOGY

The linear relationship between n-variable is:

$$\frac{A_{11}}{\sigma_1}x_1 + \frac{A_{12}}{\sigma_2}x_2 + \cdots \ldots\ldots\ldots\ldots.. + \frac{A_{1n}}{\sigma_n}x_n = 0$$

where; $A_{11},\ A_{12}....,\ A_{1n}$ are cofactor of correlation coefficient matrix and can be computed by $A_{ij}=(-1)^{i+j}M_{ij}$, σ_1 , σ_2 ,...., σ_n are standard deviation of the individual data and computed by $(\sigma_i) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$,$x_i = X_i - \bar{X}_i$

The linear relationship between 3-variable is:

$$\frac{A_{11}}{\sigma_1}x_1 + \frac{A_{12}}{\sigma_2}x_2 + \frac{A_{13}}{\sigma_3}x_3 = 0 \text{ ......(i)}$$

where; $A_{11},\ A_{12}$ are cofactors of correlation coefficient matrix, σ_1, σ_2 are standard deviation of the individual data, $x_i = X_i - \bar{X}_i$. Here, one can find the cofactors of correlation coefficient matrix, mean and standard deviation of each data and establish the linear relationship between regressed and regressor variable (Chaudhary and Mishra, 2021) [1].

The multiple regression equation using deviations from mean, zero order correlation coefficients and standard deviation.

$$(X_1 - \bar{X}_1) = b_{12.3}(X_2 - \bar{X}_2) + b_{13.2}(X_3 - \bar{X}_3)$$

Now,

$$b_{12.3} = \frac{\sigma_1(r_{12}-r_{13}r_{23})}{\sigma_2(1-r_{23}^2)}$$

$$b_{13.2} = \frac{\sigma_1(r_{13}-r_{12}r_{23})}{\sigma_3(1-r_{23}^2)}$$

SOURCE OF DATA

The Ministry of Finance is the main source of data, with the help of n-variable regression model and multiple regression equation using deviations from mean, zero order correlation coefficients and standard deviation, a model will be generated.

RESULT AND DISCUSSION

To verify the n-variable linear model both the method is applied for the same data and estimated the best fit of the line and checked after applying the correlation coefficient matrix method and multiple regression equation using deviations from mean, zero order correlation coefficients and standard deviation, do the fluctuation existed in the line or not is estimated (Table 1).

Under calculation:

Mean $(\bar{X}_1) = \frac{\sum X_1}{No.of\ samples\ in\ services} = \frac{21.6}{5} = 4.32$

Similarly, Mean $(\bar{X}_2) = 6.678$ and Mean $(\bar{X}_3) = 2.896$

Standard deviation $(\sigma_1) = 4.34284$
Similarly, (σ_2) =8.40522 and (σ_3) =10.3202

Table 1. Annual growth rate of GDP by construction and manufacturing.

Year	GDP (X_1)	Construction (X_2)	Manufacturing (X_3)
2015/16	0.43	0.12	–9.51
2016/17	8.98	18.68	16.83
2017/18	7.62	12.1	9.21
2018/19	6.66	7.48	6.52
2019/20	–2.09	–4.99	–8.57

Source: Ministry of Finance [6]

A person's correlation coefficient whose value lies between –1 and 1 inclusively can be computed by the formula [7]

$$r = \frac{n\sum XY-(\sum X)(\sum Y)}{\sqrt{n\sum X^2-(\sum X)^2}\sqrt{n\sum Y^2-(\sum Y)^2}}$$

Using the formula for computation, one can find all the associative values,

$r_{11} = 1$, $r_{22} = 1, r_{33} = 1$, $r_{12} = r_{21} = 0.96432$,$r_{23} = r_{32} = 0.96726$ and $r_{13} = r_{31} = 0.9639$

Let us represent these in matrix form say that matrix as correlation coefficient matrix as

$$A = \begin{bmatrix} 1 & 0.96432 & 0.9639 \\ 0.96432 & 1 & 0.96726 \\ 0.9639 & 0.96726 & 1 \end{bmatrix}$$

and its cofactor matrix is

$$A = \begin{bmatrix} 0.064 & -0.032 & -0.031 \\ -0.032 & 0.071 & -0.038 \\ -0.031 & -0.038 & 0.07 \end{bmatrix}$$

Substituting all the necessary values in (i) we get

$$\frac{A_{11}}{\sigma_1}(X_1 - \bar{X}_1) + \frac{A_{12}}{\sigma_2}(X_2 - \bar{X}_2) + \frac{A_{13}}{\sigma_3}(X_3 - \bar{X}_3) = 0$$

$$\Rightarrow \frac{0.064}{4.34}(X_1 - 21.6) + \frac{-0.03}{8.41}(X_2 - 6.68) + \frac{-0.03}{10.32}(X_3 - 2.89) = 0$$

$$\Rightarrow 0.015X_1 - 0.004X_2 - 0.003X_3 - 0.319 + 0.024 + 0.008 = 0$$

$\Rightarrow 0.015X_1 - 0.004X_2 - 0.003X_3 - 0.287 = 0$............(A)

Let us test with the help of multiple regression equation using deviations from mean, zero order correlation coefficients and standard deviation. Under this method suppose the model is

$(X_1 - \bar{X}_1) = b_{12.3}(X_2 - \bar{X}_2) + b_{13.2}(X_3 - \bar{X}_3)$.......(ii) [8]

Now,

$$b_{12.3} = \frac{\sigma_1(r_{12} - r_{13}r_{23})}{\sigma_2(1 - r_{23}^2)} = \frac{0.064(0.96432 - 0.9639 \times 0.96726)}{8.41(1 - (0.96726)^2)} = 0.004$$

$$b_{13.2} = \frac{\sigma_1(r_{13} - r_{12}r_{23})}{\sigma_3(1 - r_{23}^2)} = \frac{0.064(0.9639 - 0.96432 \times 0.96726)}{10.32(1 - (0.96726)^2)} = 0.003$$

Then (ii) reduces to

$(X_1 - 4.32) = 0.004(X_2 - 6.678) + 0.003(X_3 - 2.896)$

$\Rightarrow X_1 - 0.004X_2 - 0.003X_3 - 4.32 + 0.027 + 0.009 = 0$

$\Rightarrow X_1 - 0.004X_2 - 0.003X_3 - 4.284 = 0$ ....(B)

Equations (A) and (B) are nearly identical but some fluctuation is seen between (A) and (B), the reason behind it is due to decimal, If any of variance of residual occurred then it can be computed with the help of $\sigma_i^2 \frac{A}{A_{ii}}$, Where A is the determinant of correlation coefficient matrix, A_{ii} is cofactor of a_{ii} position and σ_i^2 is variance of the dependent variable [9].

CONCLUSION

The estimation of dependent variable with the help of independents variables can be estimated through the regression model. In our daily life, lot of independents variables affects dependent variable. To estimate it we have new model i.e. $\frac{A_{11}}{\sigma_1} x_1 + \frac{A_{12}}{\sigma_2} x_2 + \dots\dots\dots\dots\dots.. + \frac{A_{1n}}{\sigma_n} x_n = 0$. If variance of residual (error) occurred then it can be computed with the help of $\sigma_i^2 \frac{A}{A_{ii}}$.

Acknowledgement

The article cannot be done without the support of MBMAN students. Special thanks to Bhawana Basnet and Sunil Adhikari, BE-Computer of fifth semester and thanks to Librarian Kamala Dahal for providing me books and necessary material on completing this article. Thanks to Asst. Prof. Rashmi Thakur for motivation.

REFERENCES

1. Chaudhary KK et al. Development of n-variable regression model. 2021; 6(1&2): 1–3. Available at: https://doi.org/10.24321/2455.7021.202101.
2. Chaudhary KK et al. Analysis of GDP using the n-variable regression Model. 2021; 6(1): 170–175. Available at: https://doi.org/10.5281/zendo.4772970.
3. Chaudhary KK et al. Impact of agriculture on economic development of Nepal using statistical model. 2021; 8(2). Available at: https://doi.org/10.24321/2455.3039.202101.
4. Aldrich J. Correlations genunine and spurious in Pearson and Yule. Statistical Science. Nov 1995; 10(4): 364–376. Available at: http://uk.jstor.org/cgibin/jstor/printpage/08834237/di984050/98p004.
5. Yule GU. Notes on theory of association attributes in statistics. Biometrika. 1903; 2; 121–134.
6. Central Bureau of Statistics, 2020. Available at: www.cbc.gov.np. Accessed on May 2021.
7. KK Chaudahry. Easy Teaching Text Book of Integrated Probability and Statistics. Princeton Press-Publishing Company, United States of America. 2021. pp. 216. Available at: https://doi.org/10.5281/zenodo.5727953.
8. Krishna Kumar et al. Existence of cross product and application. Scholars Journal of Physics. Mathematics and Statistics. 2020; 7(10): 211–218. Available at: https://doi.org/10.36347/sjpms.v07i10.001.
9. Chaudhary KK et al. Equation of existence of motion of three body and four body problem. Journal of Advanced Research in Applied Mathematics and Statistics. 2019; 4(3&4): 24–27. Available at: https://doi.org/10.24321/2455.7021.201901.

