International Journal of Innovative Mathematics, Statistics & Energy Policies
11(1):53-59, Jan.-Mar., 2023

© SEAHI PUBLICATIONS, 2023 www.seahipaj.org ISSN: 2467-852X

Application Of Determinant In Multiple Regression Analysis

Usman Kabir Kurfi

Department of Mathematics and Statistics,
Hassan Usman Katsina Polytechnic, Katsina, Nigeria
Kabirusman02@gmail.com/ +2348038577141

ABSTRACT

Regression analysis is a collection of statistical techniques that sever as a basis for drawing inference about relationships among interrelated variables. Regression analysis is one of the method used in data analysis which can predict future value. The study was to generate a multivariate data with more than one independent variables, apply Cramer's rule to compute the coefficients and the estimate of the dependent variable Y and forecast the data and compare the forecasted result with the original data. There are so many methods of computing regression coefficient when dealing with more than three equations which includes formula, simultaneous and determinant methods, in this case determinant method using Cramer's rule can be used. Cramer's rule is a theorem in linear algebra, which gives the solution of a system of linear equations or corresponding square matrices in terms of determinants. It is a method of determinants to solve systems of equations. Secondary data related to three variables including weight as dependent variable height and age as the two independent variables was used to test the procedure. Multiple regression equation was computed using the adopted procedure and tested using statistical software. The height and age were sampled for the estimate of weight computed and compared with the original data. The result shows that determinant method using Cramer's rule was the best.

Keywords: Regression analysis, Cramer's rule, determinant, simultaneous

INTRODUCTION

Regression analysis is a statistical method that deals with the formulation of mathematical model that describe relationship among variables and the use of this relationship for the purpose of predictions and other statistical inferences. It can predict future value. Regression analysis is very useful in social sciences, economic, marketing etc. It is also common in practical application. There are many methods to solve it. The least square method is the most commonly used method to solve the model formula. In 1806, French scientist Legendre invented the "least square method" independently, but it was unknown to the world (Yang, 2018). Linear regression is used to study the linear relationship between a dependent variable Y (either blood pressure or height) and one or more independent variables X (age, weight, sex).

The dependent variable Y must be continuous, while the independent variables may be either continuous (age), binary (sex), or categorical (social status). The initial judgment of a possible relationship between two continuous variables should always be made on the basis of a scatter plot (scatter graph).

Linear regression can be used to estimate the weight of any persons whose height lies within the observed range (1.40 m to 1.73 m). The data set need not include any person with this precise height. Mathematically it is possible to estimate the weight of a person whose height is outside the range of values observed in the study. However, such an extrapolation is generally not useful.

Multiple regression analysis is a type of analyzes it in such away that amount of variance explained in a dependent variable by more than one predictor variable. The variables are interval, but sometimes they may have continuous predictors, meaning they are not discrete. For example, weight is a continuous variable, since someone may weigh 125.6 pounds (Amanda, 2017). The degree of relationship existing between two or more independent variables is called multiple regression. The multiple linear regression model is expressed as follows:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \hat{b}_4 X_4 + \dots\dots\dots + \hat{b}_k X_k$$

When sample data are analyzed, the sample regression co-efficient $(b_0,\ b_1,\ b_2 \dots b_k)$ are used as estimate of the true parameter $(\beta_0,\ \beta_1,\ \beta_2 \dots \beta_k)$.

The value of sample regression coefficients can be obtained using formula or using equations. Most companies use regression analysis to describe a phenomenon they want to understand (e.g. why did customer service calls drop last three month?); predict things about the future (e.g. what will sales look like over the next ten months?); or to decide what to do (e.g. should we go with this promotion or a different one?). A note about correlation is not causation: Whenever you work with regression analysis or any other analysis that tries to explain the impact of one factor on another, you need to remember the important adage: Correlation is not causation. This is critical and here's why: It's easy to say that there is a correlation between rain and monthly sales. The regression shows that they are indeed related. But it's an entirely different thing to say that rain caused the sales. Unless you're selling umbrellas, it might be difficult to prove that there is cause and effect (Cham, 2021).

LITERATURE REVIEW

Regression analysis showed that inhibin B was also an independent regulator of serum LH across the follicular phase of the female menstrual cycle. While this relationship has been supported by in vitro studies, little evidence has been observed in vivo. This is one of the first report (Makanji, 2011). These studies also shows that possible role for AMH in the regulation of FSH. Studies showed analyzes of AMH hormone inhibitors independently and inversely with FSH. There is currently some little evidence that supports the pituitary gland for AMH; is there a course related to the preservative procedure on the creation of inhibitors by the proven procedure to remove the pituitary gland? Finally, the apparent biological activity of inhibin B cannot be translated into other biological systems. For example, in the menopausal phase of the female menstrual cycle, there is more enzyme than B inhibitors, but inhibitors play a greater role in the regulation of the hormone FSH. In conclusion, the regulation of FSH by ovarian inhibitors is a multi-step process with coenzyme B appearing to be the main enzyme involved. However, it is clear that structurally different forms of inhibitors related to post-translational changes and other factors such as ovarian steroids and AMH contribute, but these aspects are less well defined.

Regression analysis is one of the most widely used techniques in statistical analysis. The term regression was first introduced by Sir Francis Galton. Galton was Charles Darwin's cousin and he loved science and biology in particular. Sir Francis Galton published some articles on heredity where he observed a relationship between paternal height and offspring growth. The person that first found regression model was Carl Friedrich Gauss, and he was one of the most famous mathematicians. Gauss created the least squares method for estimating planetary orbits using astronomical observations. Least squares estimation is based on the sum of squares. The number of squares was further developed during the 1920s, thanks to Sir Ronald Fisher. Published a formal method to apply perhaps the most popular statistical modeling framework called analysis of variance (ANOVA) in the popular book Statistical Methods for Research Professionals, (Källberg, 2014).

Regression analysis is the oldest and perhaps the most diverse method used in social sciences. Unlike traditional methods, regression is an example of dependent analysis in which the variables are not treated symmetrically. In regression analysis, the object is to obtain a prediction of one variable, based on the values of others. To achieve this change of perspective, a different word and note is used. The predicted variable is denoted by y and the predicted variables by x with the addition of subscripts to separate them from each other (Bartholomew, 2010)

In multiple linear regression, we assume for a linear combination of predictors (often called regression variables). For example, in education research, we can find out how school performance can be predicted by household conditions, age or scores in the previous period. In practice, regression models are estimated by the power of two least squares using appropriate software. Practical matters are important to choose the best option to change back, test the interest of the expert, and establish confidence in the forecast.

A mathematical relationship that permit us to predict the approximate power of the main engines of new container ships, found on data on ships built in 2005-2015. The estimates presented allow him to estimate engine power based on the length between the perpendiculars and the number of containers the ship will carry. Estimates were expand using simple linear regression.

Daniel (2020), explains regression analysis as part of statistical therapy. He is talking about how to work linear regression, and the difference between simple and multiple regressions, which helps the reader to know what linear regression what least-squares regression means.

Thus, a regression was determined to clarify the relationship between the dinoseb adsorption parameter and two soil factors: soil organic carbon and clay. An Investigates the influence of characteristic soil factors on the dinoseb adsorption parameter using various statistical methods (Yiging, 2013).

Aim And Objectives

The aim of this research is to use matrices in solving complicated problems related to multiple regression through the following objectives

- To generate or collect a multivariate data with more than one independent variables
- To apply Cramer's rule to compute the coefficients and the estimate of the dependent variable Y.
- To forecast the data and compare the forecasted result with the original data.

Scope And Limitation

The scope of this research paper is to cover the Cramer's rule application in computing the independent variables in multiple regression analysis. This paper is limited to application of multiple regression data that contains the dependent variable Y and two independent variables X_1 and X_2.

MATERIAL AND METHODS

In this research, data with more than one independent variable will be collected and presented.
The following equations will be used to compute the coefficients.
Given a data series of data with

With the estimated equation
The following equations can be generated

Multiplyˆequation by $\sum$ then by $\sum X_1$ then by $\sum X_2$ then by $\sum X_3$

$\sum \hat{Y} = n\hat{b}_0 + \hat{b}_1 \sum X_1 + \hat{b}_2 \sum X_2 + \hat{b}_3 \sum X_3$ ………………………………... (1)

$\sum X_1\hat{Y} = \hat{b}_0 \sum X_1 + \hat{b}_1 \sum X_1^2 + \hat{b}_2 \sum X_1X_2 + \hat{b}_3 \sum X_1X_3$ ……………..……... (2)

$\sum X_2\hat{Y} = \hat{b}_0 \sum X_2 + \hat{b}_1 \sum X_1X_2 + \hat{b}_2 \sum X_2^2 + +\hat{b}_3 \sum X_2X_3$ ……………..…… (3)

$\sum X_3\hat{Y} = \hat{b}_0 \sum X_3 + \hat{b}_1 \sum X_1X_3 + +\hat{b}_2 \sum X_2X_3 + \hat{b}_3 \sum X_3^2$ ……………..……. (4)

The coefficients $\hat{b}_0, \hat{b}_1, \hat{b}_2$ and $\hat{b}_3$ of the regression equation above can be can be computed using Cramer's rule

The Cramer's rule is a theorem which gives the solution of a system of linear equations or corresponding square matrices in terms of determinants Cramer's rules uses a method of determinants to solve systems linear equations. Starting with equation below.

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n
\end{aligned}
$$

The rules of $3\ by\ 3$ are similar. Given as follows:

$$ax + by + cz = j,$$

$$dx + ey + fz = k,$$

$$gx + hy + iz = l,$$

Which in matrix format is

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} j \\ k \\ l \end{bmatrix}$$

$x, y\ and\ z$ maybe found as follows:

$$x = \frac{\begin{vmatrix} j & b & c \\ k & e & f \\ l & h & i \end{vmatrix}}{\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}}, \quad y = \frac{\begin{vmatrix} a & j & c \\ d & k & f \\ g & l & i \end{vmatrix}}{\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}}, \text{ and } z = \frac{\begin{vmatrix} a & b & j \\ d & e & k \\ g & h & l \end{vmatrix}}{\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}}.$$

The value of *x*, *y* and *z* can be substituted to Y to obtain the estimated regression equation

$\hat{Y} = \hat{b}_0 + \hat{b}_1X_1 + \hat{b}_2X_2 + \hat{b}_3X_3$ for modeling and forecasting.

RESULTS AND DISCUSSION

The dataset below contains 20 records of heights and weights for some current Katsina United Players. These data were obtained from different resources.

The data includes the following variables:

- **Weight(pounds) Y**: Player weight in pounds
- **Height(inches) X_1**: Player height in inches
- **Age X_2**: Player age at time of record (including remaining days < 1 year as decimals)

Table 1. Data for Weights, heights and ages of players

	Y	**X_1**	**X_2**		**Y**	**X_1**	**X_2**
S/No	**Weight(pounds)**	**Height(inches)**	**Age**	**S/No**	**Weight(pounds)**	**Height(inches)**	**Age**
1.	180	74	22.99	2.	188	73	23.88
3.	215	74	34.69	4.	180	73	26.96
5.	210	72	30.78	6.	185	74	23.29
7.	210	72	35.43	8.	160	74	26.11
9.	188	73	35.71	10.	180	69	27.55
11.	176	69	29.39	12.	185	70	34.27
13.	209	69	30.77	14.	197	72	30
15.	200	71	35.07	16.	189	73	27.99
17.	231	76	30.19	18.	185	75	22.38
19.	180	71	27.05	20.	219	78	22.89

Slope and intercept computation using Cramer's rule

Regression computation:

$$\sum \hat{Y} = n\hat{b}_0 + \hat{b}_1 \sum X_1 + \hat{b}_2 \sum X_2 \quad \ldots\ldots (1)$$
$$\sum X_1\hat{Y} = \hat{b}_0 \sum X_1 + \hat{b}_1 \sum X_1^2 + \hat{b}_2 \sum X_1X_2 \quad \ldots\ldots (2)$$
$$\sum X_2\hat{Y} = \hat{b}_0 \sum X_2 + \hat{b}_1 \sum X_1X_2 + \hat{b}_2 \sum X_2^2 \quad \ldots\ldots (3)$$

Equations 1, 2 and 3 above will be replaced to generate the matrix equation for solving simultaneous equation using matrix.
Substituting we have:

$$3867 = 20\hat{b}_0 + 1452\hat{b}_1 + 577.39\hat{b}_2 \quad \ldots\ldots (4)$$
$$280997 = 1452\hat{b}_0 + 105522\hat{b}_1 + 41832.23\hat{b}_2 \quad \ldots\ldots (5)$$
$$112155.4 = 577.39\hat{b}_0 + 41832.23\hat{b}_1 + \hat{b}_2 17056.39 \quad \ldots\ldots (6)$$

Rewriting into matrix form, we have:

$$\begin{bmatrix} 20 & 1452 & 577.39 \\ 1452 & 105522 & 41832.23 \\ 577.39 & 41832.23 & 17056.39 \end{bmatrix} \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 3867 \\ 280997 \\ 112155.4 \end{bmatrix}$$

The Cramer's Rule says that for any simultaneous equation with the form

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} j \\ k \\ l \end{bmatrix}$$

$$\hat{b}_0 = \frac{\begin{vmatrix} j & b & c \\ k & e & f \\ l & h & i \end{vmatrix}}{\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}} == \frac{\begin{vmatrix} 3867 & 1452 & 577.39 \\ 280997 & 105522 & 41832.23 \\ 112155.4 & 41832.23 & 17056.39 \end{vmatrix}}{\begin{vmatrix} 20 & 1452 & 577.39 \\ 1452 & 105522 & 41832.23 \\ 577.39 & 41832.23 & 17056.39 \end{vmatrix}} = \frac{-120250367}{678650.6} = -177.19$$

$$\hat{b}_1 = \frac{\begin{vmatrix} a & j & c \\ d & k & f \\ g & l & i \end{vmatrix}}{\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}} = \frac{\begin{vmatrix} 20 & 3867 & 577.39 \\ 1452 & 280997 & 41832.23 \\ 577.39 & 112155.4 & 17056.39 \end{vmatrix}}{\begin{vmatrix} 20 & 1452 & 577.39 \\ 1452 & 105522 & 41832.23 \\ 577.39 & 41832.23 & 17056.39 \end{vmatrix}} = \frac{2851094.65}{678650.6} = 4.201123$$

$$\hat{b}_2 = \frac{\begin{vmatrix} a & b & j \\ d & e & k \\ g & h & l \end{vmatrix}}{\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}} = \frac{\begin{vmatrix} 20 & 1452 & 3867 \\ 1452 & 105522 & 280997 \\ 577.39 & 41832.23 & 112155.4 \end{vmatrix}}{\begin{vmatrix} 20 & 1452 & 577.39 \\ 1452 & 105522 & 41832.23 \\ 577.39 & 41832.23 & 17056.39 \end{vmatrix}} = \frac{1540656.82}{678650.6} = 2.270177$$

Substituting in to the regression equation $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2$ we have

$$\hat{Y} = -177.19 + 4.2011X_1 + 2.2701X_2$$

Intercept SPSS output

The resulting coefficients were also verified using statistical software as shown in table 1 below.

Table 2 SPSS output showing intercept and slope of height, weight and age data

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
	B	Std. Error	Beta			Zero-order	Partial	Part
(Constant)		124.369		-1.425	.172			
1 HEIGHT_X1	-177.179	1.550	.569	2.710	.015	.321	.549	.516
AGE_X2	4.201 2.270	.814	.586	2.789	.013	.345	.560	.531

a. Dependent Variable: WEIGHT_Y

Forecasting and comparison

The weight (pounds) Y of a player can be estimated when the height (inches) X_1 and age X_2 were known. Given the players height (inches) X_1 = 70 inches and the age X_2 = 30.00 years, the weight of the player is given by:

$$= -177.19 + 4.2011X_1 + 2.2701X_2$$
$$\text{weight } = -177.19 + 4.2011(70) + 2.2701(30)$$
$$\text{weight } = -177.19 + 294.077 + 68.103$$
$$\text{weight } = 184.99 \approx 185 \text{ pounds}.$$

Given the players height (inches) X_1 = 80 inches and the age X_2 = 23.00 years, the weight of the player is given by:

$$\hat{Y} = -177.19 + 4.2011X_1 + 2.2701X_2$$
$$\text{weight } = -177.19 + 4.2011(80) + 2.2701(23)$$
$$\text{weight } = -177.19 + 336.088 + 52.2123$$
$$\text{weight } = 211.1103 \approx 211 \text{ pounds}.$$

To forecast the weight of a player using a sample from the original data, player one was selected as shown in the table below.

Table 3

	Y	**X_1**	**X_2**
S/No	**Weight(pounds)**	**Height(inches)**	**Age**
1.	180	74	22.99

The original weight of the player in table 4 can be forecasted by substituting X_1 with 74 inches and X_2 = 22.99 years as follows:

$$= -177.19 + 4.2011X_1 + 2.2701X_2$$
$$\text{weight } = -177.19 + 4.2011(74) + 2.2701(22.99)$$
$$\text{weight } = -177.19 + 310.8814 + 52.1896$$
$$\text{weight } = 185.88 \approx 186 \text{ pounds}.$$

The estimated weight (186 pounds) when compared with the original weight (180 pounds) contains 6 pounds which shows that the estimator was good and reliable.

Discussion of the results: The purpose of this research paper was to generate or collect a multiple data with more than one independent variable in which we used one dependent variable weigh and two independent variable height and age of some selected football players. Secondly we want to apply Cramer's rule to compute the coefficients and the estimate of the dependent variable Y where we obtain the intercept and the two slopes representing height and weight and the estimated equation was obtained. This estimated equation was used to forecast the data using some selected heights and ages of the players. Lastly the height and age of one player were sample and the estimated weight was computed and compared with the original data. The result shows that determinant method using Cramer's rule was the best and simplest and the procedure is reliable as it gives the actual parameters obtained using SPSS output and when estimating the weight of a player given height and age.

CONCLUSION

This study shows very important role of statistics play in the collection, organizing, analyzing and interpretation of data for decision making, whereby the data used for analysis in this research work have clearly viewed that readers of this research will focus their attention on the method used to compute multiple regression coefficient using Cramer's rule. It is concluded that Cramer's rule was the best, simplest and reliable procedure for computing multiple regression as it gives the actual parameters obtained using statistical software. Cramer's rule was among the simplest method of determining the estimates of the dependent variables. **,** Determinant of a matrix plays a role in finding the intercept and the slope of a multiple regression problem., There procedure was very accurate with other procedures hence it is reliable and applicable to multivariate data.**,** There is the need to apply the method in more than two independent variables.

REFERENCES

Amanda R., Willson, V.L. (2017).Multiple Regression with Two Continuous Predictors and the Interactions Between Them.

D. J. Bartholomew, in International Encyclopedia of Education (Third Edition), Analysis and Interpretation of Multivariate Data

Daniel J. D. (2020). Simple and Multiple Linear Regression

Cham, H.(2021). In Reference Module in Neuroscience and Biobehavioral Psychology, 2021 Quasi-Experimental Designs.

Yumao Li. Theory and Application of Univariate Linear Regression Method [J]. Journal of Chifeng University (Natural Science Edition), 2017,33 (15): 1-2.

Yang, Y. (2018) Research of a New Method for Solving Linear Regression. Advances in Intelligent Systems Research, volume 161. International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018).