**J West Bengal Univ Health Sci. 2020; 1(2):44-52.**

**UNDERSTANDING THE APPLICATION OF MULTIPLE LINEAR REGRESSION MODEL IN HEALTH CARE RESEARCH USING SIMULATION WITH COMPUTER GENERATED DATA.**

**Keshab Mukhopadhyay1, Ritesh Singh2, Chanchal Kumar Dalai1, Shah Newaz Ahmed1C:\Users\admin\Desktop\20061.png,Kushal Banerjee3**

**Author Information**

1. Dept. of Pharmacology, College of Medicine and JNM Hospital, The West Bengal University of Health Sciences, Kalyani, Nadia, West Bengal

2. Dept. of Community and Family Medicine, All India Institute of Medical Sciences, Kalyani, Nadia, West Bengal.

3. Dept. of Pharmacology, Medical College and Hospital, Kolkata, West Bengal

email: shahnewazpharmacology@gmail.com



**ABSTRACT**

Background: The exponential increase in medical information in contemporary science warrants the use of computational tools to simplify and ameliorate patient care. Multiple linear regression modelling is a statistical method that has wide applications in analysis and interpretation of clinical data. In this article, we describe the method of developing a multiple linear regression model using simulation of computer generated data.

Methods: Data was generated for a sample size of 40, for one dependent variable (Y) and four independent variables (X1, X2, X3 and X4). In the first step, bivariate correlation was used to find the individual strength of correlation (R) between the dependent and independent variables. In the second step, the significant variables were added in the model in order of decreasing value of R. Variables which remained statistically significant (p<0.1) in the model were retained while insignificant and multicollinear variables were removed. The final best fit model was conceived with the significant predictors.

Results: The R value for variables X1 to X4 was 0.933, 0.911, 0.725 and 0.148 respectively. X1, X2 and X3 were statistically significant (p<0.001) while X4 was non-significant (p=0.36). X2 and X4 were not included in the best fit model because of multicollinearity and statistical non-significance respectively. The best fit model was represented by the equation Y=0.692\*X1+0.218\*X3+2.003 where 0.692 and 0.218 were the unstandardized coefficients for X1 and X2 respectively and 2.003 was the constant.

Conclusion: Multiple linear regression modelling can be a useful tool for studying the simultaneous effect of multiple variables on a single dependent variable.

Keywords: mathematical modelling, multiple linear regression, simulation, statistical method