IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IDENTIFICATION AND CONSTRUCTION OF BEST LINEAR MODEL USING ADJUSTED SQUARED R

Rajesh Anand B[1] and Khadar Babu SK[2]

1, Department of Mathematics, Sri Venkateswara University,Tirupathi.

2,Department of Mathematics, School of Advanced Sciences,VIT University,Vellore.

Abstract: In big data analytics, the major problem is to select and identification of best linear model for prediction and forecasting of future observations with effect from the studied in a particular area of concerned applications. Data analytics plays a vital role on the analysis of big data for different parameters study. The present paper shows that to identify and make a best linear regression model using adjusted squared R for more parameter study. The data collected from the different patients suffering from different chronic diseases with several parameters. The methodology adopted for model building and model validation is squared R and adjusted squared R. Usually squared R is useful for all types of the models but some cases if more parameters are involved in the model applied adjusted squared R. For model building, used OLS and MLE methods for estimation of the parameters.

Keywords : regression, linear regression, the coefficient of determination, squared R .

I. INTRODUCTION

In statistical analysis, the major role is to build model by using ordinary least squares and generalised least squares methods. In OLS, usually adopted the methodology to estimate regression coefficients. After the estimation of the regression coefficients, then the second step of the process is model validation by using standard model methodologies. In Statistics and mathematical modelling, many of the measures are feasible to apply the concept of the validation for the adopted model and validate it. The present scenario, many of the situations we are struggle for making standard mathematical model for feasibility of the study area. Therefore, mainly focuses on the present paper is on two things , one is making a suitable model for data analysis and the second is to validate the model. These are the most important for the analysis. In data analysis , identify and analyse the standard explanatory variables in the model and do the analysis. In three variable and more variable linear models , the model having the problem of auto correlation, problem of heteroscedasticity and the problem of multicollinearity are presented at the time analysis. For all these problem identification and validation ,need some standard technique to develop the methodology. One of the main validation measures, squared R and adjusted squared R are the simple and best measures for these problems , which are identifies earlier in the model.

Model building is an essential methodology to construct best model and pronounce conclusions for any area of application. Usually in regression analysis ,in multiple regression many variables play a role to give final results. In this case we have to apply three types of methods to estimate the parameters of the model.

They are (i) ordinary least square (OLS) (ii) weighted least squares (iii) Generalised least squares .

At present, applied OLS for estimating the regression coefficients. But if many explanatory variables are added in the model , the validation process again changes and squared R may give the conclusion . In this situation we can apply adjusted squared R should use for validation and better conclusions.

II. Review of literature

The generalization of R^2 proposed by Cox and Snell (1989) and applied on different agricultural data sets. NagelKerke(1991) has given a genaralization of the coefficient of determination (R^2)applied on general liner regression model and proposed a modification for earlier definition.

Khadar Babu etal in the year 2016 and 2017 studied about the differentiate the different datasets and also applied RMSE approach on different regression models and given conclusions . Durhan JL and L.Stockburger(1986) has applied the concept of the coefficient of determination for his research work and has given complete results in work.

III. Methodology

In data analysis , for modelling and handling data many measures are give standard conclusions and inferences about the hypothesis . In hypothesis , the statements that are applied by taking null difference between the effects of the variables . which means that there is no effect to change of one variable on to the other variable. Some times , make a statement that there is no significant effect of coefficients on regression equation. To test the hypothesis we can apply testing procedure. But model validation methodology gives direct conclusions about the models.

One of the best measures to validate the model is the coefficient of determination. It is mostly useful for two variable linear model but the present paper analysis that to study about many explanatory variables . If the study for more variables than the observations we can apply adjusted R^2 and give valid conclusions. The formula also different than the simple R^2.

The measure R^2 depends on the many simple general measures like SSE and SST.

These are standard measures of the model residual analytical solutions., Which are made after building the standard linear regression model.

Let the model defines as follows:

$K_i = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \rho$

The above model , four coefficients and three explanatory variable .

The error variable follows normal distribution with mean μ an variance $\sigma^{2.}$

Usually the error term should follow the standard assumptions but some cases violations of the model assumptions we can get some problems that are (i) Problem of auto correlation (ii) Problem of heteroscedasticity (iii) Problem of auto correlation.

For the above three situations , the squared R measure give conclusions . Due to this the present study illustrates about the analytical conclusions using R^2 and adjusted $R^{2/}$

$R^2 = 1- SSE/SST$

Adjusted $R^2 = 1- (1-R^2) (n-p)/ (n-p)$

The above measures gives the solutions of the model even it suffers the problem of multicollinearity.

IV. Statistical Analysis

The data collected from the cement factory of its composition and heat generated data . In this composition used three types of components mixed in cement and the response shows that the heat in calories.

The fitted model for the data is ae follows:

$K_i = 51.98625_1 + 1.61920X_2 + 0.66800X_3 – 0.03632X_4 + \rho$

Where ρ is the residual term for analysis of the data.

Table 3.1: Multiple regression coefficients

constants	constants	SSE	SE	P value
Intercept	51.98625	22.38709	2.322	0.10288
X_1	1.61920	0.15871	10.202	0.00201
X_2	0.66800	0.30284	2.206	0.11456
X_3	-0.03632	0.26103	-0.139	0.89815

Table : 3.2 : squared R measures

Residual standard error	R^2	Adjusted R^2	F statistic	P value
2.458	0.9875	0.9749	78.78	0.002374

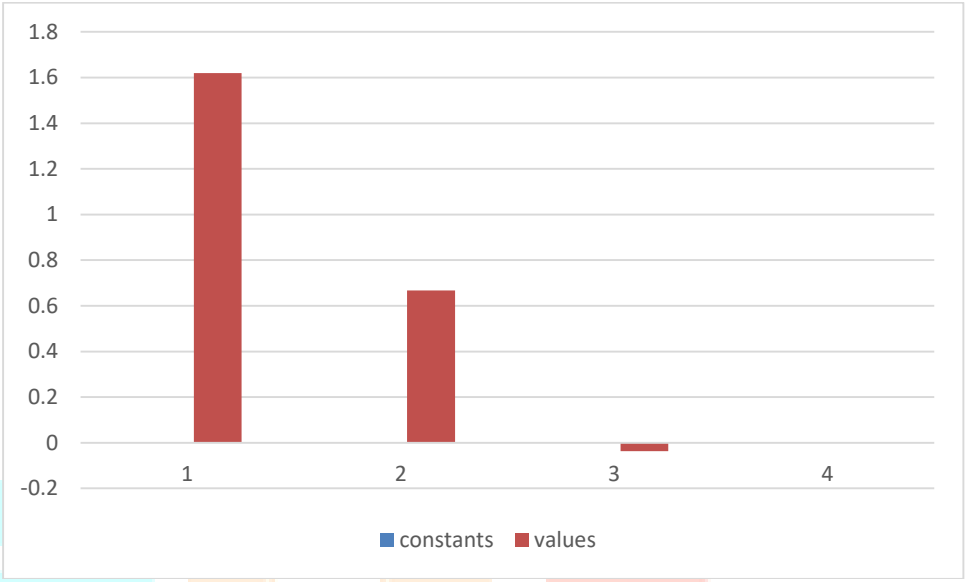

Figure 3.1: constants in multiple regression

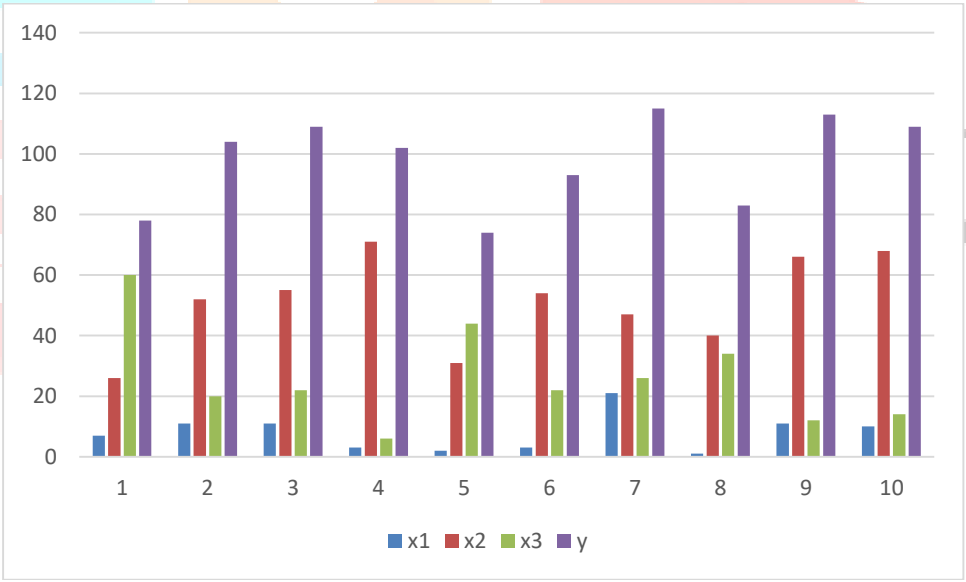

Figure 3.2: Observed data analysis

V. Conclusions

From the model analysis , the R^2 = 0.9875 and adjusted R^2 = 0.9749 and its p value is 0.002374.

For the above model we can conclude that 98.75 % of the data used to built the model and based on the adjusted R^2, 97.49 % of the data used to build the model.

Finally we conclude that the model is perfectly suitable for prediction and forecasting of the heat generated by using three mixtures.

VI. References

[1]. Cox, D.R.,partial likelihood, Biometrika,62(2):269-276(1975).

[2]. CoX, D.R ., Regression Models and Life Tables , Journal of the royal statistical society,series B (Methodological),342(2):187-220(1972).

[3]. S.Satish and SK Khadar Babu(2016),A robust control chart for variability with modified trimmed mean and standard deviation, 9(28),,471-475(2016).

[4]. Nagelkerke, N.J.D.,A note on a general definition of the coefficient of determination, Biometrika,78(3):691-692(1991)

[5]. Durhan, J L and Stockburger, L Nitric acid- air diffusion coefficient experimental determination,Atmospheric Environment,
20(3):559-563(1986).

[6]. S.K.Khadar Babu ,M.Sunithaand M.V.Ramanaiah, Mathematical Modelling of RMSEApproachon Agricultural Financial
data sets, International journal of pure and applied bio sciences, 5(6):942-947(2017).