IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TRANSFORMATION METHODOLOGY FOR BEST LINEAR REGRESSION MODEL

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

1, Associate Professor, Department of Mathematics, VIT University,vellore,Tamil Nadu

2,Assistant Professor,Department of Mathematics, Sri Venkateswara University,Tirupathi

Abstract: For construction and identification of the best linear model , we can adopt some standard applied probabilistic methods for making and construction of the standard regression models. In some situations, we can take generalized linear model methodology for making a best linear regression model with respect to the relationship between variance and expected values of the random variable X are used. When we can transform the variables into others shows that the transformation of the variable strictly followed the variance stabilizing transformations. The present paper used to identify the best transformation with respect to the relation between variance and mean of the explanatory variables. After the model building, adopt model testing statistical measures to select the best optimized standard regression model for statistical analysis.

Key words: Variable transformation methodology, best linear model, the coefficient of determination, mean squared error etc.

I. INTRODUCTION

The general linear methodology, the model should follow the standard assumptions and one of the best assumptions is that $v(u_i) = \sigma^2$, for every I value. The violation of the above assumption symptomizes that the problem of heteroscedasticity may be involved in the model. This is the standard conclusion about the data sets. In such a case model also shows that the model is a moderate and less than moderate model behaviour identified. For example if the study variable y in the model is poisson random variable in a simple linear regression model, then its variance is same as the mean. Since mean of the y variable is related to the explanatory variables x,the variance of y will be directly proportional to x, in such a case variance stabilizing transformation is useful to build the best linear model.

In some cases the study variable proportional to mean and variances such that can apply different transformations choose and make a best linear model.Usually for big data analytics , for model building can take 70% of the data and 30% of the data for model selection. In regression modelling consider 80% of the data for model building and 20% of the data for model selection . In R programming, we have many classification methods to analyse the model adequacy techniques., which are useful to analyse the model validation.

II. Review of Literature

Khadar Babu et al (2011 and 2017) studied about the different regression models based on poison and gaussian probability distributions and also given conclusion to apply auto regressive models for generation of time series data on different studies. Chen y (2012) applied least square support vector Fuzzy regression on different applications financial and weather related time series data for prediction and forecasting.

Guangyong Zon, introduced a modified poisson regression concept for prospective binary data on epidemiological studies to predict and forecasting of the stage of the chronic diseases. Ruiz and Benefa (2012) and Nearing G S etal(2014) applied different types of regression models for facial expression dynamics used Gaussian probability regression model approach.

III. Model Methodology

Let the model for analysis

$H_i = \alpha + \beta x + u_i$, where α is the intercept and the parameter β slope,

Where u_i follows normal distribution with mean zero and variance σ^2.

Usually, u_i is called error term and may useful to construct the model also. Here there is no relationship between y and u and also no relation between x and u.

In such cases, we can observe that there is a relationship between y and x interms of mean and variance. Some of the special relations that variance of y is proportional to mean of Y and may be the variance is proportional to E(y)*E(1-E(y)). In these cases, we transform the study variable into another type like sqrt(y) and 1/sin(y). Transform entire data into either sqrt(y) or 1/sin(y), then make best linear regression model to analyse the solutions and results of the study variable with effect from explanatory variable.

Consider $Y^* = Sqrt(Y)$ and the data transformed in this form.

Make model for $Y^* = \alpha + \beta x + u_i$.

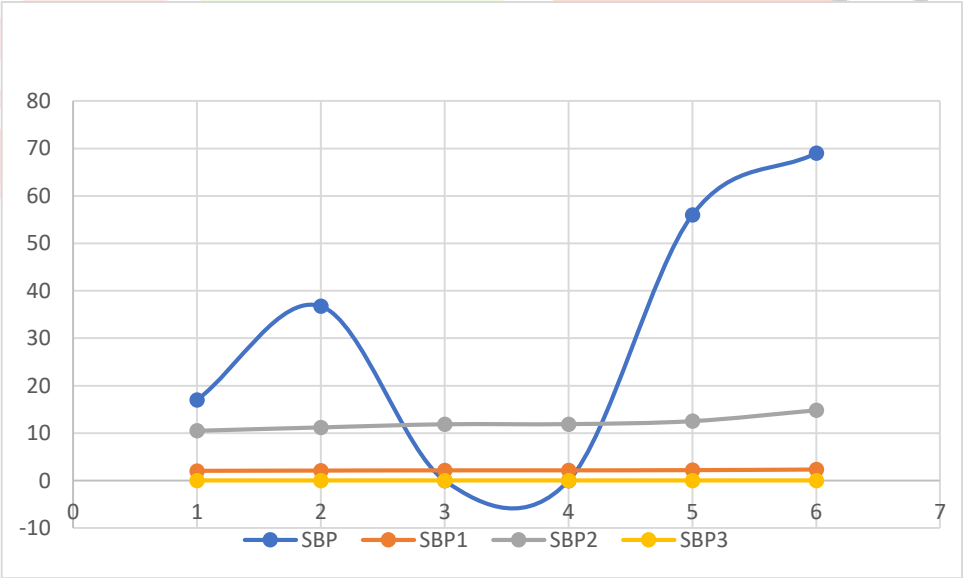
For the model find the R^2 value and take a decision about the regression coefficients plays a role or not.

IV. Statistical Analysis

Data collected from different patients suffering from standard blood pleasure and taking medication frequently. For the data, we can fine model summary statistics and estimation of the squared error and mean squared error in tabular form.

Table 4.1 ; Summary statistics

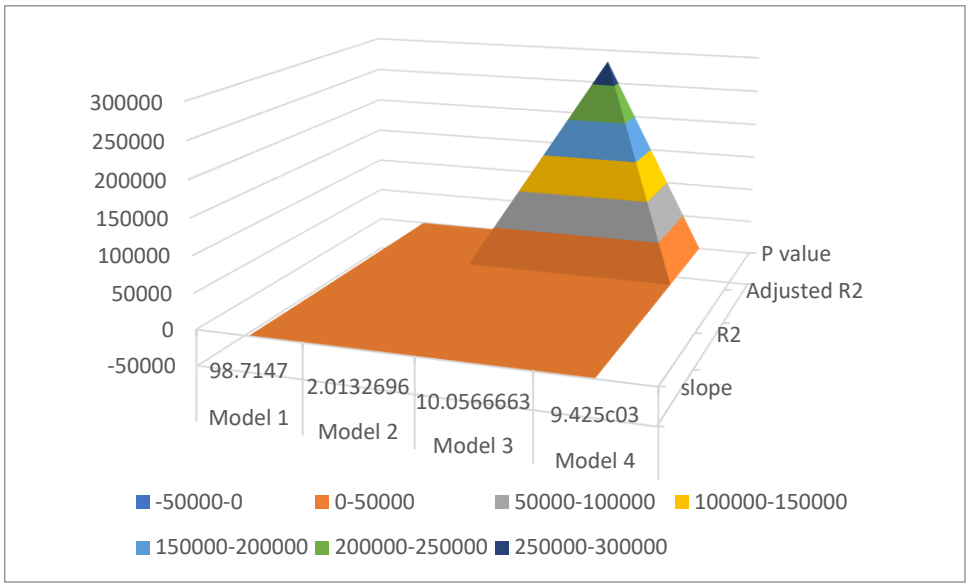
measure	SBP	SBP1	SBP2	SBP3
Minimum	17	2.041	10.49	0.004545
Ist quartile	36.75	2.099	11.21	0.006370
Median	45,50	2.149	11.87	0.007093
mean	45,13	2.149	11.90	0.007168
3 rd quartile	56.00	2.196	12.53	0.007953
Maximum	69.00	2.342	14.83	0.009091



Graph 4.1: structure of measures

Table 4.2:

measures	Model 1	Model 2	Model 3	Model 4
Intercept	98.7147	2.0132696	10.0566663	$9.425c^{03}$
slope	0.9709	0.0030092	0.040952	$-4.999e^{-05}$
R^2	0.4323	0.5056	0.471	0.5606
Adjusted R^2	0.4121	0.488	0.4521	0.5449
P value	$7.86e^{-05}$	1.062e-05	$2.828e^{05}$	$1.949e^{-06}$



Graph 4.2: structure model measures

V. Results and Discussions

For conclusions, the table gives the complete information about the models.

Model	Mean	Variance	R^2
I	142.5	509.9126	0.4324
II	2.149	0.004189067	0.5056
III	11.90	0.8328992	0.471
IV	0.007168	$1.04296e^{-06}$	0.5606

From the standard statistics , model IV R^2 value is high and the maximum compare with all models . So that Model IV is the best linear model for forecasting and prediction of the time series data of medical observations.

VI. References

[1]. Chen Y(2012), Least squares support vector Fuzzy Regression, Energy Procedia 17,711-716. S

[2]. Guangyong Zon ,A modified poisson regression approach to prospective studies with binary data, American journal of epidemiology,159(7),pp702-706(2004).

[3]. S.Sathish and SK.Khadar Babu , Markovian prediction of future values for food grains in the economic survey, IOP conference series,Material science and engineering,263(2017).

[4]. Khadar Babu SK ,Sunitha, M and Ramanaiah M V (2017),Mathematical Modelling of RMSE approach on agricultural and Financial data sets, International journal of Pure and Applied bio sciences.

[5]. Khadar Babu SK ,Karthikeyan K , Ramanaiah M V and Ramana D(2011),Prediction of rain fall flow timeseries using auto regressive models , Pelagia research library, Advances in Applied Science Research, 2(2),128-123.

[6]. Ruiz and Benefa (2012) ,Modelling facial expressions dynamics with Gaussian process regression , Frontiers in Artificial Intelligence and Applications,248,91-100.

[7]. Nearing G.S Gupta,H.V and Crow,WT (2014) , Information loss in approximately Bayesian estimation technique: A Comparison of generative and discriminative approaches to estimating agricultural productivity ,Journal of hydrology , 507(12),163-173.