Engineering Research Models Evaluation Methods

Awari Mahesh Babu
Professor
Civil Engineering Department,
Tirumala Engineering College,
Hyderabad.TS. India.
E-mail: maheshbabuawari@gmail.com

Abstract: The objective of several records analysis is in the direction of extorts since unrefined in sequence the precise evaluation. Individual of the mainly significant as well as frequent difficulty regarding but here is statistical correlation among a retort unpredictable Y and instructive unpredictable Xi. An alternative in the direction of response this problem is just before take up regression analysis in order to model its correlation. There are different kinds of regression analysis. The category of the regression model depends scheduled the category of the distribution of Y. In modeling it is to anticipate the result Y stand scheduled standards of a set of unpredictable Xi. In research problems and models regression analysis and evaluation of methods are very important In this paper it is focused in regression and evaluation methods for validation of models.

***Keywords**: Regression analysis, models, variance, covariance, evaluation*

*****

I. INTRODUCTION

In *statistical* modeling, regression analysis is a statistical method for estimate the relations between unpredictable. It contains diverse methods on behalf of modeling moreover evaluate a number of unpredictable, Further overtly, regression analysis facilitate one identify how the attribute assessment of the reliant unpredictable changes while several individual of the autonomous unpredictable is mottled, whereas the other autonomous unpredictable are apprehended predestined In regression analysis, it is besides of significance toward distinguish the deviation of the reliant unpredictable about the regression role which know how to be described with a probability distribution. Regression analysis and evaluation methods experimental values of SPEED and Stopping Sight Distance (SSD) are indentified.

II. REGRESSION ANALYSIS

Various regressions give an equation that calculate one unpredictable from two or more autonomous unpredictable,

$Y = \beta_0 + \beta_1X_1 + \beta_2X_2$ ……..---------- (1)

Important steps concerned in performing a regression analysis

Step-1: Creation of the regression model

The construction of an instructive model is a fundamental step in the regression analysis. It has to be described through position to the action theory of the intervention. It is expected that a number of category of unpredictable exist. An unpredictable might also correspond to an observable attribute or an unobservable one.

The model may imagine that a particular unpredictable evolves in a linear, logarithmic, exponential or other way. All the instructive models are constructed on the base of a model, such as the subsequent, for linear regression:

$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots. + \beta_kX_k$, where -------- (2)

Y is the differ that the program is principally theoretical to make

$X_{1\text{-}k}$ are autonomous unpredictable likely to explain the change.$\beta_{0\text{-}k}$ are constants.

Step-2: Construction of a model

To apply multiple regressions, a large model is generally essential.

Step-3: Records collection

Consistent information have to be collected, any from a monitoring scheme, from a questionnaire survey or from field survey.

Step-4: Computation of coefficients

Coefficients know how to be premeditated rather easily, using statistical software that is both reasonable and available to PC users.

Step-5: Valuation of the model

The model plan to elucidate as much of the variability of the experimental changes as possible. Assessment can be done from graphical and computational analysis and significance tests. The percentage of the variability in the y variable can be explained by the x variable.

Flowchart showing the step by step procedure of Regression analysis is presented in Figure 1

ISSN: 2321-8169
158 - 161

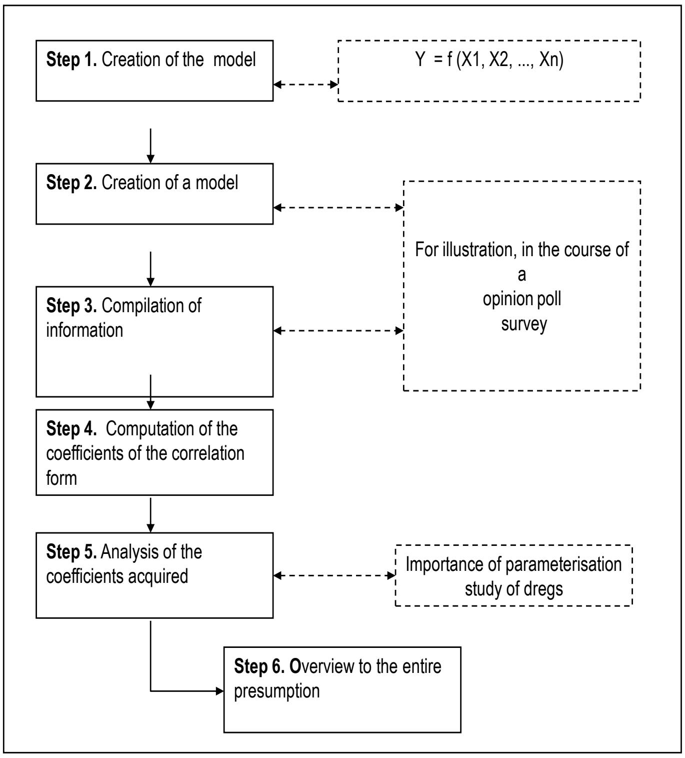

Fig 1 Flow Chart for Regression analysis

III. EVALUATION OF MODELS

For analysis of the data it has to be fitted with one or more models. The analysis deals numerical fit analysis, graphical fit analysis and significance tests. The graphical measures deals with the residuals and prediction bounds and the numerical measures deals with the goodness of fit statistics as given below

1. Confidence and Prediction bounds

Confidence bounds pro the integral coefficients with prediction bound for new observations or the integral task know how to be considered. The confidence bounds are numerical which are displayed through entity polynomial coefficients, while the prediction bounds are displayed graphically. Confidence and prediction bounds delineate the subordinate and superior values of the connected period, and describe the size of the interval.

3. Sum of squares due to errors (SSE)

The above statistic measures the total deviation of the response values from the fit to the response values. A value closer to zero indicates a better fit. Mathematically it is represented as

$SSE = \sum_{i=1}^{n} w_i (yi - \hat{y}_i)^2$ ------------- (3)

Where y_i is the response value and $\hat{y}_i$ is the predicted value

4. R²

These statistics dealings how successful the fit is in explaining the variation of the data. It is the square of the correlation between the response values and the predicted response values.

SSR is defined as

$$SSR = \sum_{i=1}^{n} w_i (\hat{y}_i - \bar{y})^2 \quad \text{--------------- (4)}$$

Where $\bar{y}$ is the response vector

SST, which is furthermore called as squares about the mean, is defined as

$$SST = \sum_{i=1}^{n} w_i (y_i - \bar{y})^2 \quad \text{--------------- (5)}$$

Where SST = SSR+SSE

Therefore R² can be expressed as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \text{ -------------- (6)}$$

The coefficient of determination, R², is constructive since it gives the quantity of the discrepancy (fluctuation) of one unpredictable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph. The coefficient of determination is the ratio of the explained variation to the total variation.

The coefficient of determination is such that $0 < R^2 < 1$, and denotes the strength of the linear association between x and y.

5. *Degree of freedom (DFE)*

The residuals degree of freedom is distinct as the quantity of response values n less the quantity of fitted coefficients m predictable from the response values.

$$v = n - m \quad \text{--------------- (7)}$$

v designate the quantity of independent pieces of information on which the approximation is based and is known as Degree of Freedom (DFE) of an estimate. A lesser amount of the degree of freedom less is the complication and degree of freedom also helps to evade the over fitting of the data.

6. Root Mean Square Error –RMSE. It is also known as Standard Error

This statistic is also identified as the standard error of the regression

$$RMSE = s = \sqrt{MSE} \quad \text{----------- (8)}$$

Where MSE is the mean square error or the residual mean square

$$MSE = \frac{SSE}{v} \quad \text{------------- (9)}$$

Standard Error (RMSE) value closer to 0 indicates a better fit.

7. Multiple correlation coefficients R

It is the square root of R^2

Linear correlation coefficient,

r = Σ (xy) sqrt [(Σ x2) * (Σ y2)]---(10)

The quantity r, called the linear correlation coefficient, measures the potency and the direction of a linear correlation between two variables. The value of r is such that -1 < r < +1. The + and – signs are used for positive linear correlations and negative linear correlations,

International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 2 158 - 161

respectively. – ve sign indicates that one variable increases and other will decreases and vice versa.

IV. REGRESSION ANALYSIS AND EVALUATION OF MODELS

For case study transportation engineering problem is premeditated. The following table values are experimental values of speed and stopping sight distance

Table 1

Nodes/ Criterions	SPEED	SSD
1	32.006	49.109
2	33.541	50.522
3	32.368	41.615
4	32.314	44.423
5	28.007	40.061

Computation of good fit for regression models

Computation of good fit for SPEED and SSD (Experimental values)

Linear model Poly3

$$f(x) = n1x^3 + n2\,x^2 + n3x + n4 \text{--------} \quad (11)$$

Where x is normalize by mean 45.15 and std 4.568.

Coefficients (with 95% confidence bounds):

n1 = 3.864 (-14.47, 22.19), n2 = -1.384 (-11.85, 9.083)

n3 = -2.596 (-23.56, 18.37), n4 = 32.42 (22.42, 42.42)

The Strength and Significance coefficients:

- Multiple correlation coefficient -MCC= 0.98, Sum of squares due to errors SSE = 0.76,coefficient of determination R^2= 0.96, Standard Error-SE = 0.87, Degree of freedom (DFE)=1

Linear model Poly4:

$$f(x) = n1\,x^4 + n2x^3 + n3\,x^2 + n4x + n5 \text{---------} \quad (12)$$

Where x is normalized by mean 45.15 and std 4.568,Coefficients:

n1 = -2.039, n2 = 4.357, n3 = 1.555, n4 = -3.068 n5 = 31.81

The Strength and Significance coefficients:

MCC=1, SSE =2.52, R^2= 1, SE= NaN, DFE =0

Table 2 Comparison of risk parameters obtained from above analysis

Polynomial degree	3rd degree polynomial				4th degree polynomial				Good fit polynomial
Parameter	SSE	R^2	RMSE	DFE	SSE	R^2	RMSE	DFE	
SSD	0.76	0.96	0.87	1	2.52	1	NaN	0	3rd

iv) Comparison of parameters to get good fit graphical analysis for SPEED and SSD

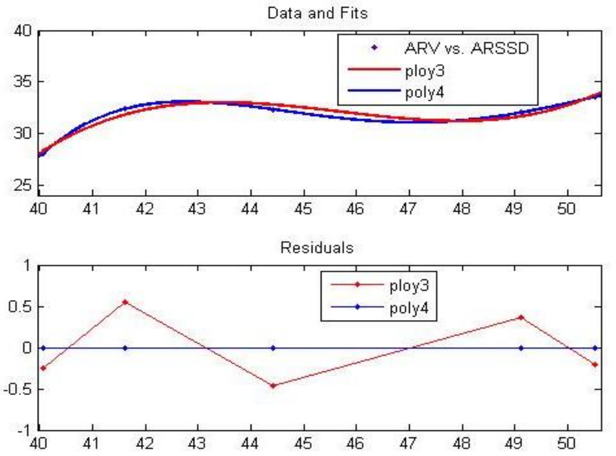

Fig 2 Curves for good fit of SPEED and SSD

The above information revealed for absolute ideals of Speed and Stopping sight distance. In Residual graph, residual data of 3rd degree polynomial data become visible arbitrarily which suggests the model fits the data well.

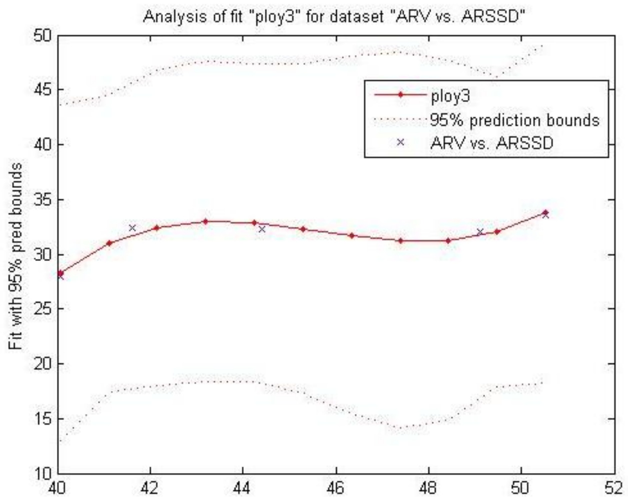

Fig 2.2 Prediction bounds for fit using poly3

 ISSN: 2321-8169
158 - 161

V. CONLUSION

It is observed that 4th degree polynomial model gives R^2 significance identical to 1 however it more than fits the data and amplify the computational intricacy. From graphical analysis it can be measured that 3rd degree polynomial fits the model and it is also observed that 3rd degree polynomial SSE is less than 4th degree polynomial which indicates better fit. The deterioration analysis contrivance is advanced tools that be capable of categorize how unlike variables in a method are linked. The regression tool will inform if one or multiple variables are unified among a method output. This in order know how to categorize where in the method direct is desired or what factors is the unsurpassed first point for a method upgrading task.

REFERENCES

[1] M. H. Kutner, C. J. Nachtsheim, and J. Neter (2004), "Applied Linear Regression Models", 4th ed., McGraw-Hill/Irwin, Boston (p. 25)

[2] N. Ravishankar and D. K. Dey (2002), "A First Course in Linear Model Theory", Chapman and Hall/CRC, Boca Raton (p. 101)

[3] Steel, R.G.D, and Torrie, J. H., *Principles and Procedures of Statistics with Special Reference to the Biological Sciences.*, McGraw Hill, 1960, page 288.

[4] Chiang, C.L, (2003) *Statistical methods of analysis*, World Scientific. ISBN 981-238-310-7 - page 274 section 9.7.4 "interpolation vs extrapolation"

[5] *Good, P. I.; Hardin, J. W. (2009). Common Errors in Statistics (And How to Avoid Them) (3rd ed.). Hoboken, New Jersey: Wiley. p. 211. ISBN 978-0-470-45798-6.*

[6] *Tofallis, C. (2009). "Least Squares Percentage Regression". Journal of Modern Applied Statistical Methods* ***7****: 526–534. doi:10.2139/ssrn.1406472.*

[7] *YangJing Long (2009). "Human age estimation by metric learning for regression problems" (PDF). Proc. International Conference on Computer Analysis of Images and Patterns: 74–82.*

[8] Rothman KJ, Greenland S. Modern Epidemiology. 2nd ed. Lippincot-Raven; 1998.

[9] Altman DG. Practical Statistics for Medical Research. Chapman & Hall/CRC; 1991. .

[10] Draper NR, Smith H. Applied Regression Analysis. Wiley Series in Probability and Statistics; 1998. Applied Regression Analysis.

[11] Munro BH. Statistical Methods for Health Care Research. 5th ed. Lippincott Williams & Wilkins; 2005.