IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728.Volume 5, Issue 5 (Mar. - Apr. 2013), PP 53-56
www.iosrjournals.org

Regression Models Use In Data Splitting Techniques

Rajeev K. Gupta
Professor & Head (Applied Mathematics Department)
NRI-College of Engineering & Management, Gwalior (M.P.), India

Abstract: *Model validity is the stability and reasonableness of the regression coefficients, the plausibility and usability of the regression function and ability to generalize inference drawn from the regression analysis. Model validation is an important step in the modeling process and helps in assessing the reliability of models before they can be used in decision making. This research work therefore seeks to study regression model validation process use in data splitting techniques.*

We review regression model validation by comparing predictive index accuracy of data splitting techniques. Various validation statistic such as the mean square error (MSE) and R2 were used as criteria for selecting the best model and the best selection procedure for each data set. The data splitting techniques provides the most precise estimate of R^2 which reduce the risk over fitted models than in data splitting techniques.

Keywords: *Data splitting technique, coefficient of determination, piecewise regression and validation.*

I. Introduction

Model selection and validation are critical in predicting a dependent variable given the independent variable. The correct selection of variables minimizes the model mismatch error while the selection of suitable model reduces the model intimation error. Models are validated to minimize the model prediction error. A more flexible model can better represent the data may also more easily lead the user astray by noise in the data. Determining the right form of the model in order to reduce model mismatch error is accomplished during model construction phase, whereas determining the correct model parameter can be achieved at the model selection and validation.

Once a regression model has been constructed, it is important to confirm the goodness of fit of the model and the statistic significance of the estimated parameters, commonly used are check of goodness of fit include analysis of the pattern of residuals and hypothesis testing, statistically significance checked by an f-test of the overall fit, followed by t-test of individual parameters interpretation of these diagnostic tests.

Validation is an essential part of model building, its application and levels of confidence in usage are highly important. It entails checking the R^2 statistic from the regression fit, carrying out a diagnostic of the residual either through exploratory statistic, checking the mean confirmatory statistics, checking the mean square error.

Model validity refers to stability and reasonableness of the regression coefficients, the plausibility and usability of the regression function and ability to generalize inferences drawn from the regression analysis. Validation is a useful and necessary part of the model building process. A good fit of a model to the data set is not an only goal of model validation but also to get a perfect fit i.e. its predictive accuracy of the model is how the model validates a new dataset.

Model validation requires checking the model against independent data to see how well is predicts. Several researchers have work extensively on model validation using Data splitting techniques. A drawn back of cross validation is the choice of the number of observation to hold out each fit. Also cross validation may not fully represent the variability of variable selection. The major disadvantages of data splitting techniques in model validation is that different investigators using the same data could split the data differently generate different models, hence obtain different validating result. Snee (1997) researched extensively on method of Validation, Neumann et al (1977) and Shapiro (1984) have employed Monte Carlo testing to estimate artificial predictability in tropical Cyclane prediction models. Reinduer and Run (1980) and Lanzante (1984) carried out set of Monte Carlo test to examine false predictability and the inflation of $\mathbf{R^2}$ as a function of sample size, size of the predictor and number of predictor selected.

Model validation is an important step in the modeling process and helps in assessing the reliability of models before they can be used in decision making (Jannath and Tsuchido 1998). Hall and Wilson (1991),Davison and Hinkely (1997).

These research works examine the validation of regression by comparing the predictive accuracy of data splitting techniques. This work proposes a procedure for construction, selection and validation of regression models.

II. Material And Methodology

Validating regression model was implemented in this work using the technique of data splitting. In data splitting, three different regression procedures were used to fit regression model to two different data sets. The data sets have many variables predicting the response variable. In data splitting, we split the data sets into two separate samples using one part for modeling and the other for testing the model. We also hope to see if the peculiarities of the original set will be seen in the split modeling set.

The first data set is a institutional growth of prior1993 and another is after 1993. In data splitting techniques we employed the approach of stepwise regression procedures in selecting variables into a regression model. The include Forward Selection, Backwards Elimination and Best subset Regression. They add or remove variable one at a time until some stopping rule is satisfied.

The forward selection regression procedure sequentially adds variables to the model one at a time. Starts with an empty model and adds the variable has the smallest p value usually less than 0.05 or 0.10 the model.

Aside the p-value criterion, at any stage in the selection process, forward selection adds the variable that has the highest partial correlation, increases $\mathbf{R^2}$ the most, and gives the largest absolute t of F statistic to the model. This procedure is a model reduction method. The Backward Elimination regression procedure starts with all the predictors in the model and sequential deletes variable from the model. At any stage, in the selection process, it deletes the variables with the smallest $\mathbf{R^{2,}}$ Backward Elimination procedure gives adequate model since the procedure involves starting the model building with all the variables and deleting the variables that add nothing to the model.

Best subset regression examiner all possible models and chooses the one with the most favorable value of some summary measure such as large adjusted; $\mathbf{R^2}$ and smallest standard error. All possible regression has a large advantage over stepwise procedures in that it can let the analyst see competing models, models that are almost as good as the best.

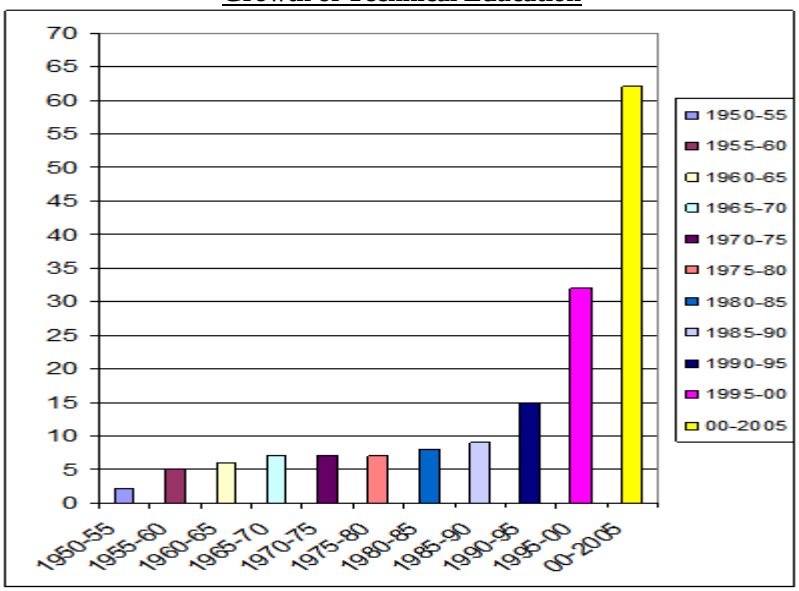
Data splitting has the advantage of allowing hypothesis tests to be confirmed in the test sample however, the major disadvantages it has is that different investigators using the same data could split the data differently and generate different models, hence obtaining different validating results.

The analytical technique used for analyzing the data was the trend analysis of the number of engineering institutions in the state. As already discussed, there has been a very rapid increase in both the number of engineering colleges the enrollment in engineering education after 1993; the trend in the number of institutions has been different in the period prior to 1993 and the period after 1993. In order to capture the two trends in the predictive model, the piecewise linear regression a technique was applied to analyses the time trend. The piecewise linear regression analysis uses a dummy variable to differentiate the trend before 1993.

GROWTH OF TECHNICAL EDUCATION IN MADHYA PRADESH

The graph and table below show the summary of the growth of technical education in India particularly in Madhya Pradesh:

Growth of Technical Education



III. Result Of Piecewise Regression Analysis Of Growth Of Colleges

$R^2 = 0.879$, Adj. $R^2 = 0.874$

ANOVA

	SS	df	MSS	F	P
Regression	51.384	2	25.692	189.001	0.000
Residual	7.059	52	0.136		
Total	58.452				

1.1 (a)

Regression

	B	SE (B)	β	t	P
Constant	0.448	0.110		4.066	0.000
Time	0.054	0.004	0.833	12.194	0.000
Total	0.038	0.018	0.140	2.052	0.045

1.1 (b)

INTERPRETATION OF PIECEWISE REGRESSION OF THE GROWTH OF COLLEGES

It is possible to analyze the trend in a number of engineering colleges over time through the application of piecewise regression. This approach particularly suited in the present case as the trend in the increase in engineering colleges has been different from the period 1951-93 and 1993-2004.

Results of the piecewise regression analysis are given in table1.1 (a) &1.1(b). The model provides good fit to the data as the regression coefficient is significant and in expected direction and the model explains more than 87 percent of the variation in the original dataset.

The regression analysis suggests that the number of engineering colleges in the state increased at the rate of 5.5% during 1951-93 but at the rate of almost 10% per year during 1993-2004.

IV. Conclusion

To conclude it can be said that the present study of engineering and technical education in India (as well as of Madhya Pradesh) reveals that the leading institutions have adopted study programs. Some of the programs are innovative in nature and offer tremendous benefits to students, industries and universities.

The chief advantages to the students can be summarized as: gaining confidence in decision-making, relating theory with practice, increasing job opportunities, and opportunities to work with modern equipment and on problems of current importance.

Finally, it can be concluded that other universities and institutions should adopt more job and object-oriented engineering education curricula linked with industries and research organizations to meet the present and future challenges of rapid technological changes and industrial development in India.

References

[1] Agrawal R.C., Agrawal D.P.: Need and Ways of Collaboration between Institute and Industry, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 207-210.

[2] Annual Report 2000-2001: All India Council for Technical Education, New Delhi, 2001.

[3] Annual Technical Manpower Review: Madhya Pradesh 2005 (Engineering) Institute of Applied Manpower Research, New Delhi, 2005.

[4] Annual Technical Manpower Review: Madhya Pradesh 2006 (Engineering) Institute of Applied Manpower Research, New Delhi, 2006.

[5] Arora S.S., Khare V.K.: Industry Institute Partnership for Overall Economic Growth by Increasing the Duration of Engineering Degree Course, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 185-188.

[6] Basavanna B.M., Shivshankar H.N.: Formal Technical Education in Universities– Present Trends and Future Directions, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 340-347.

[7] Chawala Meenu, Shrivastava Ritu, Tiwari A.K.: Changing Technology and Technical Education, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 140-145.

[8] Choukse R.C., Dishoriya H.P.: Autonomy to Technical Institution, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 97-100.

[9] Christensen Ronald: Linear Models for Multivariate Time Series and Spatial Data, Springer Verlag, New York, 1991.

[10] Compendium on Technical Education 1947-1997-1999: All India Council for Technical Education, New Delhi, 2000.

[11] Cox D.R.: Regression Models and Life tables (with Discussion), Journal of Royal Statistical Society Series B, 1972, 34 (187-220).

[12] Cullagh M.C., Nelder J.A.: Generalized Linear Models, Second Edition, Chapman and Hall, London, 1989.

[13] Dasgupta R.: Strategies for Imparting Quality in Technical Education, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 629-636.

[14] David A. Harville: Matrix Algebra from a Statisticians Perspective, Springer Verlag, New York, 1997.

[15] Dominic J., Nirmala P.J.: Digital Libraries: World Wide Implications, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 311-318.

[16] Dongre Ashish: Performance of Teacher in The Class Room and Teacher – Taught – Output Relationship From the View Point of Students in Reference to T.Q.M. in Technical Education, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 620-628.
[17] Emanuel Parzen: Modern Probability Theory and Its Application, Wiley Eastern Pvt. Ltd., New Delhi, 1972.
[18] Gaur Anand Niketan: How to Make Technical Education Complete "Wholistic Knowledge" of Concerned Areas to Fulfill the Challenges of 21 Century, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 411-414.
[19] Goel Aditya, Somkunwar Ajay: Technical Engineering Education: Existing and Futuristic Scope, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 284-288.
[20] Grant W., Coombs C. F.: Handbook of Reliability Engineering and Management, McGraw Hill, New York, 1988.
[21] Guidelines and Formal 2001 to 2002: All India Council for Technical Education, New Delhi, 2001.
[22] Institute Profile: Information Engineering and Architecture Colleges in Madhya Pradesh, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, 2002-03.
[23] Institute Profile: Information Engineering and Architecture Colleges in Madhya Pradesh, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, 2003-04.
[24] Jain A.K., Agnihotri G.: Quality Assurance in Technical Education, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 72-81.
[25] Jain A.K., Mehra Chanchal: Rural Industrialization- Role of Community Polytechnics, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 673-681.
[26] Jain J.K., Singhai Jyoti: Planning Technical Education for 21st Century with Effective Involvement of User Agencies, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 146-151.
[27] James K. Lindsev: The Analysis of Stochastic Processes using GLIM, Springer Verlag, New York, 1992.
[28] Kapur J.N., Saxena H.C.: Mathematical Statistics, S. Chand & Company Ltd., New Delhi, 1981.
[29] Kawitkar R.S..: An Approach for Advanced Curricula to Meet Challenges in 21st Century, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 334-339.
[30] Keeping E.S.: Introduction to Statistical, Inference. Van Nostrand Company, Canada, 1964.
[31] Kenney J.F., Keeping E.S.: Mathematics of Statistics, Van Nostrand 1951, East-West 1964.
[32] Khan Gazala, Khan Arshad: Computer Based Technical Education. Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 270-275.
[33] Khanna kumar Indra: Strategic Planning for Technical Education, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 425-432.
[34] Khare A.K., Singh Onkar: Engineering Education in India- Present Scenario and Challenges Ahead, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 348-354.
[35] Kshirsagar P.H., Thete A.R.: New Strategies for Technical Education in 21st century, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 403-410.
[36] Kumar Ranjan: Industry Institute Partnership: Role of Small Scale Industry, Proceedings of Technical Education in 21st Century (ICTE-21), Department of Manpower Planning, Government of Madhya Pradesh, Bhopal, 1998, 230-233.