ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2019; 5(4): 52-54
www.allresearchjournal.com
Received: 01-02-2019
Accepted: 29-03-2019

Gaurav Kumar
Associate Professor, Department of Mathematics, N.A.S. College, Meerut, Uttar Pradesh, India

Corresponding Author:
Gaurav Kumar
Associate Professor, Department of Mathematics, N.A.S. College, Meerut, Uttar Pradesh, India

An overview of data analysis methods for environment valuation

Gaurav Kumar

Abstract
Environment degradation has been a major issue in industrial areas. There is need to check this degradation. Policy makers need data to formulate better policy to tackle this issue. While the importance of data collation cannot be undermined but it is the analysis of data which provides the insight into the problem. Therefore, data analysis has major role to play. There are various techniques of data analysis. In the present work, some of the methods of data analysis are presented.

Keywords: Data analysis, regression, fractional polynomial, artificial neural network

1. Introduction
For last two decades, there has been problem of environment degradation. Environment degradation has become a major problem not only for India but also for the world. The degradation has happened rapidly after the industrial revolution. While this revolution has given lot of comforts to the human beings and have made our life simpler and better but the cost has been paid by the environment. Tree cutting has been a major activity so as to spread the fruit of industries to all the corners of the world. This activity has heavily impacted the environment in a negative manner. Air pollution has come out as a very big problem for the mankind. This has resulted into number of untimely deaths of the people across nations. WHO has recognized as a major problem to be tackled. It is to be mentioned here that not only air pollution but also noise and water pollution have been byproducts of industrial growth. There have been attempts by different governments and non-government organizations to curb this menace of pollution. On one side various rules and regulations have been put in place to check the same and on another side there have been made efforts to sensitize the people about the environment. There have been positive results of such exercises but not sufficient. Therefore, it has become necessary to put some value to the environment. Hence environment valuation has become an important tool to reduce the pollution and improve the quality of the environment. There have been efforts from scientists, economists and mathematicians to come out with models for environment valuation. In these models, data analysis has taken an important place and hence methods for data analysis for the valuation have been formulated. An overview of some of such methods is presented in this work.

2. Related Work
Ridker and Hening (1967) [7] used regressing analysis with air pollution as independent variable vis-à-vis property price as a dependent variable. Rosen (1974) [9] extended this theory of impact of air quality on property prices. A theoretical model of willingness to pay was developed by MacDonald *et al.* (1987) to study the impact of probability of flooding on residential location. Jeff E. Brown *et al.* (1995) [1] worked on various functional forms of multiple regression for better data analysis. Jorge Rogat (1998) [8] used multiple regression for the valuation of air quality in Santiago De Chile. Kauko *et al.* (2002) [3] examined neural network modelling with an application to the housing market of Helsinki, Finland. Murty, M N (2003) and Das gupta Purnamita (2004) [4] estimated health benefits from improved air and water quality in India using different mathematical forms of multiple linear regression. Murty MN and SC. Gulati (2005) [5] used the this method to estimate household marginal Willingness to Pay function for urban air quality improvement in the cities of Delhi, Kolkata

and Hyderabad. Thrane (2007) [11] used this methodology to value hotel attributes that affect room rates which include location, facilities and amenities, service quality, star rating, atmosphere, etc. Guozheng Zhang (2008) applied neural network method to analyse data from a China company customer satisfaction survey. Hasan Selim (2009) [10] compared multiple linear Regression with artificial neural network method to determine house prices in Turkey. Qeethara *et al.* (2011) used neural network method in medical diagnosis. Francisco J (2013) used this method for predicting real estate prices in medium size city in South of Spain and compared the same with multiple linear regression method. Felipe *et al.* (2011) estimated the implicit prices of the crime rate and airborne pollution in Chile using regression model. Ebubekir *et al.* (2014) [2] employed regression model for the valuation of apartments in complexes of metropolitan Izmit area of Turkey. They found that structural characteristics of an apartment as well as amenities provided in an apartment complex are important determinants or price and Air Quality has a major effect on price as well. Valencia *et al.* [6] used a general regression neural network for modeling the behavior of PM10 concentration level in Santa Marta, Columbia.

3. Linear Regression Method

In this method, a dependent variable is studied with respect to one independent variable. The independent variable is called explanatory variable or regressor. So to formulate this model, one has to measure environment quality as a variable. Let this variable be *y* and the explanatory variable be *x*. Then *y* is a function of *x*. To formulate the model, data is collated around these two variables. There are different scientific methods for data collation so that there is minimum error and biasedness in this activity. The dependent variable is then regressed vis-à-vis independent variable. The equation for linear regression is given by –

$$y = \alpha_0 + \alpha_1 x + \varepsilon \quad (1)$$

Where,

α's are constant to be determined and ε is the error of model.

In this method, the model is formulated as a linear line between *y* and *x*. The aim is to minimize the error ε. This error is minimized using least square method. The problem with this method is that here we have only one independent variable while the environment quality is found to be dependent on various parameters. Therefore, this method is not of much use as far as environment valuation is concerned.

4. Multiple Regression

In this method, we have the number of independent variables are more than one. The environment quality is taken as dependent variable represented by *y* and let there be *n* number of independent variables. Then the equation for the model is given by-

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots . \alpha_n x_n + \varepsilon \quad (2)$$

Where ϵ is the random error in the model? Here the parameters $\alpha_0, \alpha_1, \alpha_2, \ldots . \alpha_n$ are linear in nature and therefore it is a linear model. Here it is assumed that the errors reprreesented by α's are normally distributed. One of the condition for formulation of this model is that the independent variables should not be related. In this method also, the values of parameters α's are estimated using ordinary least square method. The total error in dependent variable *y* is given by:

$$SST = SSR + SSE \quad (4)$$

Where,

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

And

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y})^2 \quad (3)$$

The coefficient of determination is denoted by R^2. It evaluates the goodness of the fitted model and is given by:

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

The value of R^2 lies between 0 and 1 i.e. $0 \le R^2 \le 1$

The advantage of this method is that this model considers number of independent variables and therefore the dependent variable environment quality can be studied with respect to different variables. But this disadvantage of the method is that the model considers only linear relationship between the dependent and independent variables. The model does not produce better results in case when the relationship is non-linear.

5. Fractional Polynomial Method

In this method, the non-liner relationship between dependent and independent variables is taken care of. This method can be built not only for one independent variable but also for more than one. Therefore, it can be uni-variate or multi-variate method. It is denoted by FP. An FP1 model is defined by –

$$y(x; p) = \alpha_0 + \alpha_1 x^p$$

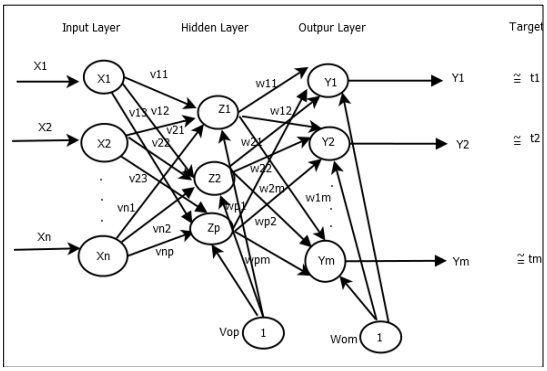
Where *x* is a positive argument and *p* is element of the set –

$$P = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$$

The FP model fitting is done by maximum likelihood. Here the value of α_1 is found using maximum likelihood estimation and this is done using each possible value of *p* from the set *P*.

6. Artificial Neural Network Method

An Artificial Neural Network is like a system which has capability of processing the information which is common with biological neural networks. Since they are nonparametric, therefore they have usage in data analysis, pattern recognition and speech recognition. An artificial neuron is a computational or mathematical model just like a natural neuron in human brain. Training of a neural network model plays an important role in formulation of model. It is done with the help of historical data. Architecture of Artificial Neuron consists of inputs which are multiplied by weights and computed by a mathematical function which gives the activation of the neuron. There are different models of neural network. One of the most used model is Feed Forward Neural Network model which is used with back propagation as learning method.



The inputs are fed to the backpropagation network and the output obtained from it could be either binary (0, 1) or bipolar (-1, +1). The activation function could be any function which is monotonically increasing and is differentiable. The backpropagation network implements delta rule. It is a gradient descendent method which minimizes the total squared error of the output of the network. The accuracy of ANN models is evaluated using three parameters *viz* Sum of Square Error (E), R Square (R^2) and Root Mean Square Error (RMSE). These three parameters are given as:

$$E = 0.5\sum_{k=1}^{n}(t_k - o_k)^2 \tag{7}$$

Where t_k is target output, o_k is the estimated output and n is the number of observations.

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(t_k - \hat{o}_k)^2}{\sum_{k=1}^{n}(t_k - \bar{o})^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(t_k - \hat{o}_k)^2}$$

The model with a higher R^2 and lower *RMSE* will be considered a better model.

7. Conclusion

In this work, fundamentals of different techniques of data analysis for environment valuation are covered. Regression, Fractional Polynomials, Artificial Neural Network methods are discussed. All the methods have some advantages and disadvantages. It depends on the data and the condition which decides the method to be adopted. These methods help policy makers to conduct cost benefit analysis before implementing a policy to improve the environment. A method called hedonic method for environment valuation is based on multiple linear regression and its various forms. There is scope to develop models based on fraction polynomial and neural networks to further enhance the valuation models.

8. References

1. Brown Jeff E, Ethridge Don E. Functional Form Model Specification: An Application to Hedonic Pricing, Agricultural and resource economics review: ARER. - Newark, Del., ISSN 1068-2805, ZDB-ID 11999251. 1995;24(2):166-173.
2. Ebubekir Ayan, Cenk Erkin H. Hedonic Modeling for a Growing Housing Market: Valuation of Apartments in Complexes, International Journal of Economics and Finance. 2014;6(3):188-199.
3. Kauko T, Hooimeijer P, Hakfoort J. Capturing housing market segmentation: An alternative approach based on neural network modeling. Housing Studies. 2002;17(6):875-894.
4. Murty MN, Gulati SC, Banerjee A. Hedonic Property Prices and Valuation of Benefits from Reducing Urban Air Pollution in India, 62/2003, Institute of Economic growth, 2003.
5. Murty MN, Gulati SC. A Generalized Method of Hedonic Prices: Measuring Benefits from Reduced Air Pollution", E/254, Institute of Economic Growth, 2005.
6. Valencia R, Sanchez G, Diaz I. A General Regression Neural Network for modeling the behavior of PM10 concentration level in Santa Marta, Columbia, ARPN Journal of Engineering and Applied Sciences, 2016, 11(11).
7. Ridker RG, John A. Henning. The determinants of Residential Property Values with Special Reference to Air Pollution, The review of Economics and Statistics. 1967;49(2):246-57.
8. Rogat Jorge. The Value of Improved Air Quality in Santiago De Chile", Printed in Sweden Kompendiet-Göteborg, ISBN 91-88514-36-6. 1998.
9. Rosen S. Hedonic prices and implicit markets: product differentiation in pure competition, Journal of Political Economy. 1974 Jan- Feb;82(1):34-55.
10. Selim Hasan. Determinants of House Prices in Turkey: Hedonic Regression versus Artificial Neural Network", Expert Systems with Applications. 2009;36:2843-2852.
11. Thrane C. Examining the Determinants of Room Rates for Hotels in Capital Cities: The Oslo Experience, Journal of Revenue and Pricing Management. 2007;5: 315-323.
12. Zhang W, Cao Q, Schniederjans MJ. Neural Networks earning per share forecasting models: A Comparative Analysis of Alternative Methods", Decision Sciences. 2008;35(2):205-237.