Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.5802

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **5 (6):** 942-947 (2017)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Mathematical Modelling of RMSE Approach on Agricultural Financial Data Sets

S. K. Khadar Babu[1*], M. Sunitha[2] and M. V. Ramanaiah[3]

[1]Assistant Professor Senior, Department of Mathematics, SAS, VIT University, Vellore, TN, India

[2]Department of Statistics, SPMVV, Tirupathi, A.P., India

[3]Professor, Department of Statistics, Sri Venkateswara University, Tirupati, A.P, India

*Corresponding Author E-mail: khadarbabu.sk@vit.ac.in

Received: 28.09.2017 | Revised: 26.10.2017 | Accepted: 30.10.2017

ABSTRACT

In Statistics, Root Mean Square Error Approach (RMSE) plays an important role for Model fitting of agriculture data sets. The Present paper gives the whole information about the RMSE statistical analysis for different heterogeneous areas of data sets with standard parameters in agricultural applications. Initially, we fit a model for crops financial data sets and give complete statistical analysis through the RMSE approach. Finally, It attempts that it is the best measure to compare with the coefficient of determination.

Key words: *Model fit, squared deviations, chi square test.*

INTRODUCTION

Agricultural science is a multidisciplinary field of broad areas of biology that emphasis the parts of exact, natural, economic and social sciences that are used in the practice and understanding of agriculture. Evaluation of association and independence between two categorical factors is a classic topic of interest in statistical inference. Pearson celebrated goodness of fit test yielded the chi square test in the analysis of contingency table. The Chi square test begins by hypothesizing that the distribution of variable behaves in a particular manner. Usually, every data set should follow a specific mathematical model in nature.

Suppose that a variable has a frequency distribution with k categories into which the data has been grouped. The frequencies of the occurrence of the variable for each category are called the observed values. The manner in which the chi square goodness of fit test works is to determine, how many cases there would be in each category if the sample data were distributed exactly according to the claim. These are termed the expected number of cases for each category. The total of the expected number of cases is always the made equal to the total of the observed number of cases.

Linear regression is an approach for modelling the relationship between a scalar dependent variable and one or more explanatory variable. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable is called multivariate linear regression.

Cite this article: Babu, S.K.K., Sunitha, M. and Ramanaiah, M.V., Mathematical Modelling of RMSE Approach on Agricultural Financial Data Sets, *Int. J. Pure App. Biosci.* **5(6):** 942-947 (2017). doi: http://dx.doi.org/10.18782/2320-7051.5802

Ordinary least square method is used to estimate the unknown parameters in a linear regression model. In the regression model developed, the response of the previous reading of the time period has become the predictor. In this case, the order of an auto regression is the number of the immediate preceding values in the series that are used to predict the values at the present time. This model is called First order auto regressive model AR-1. Actually, the mathematical model depends on the yield variable of past qualities.

The Root Mean Squared Error Approach (RMSE) is a vital content in the analysis of different data homogeneous and heterogeneous data applications in science and engineering. In a mathematical modelling, we should obtain the fluctuations using the standardized error approach. In error analysis we consider the residual clustering techniques for model fitting in mathematical sciences. Actually, it illustrates the complete analytical solutions through the method of least squares. Especially, in agricultural sciences, the coefficient of determination only the method for statistical analysis to obtain the percentage of variations occurred in the data sets. In this case, we are not aware to overcome the identification of the percentage of fluctuations raised by the explanatory variable in bulk postings.

The present article supports that the method of RMSE approach is comparatively good enough in agricultural and medical sciences. We can identify and known error of the variables in any number of data sets and we can easily predict the error by taking the different explanatory. The analysis of the statistical data can easily analyzed by using the approach.

REVIEW OF LITERATURE

The generalization of squared R was proposed by Cox and Snell (1989, 208-209) and, apparently independently, by Magee (1990), but had been suggested earlier for binary response models by Maddala (1983).

Pradip Saha et al., in tea quality prediction by autoregressive modelling of electronic tongue signals have used an autoregressive model to predict tea quality. An autoregressive model was developed based on the responses collected from the electronic tongue (ET) sensors and AR moving averages were also implemented. The responses obtained from the sensor are characterized with the tea samples from different qualities. Coefficients of the autoregressive model are used as the characteristics of the ET response corresponding to the different samples of tea quality. Proposed method is useful for the prediction of tea quality.

Shota Tanaka et al., in Footstep Modelling and Detection using locally stationary autoregressive model developed a footstep based surveillance system. Accurate footstep analysis would be powerful tool in various applications.

Lawrence Lin and Lynn. D (1996) developed by the new statistic, the Coefficient of Accuracy, Ca, for methods comparison. When an old measurement method is compared to a new measurement method or if the same method is compared in two laboratories, the Coefficient of Determination, r2, is typically used to measure the relationship. However, r2 only measures the precision of the relationship. The newly developed statistic, Ca, measures the accuracy of the relationship. When these two statistics are combined together, they form a single statistic for both accuracy and precision called the Concordance Correlation Coefficient, etc.

Nagelkerke (1991), gives a generalization of the coefficient of determination squared R to general regression model is discussed and proposed a modification of earlier definition to allow.

Durhan J L and L. Stockburger (1986) has used the coefficient of determination for his research work entitled Nitric Acid – air diffusion coefficient experimental determination.

METHODS AND DISCUSSIONS

Actually, here we define two variables one is Explanatory and the other is explained. Let us consider X and Y are the variables denotes explained and explanatory in this case. Present

article gives the importance of the concept root mean square error approach. In model fitting, the squared R plays a role but it had some limitations to analyze the error in terms of percentage fluctuations. The RMSE directly gives the standard error involved in the data and it is useful to infer that the model was best fit.

The data for analysis was taken from the Central agricultural statistical Bulletins. Specifically, the two variables, one is year and the other is subsidy given to farmers for crops. For this data fit a model and illustrate the RMSE and squared R. Finally, compare the values and give appropriate findings. Every year Government of India gives rupees in crores for purchase of crops through subsidy mode for upliftment of the agriculture in India. Now we are taking consecutive years of 2013,14,15,16, and 17 and their corresponding subsidies in rupees given by the Government for model fitting. The following diagrams show the appearance of the model in the data.

Fitted model for the data is Y=7602.1X-2E+07, where X is the years and Y subsidiaries. And squared R for model fit was 0.0112 and exponential smoothing Y= exp (0.0057X) and its squared R 0.1221. By RMSE, we can get exact error occurred in the data sets and give the possible conclusions for any specified problem. In the present example squared r value is 0.8357, which means that the data contains 83% of the error occurred in the data due to the effect of the given explanatory by taking exponential smoothing curve. By seeing the linear trend, the squared R value is 0.0112, which means that contains very small error between the explained and explanatory variables. And the trend equation is given on the diagram. All possible and useful diagrams are drawn for writing conclusions.

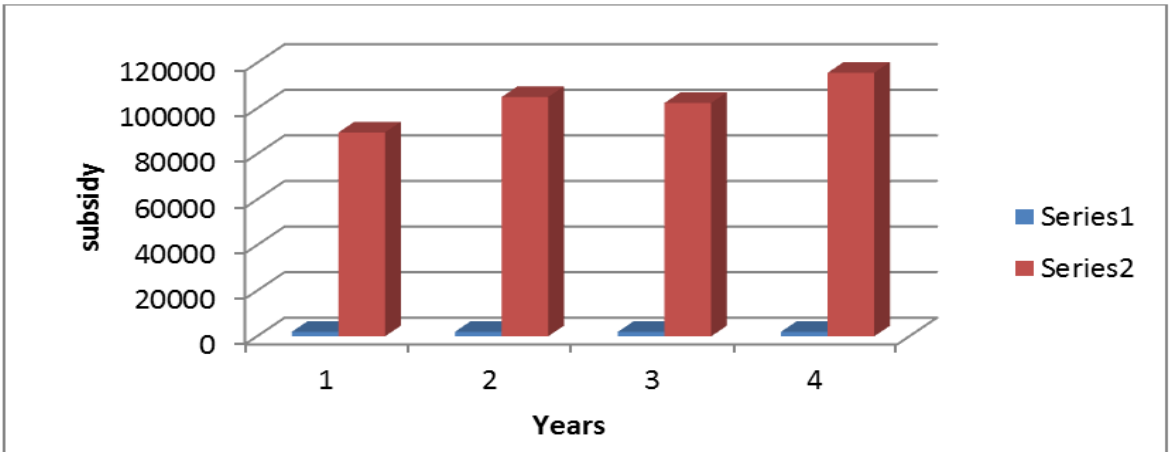

Fig. 1: Bar diagram for subsidiaries

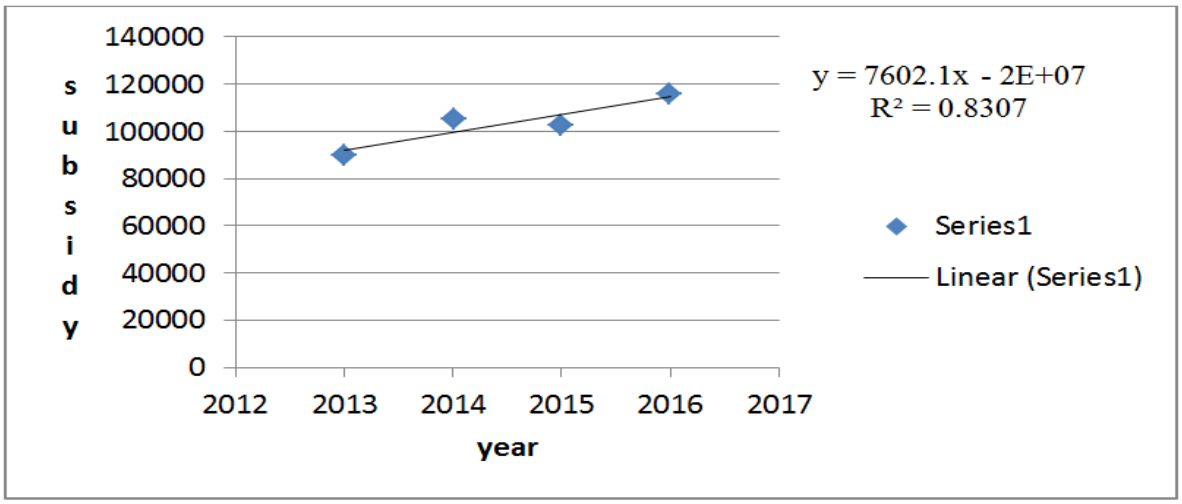

Fig. 2: Trend line with squared R

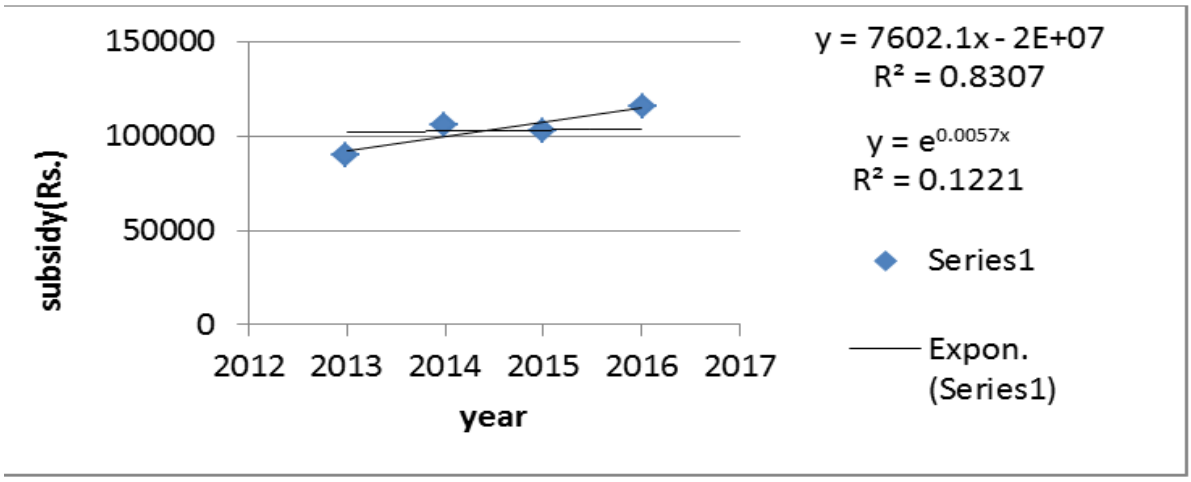

Fig. 3: Exponential smoothing for subsidy in Rs.

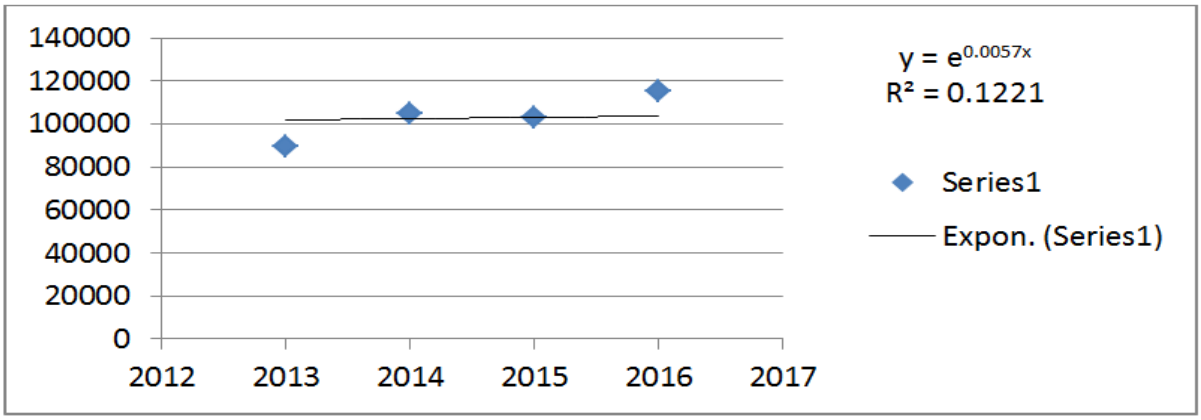

Fig. 4: Bar diagram for subsidy

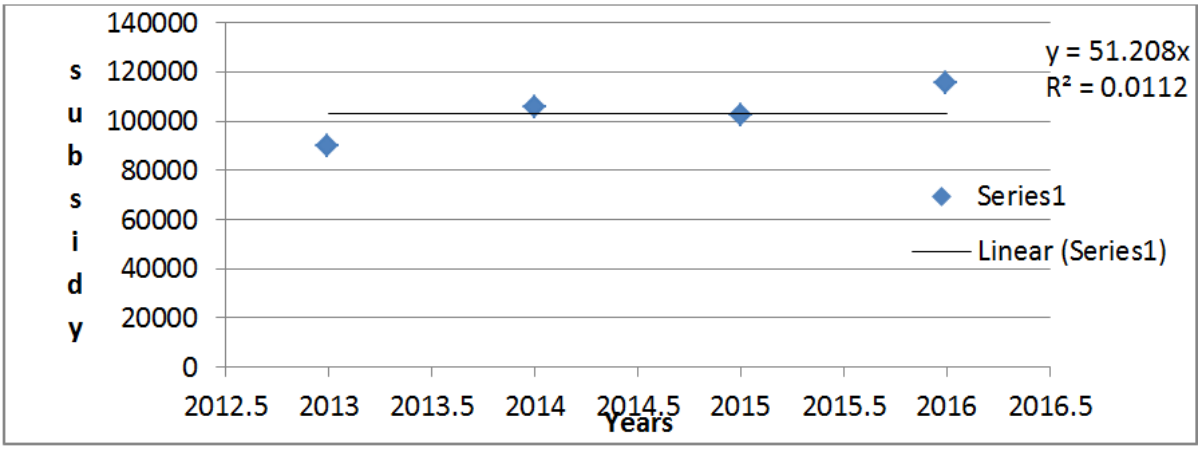

Fig. 5: Exponential smoothing

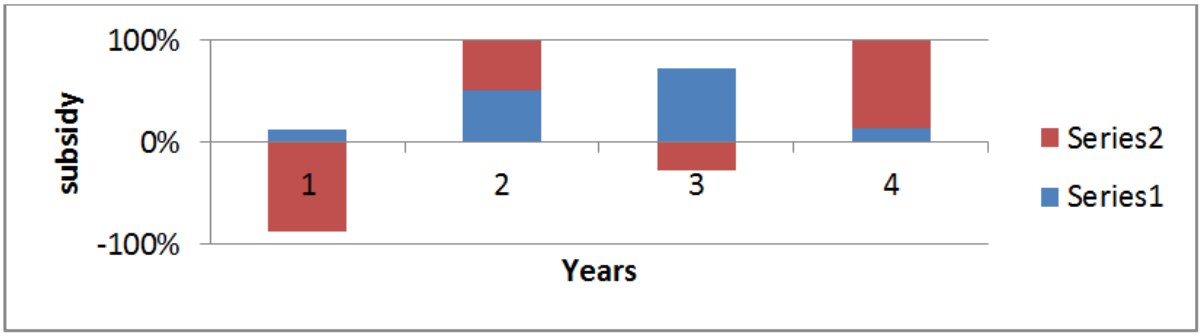

Fig. 6: Linear trend with R squared

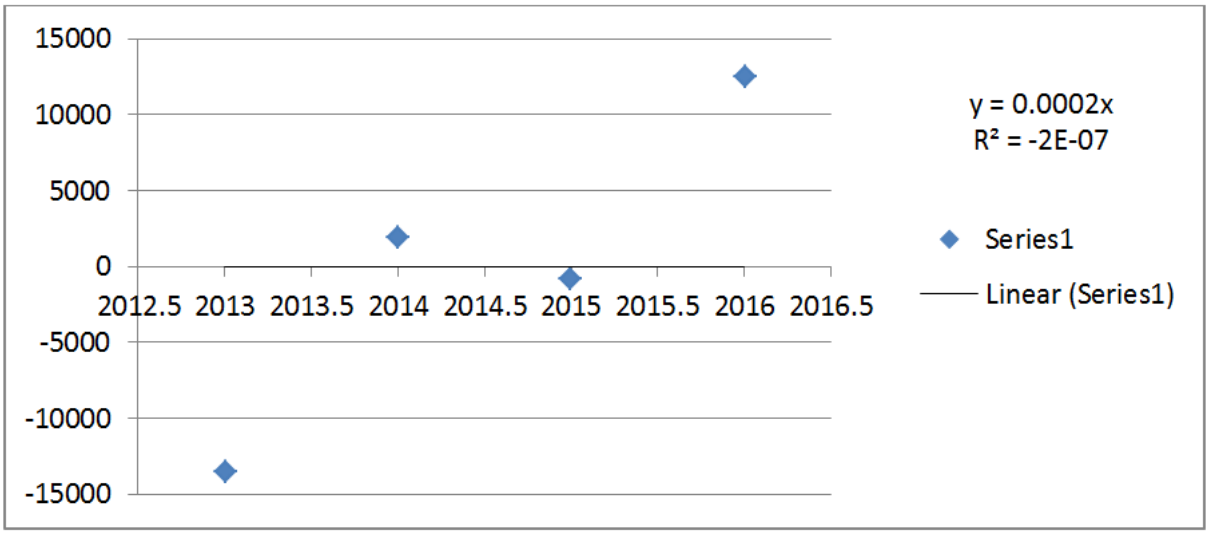

Fig. 7: Bardiagram for residuals

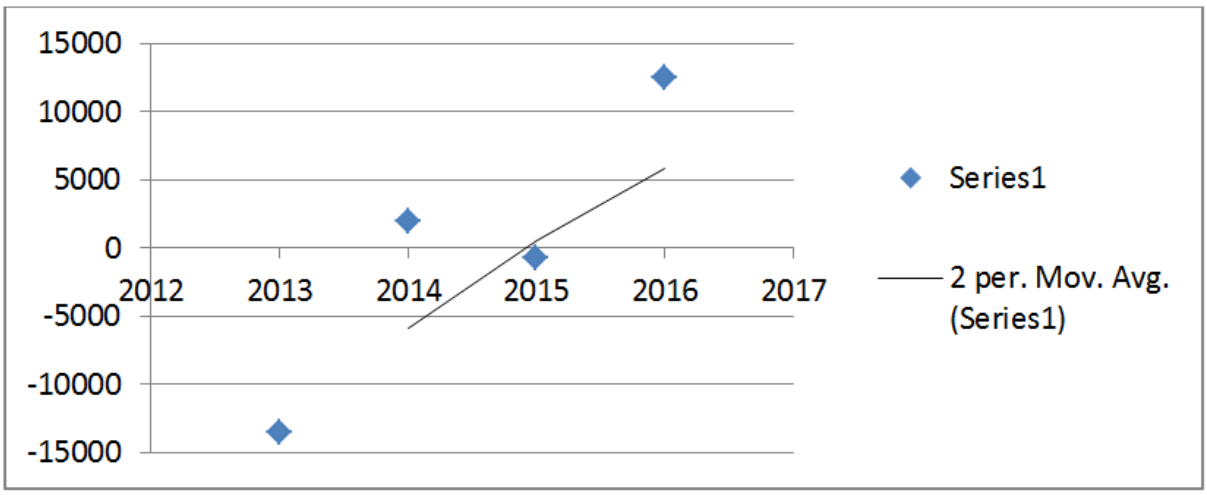

Fig. 8: residual Linear trend line

The fitted regression line Y = 51.208 X, where X is explained variable and Y is Explanatory.

The RMSE = SQRT $\sum (Y_i - Y_i^{\delta})^2$

= 18546.23

The corresponding coefficient of variation is 0.0112, It gives the percentage of fluctuations.

CONCLUSIONS AND FINDINGS

In the present article. We estimate the values of squared R and RMSE for data taken from the statistics of the Government of agriculture, India. We observe that the values obtained through formulas of squared R and RMSE approach. RMSE gives the exact error occurred in the data but squared R does not give any error value but it give the value between 0 to 1 and convert it in to the percentages, then only possible to give percentage of the fluctuations occurred in the data. Ultimately, by comparison of the RMSE and squared R, slightly RMSE is comparatively superior to the squared R. In agriculture sciences, the people have aware about the squared R only. In engineering applications, they need exact error and then they have a easy way to reduce error. Engineering, the research was developed broadly compare with other areas. In agriculture, they are unable to think about to reduce the error but they need only the percentage of error occurred in the data sets.

REFERENCES

1. Tanaka, S., Mizuno, K. and Itai, A., Footstep modeling and detection using locally stationary autoregressive model, *10th International Conference on Information, Communications and Signal*

Processing (ICICS), Singapore. 1-5 (2015).

2. Durhan, J.L. and Stockburger, L., Nitric acid-air diffusion coefficient experimental determination, *Atmospheric Environment*, **20(3):** 559-563 (1986).
3. Lawrence, L., and Lynn, D., Torbeck, Coefficient of Accuracy and Concordance Correlation Coefficient: New Statistics for Methods Comparison, *PDA Journal of Pharmaceutical Science And Technology*, **52(2):** 55-59 (1996).
4. Nagelkerke, N.J.D., A note on a general definition of the coefficient of determination, *Biometrika*, **78(3):** 691-692 (1991).
5. Cox, D.R., Regression models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological),* **34(2):** 187-220 (1972).
6. Cox, D.R., Partial likelihood, *Biometrika*, **62(2):** 269-276 (1975).
7. Maddala, G.S., Limited-dependent and qualitative variables in econometrics, Cambridge University Press, New York, 1983.