



RESEARCH PAPER

A study on cross validation for model selection and estimation

Fehim Jeelani Wani*, S.E.H. Rizvi¹, Manish Kumar Sharma¹ and M. Iqbal Jeelani Bhat¹

Division of Agricultural Statistics (SKUAST-K), SHALIMAR (J&K) INDIA

(Email : faheemwani149@gmail.com)

Abstract : In the present study, k -fold cross validation method was examined for performance evaluation of different regression models. A multistage sampling technique was adopted for the selection of samples in which districts, villages within districts and fodder trees in the selected village formed the first stage, second stage and third stage units, respectively. A total number of 10 trees were randomly selected from each village so as to constitute a predetermined total sample size of 60 trees. Primary data on height, bole height, diameter at breast height (dbh), no. of primary branches, secondary branches, average no. of leaves per secondary branch, age, canopy diameter and green fodder yield (dependent variable) for each selected tree were collected through visiting farmers field in the selected area and by adopting standard forest mensuration procedures. Regression analysis was used to study the relationship between fodder yield (dependent variable) and other parameters. Different regression models were tried and on the basis of adj. R^2 , the best five models were selected. Goodness of fit of the selected models was tested by applying chi-square test. The chi-square test results came out to be insignificant indicating thereby that the models under study were qualified for goodness of fit and could be used for further study. The models were validated for its adequacy through different criteria, namely, adj. R^2 , bias, variance, root mean square error and coefficient of dispersion. On the basis of set criteria, the models were ranked. After applying the Wilcoxon signed rank test on fitting data set, one can arrive at the final ranks by considering ranks of both fitting (R_p) and validating (R_v) data sets. Finally, on the basis of all the criteria adopted in the present investigation, the regression model obtained as $\hat{Y} = 8.480 + 0.000004L^2S$ ranked first, where \hat{Y} = estimated fodder yield, L = avg. no. of leaves per secondary branch (S) and hence, recommended for fodder yield prediction of *Grewia optiva* for the present study area.

Key Words : Cross validation, Regression analysis, Goodness of fit, *Grewia optiva*

View Point Article : Wani, Fehim Jeelani, Rizvi, S.E.H., Sharma, Manish Kumar and Bhat, M. Iqbal Jeelani (2018). A study on cross validation for model selection and estimation. *Internat. J. agric. Sci.*, **14** (1) : 165-172, DOI:10.15740/HAS/IJAS/14.1/165-172.

Article History : Received : 16.03.2017; Revised : 26.11.2017; Accepted : 09.12.2017

* **Author for correspondence:**

¹Division of Statistics and Computer Science, Faculty of Basic Science (SKUAST-J), Main Campus, CHATHA (JAMMU) INDIA