DOI: 10.15740/HAS/AU/12.TECHSEAR(3)2017/635-640 *Agriculture Update*
Volume **12** | TECHSEAR-3 | 2017 | 635-640
Visit us : *www.researchjournal.co.in*

RESEARCH ARTICLE :

Discriminant analysis for prediction and classification of farmers based on adoption of drought coping mechanisms

■ G.R. HALAGUNDEGOWDA, ABHISHEK SINGH AND H.K. MEENAKSHI

ARTICLE CHRONICLE :
Received :
10.07.2017;
Accepted :
25.07.2017

SUMMARY : The study was carried out to develop the classificatory statistical model to predict and classify the farmers into adopters and non-adopters in Kolar district of Karnataka for the year 2013. Linear discriminant analysiswas carried out by considering the various socio-economic characteristics of farmers as predictors and adoption behaviour of the farmers as response variable in order to assess the factors influencing on adoption of drought coping mechanisms. The result shows that the Box's M test is 161.3 with their F approximation 1.83 is non-significant (0.19) at 5% level of significance, Eigen value (2.51) of the first function explains 100% of variations in the data which is potential enough to classifying the groups, wilk's lambda associated with the function (λ=0.28) is transforms to a chi square of 140.82 with 12 DF, which is statistically significant and the following variables such asFarm Size (0.552), Extension Visits (0.574), Crop Diversification (0.321) and Crop Insurance (0.368) are relatively more important and positively influencing on discrimination of farmers group. Whereas the variable like Age (-0.516) negatively influencing on discrimination of adopters and non-adopters.

KEY WORDS :
Farmers, Coping mechanisms, Crop diversification

How to cite this article : Halagundegowda, G.R., Singh, Abhishek and Meenakshi, H.K. (2017). Discriminant analysis for prediction and classification of farmers based on adoption of drought coping mechanisms. *Agric. Update*, **12**(TECHSEAR-3) : 635-640; **DOI: 10.15740/HAS/AU/12.TECHSEAR(3)2017/635-640.**

Author for correspondence :
G.R. HALAGUNDEGOWDA
Section of Agricultural Statistics, Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, VARANASI (U.P.) INDIA
Email : hgowda8127@gmail.com
See end of the article for authors' affiliations

BACKGROUND AND OBJECTIVES

The State of Karnataka has 114 lakh hectare cultivable lands and 72 per cent of the cultivable area is rainfed; only 28 per cent is under irrigation (GOK News, 2015). The State is the second largest in terms of arid region and it ranks second, next only to Rajasthan in India, in terms of total geographical area prone to drought. Drought is a common phenomenon in State of Karnataka. The State faced consecutive droughts during the years 2001-02, 2002-03 and 2003-04 resulted in sharp decline of agricultural output(Nagaratna and Sridhar, 2009).

Drought stress is the major limiting factor for rice production and yield stability under rainfed crop eco system. Karnataka faces high risk of moisture stress at maximum tillering and reproductive stages of crop, which may lead to yield loss of 25 to 100 per cent (Hanamaratti *et al.*, 2008).

Adoption of drought coping mechanisms :

Drought is defined as "when a region receives below average precipitation, resulting in prolonged shortages in its water supply, whether atmospheric, surfaceor ground water. It can have a substantial impact on the ecosystem and agriculture of the affected region". As drought occurs in a particular area obviously its affects the crop and livestock production, in order to reduce the effect of drought on farm production and to stabilize the farm income, farmers have to take some systematic measures such measures are called drought coping mechanisms.

Classification techniques:

To classify objects based on their features and characteristics is one of the most important and primitive activities of human beings. Objects displaying similar features and properties based on certain pre-specified criteria are classified into the same group or category. For example in agriculture, crops are classified into different groups based on growing cycle (temporary/ permanent), crop variety (hybrid/ordinary), season (winter/spring crop), forewarning of incidences (say, presence or absence) of crop pests and diseases, etc.Classification has two distinct meanings. We may be given a set of observations with theaim of establishing the existence of classes or clusters in the data. Or we may know forcertain that there are so many classes, and the aim is to establish a rule whereby we canclassify a new observation into one of the existing classes. The former type is knownas *Unsupervised Learning* (or *Clustering*), the latter as *Supervised Learning.*

Research problem:

One of the problems of classification lies in the use of appropriate methods to fit the model depending on the nature of data. It is well known that, most of data related to adoption of any agriculturetechnology (Agriculture Extension Survey data) are having qualitative response variable with two or more categories, which is a problem when using the traditional statistical methods, such as linear regression analysis because of not satisfying the assumptions of quantitative regresend in classical linear regression model. In such case to measure the farmer's perception towards adoption of particular agriculture technology, we can think of qualitative response models such as logit model, probit model, tobit model, poison regression and multivariate techniques like linear discriminant analysis.

RESOURCES AND METHODS

The specific reason for choosing this study area was that, Kolar district belongs to Eastern Dry Zone of Karnataka and most of farmers were involved in rainfed agriculture because of shortage of rainfall and drought affected area. Hence adoption of certain coping strategies against drought is the major solution to stabilize the farm income during the drought period. The specific reason for choosing this study was toknow the factors influencing on adoption of any strategies against drought and its impact of agriculture policyon Karnataka agricultural cropping pattern and how it's fluctuating from period to period and area to area when the drought occurs.

Nature and source of data :

The current study utilizes both classification and prediction techniques. The household secondary data was used to fit the classificatory statistical models and the data were recorded on Socio-characters of farmers of Kolar districts of Karnataka (India). The data is mainly related to coping strategies implemented against drought by the farmers of this region and was collected by employing the multi stage sampling design during the year 2013-14, the department of Agricultural Economics (CARDS), Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu (India).

Table A : Variables encoding summary

Code	Variables	Measurement
Y	Adoption behaviour	Y= 0 for Non-Adopters = 1 for Adopters
X_1	Age of the farmer	Number of years
X_2	Education of the farmer	Formal years of education
X_3	Household size	Number of family members
X_4	Farm size	Number of acre's
X_5	Farming experience	Number of years
X_6	Animal husbandry	Number of farm animals and poultry birds
X_7	Media exposure	Number of sources exposed frequently
X_8	Extension visits	Number of Visits made to an research organisations
X_9	Crop diversification	Number of crops grown in that year
X_{10}	Income status	In Rupees (Rs.)
X_{11}	Worth of liquidating assets	In Rupees (Rs.)
X_{12}	Crop insurance got by the government	In Rupees (Rs.)

Discriminant analysis :

Discriminant analysis is a multivariate technique concerned with classifying distinct set of objects and allocating new objects or observations to the previously defined groups. It involves deriving variates which are combination of two or more independent variables that will discriminate best between a priori defined groups.

Linear discriminant function :

If the population covariance matrices are equal then linear discriminant function for classification is used, otherwise quadratic discriminant function is used for this purpose. The maximum number of discriminant functions that can be computed is equal to minimum of G-1 and p, where G is the number of groups and p is the number of variables. Suppose the first discriminant function is

$$Z_1 = W_{11}X_1 + W_{12}X_2 + \ldots\ldots W_{1p}X_p,$$

where, the W_{1j} is the weight of the j^{th} variable for the first discriminant function. The weights of the discriminant function are such that the ratio

$$\lambda_1 \; Z \; \frac{\text{Between groups SS of } Z_1 \text{ is maximized}}{\text{Within groups SS of } Z_1}$$

Suppose the second discriminant function is given by,

$$Z_2 = W_{21}X_1 + W_{22}X_2 + \ldots\ldots W_{2p}X_p$$

The weights of above discriminant function are estimated such that the ratio

$$\lambda_2 \; Z \; \frac{\text{Between groups SS of } Z_2}{\text{Within groups SS of } Z_2}$$

Is maximized subject to the constraint that the discriminant scores Z_1 and Z_2 are uncorrelated. The procedure is repeated until all possible discriminant functions are identified. Once the discriminant functions are identified, the next step is to determine a rule for classifying the future observations.

OBSERVATIONS AND ANALYSIS

Table 1 display the result of Box's test of equality of covariance's matrices in the form of value of test statistics, degrees of freedom and their significance level. One of the major assumptions of multivariate analysis is the equality of population covariance's matrices. It tests whether the covariance matrices are same among the two adoption groups of the dependent variables and the null hypothesis stated as the equality of population covariance matrices. If the test is not significant then there is equality of covariance matrices across the category otherwise the assumption is violated.

Table 1 : Box's test of equality of covariance's matrices

Box's M		161.302
F	Approx.	1.837
	df1	78
	df2	35362.583
	Sig.	0.190

$\mathbf{H_0}$: $p_1 = p_2 = p_3 = \ldots\ldots\ldots\ldots = p_p$

where, Σ's are the Covariance matrices of the p-populations.

In the above table Box's M test is 161.3 with their F approximation 1.83 is non-significant (0.19) at 5% level of significance. Which states that the equality of population covariance's matrices across the category and gives a clue to proceed the analysis. If *N's* for the both groups of adoption (Dependent variable) are approximately equal, then the Box test should be ignored.

Table 2 : Canonical discriminant co-efficient summary

Function	Eigen value	% of variance	Cumulative %	Canonical correlation
1	2.516*	100.0	100.0	0.846

* First Canonical Discriminant functions were used in the analysis

There are many ways to decide the numbers of functions which are sufficient to classify the groups, among them the Eigen value and associated % of variance is major one to decide the number of functions. The eigenvalue of each discriminant function reflects the ratio of importance of the dimensions which classify cases of the dependent variable. It meansbetween-groups sums of squares divided by within-groups sums of squares. A large eigenvalue is associated with a strong function.

Table 2 explains that the Eigen value and corresponding variance explained by the discriminant function from the whole data. An Eigen value represents the amount of variance associated with the function. If there is more than one discriminant function only the function with Eigen values greater than 1.0 is retained and other functions which are less than one are discorded out of the model. In the above table Eigen value (2.51) of the first function explains 100% of variations in the data which is potential enough to classifying the groups. The present canonical correlation is 0.846, which indicates strong association and the selected function strong enough to discriminate the groups.

Table 3 : Wilk's Lambda

Test of function	Wilks' Lambda	Chi-square	df	Sig.
1	0.284	140.824	12	0.000

The significance level is estimated based on the chi square transformation of statistic. A Chi square test based on lambda indicates whether the variability is systematically related to group differences is statistically significant or not. In the above table wilk's lambda associated with the function (λ=0.28) is transforms to a chi square of 140.82 with 12 DF, which is statistically significant (sig=0.000) at 5% (<0.05) level of significance. If the null hypothesis is rejected at 5% means the selected discriminant function is statistically significant and it is potential enough to discriminate the groups. That is the model is good fit.

Table 4 provides the standardized canonical discriminant co-efficients which explain the relative importance of the each predictor on discrimination of adopters and non-adopters. The sign indicates the direction of the relationship and magnitude indicates extent of contribution to the group discrimination. The predictors like Farm Size (0.552), Extension Visits (0.574), Crop Diversification (0.321) and Crop Insurance (0.368) are relatively more important and positively influencing on discrimination of groups. Whereas the variable like Age (-0.516) negatively influencing on discrimination of adopters and non-adopters.

Table 4 : Standardized canonical discriminant function co-efficients

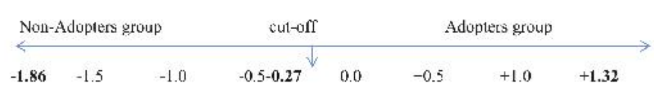
Variables	Function 1
Age	-.516
Education	.282
Household size	-.111
Farm size	.552
Farming experience	-.186
Animal husbandry	-.015
Media exposure	.238
Extension visits	.574
Crop diversification	.321
Income status	.296
Liquidating assets	.085
Crop insurance	.368

Table 5 represents the group centroids, the group centroids indicates the mean discriminant scores of the members of a group on a given discriminant function. For classification and prediction purposes, the discriminant score of each group case (each individual) is compared to each group centroid and the probability of group membership is calculated. The closer a score is to a group centroid, then the greater the probability the case belongs to the group.

Table 5 : Functions at group centroids

Adoption groups	Function1
Non adopters	-1.861
Adopters	1.329

Group centroids reveal how much and in what ways the groups are differentiated on each function. The absolute magnitude of the group centroids indicates the degree to which a group is differentiated on a function and the sign of the centroid indicates the direction of differentiation. In the above table the function 1 is discriminates the adopters from the non-adopters groups. Adopters scored at the positive end (1.329) on the bipolar function and non-adopters at the negative end (-1.861) of the function. Most of classification and prediction purposes first fit the discriminant function by considering the unstandardized co-efficients and then find out the cut off value by using these two extreme centroid values. In our case -0.25 is the cut off value for classification purpose.



Predict the outcome for every individual respondent in the sample by using discriminant function. If the predicted outcome is below the cut-off value or mid value (<-0.27) then such respondents are grouped as Non-Adopters, otherwise the predicted outcome is above the cut-off value (>-0.27) then such respondents are grouped as Adopters.

Table 6 provides the classification table for both training data set and validation data set, the classification table helps to assess the performance of the model by cross tabulating the observed response categories with

Table 6: Classification matrix

Sample	Observed	Predicted		
		Non adopters	Adopters	Per cent correct
Training set	Non adopters	48	2	96.00%
	Adopters	4	66	94.30%
	Overall per cent			95.15%
Testing set	Non adopters	11	1	91.70%
	Adopters	2	16	88.90%
	Overall per cent			90.30%

the predicted response categories and shows how well our full model correctly classifies the cases. Classification Results is a simple summary of number and per cent of subjects classified correctly and incorrectly. The leave-one out classification is a cross-validation method, of which the results are also presented.

The rules works for each case such that the predicted response category treated as adopters, if that category's predicted outcome is greater than the user-specified cut-off (>-0.27), otherwise it's treated as 0. In case of training data set out of 50 non-adopters cases 48 cases are correctly categorised as non-adopters and remaining 2 cases are wrongly classified in to adopters section which means 96 % accuracy for non-adopters and in the same way out of 70 adopters cases 66 respondents are correctly categorised as adopters and only 4 respondent is mismatched and grouped in to non-adopters section, which means 94.3% accuracy for adopters. Themain information is the overall percentage in the lower right corner which shows our discriminant function is 95.15% accurate andit is excellent.

Moving to the validation data set, 11 non-adopter cases out of 12 instances are correctly categorised as non-adopters and only 1cases were wrongly classified in to adopters section which means 95.15% accuracy for non-adopters and in the same way out of 18 adopters cases 16 respondents are correctly categorised as adopters and only 2 respondent is mismatched and grouped in to non-adopters section, which means 88.9 % accuracy for adopters. The overall percentage of accuracy is 90.30%. The result of both training and validation classification tables facilitate to compare the efficiency and classification ability of the model.

Comparison made with consideration of the overall percentage of accuracy and the result of training data is quite good. Comparatively the training data shows 95.15% accurate than the validation data set (90.30%) and it may due to sample efficiency. When we take in to consideration of all factors and in broad view there is not much difference in the result of training and testing set. The model behaves good way for both training and testing data set in terms of effective prediction and classification.

Authors' affiliations :

ABHISHEK SINGH, Section of Agricultural Statistics, Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, VARANASI (U.P.) INDIA

H.K. MEENAKSHMI, Department of Agricultural University, Tamil Nadu Agricultural University, COIMBATORE (T.N.) INDIA

REFERENCES

Aquilanti, L.,Santarelli, S., Babini, V., Osimani, A and Clementi, F. (2013). Quality evaluation and discrimination of semi-hard and hard cheeses from the Marche region (Central Italy) using chemo metric tools. *Internat. Dairy J.*, **29** : 42-52.

Arafat, S.M., Aboelghar, M.A and Ahmed, E.F. (2013). Crop discrimination using field hyper spectral remotely sensed data. *Adv. Remote Sensing*, **2** : 63-70.

Gayathiri, G.R. (2015). Discriminant analysis of conventional and Sri Method in thanjavur district.Zenith.*Internat. J. Business Econ. & Mgmt. Res.*, **5** (8): 68-77.

Gwary, M.M., Gwary, T.M and Mustapha, S.B. (2012). Discriminant Analysis of the Influence of Farmers' Socio-Economic Characteristics on their Participation in Research and Extension Activities in Borno State, Nigeria.*Internat. Res. J. Soc. Sci*., **1** (4) :1-6.

Gwaza, D.S., Tor, N.E.T. and Wamagi, T.I. (2013). Discriminant analysis of morphological traits in selected population of the Tiv Local Chicken Ecotype in the Derived Guinea Savanah of Nigeria.*J. Agric. & Vet. Sci*., **3** (6): 60-64.

Hanamaratti, N.G., Prashanthi, S.K., Salimath, P.M., Hanchinal, R.R., Mohankumar, H.D., Parameshwarappa, K.G., and Raikar, S.D. (2008). "Traditional Land Races of Rice in Karnataka: Reservoir of valuable traits", *Curr. Sci.*, **94**(2): 242-247.

John, N., Iheanacho, A.C and Irefin, D. (2011). Effects of Scio-

Economic Characteristics of food crops farmers on the selection of coping strategies against drought in Bourno state, Nigeria. *Lincoln University J. Sci.*, **2** (1): 13-18.

Kusbach, A., Shaw, J.D. and Long, J.N. (2014). Discriminant analysis reveals limited association between forest habitat types and the environment inWestern United States land classification. *Applied Veg. Sci*., **7** (8): 2-10.

Lekshmi, P.S., Chandrakandan, K. and Balasubramani, N. (2013). Discriminating Factors of High and Low Adopters of Shrimp Farmers in Nagapattinam of Tamil Nadu. *Indian Res. J. Extn. Edu.*, **13** (1): 120-125.

McDonald, L.S., Panozzo, J.F., Salisbury, P. A and Ford, R. (2016). Discriminant Analysis of Defective and Non- Defective Field Pea (*Pisumsativum* L.) into Broad Market Grades Based on Digital Image Features.*PLOS ONE Online J.*,**11**(5): e0155523.

Meng, L., Ford, T and Guo, Y. (2016).Logistic regression analysis of drought persistence in East China. *Internat. J. Climatology,* **5**(4):2-9.

Nagaratna, Biradhar and Sridhar (2009). A study on Consequences of 2003 Drought in Karnataka with reference to Livestock and Fodder", *J. Human Ecol.*, **26**(2): 123-130.

Ogbanje, C.E., Chedibelu, S. and Nweze, N.J. (2014). Discriminant function analysis of factors affecting off-farm diversification among small-scale farmers in North Central Nigeria. *J. Econ. & Sustainable Development*, **5** (13): 127-135.

Olaleye, O.L. (2010). Drought Coping Mechanisms: A Case Study of Small Scale Farmers in Motheo District of the Free State Province. Master of Science thesis, University Of South Africa.SA.

Piraiesh, Abotaleb, Ebrahimi, Mohammad,S., Abedi-Koupai and Jahangir (2015).Discriminant analysis to identify the farmers for develop the sprinkler irrigation systems in Iran. *J. Reports & Opinions*.**7** (7): 74-77.

Rakgase, M. and Norris, D. (2014).Factors that Influence Choice of Drought Coping Strategies in Limpopo Province, South Africa.*J. Human Ecol*., **47**(2): 111-116.