International Journal of Agriculture Sciences
ISSN: 0975-3710&E-ISSN: 0975-9107, Volume 10, Issue 6, 2018, pp.-5593-5597.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000217

Research Article

STATISTICAL MODELS FOR CLASSIFICATION OF GENOTYPES FOR YIELD OF LITTLE MILLET

NAGARAJA M.S. AND SINGH ABHISHEK

Section of Agricultural Statistics, Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, 221005, Uttar Pradesh
*Corresponding Author: Email-msn8129@gmail.com

Received: March 13, 2018; Revised: March 27, 2018; Accepted: March 28, 2018; Published: March 30, 2018

Abstract- Use of statistical models such as Ordinal Logistic Regression Model and Multiclass Discriminant Model for classification of genotypes or creation of genetic variability is undergoing an outpouring in interest among research workers. These models were fitted to data recorded on yield and yield attributing characters of 722 genotypes of little millet and the data has been collected from Project coordination cell, All India Coordinated Small Millets Improvement Project (AICSMIP), ICAR, Bengaluru. Classification ability measures such as Accuracy Rate, Kappa Statistics, Avgprecision, and Avgrecall were used for testing samples. Days to fifty percent flowering, Plant height, Number of basel tillers, Flag leaf length, Flag leaf width were considered to be important attributing characters for classification and Ordinal Logistic Regression Model (56.55%) was performed better than Multiclass Discriminant model (53.79%) for classification of genotypes for different classes of yield of little millets.

Keywords- Ordinal Logistic Regression Model, Ordinal Logistic Regression Model, Attributing, Classification and Significant.

Citation: Nagaraja M.S. and Singh Abhishek (2018) Statistical Models for Classification of genotypes for yield of Little Millet. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 10, Issue 6, pp.-5593-5597.

Academic Editor / Reviewer: T Kumareswari

Introduction

Millets are small-seeded grasses that are hardy and grow well in dry zones as rain-fed crops, under marginal conditions of soil fertility and moisture. Millets have been known to humans since long period of time and the first cereal grain to be used for domestic purposes. Which are also unique due to their short growing season. They can develop from planted seeds to mature, ready to harvest plants in short period of time. Little Millet (*Panicum sumatrense)* is one of the small millets is indigenous to Indian subcontinent. The crop is known by different names as kutki in Hindi, same in Kannada, samai in TamilNadu and samulu in telagu. The crop is well known in Tamil Nadu, Karnataka, Andra Pradesh, Maharashtra, Madhya Pradesh, Jharkhand, Chhattisgarh, Orissa and Gujarat. It can be harvested within 70-75 days marking it an ideal as excellent catch crop in multiple and relay cropping. Little millet is well known for its drought tolerance and is considered as one of least water demanding crop [1]. The classification of genotypes for different classes of economically important characters is one of the way to separation or partition of genotypes among themselves, which helps to provide information on genetic variability and parentage for crossing in crop breeding programmes. In plant breeding programme, to increase the productivity breeder needs to maintain a pool of diverse desirable donor parents [2]. Moreover, evaluation of genetic diversity is important to know the source of genes for a particular trait within the available germplasm [3]. Hence, the present study was aimed to classification of genotypes of little millet for different classes of yield using statistical model.

Material and Method

The secondary data was collected on yield and yield attributing characters of little millet such as Yield, Days to 50 per flowering (Maturity), Peduncle length, Flag leaf Length, Flag leaf Width, Flag leaf sheath length, Number of Basel tillers, Length of inflorescence, 1000 grain weight and Plant height from Project coordination cell, All India Coordinated Small Millets Improvement Project (AICSMIP), ICAR, Bangalore. The yield of little millet has been classified as low, medium and high yield classes and all other yield attributing characters were considered as independent variable(predictors).The data set is divided randomly into training data consists of 80% of data (577 genotypes) and test data consists remaining 20% data (145 genotypes).The genotypes having a less than 10 gram per plant grain yield is considered as low yield genotype, 10 gram to 15 gram per plant grain yield as medium yield genotype, greater than 15 gram per plant grain yield genotype as high yield genotype. Ordinal Logistic Regression Model and Multiclass Discriminant Model were fitted to data for classification of genotypes, which were used to classify and predict the classes of yield of little millet, data has been analyzed by using the R version 3.3.1 statistical package and SPSS 22.0 statistical package respectively. Ordinal Regression model (also known as Ordinal Logistic Regression Model) is another extension of binomial logistics regression model. Consider the following simple ordinal logistic regression model with single predictor variable and a response variable:

$$Y_i=\beta_0+\beta_1X_i+\varepsilon_i, \qquad i = 1, 2, \ldots, n \qquad [3.1]$$

$$logit(p_1) = log\frac{p_1}{1-p_1} = \alpha_1 + \beta'x \qquad [3.2]$$

$$logit\ (p_1 + p_2) = log\frac{p_1 + p_2}{1 - p_1 + p_2} = \alpha_2 + \beta'x$$

$$logit(p_1 + p_2 + \cdots + p_k) = log\frac{p_1+p_2+\cdots+p_k}{1-p_1+p_2+\cdots+p_k} = \alpha_k + \beta'x \qquad [3.3]$$

$$p_1 + p_2 + \cdots + p_{k+1} = 1$$

The maximum likelihood estimates method has been used to estimate parameter of model and the Pearson chi-square statistics was used to test goodness of fit model. Which compare the observed distribution to an expected distribution, in a situation where we have two or more categories. In other words, it compares multiple observed proportions to expected probabilities. The null hypothesis for goodness of fit test for multiclass distribution is that the observed frequency f_i s equal to an expected count e_i in each class. It is to be rejected if the p-value of the following chi-squared test statistics is less than a given significance level α.

$$x^2 = \sum_i \frac{(f_i - e_i)^2}{e_i} \qquad [3.4]$$

The p-value of the test is greater than the significance level alpha (0.05) we can conclude that the observed proportions are not significantly different from the expected proportions (classes), then model fit the data very well. Discriminant analysis is a multivariate technique concerned with classifying distinct set of objects (or set of observations) and with allocating new objects or observations to the previously defined groups. In other words, it is used to facilitate the interaction of dependent variables (having multiple ordered levels) with one or more independent variables. 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

$$Z_1 \,\square\, W_{11}X_1 \,\square\, W_{12}X_2 \,\square\, ...W_{1p}X_p, \qquad [3.5]$$

Where, the W1j is the weight of the jth variable for the first discriminant function. The weights of the discriminant function are such that the ratio

$$\lambda_1 = \frac{\text{Between groups SS of } Z_1 \text{ Maximized}}{\text{Within groups SS of } Z_1}$$

Suppose the second discriminant function is given by,

$$Z_2 \,\square\, W_{21}X_1 \,\square\, W_{22}X_2 \,\square\, ...W_{2p}X_p \qquad [3.6]$$

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

$$\lambda_2 = \frac{\text{Between groups SS of Z2 Maximized}}{\text{Within groups SS of Z2}}$$

Is maximized subject to the constraint that the discriminant scores Z1 and Z2 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. Classification procedure involves the division of the discriminant space in g mutually exclusive and collectively exhaustive regions. Which was mainly consists of Tests of Equality of Group Means, Tests of Covariance's Matrices, Wilk's Lambda, Standardized Canonical Discriminant Function, Structural Matrix, Unstandardized Canonical Discriminant Function and Classification.

Classificatory ability of the models

Classification performance of the different models is measured using Accuracy rate and Kappa statistics given respectively in [Eq-3.27 and 3.28].

$$Accuracy\ rate = \frac{\#\ correctly\ classfied\ data}{\#\ total\ data} \times 100 \qquad [3.7]$$

$$Kappa\ Statistic\ (\kappa) = \frac{N \times \sum_{i=1}^{l} x_{ii} - \sum_{i=1}^{l} x_{i_r} x_{i_c}}{N^2 - \sum_i^l x_{i_r} x_{i_c}} \qquad [3.8]$$

Where x_{ii} the count of cases is in the main diagonal of confusion matrix, N is the number of examples, and x_{i_r}, x_{i_c} are the rows and columns total counts, respectively. Larger the value of Accuracy rate and Kappa statistics better the classification ability of model. Average precision (Avgprecision) and Average recall (Avgrecall) are also used for comparison classification ability of different models for various classes of yield of little millets. Which were calculated with help of below Confusion Table.

Table- *Confusion matrix*

	A	B	C
A	AA	AB	AC
B	BA	BB	BC
C	CA	CB	CC

Where A, B and C are three classes, AA, BB and CC represent the correct prediction number of samples, the remaining number of samples is representative of the error prediction. AA represents the number of samples correctly identified as samples A. AB are predictive number that original Sample A which is incorrectly predicted as Sample B. The remaining items have the same meaning. Precision is the fraction of retrieved instances that are relevant. Precision reflects the classification accuracy. In practical applications, the average precision are often used to evaluate multi-classification (taking categories as example) model, which is calculated as follows

Avg precision = ((AA/ (AA+AB+AC)) + ((BB/ (BA+BB+BC)) + ((CC/ (AC+BC+CC))

Recall is the fraction of relevant instances that are retrieved. Recall reflects the classification comprehensiveness. In practical applications, the average recall are often used to evaluate multi-classification (taking categories as example) model, which is calculated as follows.

Avgre call = ((AA/(AA+ BA+ CA)) +((BB/(AB+ BB+ CB))+ ((CC/(CA +CB+ CC))

These Criteria were used to choose a best model for classification of various classes of yield of Little Millet.

Results and Discussion

The Ordinal logistic regression model and Multiclass Discriminant model were fitted well to research data and the results of these models were discussed in details as below. The model fitting information for ordinal logistic regression model as given in [Table-4.1]. Final model was statistically significant at 1 percent of level of significance with chi-square values (211.88) and p-values (0.00), it indicates that fitted model is more suitable for data for classification of genotypes for different classes of yield of little millet as compare to intercept only model.

Table-4.1 *Model fitting information for yield of little millet.*

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1577.878			
Final	1365.995	211.884	27	.000

[Table-4.2] shows the predictor variables considered in the model with their maximum likelihood estimates (B), their standard errors, Wald test statistic associated with corresponding probability values. The Table helps to know the effect and quantify the influence of each independent variable on classification of different classes of yield of little millet. Logistic regression coefficient (B) or the odds ratio is the predicted change in odds ratio of being different categories of little millet yield for a one unit change in the predictor. Plant Height (Wald= 44.40, P=0.00) is statistically significant at 1% level of significance which indicates if one unit increase in plant height then on an average estimated odds ratio of being high yield class of genotype verses low or medium yield class will be increased by 0.071 times. The Number of Basel tillers (Wald= 6.82, P=0.00) is statistically significant at 1% level of significance, which indicates if one unit increase in number of basel tillers then on an average estimated odds ratio of being high yield class of genotype verses low or medium yield class will be increased by 0.046 times.

Table- 4.2 *MLE of OLR for Yield of little millet*

Variable	B	Std.Error	Wald	p-value
Days to 50 percent flowering	0.017	0.015	1.225	0.268
Plant Height	0.071	0.011	44.401	0.000**
Number of Basel tillers	0.046	0.018	6.821	0.009**
Flag leaf length	0.002	0.007	0.115	0.735
Flag leaf width	1.087	0.742	2.143	0.143
Flag leaf sheath length	-0.019	0.012	2.490	0.115
Length of peduncle	-0.009	0.009	1.177	0.278
Length of inflorescence	-0.005	0.005	1.059	0.303
1000 grain weight	0.003	0.007	0.143	0.706
Low | medium	7.555	1.019	54.989	0.000**
Medium | High	9.529	1.053	81.940	0.000**

Intercepts for low or medium class verses high class of yield (1/2) and medium or high class verses low yield class (2/3)) are statistically significant at 1% level of significance, it indicates that keeping all predictor variables constant as result an average estimated odds ratio of being low or medium yield class verses high yield class will be increased by 7.55 times and an average estimated odds ratio of being medium or high class verses low yield class will be increased by 9.529 times respectively. Days to 50 percent flowering (Wald= 1.225, P=0.268), Flag leaf length (Wald=-0.115, P=0.73), Flag leaf width (Wald= 2.14, P=0.143), Flag leaf sheath length (Wald= 2.49, P=0.115), Length of peduncle (Wald=1.17, P=0.278), Length of inflorescence (Wald= 1.05, P=0.303), 1000 grain weight (Wald= 0.143, P=0.706) are statistically non-significant effect on the response variable for classification of genotypes for yield of little millet.

Table- 4.3 *Classification Matrix of Ordinal Logistic Regression for Yield of Little Millet*

Sample	Observed	Predicted			
		Low	Medium	High	Percent Correct
Training	Low	227	52	6	79.64%
	Medium	131	57	8	29.08%
	High	30	33	33	34.37%
	Overall Percent	67.24%	24.61%	8.14%	54.93%
Testing	Low	58	4	1	92.06%
	Medium	41	10	3	18.51%
	High	7	7	14	50%
	Overall Percent	73.10%	14.48%	12.41%	56.55%
Dependent Variable: Yield Class					

Classification of genotypes for different classes of yield of little millet according to Ordinal Logistic Regression model was explained in [Table-4.3]. The cells on the diagonal of the cross classification of cases are correct predictions for each sample of both training and testing data set. The cells off the diagonal of the cross classification of cases are incorrect predictions of the cases used to create the model. In training data set, 227 out of the 285 low yield genotypes are correctly classified with 79.64% of accuracy, 57 out of the 196 Medium yield genotypes are classified correctly with 29.08 % of accuracy, 33 out of 96 High yield genotypes are correctly classified with 34.37% of accuracy and Overall, 54.93 % of the training cases are classified correctly. A better model should correctly identify a higher correct percentage of the cases. In testing data set, 58 out of 63 low yield genotypes are classified correctly with 92.06% accuracy, 10 out of 54 medium yield genotypes are correctly with 18.51 % of accuracy, 14 out of 28 High yield genotypes are correctly with 50.00 % of accuracy and Overall, 56.55% of the testing cases are classified correctly. The testing sample helps to validate the model; here 56.55 % of these cases were correctly classified by the model. This suggests that overall model is in fact correct and efficient in prediction and classification.

Multiclass Discriminant Model

Discussion of results of multiclass Discriminant model for yield of little millet mainly consists of Tests of Equality of Group Means, Tests of Covariance's Matrices, Wilk's Lambda, Standardized Canonical Discriminant Function, Unstandardized Canonical Discriminant Function and Classification. [Table-4.4] explains the results of equality of group means of yield of little millet, which is comprise of Variables, Wilk's lambda, F statistics, degrees of freedom for discriminant functions and their probability level. The Predictors having larger value of F statistics are statistically significant at different level of significance and which indicates that effect or contribution of independent variables on group mean of dependent variable. The variable such as Fifty percent flowering (F=43.62, p=0.00), Plant Height (F =71.44, p=0.00), Number of Basel tillers (F=4.89, p=0.00), Flag leaf length (F=4.28, p=0.00), Flag leaf width (F=23.71, p=0.00) are statistically significant at 1% level. These predictors are main contributors for differences in the means of three groups and they are capable enough to discriminate the groups and Flag leaf sheath length (F =0.41, p=0.65), Length of peduncle (F=0.58, p=0.55), Length of inflorescence (F=0.19, p=0.82), 1000 grain weight (F=1.67, p=0.18) are statistically non-significant.

Table -4.4 *Tests of Equality of Group Means of yield of little millet*

Variables	Wilks' Lambda	F	df1	df2	Significance
fifty percent flowering	0.868	43.62	2	574	0.00**
Plant Height	0.801	71.44	2	574	0.00**
Number of Basel tillers	0.983	4.89	2	574	0.00**
Flag leaf length	0.985	4.28	2	574	0.01*
Flag leaf width	0.924	23.71	2	574	0.00**
Flag leaf sheath length	0.999	0.41	2	574	0.65
Length of peduncle	0.998	0.58	2	574	0.55
Length of inflorescence	0.999	0.19	2	574	0.82
1000 grain weight	0.994	1.67	2	574	0.18

Box's M Test has been used to test the equality of Covariance matrices with postulated null hypothesis that covariance matrices are same among the different groups of yield of little millet. If the test is not significant then there is equality of covariance matrices across the groups otherwise the assumption is violated. If Box's M Test is significant, then, we need to proceed with the analysis using separate covariance matrices for each group instead of the pooled within group covariance matrix.

Table-4.5 *Box's Test of Equality of Covariance's Matrices*

	Box's M	3802
F	**Approx.**	292
	df1	90
	df2	28039
	Sig.	0.19

Box's M test gives the values of 3802 with their F approximation 292 is non-significant (0.19). Which conclude that the equality of population covariance's matrices across the groups of dependent variables and allows to proceed the analysis as given in [Table-4.5].

Table-4.6 *Canonical Discriminant Coefficient Summary for yield of little millet*

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.289	88.5	88.5	0.473
2	0.038	11.5	100.0	0.190

* First 2 Canonical Discriminant functions were used in the analysis.

In Case of multiple group discriminant analysis, if there are G groups, G-1 discriminant functions can be estimated if the number of predictors larger than this quantity. Suppose study with G groups and K predictors, it is possible to estimate up to the smaller of G-1 or K discriminant functions, so, in present research two Discriminant functions are considered for analysis. The first function has the highest ratio of between groups to within groups' sum of squares. The second function uncorrelated with the first and has second highest ratio and so on., as we have only three groups then two discriminant functions are much enough to classify the groups. [Table-4.6] 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. In the above table shows two discriminant functions, first function has eigen value 0.289 and it explained 88.5 percent of variation, second function has Eigen value 0.038 and it explained 11.5 percent of variation. Which is potential enough to classifying the groups of maturity of little millet?

Table-4.7 *Wilk's Lambda*

Test of Function	Wilks' Lambda	Chi-square	df	Sig.
1	0.748	165.591	18	0.00
2	0.964	21.023	8	0.00

[Table-4.7] shows Wilk's Lambda value, its Chi square statistics with corresponding significance. It indicates the statistical significance of the estimated discriminant functions and we need to test the statistical significance with stated null hypothesis that the means of all discriminant functions in all groups are equal. The Chi-square test is used to test the statistical significance of lambda value of different discriminant functions, In the above table Wilk's lambda associated with the first function (λ=0.748) transforms to a chi square of 165.59 with 18 df and second function has (λ=0.964) transforms to a chi square of 21.02 with 8 df, which are statistically significant at 1% (<0.01) level of significance. If the null hypothesis is rejected at 1% means the selected discriminant function is statistically significant and it has enough to discriminate the groups, model is good fit to study data.

Table-4.8 *Standardized Canonical Discriminant Function Coefficients*

Variables	Function 1	Function 2
Fifty percent flowering	0.249	-0.390
Plant Height	0.923	0.016
Number of Basel tillers	0.258	0.619
Flag leaf length	0.061	-0.208
Flag leaf width	0.011	0.772
Flag leaf sheath length	-0.162	-0.027
Length of peduncle	-0.068	-0.306
Length of inflorescence	-0.125	0.058
1000 grain weight	-0.008	0.370

To eliminate scaling differences among the discriminator variables, standardised discriminant coefficients of discriminant functions are generally converted to Z scores (Mean=0, SD=1). Which helps to determine the degree to the absolute magnitude of standardized discriminant coefficients and the relative importance of each discriminator variables to group discrimination. Large the value of standardized coefficients more the discriminating power of the functions as compared with the predictor with smaller coefficients as given in [Table-4.8]. [Table-4.8] explain the relative importance of each predictor on discrimination of groups of yield of little millet. The sign indicates the direction of the relationship and magnitude indicates extent of contribution to the group discrimination by different discriminant functions. According to first discriminant function the predictors such as Fifty percent flowering (0.249), Plant Height (0.923), Number of Basel tillers (0.258), Flag leaf length (0.061) and Flag leaf width (0.011) are relatively more important and positively influencing on discrimination of groups. Whereas the variable like Flag leaf sheath length (-0.162), Length of peduncle (-0.068), Length of inflorescence (-0.125) and 1000 grain weight (-0.008) are negatively influencing on discrimination of different groups of yield of little millet. In second discriminant function the predictors such Plant Height (0.016), Number of Basel tillers (0.690), Flag leaf width (0.772), Length of inflorescence (0.058) and 1000 grain weight (0.370) are relatively more important and positively influencing on discrimination of groups, where predictors such as Fifty percent flowering (-0.390), Flag leaf length (-0.208), Flag leaf sheath length (-0.027) and Length of peduncle (-0.306) negatively influencing on discrimination of different groups of yield of little millet.

The unstandardized coefficients (b) are used to fit the discriminant function (equation) for prediction and classification purpose. However, the unstandardized coefficients cannot be used to compare of contribution of predictors on classification ability of groups of dependent variable and function will predicts and classify the members in to mutually exclusive groups. Unstandardized coefficient (b) for different predictors for each discriminant function is as given in [Table-4.9].

Table-4.9 *Unstandardized Canonical Discriminant Function Coefficients*

Variables	Function 1	Function 2
Fifty percent flowering	0.580	0.574
Plant Height	0.208	0.242
Number of basel tillers	0.696	0.740
Flag leaf length	-0.009	-0.013
Flag leaf width	35.582	37.752
Flag leaf sheath length	-0.061	-0.072
Length of peduncle	0.056	0.045
Length of inflorescence	0.013	0.012
1000 grain weight	0.009	0.024
Constant	-44.474	-49.234

Table-4.10 *Functions at Group Centroids*

Groups	Function1	Function 2
Low	-0.402	-0.131
Medium	0.038	0.269
High	1.115	-0.159

[Table-4.10] represents the group centroids for different groups of yield of little millet, it indicates the mean discriminant scores of the members of a group on an each discriminant function. The discriminant score of each group case is compared to each group centroid and the probability of group membership is calculated for classification and prediction purposes. The individual having closer to score of a group centroid, then the greater the probability the case belongs to that group. 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. The discriminant function 1 and discriminant function 2 were jointly considered on two-dimensional scale to find out the group centroids for the membership of different categories. The high yield class has group centroid value between 1.115 and -0.159 for discriminant function 1 and discriminant function 2 respectively, the medium yield class has group centroid value between 0.038 and 0.269 for discriminant function 1 and discriminant function 2 respectively and the low yield class has group centroid value between -0.402 and -0.131 for discriminant function 1 and discriminant function 2 respectively

Table-4.11 *Classification Matrix of Discriminant Analysis for Yield of little Millet*

Sample	Observed	Predicted			
		Low	Medium	High	Percent Correct
Training	ow	170	72	43	59.65%
	ledium	66	95	35	48.47%
	ligh	12	26	58	60.42%
	Dverall Percent	42.98%	33.45%	23.57%	55.98%
Testing	ow	47	11	5	74.60%
	ledium	31	9	14	16.67%
	ligh	3	3	22	78.57%
	Dverall Percent	55.86%	15.86%	28.28%	53.79%

[Table-4.11] shows, In Training data set, 170 of the 285 Low yield genotypes are correctly classified with 59.65 % of accuracy, 95 of the 196 Medium yield genotypes are classified correctly with 48.47 % of accuracy, 58 out of 96 High yield genotypes are correctly classified with 60.42 % of accuracy and Overall, 55.98 % of the training cases are classified correctly. A better model should correctly identify a higher percentage of the cases. In testing data set, 47 out of 63 low yield genotypes are classified correctly with 74.60% accuracy, 9 out of 54 medium yield genotypes are correctly with 16.67 % of accuracy, 22 out of 28 High yield genotypes are correctly with 78.57% of accuracy and Overall, 53.79 % of the testing cases are classified correctly, the testing sample helps to validate the model; here 53.79% of these cases were correctly classified by the model. This suggests that overall model is in fact correct and efficient in classification.

Table-4.12 *Classification ability measures of models for yield of little millet*

Measures	Ordinal Logistic Regression	Discriminant Analysis model
Accuracy Rate	56.55	53.79
Kappa statistics	0.13	0.07
Avgprecision	1.61	1.70
Avgrecall	1.80	1.51

Summary and Conclusion

The models having high value of Accuracy Rate, Kappa Statistics Avgprecision, Avgrecall were considered as best models for classification. The Ordinal logistic regression model ((Accuracy Rate =56.55), (Kappa statistics=0.13), (Avgprecision=1.61), (Avgrecall=1.80)) was performed better than the Multiclass Discriminant model ((Accuracy Rate =53.79), (Kappa statistics=0.07), (Avgprecision=1.70), (Avgrecall=1.51)) as it has larger values of classification ability measures for classification of genotypes for different classes of yield of little millet as given in [Table-4.12]. Plant height, Yield, Days to 50 percent flowering, Number of basel tillers, flag leaf width were considered to be important contributing predictor for classification of genotypes for different classes of yield of little millet as per the ordinal logistic regression model. Fifty percent flowering, Plant height, Number of basel tillers, Flag leaf length, Flag leaf width were considered to be important contributing predictor for classification of genotypes for classes of yield of little millet as per the Multiclass Discriminant model.

Application of research: The Ordinal logistic regression model has better classification ability as compare to that of Multiclass Discriminant model for classification of genotypes for different classes of yield of little millet and also it can be used for classification of genotypes for different attributing characters of different crops of millets.

Research Category: Agricultural Statistics

Acknowledgement / Funding: Author thankful to Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, 221005, Uttar Pradesh

Research Guide or Chairperson of research: Dr Abhishek Singh
University: Banaras Hindu University, Varanasi, 221005, Uttar Pradesh
Research project name or number: Nil

Author Contributions: All author equally contributed

Author statement: All authors read, reviewed, agree and approved the final manuscript

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Small Millets in India: Current status and future thrusts. All India Coordinated Small Millets Improvement Project. Report.
[2] Joshi M., Verma S.K., Singh J.P. and Barh A. (2013) *The Bioscan*., 8(4), 1529 - 1532.
[3] Tomooka N. (1991) Genetic diversity and landrace differentiation of mungbean, Vigna radiate (L) Wilcsek and evaluation of its wild relatives as breeding materials. Tech. p. 1. Bull. Tropical Research Centre, Japan. No. 28. Ministry of Agriculture. Forestry and Fisheries. Japan.
[4] Teklebrhan T., Urge M and Mekas Y. (2013) *African Journal of Agricultural Research,* 8(39), 4918-4921.
[5] Gobu R., Harish Babu B.N., Kailash Chandra Shankar M. and Omprakash (2017) *Int. J. Curr. Microbiol. App. Sci*., 6(3), 749 - 760.
[6] Das S. and Rahman R. (2011) *Nutrition Journal*, 10,124.
[7] Muhammad M.S. and Tuti P.S. (2013) *International Journal of Statistics and Applications*, 3(1), 1-8.
[8] Praveen S. and Gayatri V. (2005) *Indian J. Agric. Res*., 39 (3), 203 – 207.
[9] Teklebrhan T., Urge M. and Mekas Y. (2013) *African Journal of Agricultural Research,* 8(39), 4918-4921.
[10] Yay M. and Eylem D. (2009) *Cypriot Journal of Educational Sciences,* 4, 58–69.
[11] Adeleke K.A. and Adepoju A. A. (2010) *Journal of Mathematics and Statistics*, 6(3), 279-285