Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.6252

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **6 (6):** 248-258 (2018)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Use of Ordinal Logistic Regression and Multiclass Discriminant Model for Classification of Genotypes for Maturity of Little Millet

Nagaraja M. S.[1*] and Abhishek Singh[2]

[1]Assistant Professor, Department of Statistics, Christ (Deemed To Be University), Hosur Road, Bengaluru-560029

[2]Assistant Professor, Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, Uttar Pradesh-221005

*Corresponding Author E-mail: msn8129@gmail.com

Received: 14.02.2018 | Revised: 23.03.2018 | Accepted: 3.04.2018

ABSTRACT

Classification in agricultural systems are quite useful for Planning purposes for which various subjective and objective approaches are in vogue, classification of genotypes or germplasm based on yield and yield attributing characters is important for an accurate measurement of the differences between populations as well as for rapid assessment of their breeding potential. In present study the effort has been made to study the statistical model such as Ordinal logistic regression model and Multiclass Discriminant model and same has been used for classification of genotypes of little millet for different classes of maturity based on yield and yield attributing characters. These models were fitted to secondary 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, and Bengaluru. Classes of Fifty percent flowering (Maturity) was considered as dependent variable and all other attributing characters as predictors. Classification ability measures such as Accuracy Rate, Kappa Statistics, Avgprecision, and Avgrecall were used for testing samples. Yield, Plant height, Number of basel tillers, Flag leaf length, Flag leaf width were considered to be important attributing characters for classification and Multiclass Discriminant model (71.72 %) was performed compare to Ordinal Logistic Regression Model (68.28 %) for both classification of genotypes for different classes of maturity of little millets.

Key words: *Ordinal logistic regression, Multiclass Discriminant model, Classification, Attributing, Accuracy.*

INTRODUCTION

Millet grains have substantial benefits as a drought resistant crop, yields good productivity in the areas with water scarcity, possesses remarkable edible & nutritive values, and ease of processing & food manufacturing. Agriculture & Food security policymakers of developing countries should give due attention in promoting the research work & projects for studying the processing, food manufacturing, improvement in nutritive values and potential health benefits of the millet grains to promote their utilization as food in respective countries.

Cite this article: Nagaraja, M.S. and Singh, A., Use of Ordinal Logistic Regression and Multiclass Discriminant model for Classification of genotypes for maturity of Little Millet, *Int. J. Pure App. Biosci.* **6(6):** 248-258 (2018). doi: http://dx.doi.org/10.18782/2320-7051.6252

Most of the developing countries have already started working in the field of improvement of edible potential of millet grains. Millet oil could be a good source of linoleic acid and tocopherols. Millet is an alkaline forming grain that is gluten-free. Millets are also rich sources of phytochemicals and micronutrients, play many roles in the body immune system. Millets have nutraceutical properties in the form of antioxidants which prevent deterioration of human health such as lowering blood pressure, risk of heart disease, prevention of cancer and cardiovascular diseases, diabetes, decreasing tumor cases etc.[4].

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.

The classification of genotypes for different classes of maturity, which helps to create genetic variability among the genotypes with respect to that particular character, which is also one of the way to develop the different parentage for breeding programme Systematic improvement of any crop depends mainly on the information on genetic variability and diversity which forms the basis for any crop breeding programme[1]. Genetic diversity in crop plants is essential to sustain the level of high productivity[3].

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 from Project coordination cell, All India Coordinated Small Millets Improvement Project (AICSMIP), ICAR, Bangalore. The maturity of little millet has been classified as early, medium and late maturity classes and all other yield attributing characters were considered as independent variable. 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 genotype having less than 45 days to fifty percent flowering considered as early mature genotype, 45 to 54 days to fifty percent flowering as medium mature genotype, more than 54 days to fifty percent flowering as late mature genotype.

Ordinal Logistic Regression Model and Multiclass Discriminant Model were fitted to data for classification of genotypes, which were used to classify the classes of maturity 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 fi s equal to an expected count ei 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

$$Z_1 = W_{11}X_1 + W_{12}X_2 + ...W_{1p}X_p, \qquad (3.5)$$

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,

$$Z_2 = W_{21}X_1 + W_{22}X_2 + ...W_{2p}X_p \qquad (3.6)$$

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 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. 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 ability performance of the different models is measured using Accuracy rate, Kappa statistics, Average precision (Avgprecision) and Average recall (Avgrecall) are given as following equations (3.7 and (3.8).

$$Accuracy\ rate = \frac{\#\ \boldsymbol{correctly\ classfied\ data}}{\#\ \boldsymbol{total\ data}} \times 100 \tag{3.7}$$

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

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 is 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

Avgprecision = ((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.

Avgrecall = ((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 maturity 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 model 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 (449.19) and p-values (0.00), it indicates that fitted model is more suitable for classification of genotypes for different classes of maturity of little millet as compare to intercept only model.

Table 4.1.Model fitting information for maturity of little millet

Model	**Model Fitting Criteria**	**Likelihood Ratio Tests**		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1535.959			
Final	1086.765	449.193	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 maturity of little millet.

Plant height (Wald= 139.310, P=0.00) is statistically significant at 1 % level of significance, which indicates if one unit change in plant height then on an average estimated odds ratio of being late maturity of genotype verses early or medium maturity will be increased by 0.145 times. The Flag leaf width (Wald= 3.580, P=0.049) is statistically significant at 5% level, which indicates if one unit change in flag leaf width then on an average estimated odds ratio of being late maturity of genotype verses early or medium maturity will be increased by 1.776 times. Intercepts for early or medium maturity verses late maturity class (1/2) and medium or late maturity class verses early maturity (2/3) are statistically significant at 1% level 1, it indicates that keeping all predictor variables constant as result an average estimated odds ratio of being early or medium maturity class verses late maturity will be increased by 12.80 times and an average estimated odds ratio of being medium or late maturity class verses early maturity class will be increased by 16.64 times respectively.

Grain yield (Wald= 0.228, P=0.633), Number of Basel tillers (Wald=0.236, P= 0.627), Flag leaf length (Wald= 1.696, P=0.193), Flag leaf sheath length (Wald= 0.728, P=0.393), Length of peduncle (Wald= 1.424, P=0.233), Length of inflorescence (Wald= 0.336, P=0.562), 1000 grain weight (Wald=0.510, P=0.475) are statistically non-significant effect on the response variable for classification of genotypes of little millet for maturity.

Table 4.2 MLE for maturity of little Millet.

Variable	B	Std.Error	Wald	p-value
Grain yield	-0.005	0.010	0.228	0.633
Plant Height	0.145	0.012	139.310	0.000**
Number of Basel tillers	0.006	0.011	0.236	0.627
Flag leaf length	0.009	0.007	1.696	0.193
Flag leaf width	1.776	0.939	3.580	0.049*
Flag leaf sheath length	0.010	0.012	0.728	0.393
Length of peduncle	-0.035	0.029	1.424	0.233
Length of inflorescence	0.003	0.004	0.336	0.562
1000 grain weight	-0.007	0.010	0.510	0.475
Early | medium	12.805	1.183	117.237	0.000**
Medium | late	16.649	1.327	157.501	0.000**

Model Selection

The residual deviance and AIC are important measures of model accuracy in ordinal logistic regression and as lower values of these indicates better model accuracy. These measures also be used for between models comparison. For maturity of little millet the ordinal logistic regression model has Residual Deviance of 699.8 and AIC of 721.8 as given in table 4.3.

Table 4.3 Selection Criteria of OLR for maturity of little Millet

Criteria	Values
Residual Deviance	699.8
AIC	721.8

Classification

Table 4.4 show that, In Training data set, 279 out of the 321 Early mature genotypes are correctly classified with 86.91 % of accuracy, 118 out of the 209 Medium mature genotypes are classified correctly with 56.45 % of accuracy, 20 out of 47 late mature genotypes are correctly classified with 42.55 % of accuracy and Overall, 72.27 % of the training cases are classified correctly. A better model should correctly identify a higher correct percentage of the cases. In testing data set, 59 out of 74 early mature genotypes are classified correctly with 79.72% accuracy, 16 out of 44 medium mature genotypes are correctly with 36.36 % of accuracy, 24 out of 27 late mature genotypes are correctly with 88.88 % of accuracy and overall, 68.27% of the testing cases are classified correctly. The testing sample helps to validate the model; here 68.27% of these cases are correctly classified by the model. This suggests that overall model is in fact correct and efficient in classification.

Table 4.4 Classification Matrix of OLR for Maturity of little Millet

Sample	Observed	Predicted			
		Early	Medium	Late	Percent Correct
Training	Early	279	41	1	86.91%
	Medium	83	118	8	56.45%
	Late	2	25	20	42.55%
	Overall Percent	63.08%	31.88%	5.02 %	72.27%
Testing	Early	59	14	1	79.72%
	Medium	23	16	5	36.36%
	Late	0	3	24	88.88%
	Overall Percent	56.55%	22.75%	20.68%	68.27%
Dependent Variable: Maturity class					

Multiclass Discriminant Model

Discussion of results of multiclass Discriminant model for maturity 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.

Tests of Equality of Group Means of Maturity of little Millet

Table 4.5 explains the results of equality of group means of maturity 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.

Table 4.5: Tests of Equality of Group Means

Variables	Wilks' Lambda	F	df1	df2	Significance
Plant Height	0.546	238.322	2	574	0.00**
Number of Basel tillers	0.920	25.052	2	574	0.00**
Flag leaf length	0.963	10.961	2	574	0.00**
Flag leaf width	0.841	54.395	2	574	0.00**
Flag leaf sheath length	0.964	10.791	2	574	0.00**
Length of peduncle	0.997	0.743	2	574	0.47
Length of inflorescence	0.976	7.181	2	574	0.00**
1000 grain weight	1.000	0.113	2	574	0.89
Grain yield	0.963	10.968	2	574	0.00**

The variable such as Plant Height (F=238.32, P=0.00), Number of Basel tillers (F=25.05, P= 0.00), Flag leaf length (F=10.96, P= 0.00), Flag leaf width (F=54.39, P=0.00), Flag leaf sheath length (F=10.79, P= 0.00), Length of inflorescence (F=7.18, p=0.00), Grain yield (F=10.96, p=0.00) are statistically significant at 1% level. These predictors are main contributors for differences in the means of three groups of maturity of little millet.

Tests of Covariance's Matrices

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 maturity of finger 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 4.6: Box's Test of Equality of Covariance's Matrices

Box's M		2180.20
F	Approx.	179.30
	df1	90
	df2	23890.60
	Sig.	0.29

Box's M test gives the values of 2180.20 with their F approximation 179.30 is non-significant (0.29). Which conclude that the equality of population covariance's matrices across the groups of dependent variable and allows to proceed the analysis as given in table 4.6.

Table 4.7: Canonical Discriminant Coefficient Summary

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.857	95.4	95.4	0.679
2	0.042	4.6	100.0	0.200

* 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 4.7 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 0.857 and it explained 95.4 percent of variation, second function has Eigen value 0.042 and it explained 4.6 percent of variation. Two functions together explained 100 percent of variation in data.

Table 4.8: Wilk's Lambda

Test of Function	Wilks' Lambda	Chi-square	df	Sig.
1	0.517	376.107	18	0.00
2	0.960	23.277	8	0.00

Table 4.8 shows Wilk's Lambda value, its Chi square statistics, degree of freedom 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. Lambda is the ratio of within-groups sums of squares to the total sums of squares. The value of varies from 0 to 1, if its 1 indicates that the almost all the variability in the discriminator variables is due to within group differences, while a small lambda occurs when within-groups variability is small compared to the total variability. A small lambda (near to zero) indicates that group means appear to differ and almost all the variability in the discriminator variables is due to within group differences.

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=0.517) transforms to a chi square of 376.10 with 18 df and second function has (λ=0.96) transforms to a chi square of 23.27 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 of maturity of little millet, model is good fit to study data.

Table 4.9: Standardized Canonical Discriminant Function Coefficients

Variables	Function 1	Function 2
Plant Height	0.992	-0.133
Number of Basel tillers	-0.035	0.502
Flag leaf length	0.065	0.336
Flag leaf width	-0.050	0.425
Flag leaf sheath length	0.074	0.168
Length of peduncle	-0.123	-0.281
Length of inflorescence	-0.003	0.519
1000 grain weight	-0.061	0.021
Grain yield	0.031	-0.405

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.

Table 4.9 explain the relative importance of the each predictor on discrimination of groups of maturity 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 Plant Height (0.99), Flag leaf length (0.065), Flag leaf sheath length (0.074) and Grain yield (0.031) are relatively more important and positively influencing on discrimination of groups. Whereas the variable like Number of Basel tillers (-0.35), Flag leaf width (-0.050), peduncle length (-0.123), Length of inflorescence (-0.003) and 1000 grain weight (-0.061) are negatively influencing on discrimination of different groups of maturity of little millet. In second discriminant function the predictors such Number of Basel tillers 0.502), Flag leaf length (0.336), Flag leaf width (0.425), Flag leaf sheath length (0.168), Length of inflorescence (0.519) and 1000 grain weight (0.021) are relatively more important and positively influencing on discrimination of groups, where predictors such as Plant Height

(-0.133), Length of peduncle (-0.281) and Grain yield (-0.405) are negatively influencing on discrimination of different groups of maturity of little millet.
In present research predictors such as plant height (0.98), Flag leaf width (0.46) and Flag leaf width (0.316), Length of inflorescence (0.52), Flag leaf length (0.42), Number of Basel tillers (0.41) are important variables according to discriminate function1 and discriminate function 2 respectively.

Table 4.10: Unstandardized Canonical Discriminant Function Coefficients

Variables	Function 1	Function 2
Plant Height	0.097	-0.013
Number of Basel tillers	-0.004	0.063
Flag leaf length	0.005	0.025
Flag leaf width	-0.361	3.080
Flag leaf sheath length	0.009	0.021
Length of peduncle	-0.009	-0.020
Length of inflorescence	0.000	0.026
1000 grain weight	-0.006	0.002
Grain yield per plant	0.003	-0.041
Constant	-6.997	-4.101

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 4.10.

Table 4.11: Functions at Group Centroids for Maturity of little millet

Groups	Function1	Function 2
Early	-0.696	-0.097
Medium	0.521	0.245
late	2.439	-0.422

Table 4.11 represents the group centroids for different groups of maturity 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 late mature class has group centroid value between 2.439 and -0.422 for discriminant function 1 and discriminant function 2 respectively, the medium mature class has group centroid value between 0.521 and 0.245 for discriminant function 1 and discriminant function 2 respectively and the early mature class has group centroid value between -0.696 and -0.097 for discriminant function 1 and discriminant function 2 respectively.

Classification of genotypes for different classes of maturity of little millet using Multiclass Discriminant Analysis

Table 4.12 explains that the cells on the diagonal of the cross classification are correct predictions for each classes of both Training and Testing data set. The cells off the diagonal of the cross classification are incorrect predictions create the model. In Training data set, 285 of the 321 Early mature genotypes are correctly classified with 88.79 % of accuracy, 93 of the 209 Medium mature genotypes are classified correctly with 44.50 % of accuracy, 36 out of 47 late mature genotypes are correctly classified with 76.60 % of accuracy and Overall, 71.75 % of the training cases are classified correctly. A better model should correctly identify a higher percentage of the cases.

Table 4.12 Classification Matrix of Discriminant Analysis for Maturity of little Millet

Sample	Observed	Predicted			
		Early	Medium	Late	Percent Correct
Training	Early	285	34	2	88.79%
	Medium	92	93	24	44.50%
	late	2	9	36	76.60%
	Overall Percent	65.68%	23.57%	10.75%	71.75%
Testing	Early	64	7	3	86.49%
	Medium	21	15	8	34.09%
	late	0	2	25	92.59%
	Overall Percent	58.62%	16.55%	24.83%	71.72%
Dependent Variable: Maturity					

In testing data set, 64 out of 74 early mature genotypes are classified correctly with 86.49% accuracy, 15 out of 44 medium mature genotypes are correctly with 34.09 % of accuracy, 25 out of 27 late mature genotypes are correctly with 92.59 % of accuracy and Overall, 71.72 % of the testing cases are classified correctly,. The testing sample helps to validate the model; here 71.72% of these cases were correctly classified by the model. This suggests that overall model is in fact correct and efficient in prediction and classification.

Table 4.13: Classification Ability of models for Maturity of little millet

Criteria	Measures	Ordinal Logistic Regression	Discriminant Analysis
Classification Ability	Accuracy Rate	68.28	71.72
	Kappa statistics	0.37	0.43
	Avgprecision	2.05	2.13
	Avgrecall	2.00	2.07

SUMMARY AND CONCLUSION

The models having high value of Accuracy Rate, Kappa Statistics, Avgprecision, and Avgrecall were considered as best models for classification genotypes for different classes maturity of little millet.

Multiclass Discriminant model was ((Accuracy Rate =71.55), (Kappa statistics=0.43), (Avgprecision=2.13), (Avgrecall=2.07)) performed better as compare to Ordinal logistic regression model ((Accuracy Rate =68.28), (Kappa statistics=0.37), (Avgprecision=2.05), (Avgrecall=2.0)) for classification genotypes for different classes of maturity of little millet as it has larger values of classification ability measures as given in table 4.13.

Plant height, Flag leaf width, Number of basel tillers, Flag leaf sheath length, Length of inflorescence, Grain yield per plant, Flag leaf length were considered to be important contributing predictor for classification of genotypes for different classes of maturity of little millet.

REFERENCES

1. Gobu, R., Harish Babu, B. N., Chandra, K., Shankar, M. and Omprakash. Genetic Variability, Heritability and Genetic Advance in Eggplant (Solanum melongena L.) Genotypes under Normal and Osmotic Stress in invitro condition. *Int. J. Curr. Microbiol. App. Sci*. **6(3):** 749 – 760 (2017).
2. Das, S. and Rahman, R., Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in Bangladesh. *Nutrition Journal*. **10:** 124 (2011).
3. Tripathi, A., Bisen, R., Ahirwal, R. P., Paroha, S., Sahu, R. and Ranganatha. A. R. G., Study on genetic divergence in sesame (Sesamum indicum L.) Germplasm based on morphological and quality traits. *The Bioscan*. **8(4):** 1387 – 1391 (2013).
4. Sarita and Singh, E., Potential of Millets: Nutrients Composition and Health Benefits. *Journal of Scientific and Innovative Research*. **5(2):** 46-50 (2016).
5. Adeleke, K. A. and Adepoju, A. A., Ordinal Logistic Regression Model: An Application to Pregnancy Outcomes. *Journal of Mathematics and Statistics*. **6(3):** 279-285 (2010).
6. McDonald, L. S., Panozzo, J. F., Salisbury, P. A. and Ford, R., Discriminant Analysis of Defective and Non- Defective Field Pea (Pisum sativum L.) into Broad Market Grades Based on Digital Image Features. *PLOS ONE online journal*. **11(5):** e0155523 (2016).
7. Mohammadi, H., Torabi, S. and Dogani, A., Application of Ordered Logit Model in Investigating the factors Affecting People's Income (A Case Study in Tehran City). *International Journal of Academic Research in Economics and Management Science*. **5(3):** 2226-3624 (2015).
8. Muriithi, D. K., Kihoro, J. and Waititu, A., Ordinal logistic regression versus multiple binary logistic regression model for predicting student loan allocation. *Jomo Kenyatta University of Agriculture and Technology*. **14(1):** 190-203 (2012).
9. Yay, M. and Eylem, D., Application of Ordinal Logistic Regression and Artificial Neural Networks in a Study of Student Satisfaction. *Cypriot Journal of Educational Sciences*. **4:** 58–69 (2009).
10. Praveen, S. and Gayatri, V., Discriminant Analysis for Rice-Wheat System Having the Same Attributing Characters towards Grain Yield. *Indian J. Agric. Res*. **39(3):** 203 – 207 (2005).