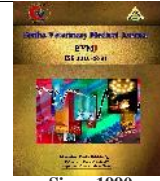




Official Journal Issued by  
Faculty of  
Veterinary Medicine

## Benha Veterinary Medical Journal

Journal homepage: <https://bvmj.journals.ekb.eg/>



Since 1990

### Original Paper

## Application of linear discriminant model to evaluate the association between milk production, reproductive performance, and calving season in dairy cattle

Hayaa A. Abd-El Rahman<sup>1,\*</sup>, Fardos A.M. Hassan<sup>2</sup>, Eman R. Kamel<sup>3</sup>

<sup>1</sup>*Biostatistics, Department of Animal Wealth Development, Faculty of Veterinary Medicine, Benha University, Moshtohor, Toukh 13736, Qalyubia, Egypt.*

<sup>2</sup>*Veterinary Economics and Farm Management, Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, 44511, Egypt.*

<sup>3</sup>*Economics and Farm Management, Department of Animal Wealth Development, Faculty of Veterinary Medicine, Benha University, Moshtohor, Toukh 13736, Qalyubia, Egypt*

### ARTICLE INFO

#### Keywords

*Dairy cattle*  
*Discriminant Analysis*  
*Milk yield*  
*Reproductive measures*

*Received* 26/11/2020

*Accepted* 13/12/2020

*Available On-Line*  
20/01/2021

### ABSTRACT

This study was aimed to determine the performance of linear discriminant analysis (LDA) for differentiation between seasons of calving on the basis of 305day- milk yield, fat %, protein %, days open (DO), days to first insemination (DFI), and number of services per conception. By considering the assumption of this method, a random sample was selected from the animals being represented by all explanatory variables. The discrimination between seasons of calving was depended on the significance of coefficients, classification rate, in addition to the group centroids. Results showed that LDA method selected 305day-milk yield (kg), days open (DO), days to first insemination and number of services per conception, as the significant ( $P < 0.05$ ) contributors for data classification. The total variance explained by 2 functions was 79.2% and 15.9%, respectively. So, the 1<sup>st</sup> function can do well in discrimination process than the 2<sup>nd</sup> one. The percentage of correct classification was 64.6%. In conclusion, LDA can be used effectively for classification of calving seasons, even with violation of normality assumption

## 1. INTRODUCTION

The dairy industry is a significant agricultural subsector in both developed and developing countries. Recently, milk quality features have become increasingly relevant as consumer demand for healthy diets grows, and subsequently cow's milk is being progressively documented as a valuable source of protein, energy, essential vitamins, and minerals (German and Dillard, 2006). In this context, increasing milk yield in terms of quality and quantity as well as improving reproductive performance have always been the desired goal for dairy producers (Senger, 2001). Notably, the reproductive success of dairy cattle is a fundamental indicator for the sustainability of the dairy farming system, and it has a significant relationship with the milk yield (El-Tarabany and El-Bayoumi, 2015). Poor reproductive performance of dairy cattle leads to significant economic losses due to lower returns from calves born and milk produced, cost of prolonged calving interval, increased insemination costs, and forced replacements in the event of culling (Nishida et al., 2006). In order to improve productivity and profitability of dairy cattle, it is substantial to be aware about the environmental and physiological factors affecting the performance of animals. Season of calving is one of the crucial environmental factors that determine cattle productive and reproductive efficiency (Khosroshahi et al., 2011). It was documented that severe environmental conditions in subtropical regions showing decreasing in milk yield amount and predisposes the animal to generate low-gain (Collier et al., 2006). Notably, the

influence of seasonal variation on the performance of dairy cattle have been reported (Ramadan and El-Tahawy, 2014; Hassan et al., 2017), which may be attributed to the harsh variation in environmental temperature, photoperiod, and feedstuffs availability (Zicarelli, 1997). Multivariate statistical analysis, such as discriminant analysis (DA), is an important method to evaluate the impact of various production environments and in the production of milk and milk products according to consumers' needs (Mele et al., 2016). Linear discriminant analysis (LDA) is a very important statistical technique for examination of the relationship between a categorical outcome and multiple predictor variables in the form of discriminant function (Cramer, 2003). This model can be used to explore which is the best independent variable that differentiates between two or more groups along with classification of cases into their group (Timm, 2002; Hamid, 2010). It is based on the conception of a function which could provide the identification of the subject of analysis (Kara et al., 2005). Therefore, this study aimed to apply discriminant function to determine the weight of productive (305day-milk yield, milk fat %, and milk protein %) and reproductive measures (Days to first insemination, days open, and number of services) in prediction of calving season for Holstein-Friesian cows.

## 2. MATERIAL AND METHODS

### 2.1. Sampling and data collection

A standardized data of 806 pure breed Holstein-Friesian cows was collected from reliable records of large

\* Corresponding author: hayaa93mohamed@gmail.com

commercial dairy farm, extensively monitored for research purposes, Sharkia governorate, Egypt, during the period extended from January 2018 to December 2019. The area climate is subtropical in nature with ambient temperature ranging between 35°C in summer and 7.2°C in winter and relative humidity varies from 58% in winter to over 77% in summer (El-Ramady et al., 2013). Also, the region receives more rainfall in winter with an average of 20 mm per annum. The average temperature and THI values in different seasons are illustrated in Fig.1. The data included calving season, 305-day-milk yield (kg), Fat%, protein%, days to first insemination (days), days open (days), and No. of services per conception. The dependent variable was calving season. All animals enrolled in the current study were classified according to calving season into winter calving (n= 202), spring calving (n= 201), summer calving (n= 201), and autumn calving (n=202). However, the independent (predictor) variables included 305 day-milk yield, Fat %, protein %, days to first insemination, days open, and No. of services per conception (Table 1),

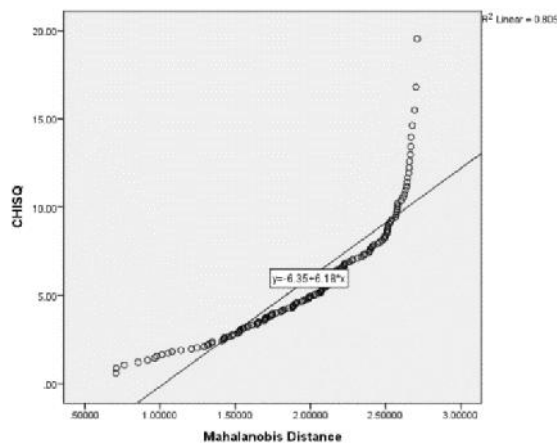


Fig. 1 Multivariate normality graph for variables within seasons of calving

## 2.2. Herd management

Whole animals presented in the dairy farm were housed in a free-stall barn provided with water splashing systems (cooling system) that operate when the ambient temperature exceeds 30°C. The cows were machine-milked three times a day, with milk yield and composition recorded at each milking. The total mixed ration (TMR) was provided for all animals twice a day, and determined based on the actual milk production and body condition score. The TMR was formulated to meet the optimized requirements of energy, protein, vitamins, and minerals. Monthly, TMR was sampled and analyzed using wet chemistry methods. The primary analysis of TMR includes crude protein (16.55%), net energy for lactation (Mcal/kg = 1.79), and neutral detergent fiber (24.74%). The main utilized forage is Alfalfa hay. All animals were regularly vaccinated against foot and mouth disease (FMD), hemorrhagic septicemia, and brucellosis. For mastitis, lactating cows were vaccinated (Rotatec J5) every 4 months, while heifers and dry cows were vaccinated at 60 and 30 days before calving, respectively. Cows estimated in heat by visual inspection and/or excessive activities recorded by the pedometer, they were introduced to insemination 10 to 16 hours later.

The productive and reproductive data were tracked by a commercial on-farm software program (AfiFarm version 4.1).

## 2.3. Model Specification

### 2.3. a. Assumptions of Linear Discriminant Analysis:

Predictor variables were tested for univariate normality by using histogram and normal Q-Q plot, while multivariate normality was assessed by chi-square versus Mahalanobis distance plot. Data were also tested for linearity (linear relationship) between the outcome variable and each of the explanatory variables using scatter plots according to Utts (2005). Variance inflation factor (VIF) was used to detect the multicollinearity between predictors. Homogeneity of covariance matrices were done by using Box's M test and according to Hahs-Vaughn (2017) who recommended the significance of the test at ( $P < 0.001$ ). Finally, multivariate outlier is done by using a measure called Mahalanobis distance.

### 2.3. Linear Discriminant Equation:

Linear discriminant technique was used to detect the predictor variables being discriminated well between the four seasons of calving. The initial discriminant equation was as follow:

$$D_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i}$$

Where,

$D_i$  = predicted score (discriminant score)

$x_{1i} - x_{6i}$  = independent (predictor) variables

$\beta_0$  = constant

$\beta_1 - \beta_6$  = discriminant coefficients or weights.

Depending on the estimates of the coefficients of discriminant analysis, we able to determine which independent variables would be used to differentiate between the groups. The coefficients with high measure refer to the importance of the corresponding variable in explaining the outcome. Furthermore, discriminant function gives what is known as discriminant scores, from which the predicted probabilities will be determined for each case of the categorical outcome variable. These discriminant scores along with centroids (the group means) share in the discrimination of cases into their groups (Worth and Cronin, 2003). All statistical procedures were completed using SPSS V. 23.0 (SPSS Inc., Chicago, Illinois, USA).

## 3. RESULTS

The first set of analyses in the current study was applied to examine the assumptions of linear discriminant analysis. The results revealed the violation of multivariate normality of explanatory variables, even with log transformation of data. Nevertheless, the analysis was conducted in violation of this assumption since the discriminant analysis achieves good performances in the tasks of face and object recognition, even though the normality assumption is often violated (Duda et al., 2001). Scatterplots which has been used to test the linearity revealed linear relationship between the outcome variable and each of the explanatory variables. All variables showed homogeneity of covariance matrices (Box's M = 175.64, F = 5.56, and  $P > 0.001$ ), absence of multicollinearity (VIF value close to 1), and absence of multivariate outlier (Minimum mahalanobis = 0.78, and  $P > 0.05$ ). The explanatory variables which used in discrimination between calving seasons (dependent variable) as shown in table (1).

Table 1 Descriptive statistics of independent variables

Variables	Mean	SD	Min	Max.
305 day- milk yield (kg)	7140.7	1352.5	456 1.0	11662.0
Fat (%)	3.56	0.14	3.07	4.07
Protein (%)	3.05	0.07	2.79	3.26
Days to first insemination (days)	99.8	58.9	38.0	268.0
Days open (days)	164.5	91.5	38.0	390.0
Number of services/conceptions	2.54	1.89	1.0	10.0

SD: Standard deviation

In term of determining the best set of predictors which significantly differentiate between calving seasons, the results showed that 305 day-milk yield, days to first insemination, days open, and number of services per conception have significant ( $P < 0.001$ ) contribution in data classification. As the smaller the Wilks's lambda, the more important the independent variable in discrimination process. So, the most important variables were days open ( $\lambda = 0.894$ ), followed by number of services per conception ( $\lambda = 0.944$ ), then 305day- milk yield ( $\lambda = 0.952$ ) as shown in (table 2).

Table 2 The importance of explanatory variables in explaining the outcome using linear discriminant analysis

Variables	Wilks' Lambda	F	P-value
305 day- milk yield (kg)	0.952	13.4	< 0.001
Fat%	0.996	1.2	0.319
Protein %	0.994	1.6	0.186
Days to first insemination (days)	0.980	5.5	0.001
Days open (days)	0.894	31.6	< 0.001
Number of services/conceptions	0.944	15.8	< 0.001

In Table (3), there were 3 functions produced, the first one had higher eigenvalue 0.179 and explained 79.2% of total variance in the outcome (season of calving), the second one had eigenvalue equal 0.036 and explained 15.9% of total variance, and third function had lowest eigenvalue equal 0.011 and explained 4.9% of total variance in outcome variable.

Table 3 The total variance explained by eigenvalues

Function	Eigenvalue	Percentage of variance	Cumulative %	Canonical correlation
I	0.179	79.2	79.2	0.390
II	0.036	15.9	95.1	0.187
III	0.011	4.9	100.0	0.104

The results in Table (4) revealed that 305day- milk yield and days open are important variables in discrimination between season of calving on the first function; where 305 day- milk yield showed a positive relationship ( $r = 0.417$ ) while days open and number of services per conception showed a negative relationship ( $r = -0.848$  and  $-0.554$ , respectively) with season of calving.

Table 4 Standardized canonical discriminant function coefficients

Variables	Function		
	I	II	III
305 day-milk yield (kg)	0.417*	0.340*	0.197
Fat%	0.090	0.125	0.554*
Protein %	0.166	0.171	0.221
Days to first insemination (days)	0.163	-1.451*	-0.214
Days open (days)	-0.848*	-1.919*	-0.197
Number of services/conception	-0.554*	-0.281	-0.108

In the second function, days to first insemination and days open were significant variables in discrimination between calving seasons and showed a negative relationship with season of calving ( $r = 1.451$  and  $1.919$ , respectively).

Discriminant function plot showed that summer season had highest value on function compared to other seasons (Fig. 2). Since the function was greatly associated with 305day-milk yield, days open, and number of services per conception, it is possible to arrange the four calving season based on these three variables. Function II had low role in discrimination process compared to function I, four groups membership of calving season appeared on the graph at nearly same level; no highest or lowest levels can be detected obviously

#### 4. DISCUSSION

This study aimed to assess the ability of linear discriminant analysis (LDA) for differentiation between seasons of calving on the basis of 305day-milk yield, fat %, protein %, days open (DO), days to first insemination (DFI), and number of services per conception (N/C). The results of this study revealed that (LDA) can do well in discrimination between calving seasons and among significant predictors. The most important explanatory variable that used in differentiation between calving seasons was days open, followed by number of services per conception, then 305day-milk yields. On the other side, fat% and protein% were proved to be non-significant discriminators for seasons of calving. The significant effect of days open, number of services per conception, and 305day-milk yield are in the same line with previous report that season of calving has a significant effect on reproductive and productive performance of Holstein cow (Soydan et al., 2005 and Torshizi, 2016). However, White et al. (2002) founded that calving season had no significant impact on total lactation production. Wondossen et al. (2018) showed that calving season had no significant effect on calving interval (CI) and days open (DO) ( $P > 0.05$ ). Our results revealed that there were 3 functions were used in discrimination process and according to the percentage of variance occurred in calving season, the function one had higher eigenvalue, thus the highest variance explained by it, and it can do well in differentiation between calving seasons than function II and function III. Canonical correlation revealed that there was strong association between discriminant function and dependent variable (season of calving). The discriminant function plot showed that group 1 that referred to summer season had highest value on function compared to other seasons. Because function I was greatly associated with number of services per conception, days open and 305day-milk yield, one would expect that four group's membership to be ordered on these three variables. Function II had low role in discrimination process compared to function I. The four group's membership of calving season appeared on the graph at nearly same level; no highest nor lowest levels can be detected obviously, while function III not significant in discrimination process as variables used in it not able to discriminate between seasons of calving, so we ignore it. The results of discriminant analysis revealed that days open is the most variable obviously separated groups of calving seasons. These findings are in agreement with Melendez and Pinedo (2007) who reported that calving seasons have significant effect on days open and other reproductive measures. In harmony, Soydan et al. (2005) who documented that the cows which calving in summer season

had days open of 35 days and that longer than those calving in months of winter. Siatka et al. (2017) stated that for improving fertility in dairy farms you should put calving season in your consideration as summer was the least favorable period for insemination unlike winter season was the most favorable period. Bouallegue et al. (2013) reported that the lowest level of milk production occurred for cows calved in summer than other seasons.

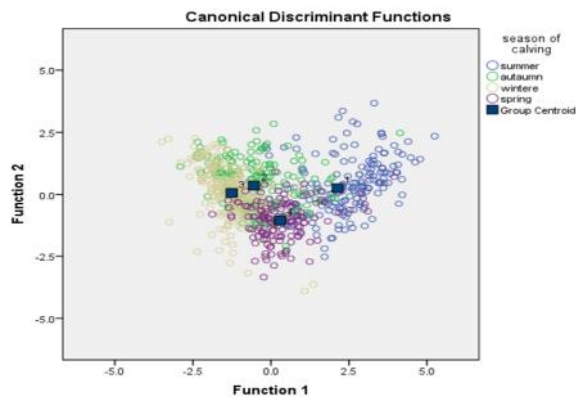


Fig. 2 Discriminant function plot

## 5. CONCLUSIONS

The results revealed that there was strong association between discriminant function and dependent variable (season of calving), that means there is a significant association between season of calving, 305 day-milk yield, and reproductive measures (days open, number of services per conception, days to first insemination).

## ACKNOWLEDGMENT

The authors wish to thank the owners of cattle herd at Sharkia governorate, Egypt, for allowing us for collecting the data.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## 6. REFERENCES

- Bouallegue, M., Haddad, B., Aschi, M.S., Ben, H.M., 2013. Effect of environmental factors on lactation curves of milk production traits in Holstein – Friesian cows reared under North African condition. *Livestock Research for Rural Development*. 25-5. Article #75. Retrieved January 21, 2021, from <http://www.lrrd.org/lrrd25/5/boua25075.htm>
- Collier, R.J., Dahl, G.E., VanBaale, M.J., 2006. Major advances associated with environmental effects on dairy cattle. *Journal of Dairy Science*. 89:1244–1253.
- Cramer, D., 2003. *Advanced quantitative data analysis*. McGraw-Hill Education (UK).
- Duda, Richard, O., Peter, E.Hart., David, G., Stork. 2001. *Pattern Classification*. New York: Wiley.
- El-Ramady, H.R., El-Marsafawy, S.M., Lewis, L.N., 2013. Sustainable agriculture and climate changes in Egypt. *Lichtfouse (Ed.), Sustainable Agriculture Reviews*, Vol. 12. DOI 10.1007/978-94-007-5961-9\_2,
- El-Tarabany, M.S., El-Bayoumi, K.M., 2015. Reproductive performance of backcross Holstein × Brown Swiss and their Holstein contemporaries under subtropical environmental conditions. *Theriogenology*. 83, 444–448.
- German, J.B., Dillard, C.J., 2006. Composition, structure and absorption of milk lipids: A source of energy-fat-soluble nutrients and bioactive molecules. *Critical Reviews for Food Science and Nutrition*. 46, 57–92.
- Hahs-Vaughn, D.L., 2017. *Applied multivariate statistical concepts*, 1st edition. Taylor & Francis.
- Hamid, H., 2010. A new approach for classifying large number of mixed Variables. *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*. 4 (10), 1355-1360.
- Hassan, F.A.M., Ali, M.A., El-Tarabany, M.S., 2017. Economic impacts of calving season and parity on reproduction and production traits of buffaloes in the sub-tropics. *Environmental Science and Pollution Research*. 24, 10258–10266.
- Kara, K., Ser, G., Arslan, F., 2005. Discriminant analysis results of some yield characteristics of sample dairy cattle corporations by considering education status of producer. *International Journal of Dairy Science*. 1, 14-17.
- Khosroshahi, Z.T., Rafat, S.A., Shoja, D., 2011. Effects of non-genetic factors in milk production and composition in East Azarbaijan native buffaloes of Iran. *Buffalo Bulletin*. 30, 202–209.
- Mele, M., Macciotta, N.P.P., Cecchinato, A., Conte, G., Schiavon, S., Bittante, G., 2016. Multivariate factor analysis of detailed milk fatty acid profile: Effects of dairy system, feeding, herd, parity, and stage of lactation. *Journal of Dairy Science*. 99, 9820-9833.
- Melendez, P. and Pinedo, P., 2007. The Association Between Reproductive Performance and Milk Yield in Chilean Holstein Cattle. *Journal of Dairy Science*. 90, 184–192.
- Nishida, M., Aziz, A., Nishida, S., Suzuki, J., 2006. Modeling number of services per conception of Japanese Black cattle by random regression. *Animal Breeding and Genetic*. 123, 56-63.
- Ramadan, E.K., El-Tahawy, A.S. 2014. Effect of calving season on the economic and production efficiency of dairy production breeds. *International Scholarly and Scientific Research & Innovation*. 8(7), 798-802.
- Senger, P.L., 2001. Fertility factors in high producing dairy cows-which ones are really important. *The Professional Animal Scientist Journal*. 17, 129-138.
- Siatka, K., Sawa, A., Kr el-Czopek, S., Pivczy ski, D., Bogucki, M., 2017. Effect of some factors on number of services per conception in dairy cows. *Journal of Veterinary Science and Technology*. 8,(5), 1-4.
- Soydan, E., Sirin, E., Ulutas, Z., Kuran, M., 2005. Calving season affects reproductive performance of high yielding but not low yielding Jersey cows. In: *Proceedings of 56th Annual Meeting of European Association for Animal Production*, Pp. 107.
- SPSS Inc ,2004. *SPSS Base User's Guide*. SPSS Inc, Chicago.
- Timm, N.H., 2002. *Applied multivariate analysis*. 2nd edition , Springer Texts in statistics.
- Torshizi, M.E., 2016. Effects of season and age at first calving on genetic and phenotypic characteristics of lactation curve parameters in Holstein cows. *Journal of animal science and technology*. 58, 8. doi: 10.1186/s40781-016-0089-1
- Utts, J., 2005. *Seeing through statistics*. 3rd Edition, Thomson Brooks/Cole, Pp: 166-167. ISBN 0-534-39402-7.
- White, S., Benson, G., Washburn, S., Green J.r, J., 2002. Milk production and economic measures in confinement or pasture systems using seasonally calved Holstein and Jersey cows. *Journal of dairy science*. 85 , 95-104.
- Wondossen, A., Mohammed, A., Enyew, N., 2018. Reproductive Performance of Holstein Friesian Dairy Cows in a Tropical Highland Environment. *Journal of Advance in Dairy Resarche* .6, 2. DOI: 10.4172/2329-888X.1000203
- Worth, A.P., Cronin, M.T.D., 2003. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. *Journal of Molecular Structure*. 622 , 97-111.
- Zicarelli, L., 1997. Reproductive seasonality in buffalo. In: *Proceedings of Third International Course of Biotechnology in Buffalo Reproduction*, Napoli 6–10 October, Italy, Suppl. *Bubalus bubalis*, 4: 29–52.