COMPARISON OF CURVE ESTIMATION REGRESSION METHODS IN PREDICTING PROTEIN AMOUNT FROM TOTAL MILK YIELD IN HOLSTEIN DAIRY CATTLE

FATMA D.M. ABDALLAH AND HAGAR F. GOUDA

Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Egypt.

ABSTRACT

Yielding of milk is of great economic importance for milk processors in dairy industry and for consumers. Also, milk composition has a major role in determining the price of milk. Protein amount is a major constituent in milk so this study focused on predicting its amount from total milk yield. Generally, the total milk yield and protein amount are linearly correlated, so it is important to study this relationship with other nonlinear models. This work attempted to: investigate the relationship between protein amount and milk production, predict protein amount from total milk yield and choose the best fit model for this purpose. Beside the linear model, ten nonlinear regression techniques were used such as power, quadratic and cubic modelling technique and others. Data of 1300 animal from lactation records of Holstein dairy cattle which belongs to Dina farms at Alexandria-Cairo desert road Egypt were used. The regression models (curve estimation regression method) were applied using SPSS software packages version 26. The goodness of fit measures for the best fit model are the highest value of R square and adjusted R square (inadequate or intuitive measures) with the lowest values of standard error of estimate and AIC values (more accurate measure). The results showed that from the 11 regression models, the power model was the best fit model to predict the amount of protein from total milk yield depending on R Square (0.856) and Adjusted R Square (0.856) that were the highest values between the models, smaller standard error of the estimates (0.230) and AIC value (-13135.84) which were the lowest values between the models. The power model could be used for prediction through this equation (protein amount = 0.130 * (total milk yield ** 0.815) after 15 iteration criteria.

Keywords: Adjusted R square, milk yield, curve estimation, protein amount, AIC value and power model.

INTRODUCTION

Milk producers concentrate on milk composition because of its economic importance and its importance to milk consumers. Milk constituents were detected from many years with appearance of Holstein breed with average 3.6 % fat, 3.2 % protein, and 4.7 % lactose (Young *et al.,* 1986).

There are many factors which have effect on milk constituents such as breed, genetic variation within breed, health, environment, management, and food.

Corresponding author: Fatma D.M. Abdallah
E-mail address: moibho_stat2021@yahoo.com
Present address: Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Egypt.

The total amount of protein in milk is detected by analyzing milk for nitrogen and multiplying by a factor of 6.38. The total protein percent of milk is about 3.5, of which 94 to 95 percent is in the form of true protein. Casein represents 80 % of the true protein approximately, whey represents 20%. Alpha-, beta-, gamma-, and kappa-caseins are different types of casein. The other proteins (whey) are beta-lactoglobulin and alpha-lactalbumin. Serum albumin, immunoglobulins, proteose peptones, lactoferrin, and transferrin which considered a smaller part (Davies *et al.,* 1983; Jenness, 1985).

As it is known that different prediction techniques are very important part in statistics. Regression methods with different types are one of the most important techniques in this purpose. These methods are applied when the dependent and explanatory factors are in form of linear or non-linear functions. The nonlinear regression methods are widely applied in studying of animal behavior and breeding (Sengül and Kiraz, 2005). Linear and nonlinear models of regression methods are widely applied to study and predict the relationship between quantitative variables (dependent and independents) in different animals researches (Cankaya, 2009).

Prediction and studying the relationship between milk yield and milk constituents of the dairy animals is very important process for the dairy managers and for human health (Nguyen *et al.,* 2020).

Numerous types of modeling techniques used for forecasting milk constituents from total milk yield with a good forecasting power (Lehmann *et al.,* 2019).

As a general rule, there were a linear relationship between fat/protein yield and milk production (they increase together slowly) (Nguyen *et al.,* 2020).

Nonlinear regression gives the curve-fitting function to detect the best fit curve shape with choosing what's known as starting values for the nonlinear algorithm.

The objective of this work was to compare between eleven different regression models (linear and non-linear) to choose the suitable models for predicting the amount of protein from total milk production, where the linear type was commonly applied in this purpose. The parameter estimates used for comparison. Linear and non-linear (inverse, s-curve, logarithmic, quadratic, cubic, power, compound, growth, exponential and logistic) were applied.

MATERIALS AND METHODS

Data source:
Data were obtained from lactation records of Dina farms at Alexandria-Cairo desert road Egypt (n= 1300) from the period 2010.Variables under study were total milk yield, protein amount.
- Independent variable is TOM (total milk yield).
- Dependent variable is protein amount.

Handling and analysis of data:
The statistical analysis process is divided into two steps: the first one is applying curve estimation regression step for choosing the best fit model. The second one is applying non-linear regression steps for forming the prediction equation which suggested by the previous step.

Statistical models:
Plotting the data is important step firstly to test if the variables are linearly related or not. If the data are not linearly related, transformation can be done or different curve estimation methods can be applied to suggest the best one by using SPSS Version 26 (SPSS, 2020) and (Hassan and Mansour, 2021).

Curve estimation is a mathematical formula or the procedures of drawing a curve which have the best fit to set of data. It is used for predicting the dependent variable from the independent with avoiding multicollinearity problem which lower the accuracy of the model (Tırınk *et al.,* 2020 and Kurnaz *et al.*, 2021).

Linear and nonlinear regression model (inverse, s-curve, logarithmic, quadratic, cubic, power, compound, growth, exponential and logistic) were utilized to study milk production and protein amount relationship.

Robust estimators:
As it is known that ordinary least squares method is not suitable in case of outliers or extremes because of large errors. Robust estimation measures help to decrease the effect outliers by identifying them to give accurate estimate (Almetwally and Almongy, 2018). Huber's M-estimator, Tukey's biweight, Hampel's M-estimator and Andrews' wave are good statistical robust estimators with least effort of computation and rapid convergence (Guo, 2003).

Statistical Hypotheses:
The first hypothesis:
Null Hypothesis:
Protein amount can't be predicted from total milk yield.

Alternative Hypothesis:
Protein amount can be predicted from total milk yield.

The second hypothesis:

Null Hypothesis:
There is no difference between linear and non-linear model in prediction of dependent from independent.

Alternative Hypothesis:
There is a difference between linear and non-linear model in prediction (Kira *et al.,* 2019).

The models mathematics:
Hassan and Mansour (2021) explained the following mathematical formulas which representing different models as follows:

1. Linear regression model:
It is $y = \beta_0 + \beta_1 x + e$, where β_0 is the y intercept of the regression line given by $\beta_0 + \beta_1 x$, β_1 is the slope of the regression line given by $\beta_0 + \beta_1 x$, and e is the deviation of the actual y value from the line (error) given by $\beta_0 + \beta_1 x$. This model assumed that: The error values are independent, normally distributed with zero mean $E(e) = 0$ and constant variance (y variance $= \sigma^2$ and is fixed for all x values) (Mason *et al.,* 2003).

2. Nonlinear regression models:
Nonlinear regression models which were used to study the relationship between amount of protein and total milk yield as in Table 1.

Table 1: Mathematical expression of the equations (curve estimation regression models).

Model	Equation
Linear	$Y = \beta_0 + (\beta_1 * x)$
Inverse	$Y = \beta_0 + (\beta_1 / x)$
S- curve	$Y = e^{**}(\beta_0 + (\beta_1/x))$ or $\ln(Y) = \beta_0 + (\beta_1/x)$
Logarithmic	$Y = \beta_0 + (\beta_1 * \ln(x))$
Quadratic	$Y = \beta_0 + (\beta_1 * x) + (\beta_2 * x^2)$
Cubic	$Y = \beta_0 + (\beta_1 * x) + (\beta_2 * x^2) + (\beta_3 * x^3)$
Power	$Y = \beta_0 * (x^{**} \beta_1)$ or $\ln(Y) = \ln(\beta_0) + (\beta_1 * \ln(x))$
Compound	$Y = \beta_0 * (\beta_1^{**}x)$ or $\ln(Y) = \ln(\beta_0) + (\ln(\beta_1) * x)$
Growth	$Y = e^{**}(\beta_0 + (\beta_1 * x))$ or $\ln(Y) = \beta_0 + (\beta_1 * x)$
Exponential	$Y = \beta_0 * (e^{**}(\beta_1 * x))$ or $\ln(Y) = \ln(\beta_0) + (\beta_1 * x)$
Logistic	$Y = \ln(\beta_0) + (\ln(\beta_1) * x)$

Where Y: protein amount in the prediction equation, β_0: it is the y intercept, β_1: the amount of change in the value of protein amount with one unit change in total milk yield, β_2: The regression factor of squared total milk yield, β_3: The regression factor of cubic total milk yield, X: The independent variable (total milk yield), X^2: square of the total milk yield, X^3: The cube of the total milk yield and In: natural logarithm.

β_1, β_2, β_3 and β_k are the regression coefficients for the k independent variables respectively.

Fitting measures of model selection:
There are many measures for suggesting the best model (goodness of fit measures) such as the coefficient of determination which considered the square of correlation coefficient, Adjusted R-squared, Akaike information criterion (AIC) and low standard error estimate.

The coefficient of determination is
R^2 = SS explained (regression) / SS Total.

SS explained: Sum of squares in regression analysis and SS Total: Total sum of squares in regression analysis.

R^2 measure is frequently applied but it is not considered a suitable measure for nonlinear models performance because of many causes (it does not explain parameters number and the full model does not contain single parameter model) so other measures criteria for model selection suggested (Wallach, 2006).

Adjusted R^2 = $(1-R^2)$ (N-1)/N-P-1 = 1- (n-1/n-p)*$(1-R^2)$.

N: Sample size. R^2: Coefficient of determination. P: Number of regression parameters.

Mean square error, MSE = SSE/(n−k), where n is the data values, SSE is error sum square and k is the parameters number.

Akiake Information Criterion:
It is a statistic for choosing suitable model after comparison of different models.
(AIC) = n*ln (SSe/n) + 2k, where n is the number of data values, k is the number of regression parameters. SSe is the error sum of squares and its small value is preferable (Akaike, 1974). AIC value is a guide for selecting better model, where its lower value is preferable than higher.

Standard error of the estimate:
It is a statistic of fitness of regression model in prediction process. Its smaller values between different models is preferable. It is the square root of the average squared deviation.

$$SEE = \sqrt{\frac{\sum(x_i - \bar{x})}{n - 2}}$$

RESULTS

Protein amount and milk yield were statistically described as in table 2. Correlation measures was applied to describe the strength of association between Protein amount and milk yield as in table 3.

Table 2: Descriptive statistics of total milk yield (independent variable) and protein amount (dependent one).

	Total milk yield (kg)	Protein amount (gm)
Mean	8230.71	198.79
Median	8161	196
Std. Dev.	3621.56	87.66
Variance	13115691.69	7684.35
Kurtosis	1.39	- 0.29
Skewness	0.56	0.23
Minimum	148	7.00
Maximum	26950	520

Table 3: Correlation between protein amount and total milk yield.

	Value	Significance (2-sided)
Pearson's correlation	0.857	0.000**
Kendall's tau	0.719	0.000**
Spearman's rank	0.886	0.000**

Huber's M-estimator and other measures of robust regression which applied for decreasing the effect of outlier for giving accurate estimates than ordinary least square method shown in Table 4.

Table 4: Robust estimators of protein amount and milk yield of 1300 animal.

	Huber's M-Estimator	Tukey's Biweight	Hampel's M-Estimator	Andrews' Wave
Protein amount	195.72 (gm)	195.36 (gm)	196.39 (gm)	195.35 (gm)
Total milk yield	8128.45(kg)	8041.16(kg)	8076.52(kg)	8038.37(kg)

There are 11 curve estimation models (inverse, s, logarithmic, linear, quadratic, cubic, power, compound, growth, exponential and logistic).

Table 5: Curve estimation regression model summaries for prediction protein amount.

	Model summary						
Model	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**	**SSE**	**F value**	**F significance**	**AIC**
Inverse	0.152	0.151	80.749	6520.41	232.881	0.000**	2100.34
S	0.441	0.440	0.453	0.205	1022.198	0.000**	-11377.32
Logarithmic	0.630	0.629	53.361	2847.39	2207.659	0.000**	1023.25
Linear	0.735	0.735	45.130	2036.74	3602.9	0.000**	587.68
Quadratic	0.770	0.770	42.041	1767.74	2175.378	0.000**	405.30
Cubic	0.781	0.780	41.080	1687.56	1539.676	0.000**	347.19
Power	0.856	0.856	0.230	0.053	7712.769	0.000**	-13135.84
Compound	0.640	0.640	0.364	0.132	2307.930	0.000**	-11949.59
Growth	0.640	0.640	0.364	0.132	2307.930	0.000**	-11949.59
Exponential	0.640	0.640	0.364	0.132	2307.930	0.000**	-11949.59
Logistic	0.640	0.640	0.364	0.132	2307.930	0.000**	-11949.59

R^2: Coefficient of determination. SSE: Sum square of error – AIC: Akaike's Information Criteria.

The coefficients of the curve estimation models for predicting dependent variable using the independent variable were shown in table 6.

Table 6: Parameter estimates (Coefficients) and T test with significance for eleven regression models for predicting protein amount.

Equation	Unstandardized Coefficients				Standardized Coefficients Beta			T test with significance			
	Constant (b_0)	b1	b2	b3	b1	b2	b3	constant	X	X^2	X^3
Inverse	214.69	-79343.61			-0.39			86.91 0.00**	-15.26 0.00**		
S	5.34	-933.18			-0.66			385.35 0.00**	-31.97 0.00**		
Logarithmic	-798.34	112.29			0.79			-37.53 0.00**	46.99 0.00**		
Linear	27.97	0.021			0.79			8 0.00**	60.03 0.00**		
Quadratic	-23.01	0.034	-7.159E-007		1.40	-0.58		-4.97 0.00**	34.26 0.00**	-14.1 0.00**	
Cubic	11.59	0.02	1.030E-006	-5.449E-011	0.77	0.83	-0.84	1.84 0.07ns	8.58 0.00**	4.55 0.00**	-
Power	0.06	.91			0.93			10.91 0.00**	87.82 0.00**		
Compound	57.70	1.00			2.23			39.92 0.00**	358969.87 0.00**		
Growth	4.06	0.00			0.80			161.89 0.00**	48.04 0.00**		
Exponential	57.70	0.00			0.80			39.92 0.00**	48.04 0.00**		
Logistic	0.017	1.000			0.45			39.92 0.00**	358969.87 0.00**		

-The dependent variable is ln(PA) in compound, S, growth, exponential and power. -The dependent variable is ln(1 / PA) in logistic. – The dependent variable in other model is PA. -The independent is TMY.

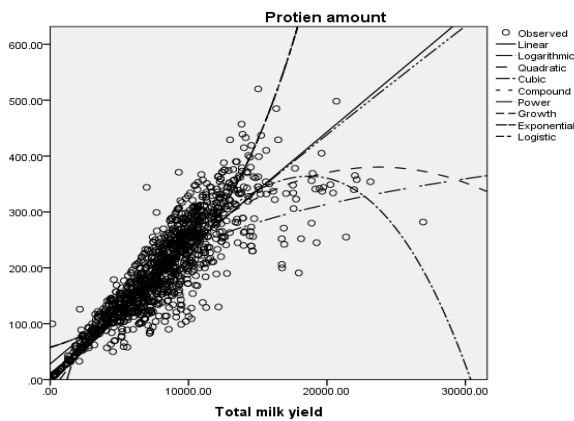The suggested models for prediction process as in the following chart.

Figure 1: Curve estimation regression for nine regression models suggested for predicting the relationship between protein amount and total milk yield.
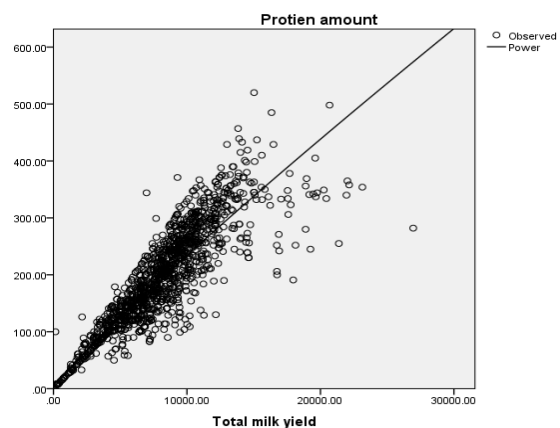
The best fit model was shown in figure 2.

Figure 2: Curve estimation regression for power regression model (the best fitted model) for predicting the relationship between protein amount and total milk yield.

After applying curve estimation step for choosing suitable models as mentioned above. The results of nine nonlinear regression procedures applied for forming the prediction equations as in table 7.

Table 7: Prediction equations of protein amount from total milk yield.

Model	Expectation equation
Power	protein amount = 0.130 * (total milk yield**0.815)
Logistic	Protein amount = ln (0.02) + (ln(0.8) * total milk yield).
Exponential	ln (Protein amount) = ln(58) + (0.0002 * total milk yield).
Growth	ln (Protein amount) = 4+ (0.0002 * total milk yield).
Compound	Protein amount = 56 * (0.5** total milk yield).
Logarithmic	Protein amount = -800 + (110 * ln(total milk yield)).
Cubic	Protein amount = 11 + (0.020 * total milk yield) + (0.0001 * total milk yield ^{2}) + (0.000* total milk yield ^{3}).
Quadratic	Protein amount = 22 + (-0.04*total milk yield) + (-0.0001* total milk yield^2).
Linear	Protein amount = 27 + (0.01* total milk yield).

- All models done after one model evaluation step. – Power model after 15 iteration criteria.

DISCUSSION

There were a high positive correlations with highly significant P value between independent and the dependent where the values of the correlation coefficient were 0.857, 0.719 and 0.886 for Pearson's correlation, Kendall's tau and Spearman's rank measures respectively as in table 3.

Huber's M-estimator, Tukey's biweight, Hampel's M-estimator and Andrews' wave, showed nearly the same results as a good indicator of avoiding the outliers effect. These results were in agreement with Okagbue *et al.* (2019).

Table 5 showed the summaries of the models, ANOVA for testing significance of the models to suggest the most suitable model. It is found that P-values for all models is less than 0.05 which means significance of the models and higher R^2 means model fitness.

Depending on R^2 values, it is found that the power, cubic, quadratic and linear models had high R^2 values, indicating fitting them to predict the amount of protein from total milk yield (0.856, 0.781, 0.770, and 0.735), respectively. It means that (%85.6, %78.1, %77, and %73.5) of the total variation is explained by the model. Adjusted R Square

were (0.856, 0.780, 770, and 0.735), respectively and smaller standard error of the estimates indicated suitability of the models, but it is wrong to depend on these measures alone (inadequate or rough measures).

S curve and inverse non-linear regression model were not suitable to predict the amount of protein from total milk yield because of low value of R^2 and adjusted R^2 (0.441 and 0.440) and (0.152 and 0.151) respectively.

The rest models predict protein amount moderately with adjusted R^2 more than 0.5.

Depending on AIC values, it is found that power model was the best fit model with the lowest AIC value (-13135.84) followed by logistic, exponential, growth and compound which their values were (-11949.59) as shown in table 5.

Based on highest R^2, Adjusted R^2 and lowest AIC value, the power model is the best fit model to predict protein amount from total milk yield.

Finally, according to the lowest AIC value that considered important measure of goodness of fit and highest R^2, the models of prediction could be arranged as follows:

Power > (logistic, exponential, growth and compound) > cubic > quadratic > linear > logarithmic.

Table 6 showed a significant effect for all parameters depending on t test and its P value which was highly significant indicating that the null hypothesis was rejected and the alternative was accepted.

Figure 1 showed the suggested models (nine from eleven) for studying the relationship and predicting purpose between dependent and independent, while Figure 2 explained that the power model fits the observed data values well. The line turned upward and most point near to the line indicating small residual.

The power model equation is $Y = \beta_0 * (x^{**}\beta_1)$.
(Protein amount = 0.130 * (total milk yield**0.815) after 15 model evaluation or iteration criteria.

The power model mentioned that the predicted values of protein amount is equal to 0.130 * (total milk yield**0.815). The resulted predicted value was positive which indicate an increase of the protein amount by the increase of total milk yield.

The other models of predictions were shown in table 7. These equations applied to predict the amount of protein from total milk yield in a manner differ from the known linear model. The results may be positive or negative according to the type of model. The negative results indicated inverse relationship between the variables.

CONCLUSION

This research concentrated on examining different regression models depending on two different methods of regression analysis to suggest which models would be suitable for predicting protein amount from total milk yield. Beside linear regression model which known for all (linear relationship between milk yield and protein amount), it is found that power model is more suitable than linear. Other models such logistic, exponential, growth and compound, logarithmic, quadratic and cubic models (non-linear regression models) were suitable also for the prediction process. S curve and inverse non-linear regression model were not suitable to fit this data.

REFERENCES

Akaike, H. (1974): A new look at the statistical model identification. IEEE Transact Autom Control, 19:716-723.

Almetwally, E.M. and Almongy, H.M. (2018): Comparison between M-estimation, S-estimation, and MM estimation methods of robust estimation with application and simulation. International journal of mathematical archive, 9(11) 55-63. Available online through www.ijma.info ISSN 2229 – 5046.

Cankaya, S. (2009): A comparative study of some estimation methods for parameters and effects of outliers in simple regression model for research on small ruminants. Trop. Anim. Health Prod., 41: 35–41.

Davies, D.T.; Holt, C. and Christie, W.W. (1983): The composition of milk. Ch. 5 in Biochemistry of Lactation, T. B. Mepham, editor., ed. Amsterdam: Elsevier.

Guo, Y. (2003): Application and computation of Huber's M-estimator in the global positioning system. A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Master of Science in Computer Science.

Hassan, K.H. and Mansour, A.S. (2021): Comparisons of some nonlinear functions to describe quail growth curves of half diallel cross progeny. Nat. Volatiles & Essent. Oils, 8(6): 2571-2578.

Jenness, R. (1985): Biochemical and nutritional aspects of milk and colostrum. Ch. 5 in Lactation, B. L. Larson, editor. , ed. Ames: Iowa State University Press.

Kira, I.; Barger, B.; Shuwiekh, H.; Kucharska, J. and Al-Huwailah, A.H. (2019): Cumulative Stressors and Traumas and Suicide: A Non-Linear Cusp Dynamic Systems Model. Psychology, 10: 1999-2018. https://doi.org/10.4236/psych.2019.10 15128.

Kurnaz, B.; Önder, H.; Piwczynski, D.; Kolenda, M. and Sitkowska, B. (2021): Determination of the best model to predict milk dry matter in high milk yielding dairy cattle. Acta Sci. Pol. Zootechnica, 20(3): 41–44.

Lehmann, J.O.; Mogensen, L. and Kristensen, T. (2019): Extended lactations in dairy production: Economic, productivity and climatic impact at herd, farm and sector level, Livestock science, 220: 100-110.

Mason, R.L.; Gunst, R.F. and J.L. Hess (2003): Statistical design and analysis of experiments. With application to engineering and science. Second edition, John Wiley & Sons Inc.

Nguyen, Q.T.; Fouchereau, R.; Fr´enod, E.; Gerard, C. and Sincholle, V. (2020): Comparison of forecast models of production of dairy cows combining animal and diet parameters. Elsevier.

Okagbue, H.I.; Adamu, P.I.; Bishop, S.A.; Obasi, E.C.M. and Akinola, A.O. (2019): Curve estimation models for estimation and prediction of impact factor and citescore using the journal percentiles: A case study of telecommunication journals. iJOE, Vol. 15, No. 14. https://doi.org/ 10.3991/ijoe.v15i14.11373.

Sengül, T. and S. Kiraz. (2005): Non-linear models for growth curves in large white turkeys. Turk. J. Vet. Anim. Sci., 29: 331–337.

SPSS (2020): Statistical package for the Social Science. New York, SPSS Inc. Available at: https://www.ibm.com/ support/knowledgecenter/en/SSLVM B_26.0.0/statistcs_mainhelp_ddita/sp ss/regression/cmd_nlr_models.html.

Tırınk, C.; Abacı, S.H. and Önder, H. (2020): Comparison of ridge regression and least squares methods in the presence of multicollinearity for body measurements in saanen kids. Journal of the Institute of Science and Technology, 10(2), 1429–1437. DOI: 10.21597/jist.671662.

Wallach, D. (2006): Evaluating crop models. In: D. Wallach et al., editors, working with dynamic crop models: Evaluations, analysis, parameterization, and applications. Elsevier, Amsterdam. p. 11–53.

Young, C.W.; Hillers, J.K. and Freeman, A.E. (1986): Production, consumption, and pricing of milk and its components. J. Dairy Sci., 69: 272.

مقارنة طرق الانحدار (تقدير المنحنى) في التنبؤ بكمية البروتين من الانتاج الكلي للحليب في أبقار الهولشتاين الحلاب

فاطمة دسوقي محمد عبدالله ، هاجر فتحي جودة

E-mail: moibho_stat2021@yahoo.com   E-mail: stathagarfathi@gmail.com.
Assiut University web-site: www.aun.edu.eg

يعتبر إنتاج الحليب ذات أهمية اقتصادية بالغة في صناعة الألبان بالنسبة لمصنعي الألبان وللمستهلكين، كما أن مكونات الحليب لها دور رئيسي في تحديد سعر الحليب. وحيث أن كمية البروتين تعتبر مكونا رئيسيا من مكونات الحليب لذلك ركزت هذه الدراسة على التنبؤ بكميته من إجمالي إنتاج الحليب.

حيث أن إجمالي إنتاج الحليب وكمية البروتين بشكل عام يرتبطان خطيا، لذلك كان من المهم دراسة هذه العلاقة باستخدام النماذج الغير خطية الأخري.

تهدف هذه الدراسة إلي دراسة العلاقة بين كمية البروتين وإنتاج الحليب، والتنبؤ بكمية البروتين من إجمالي إنتاج الحليب واختيار النموذج الأفضل لهذا الغرض.

تم استخدام نماذج الإنحدار الغير خطية بجانب النموذج الخطي حيث تم تطبيق عشرة نماذج غير خطية مثل نموذج power و quadratic و cubic ونماذج أخري.

تم استخدام بيانات ١٣٠٠ حيوان تم الحصول عليها من سجلات الحليب الخاصة بأبقار الهولشتاين الحلاب التابعة لمزارع دينا، حيث تم استخدام الحزمة الإحصائية SPSS ذات الإصدار ٢٦ لتطبيق نماذج الإنحدار وتقدير المنحني. كانت مقاييس حسن الملائمة لتحديد النموذج الأفضل هي أعلي قيمة لل R^2 و adjusted R^2 ولكنهم يعتبروا مقاييس غير كافية أو مبدئية، أيضا تم استخدام أقل قيمة للخطأ المعياري وأقل قيم AIC حيث يعتبروا مقاييس أكثر دقة.

أظهرت النتائج أنه من بين ١١ نموذج مختلف من نماذج الإنحدار كان نموذج ال power هو أفضل نموذج للتنبؤ الأمثل بكمية البروتين من كمية الحليب الكلية اعتمادا علي قيمة R^2 وقيمة adjusted R^2 حيث كانتا أعلي قيمة من بين النماذج حيث كانت تساوي (0.856) ، أيضا أقل قيم للخطأ المعياري (0.230) وأقل قيمة لل AIC وكانت تساي (13135.84-) كانت هذه القيم هي الأقل في حالة نموذج power عن باقي نماذج الإنحدار الأخري.

يمكن تطبيق نموذج power في عملية التنبؤ من خلال المعادلة: كمية البروتين = 0.130*(كمية الحليب الكلي**0.815) بعد تكرار العملية اكثر من مرة.