Biometrical Letters
Vol. 44 (2007), No. 1, 71-81

Multiple Regression Models for Lactation Curves

S.P. Pereira[1], T.A. Oliveira[2] and J.T. Mexia[3]

[1]Department of Animal Production Systems. - EZN/INIAP., Vale de Santarem, Portugal martavc@pluricanal.net
[2] UAb, Department of Exact and Technologic Sciences, R. Fernão Lopes, nº 9, 2D, Lisbon, and CEAUL, Center of Statistics and Applications, University of Lisbon, Portugal toliveir@univ-ab.pt
[3] Department of Mathematics, FCT- New University of Lisbon, Quinta da Torre, 2825 Monte da Caparica, Portugal, jtm@fct.unl.pt

SUMMARY

Several methods have been developed in order to study lactation curves. However, the lactation curves are often not well adjusted since several factors affect milk production. The usual model used to describe a lactation curve is Wood's Model, which generally uses a logarithmic transformation of an incomplete gamma curve to obtain least squares estimates of three constants: a – a scaling factor associated with average daily yield; b – associated with pre-peak curvature; and c – associated with post-peak curvature (Wood, 1976). Some disadvantages of Wood's model are strongly connected with overestimation of milk production at the beginning of lactation, with underestimation of the lactation peak; the self-correlated residuals and highly correlated parameter estimates (Scott et al, 1996). Fleischmann's Method is usually used to estimate total milk production. This method generally overestimates actual yields up to peak lactation as well as yield during the period following the last measurement, but underestimates yields for other periods (Norman et al, 1999). The total milk yield estimate according to this method considers a constant daily milk production between two records and equal to the mean of these two records, which does not describe the true variation of milk secretion during lactation.

The mentioned disadvantages led us to consider the milk curve concept as a graphical representation of milk production described by mathematical models.

In our work we considered a new approach using polynomial regression, one for each group. Polynomial curves were adjusted to daily milk records for each group and the respective hypo-graphic area was calculated to estimate total yields. An ANOVA to the comparison of these total yields was carried out and the Scheffé multiple comparison method was applied.

This approach greatly increases the power of the test, enabling work with smaller experiments, the reason for this increase being the replacement of classical replicates by time replicates, leading to a great increase in the degrees of freedom. Another advantage of this method is the use of a continuous process instead of an obligatory discrete process conversion. Differences between protein supplements and stocking rate were found using an adaptation of Scheffé's method. We concluded that a lower stocking rate and high protein content in supplement enable higher milk production.

Key words: multiple regression, lactation curve, ewe's milk production

1. Introduction

The lactation curves of ewes must be taken into consideration when selecting them. Thus the discussion and adjustment of these curves have been considered for a long time. Different statistical approaches have been proposed for lactation curves analysis. Usually these approaches are based on a theoretical pattern in which there is an initial phase of increasing yields, followed by a phase of decreasing yields. Both phases are separated by a single peak. Unfortunately the adjustments achieved by these techniques are often quite poor. Thus, starting with the classical Wood's model (Wood, 1967, 1969), several modifications have been proposed (Ruiz et al, 2000).

Macciota et al (2005) relates that the principal disadvantage of Wood's model is that it leads to adjusted lactation curves which are either concave or convex.

Another important piece of information given by the lactation curve is the total yield production estimate. Estimates are made using Fleischmann's method. This test was improved to adjust the first, second, and last test intervals for the nonlinear shape of the lactation curve, in order to prevent biases. The main disadvantages of this method are that it generally overestimates actual yields up to peak lactation as well as yield during the period following the last measurement, but underestimates yields for other periods. In this study, different applications of the Fleischmann method were compared with a new approach and the circumstances for using this method are defined.

We will suggest a new approach to lactation curves based on multiple regression designs, Mexia (1987). This approach is more flexible and enables the statistical comparison of total yields when we have randomly assigned groups of ewes to different factor levels.

We can use this method to obtain the significance of the factor.

2. Statistical methods

Let us assume that degree n polynomials are adjusted and that we have the yields of t lactation days for ℓ levels of a factor. Thus for these levels we will have the polynomials $\sum_{j=0}^{n} \beta_{i,j} x^j, i = 1,\dots,\ell$, since

$$\int_0^t \sum_{j=0}^{n} \beta_{i,j} x^j dx = \sum_{j=0}^{n} \left(\frac{\beta_{i,j}}{j+1} \right) t^{j+1} \; ; \quad i = 1,\dots,\ell$$

the total milk productions up to time h will be $\sum_{j=0}^{n} \beta_{i,j} \frac{h^{j+1}}{j+1}, \quad i = 1,\dots,\ell$.

In practice these times will correspond to lactation days, so that we have the values

$$x_{i,j} = \frac{i^{j+1}}{j+1}, \quad i = 1,\dots,\ell; \, j = 0,\dots,n$$

for the controlled variables which in this case will be time intervals (measured in number of days). If for the ℓ factor levels we have the yield vectors $\underline{Y}_i, i = 1,\dots,\ell$ we will have the adjusted coefficient vectors

$$\underline{\hat{\beta}}_i = (X^t X)^{-1} X^t \underline{Y}_i \; ; \quad i = 1,\dots\ell$$

where $X = [x_{i,j}]$.

Now, under the usual assumptions for linear regression, $\underline{\hat{\beta}}_i, i = 1,\dots,\ell$, will have the variance covariance matrix $\Sigma(\underline{\hat{\beta}}_i) = \sigma_i^2 (X^t X)^{-1}; i = 1,\dots,\ell$.

Assuming that the $\sigma_i^2; \quad i = 1,\dots,\ell$, are equal, we will have for the $\hat{\gamma}_i = \underline{c}^t \underline{\hat{\beta}}_i \; ; \quad i = 1,\dots,\ell$, the variance:

$$Var(\hat{\gamma}_i) = \sigma^2 \underline{c}^t (X^t X)^{-1} \underline{c} = \sigma^2 k$$

Moreover if we have milk yields over t days, the sums $S_i \; ; i = 1,\dots,\ell$ of squares of residues for the different regressions will be the products by σ^2 of independent central chi-squares with *t-n-1* degrees of freedom. Thus $S = \Sigma S_i$ will be the product by σ^2 of a central chi-square with $g = \ell(t - n - 1)$ degrees of freedom independent from the $\hat{\gamma}_i \; ; \quad i = 1,\dots,\ell$.

Besides this

$$Q = \sum_{i=1}^{\ell} \hat{\gamma}_i^2 - \frac{1}{\ell} \left(\sum_{i=1}^{\ell} \hat{\gamma}_i \right)^2$$

will be the product by σ^2 of a chi-square with $\ell - 1$ degrees of freedom and non-centrality parameter

$$\delta = \frac{1}{\sigma^2}\left[\sum_{i=1}^{\ell}\hat{\gamma}_i^2 - \frac{1}{\ell}\left(\sum_{i=1}^{\ell}\hat{\gamma}_i\right)^2\right].$$

Thus $F = gQ/((l-1)S)$ will have F distribution with $\ell - 1$ and g degrees of freedom and non-centrality parameter δ.

Since $\delta = 0$ if and only if the hypothesis $H_0 : \gamma_1 = ... = \gamma_\ell$ holds, the F test will be strictly unbiased.

2.1. Scheffé's multiple comparisons test

Scheffé's test is considered the most conservative test for multiple comparisons of estimable functions, but it can be used to test all mean contrasts, controlling the global level of significance (Scheffé, 1959; O'Neill and Wetherill, 1971). According to Steel and Torrie (1981), this test has low power because of the fact that it caters for a large number of comparisons. However, Mexia (1987, 1989) defends this method due to its flexibility and robustness. As it is well known, this method is intimately linked to the ANOVA F test, since there will be at least one significant contrast if the commonly used ANOVA F test is significant (O'Neill and Wetherill, 1971).

In this work, a similar procedure is used to adapt Scheffé's tests for contrasts using the MSE obtained by Mexia's approach. The critical value for Scheffé's statistics results from the fact that, with a global confidence level of (1-α)100%, every possible linear combination Ψ of k factor level means will fall in the interval:

$$\hat{\psi} - \sqrt{(k-1)\, f_\alpha\ MSE \sum_{i=1}^{k}\frac{c_i^2}{n_i}} \ \langle\ \psi\ \langle\ \hat{\psi} + \sqrt{(k-1) f_\alpha\ MSE \sum_{i=1}^{k}\frac{c_i^2}{n_i}}$$

where $MSE\sum_{i=1}^{k}(c_i^2/n_i)$ is the estimated error variance, $\hat{\psi}$ is the estimate of the contrast and f_α is the quantile of order 1-α, on an F distribution with k-1 and N-k degrees of freedom.

3. Motivating example

Over three years, from 1992 to 1994, 160 Serra da Estrela ewes from the herd at Portugal's Estação Zootécnica Nacional were observed and their milk production was recorded. Ewes were divided into two lambing seasons: September and February.

In September 1992 and September 1993, ewes were distributed among 4 groups with the following diets: (1) irrigated pasture and broken corn supplement; (2) irrigated pasture and "corn gluten feed" (CGF) supplement; (3) corn silage enriched with urea and broken corn supplement; and (4) corn silage enriched with urea and "CGF" supplement. Urea was introduced into the group 3 and 4 diets to balance crude protein differences between pasture (over 20%) and corn silage (10%) within the acceptable limits of animal nutrition (NRC, 1985). The corn silage and urea diets end up with 16% crude protein (CP). Supplements were isoenergetic – 5.6 Mj of metabolizable energy (ME) – and differed in protein quantity: 24% CP in Dry matter (DM) of CGF and 10% CP in DM of broken corn. ME was 25% of the energy required for a 55 kg live weight ewe suckling 1.5 lambs during the first 6 to 8 weeks of lactation (NRC, 1985). Rotational grazing was used, with ewes changing paddock whenever all the grass was consumed. Ewes began grazing each paddock when the grass reached a height of 25 cm.

In February 1993 and 1994, ewes were divided into 4 groups, which grazed on a medics and cocksfoot pasture subject to the combinations of two stocking rates and two supplements: (1) low stocking rate and broken corn supplement; (2) low stocking rate and "CGF" supplement; (3) high stocking rate and broken corn supplement; and (4) high stocking rate and "CGF" supplement. The quantity and quality of supplements were identical to those of September of 1992 and 1993. Rotational grazing was used. Groups had the same number of animals, while the grazing paddocks were double the size of the others. Animals changed paddocks when the leaves of medics for the group with the highest stocking rate were completely consumed. Ewes started grazing each paddock whenever the grass reached a height of 15 cm.

Milking started on the 21st day after birth, once a day at 5 pm until the 42nd day. During this phase (partial weaning) lambs were separated from their mothers between 5 pm and 8 am. On the 42nd day the lambs were totally weaned and the ewes were milked twice a day (5 pm and 5 am).

4. Results and discussion

Table 1 gives nutritional and grazing indicators for September 1992 and September 1993. It shows that ewes had access to a more nutritional pasture during the first year with an availability allowing a good intake during the period of the trial.

Table 1. Grazing indicators in September/1992 and September/1993

Season	*Sept92*	*Sept92*	*Sept93*	*Sept93*
Lactation phases	21-42 days	42-final	21-42 days	42-final
Dry Matter (kg/ha)	2206±39.8	2197±1232.0	1492±557.5	2291 ± 1076.3
Stocking rate (ewe/ha)	80 ± 6	30 ± 8	70 ± 6	20 ± 6
Crude Protein (100gr of DM)	25.5 ± 0.88	26.3 ± 5.25	21.0 ± 8.81	19.0 ± 3.37
ADF (100gr of DM)	26.4 ± 0.78	26.8 ± 1.75	24.8 ± 2.08	24.8 ± 2.78
DM Digestibility	76.7 ± 4.33	75.5 ± 0.01	73.4 ± 2.83	79.2 ± 3.42

Note: The first milking day corresponds to the 21st day of lactation.

Table 2 gives similar indicators for February 1993 and February 1994.

Table 2. Grazing indicators in February/1993 and February/1994.

Season		*Feb93*	*Feb93*	*Feb94*	*Feb94*
Lactation phases		21-42 days	42-final	21-42 days	42-final
Dry Matter (kg/ha)		2212 ± 877.2	3098 ± 271.5	1839 ± 295.0	4267 ± 643.5
Stocking rate	Lower	64 ± 35.2	230 ± 99.3	77 ± 10.4	107 ± 25.5
(ewes/ha)	Greater	138 ± 88.0	478 ± 237.3	123 ± 15.6	175 ± 78.4
Crude Protein (100gr of DM)		15.6 ± 4.19	18.9 ± 3.0	21.1 ± 3.2	13.2 ± 3.4
ADF (100gr of DM)		27.9 ± 4.22	26.7 ± 7.2	29.5 ± 5.2	56.6± 8.5
NDF (100gr of DM)		42.2 ± 7.88	39.2 ± 8.4	52.8 ± 6.5	65.9 ± 8.3
DM Digestibility		80.5 ± 6.54	74.6 ± 7.3	70.5 ± 7.8	67.4 ± 7.4

Note: The first milking day corresponds to the 21st day of lactation

Dry matter produced in February 1994 was substantially greater. Pasture sampled in February 1994 showed a decrease in quality in the final grazing period (CP - 13\% and ADF 32.6\%). This decrease was surely a consequence of lower rainfall in March and April, which lead to early maturing of the grass and reduction of its nutritional value.

The lactation curves showed that it was adequate to adjust a polynomial regression model (4th degree) for each level of the factor. The results are presented in the following graphs. As we can observe, estimated curves for

February 1993, September 1992 and September 1993 are atypical curves. They present increases at the end of the trial. We also can see that there are different shapes for the groups.

These characteristics will make Wood's estimates unfeasible, since they result in a mathematical forcing of the actual pattern into the form of a single curvature for the entire curve.(completely concave (standard shape) or convex (atypical)) (Macciota et al 2005) .

This new approach is more flexible since we can adjust several polynomial degrees.

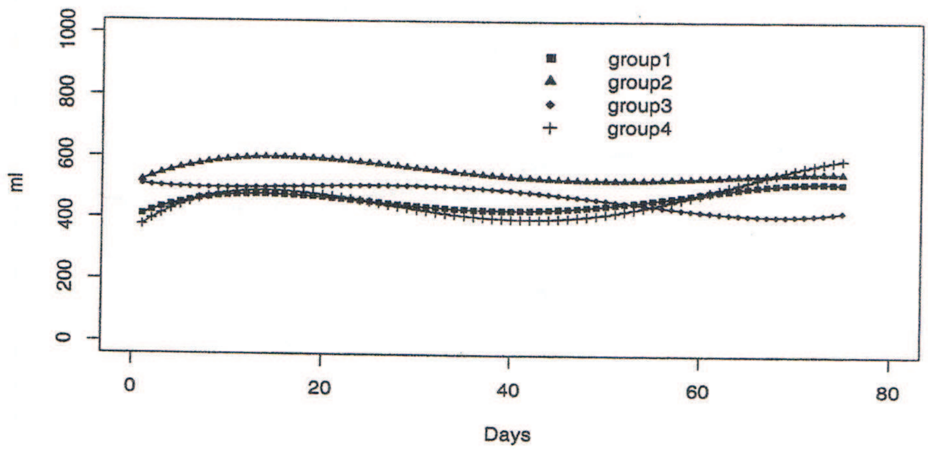

Figure 1. Estimated curves for each of the feed groups for February 1993

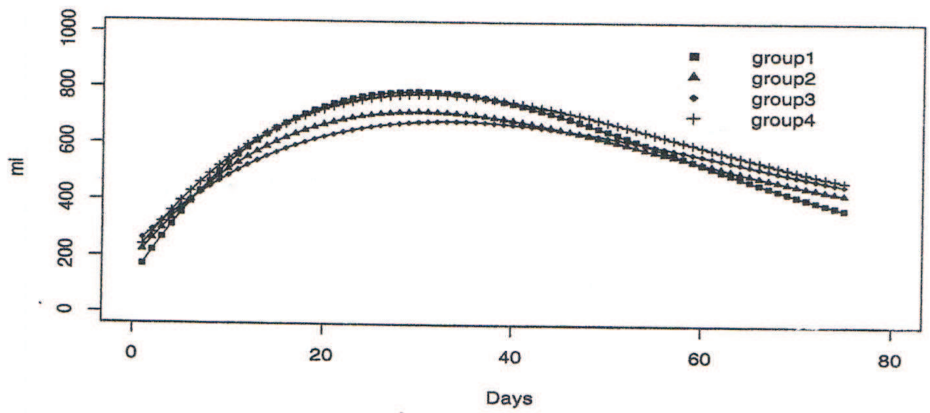

Figure 2. Estimated curves for each of the feed groups for February 1994

The estimated curves for February 93 are very different of those estimated for February 94, these results being due to the differences in pasture. The homogeneity of pasture quality during the entire trial period in February 93 led to more homogeneous curves. In 1993, the lower stocking rate with the CGF group tended to produce more milk because these ewes were given a greater availability of DM and the amount of non-degradable protein is especially important during the beginning of the lactation phase. Those effects were not so evident during the 1994 lactation.

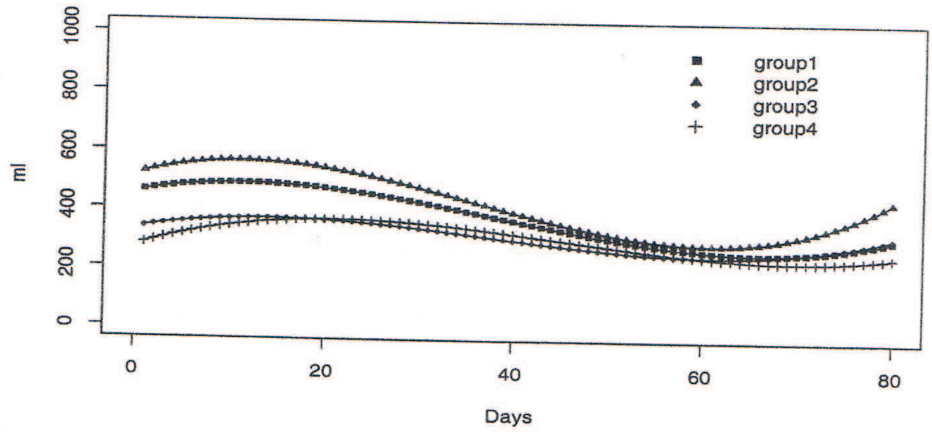

Figure 3. Estimated curves for each of the feed groups for September 1992

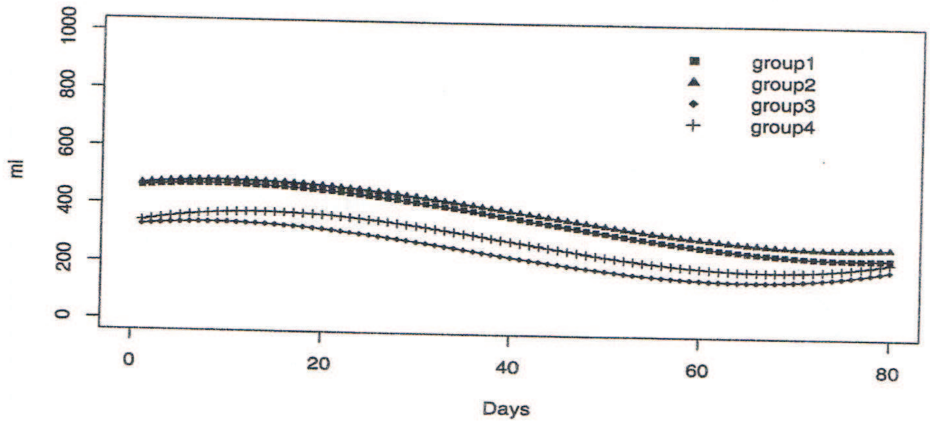

Figure 4. Estimated curves for each of the feed groups for September 1993

The curves estimated for the groups in September 92 are similar to those obtained for September 93. We can observe that group 1 and group 2 in September had higher milk productions.

The next table (Table 3) shows actual total yields and the estimates obtained by Mexia's approach, and as we can see this approach underestimates by 50 percent actual yields for all years and groups.

Table 3 : Total milk production, estimated and actual yields

		GROUPS			
		1	2	3	4
Sept 92	Estimated	29.7	33.8	24.6	24.1
	Actual	55.4	65.1	45.5	45.1
Sept 93	Estimated	28.8	30.6	19.3	22.6
	Actual	54.7	54.9	37.8	42.5
Feb 93	Estimated	34.6	41.7	35.6	34.5
	Actual	64.8	80.4	73.2	62.4
Feb 94	Estimated	44.9	43.2	42.9	47.2
	Actual	85.9	84.3	97.1	93.7

We estimated total production by Mexia's approach considering all observations, each 3 days' observations and by Fleischman's method considering each 3 days' measures. The results are presented in the next table.

Table 4. Total milk production estimated and actual yields by our method and Fleischman's method, for September 1992

	GROUPS			
	1	2	3	4
Actual	55.4	65.1	45.5	45.1
All days	29.7	33.8	24.6	24.1
	54%	52%	54%	53%
Each 3 days	28.2	33.8	24.6	24.1
	51%	495	505	50%
Fleischman's	27.5	30.7	22.3	22.8
	50%	475	495	51%

We can notice that all methods underestimated actual yield in 50, even Fleischman's method. We can conclude that this new approach is useful and precise (but not exact) for estimating total milk production.

The next table presents the ANOVAs summaries for each period.

Table 5. Variance analysis summary table

	SSE	G=i8t-j)	MSE	F
Sept 92	195628	280	52508	223
Sept 93	340166	300	85194	173
Feb 92	143302	228	47223	427
Feb 94	130659	213	46057	625

As can be seen in Table, 5 a high number of degrees of freedom are used and a high QME is estimated. This greatly increases the power of the test, since estimated MSE are very high, with very low p-levels ($p<0.000000001$).

Scheffé's method was used to compare between Low opposed to High stocking rate (Groups 1 and 2 vs. groups 3 and 4) and Broken corn opposed to CGF (groups 1 and 3 vs. groups 2 and 4) in February and Irrigation past. opposed to Corn silage (Groups 1 and 2 vs. groups 3 and 4) and Broken corn opposed to CGF (Groups 1 and 3 vs. groups 2 and 4) in September. The critical intervals obtained by this method are presented in Table 6.

Table 6. Contrasts between groups and Scheffé's Critical Intervals

	G1 and G2 vs. G3 and G4		G1 and G3 vs. G2 and G4	
	L	U	L	U
Sept 92	334.3	851.1	-376.8	-866.0
Sept 93	184.2	705.2	-216.3	-840.5
Feb 92	255.8	744.9	-168.1	-684.9
Feb 94	-160.2	463.9	-378.9	-899.9

Contrast estimates and critical values are presented in Table 6 for each test at a confidence level of 95%.

5. Conclusions

Irrigated pasture permits greater milk production than corn silage, and higher supplement protein content (CGF with 24) stimulates milk production.

No differences were found between stocking rates in February 1994. It may be concluded that grass availability for the group with the highest stocking rate made it possible for ewes to express their full productive potential. Therefore, pastures which offer 3.1 kg dry matter per day and per ewe, such as those studied, may meet the requirements of ewes in rotational grazing.

The importance of these results is the replacement of classical repeatability by time repeatability with the new approach. A great increase in degrees of freedom is achieved.

Mexia's approach has the advantage of using a continuous process instead of an obligatory discrete process conversion, and greatly increases the power of the test, enabling work with smaller experiments.

REFERENCES

Gaccione P. (2004). Nonlinear mixed effects models, a tool for analysing repeated measurements. A brief tutorial using SAS software. Proceedings of the 23rd Annual SAS Users Groups International Conference

Macciotta N. P. P., Vicario D., Cappio-Borlino A. (2005). Detection of Different Shapes of Lactation Curve for Milk Yield in Dairy Cattle by Empirical Mathematical Models. J. Dairy Sci. 88:1178-1191.

Mexia J. T. (1987). Multitreatment Regression Designs. Departamento de Matemática da Faculdade de Ciências e Tecnologia, UNL, Trabalhos de investigação nº 1.

Mexia J. T. (1987). Linear restrictions on the model and Scheffé's multiple comparisons method. Departamento de Matemática da Faculdade de Ciências e Tecnologia, UNL, Trabalhos de investigação nº 3.

Mexia J. T. (1989). Simultaneous confidence intervals: Generalization of the Scheffé's Theorem. Departamento de Matemática da Faculdade de Ciências e Tecnologia, UNL, Trabalhos de investigação nº 2.

Norman H. D., VanRaden P. M., Wright J.R., Clay J. S. (1999). Comparison of Test Interval and Best Prediction Methods for Estimation of Lactation Yield from Monthly, a.m. and p.m., and Trimonthly Testing. Journal of Dairy Science, 82, 438-444.

O'Neill R., Wetherill G.B. (1971). The present state of multiple comparison methods. Journal of the Royal Statistical Society, Series B, 33(2), 218-250.

Ruiz R., Oregui L. M., Herrero M. (2000). Comparison of models for describing the lactation curve of Laxta sheep and an analysis of factors affecting milk yield. Journal of Dairy Science, 83, 2709-2719.

Scheffé H. (1959). The Analysis of variance. John Wiley, New York.

Scott T. A., Yandell B., Zepeda L., Shaver R. D., Smith T. R. (1996). Use of lactation curves for analysis of milk production data. Journal of Dairy Science, 79, 1885-1894.

Steel R. G. D., Torrie J. H.(1981). Principles and procedures of statistics – A biometrical approach. Second Edition. McGraw Hill Book Company.

Wood P. D. P. (1967). Algebraic model of the lactation curve in cattle. Nature, 216, 164-165.

Wood P. D. P. (1969). Factors affecting the shape of the lactation curve in cattle.Animal Production, 11, 307-316.