

## **A Note on Analysis of Lactation Curves for Unequal Lactation Lengths**

Rajendra Singh , Krishan Lal<sup>1</sup> and Thomas B. Farver<sup>2</sup>  
*Indian Veterinary Research Institute, Izatnagar-243122*  
(Received : August, 2001)

### **SUMMARY**

The location and the shape of the lactation curves are affected by many factors but for better comparison of genetic and non-genetic factors, it is inequality in lactation days which must also be accounted for analysis of lactation curves. Because of unequal days in lactation, the model may become heteroscedastic model. Thus a method is proposed to analyze the lactation curves when length of lactation curves is not same for all the animals, by using General Linear Model. It has been shown that this method gives better results than the method generally used given by Wishart [21].

*Key word* : Lactation curves, GLM, Unequal lactation length.

### *1. Introduction*

It has been recognized for many years that the shape of the lactation curve is affected by environmental factors (Cerosvsky [5]), fertility (Gerdmann [9]) and season of calving (Maymone and Molossini [14]). The season of calving has perhaps the most dramatic effect owing to the stimulus to production, which occurs during the spring and early summer (Sanders [16]). For studying the factors affecting the shape of the lactation curves (repeated measures data following a trend), it is required to summarize the data by fitting a suitable response curve.

Further Wishart [21] studied the growth data and fitted linear and quadratic curves separately producing a set of estimates of parameters of growth data. These estimates were further analyzed to determine the effect of factors on growth curves. Box [4] suggested the conventional analysis of variance for the analysis of growth curves. Danford *et al.* [8], Church [6] and Snee [20] suggested different methods of analysis of equi-spaced measurements data with

---

<sup>1</sup> Indian Agricultural Statistics Research Institute, New Delhi.

<sup>2</sup> Department of Population Health and Reproduction, Haring Hall, University of California, Davis, California.

same number of measurements made on each animal. However, these methods are not adequate for the data of unequal length from repeated measurements across the time on the same experimental units. Beale and Little [1], Bhoj ([2], [3], Hartley and Hocking [10], Hocking and Mark [11] and Crepeau *et al.* [7], among others, described analysis of incomplete multivariate data on repeated measurements. Although these multivariate procedures provide sophisticated tests of significance but they are too complicated to understand and interpret by biological scientist. Thus the method given by Wishart [21] remains the method of commonly used by biological scientists, though this method of analysis is for an equal number of repeated measurements.

In this paper a general method using the Generalized Linear Model (GLM) is proposed to analyze the data on repeated measurements of unequal length, which follow certain trend. A comparison is also made of the proposed method and method of Wishart [21].

## 2. Lactation Curve Models

The data used for the present study consisted of weekly milk yield for the  $\frac{1}{2} F \times \frac{1}{2} H$  crossbred cattle collected at LPR Farm, IVRI, Izatnagar.

In literature a number of models for lactation curves are given by Kumar *et al.* [13], Ibrahim [12], Sarkar and Rana [17], and many others. Most of the best-fitted models by the above authors are modifications to the three standard models of lactation curves. These models are Incomplete Gamma (Wood [22]), Inverse Polynomial (Nelder [15]) and Exponential Parabola (Sikka [19]). Thus these three models are considered in the present study. The mathematical form of these models is, respectively, as follows

$$E(y_t) = \alpha t^\beta e^{-\gamma t} \quad (1)$$

$$E(y_t) = t (\beta_0 + \beta_1 t + \beta_2 t^2)^{-1} \quad (2)$$

$$E(y_t) = \alpha_0 \exp (\alpha_1 t + \alpha_2 t^2) \quad (3)$$

where  $E(y_t)$  is expected milk yield at weekly interval and  $\alpha, \beta, \gamma, \beta_0, \beta_1, \beta_2, \alpha_0, \alpha_1$  and  $\alpha_2$  are the response curve parameters.

Each of the above three curves was fitted to the data of individual animal. On the basis of coefficient of determination ( $R^2$ ), the inverse polynomial was found to be the best fit for the present lactation data. Thus the estimates of  $\beta_0, \beta_1$  and  $\beta_2$  were computed for each lactation shape curve and are referred to as three measures, which determine location and shape of the curve.

## 3. Analysis of Lactation Curves

Generally the interest of researcher is to examine the factors affecting the location and shape of the lactation curves through the measures of the curves.

Estimates of  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  have been worked out for  $n$  lactation curves (repeated measures) and are represented in vector form as  $\mathbf{b}_0$ ,  $\mathbf{b}_1$  and  $\mathbf{b}_2$ , each of order  $n \times 1$ .

For the analysis, let us consider the General Linear Model

$$\mathbf{b} = \mathbf{M}\boldsymbol{\gamma} + \mathbf{e}, \quad E(\mathbf{e}) = \mathbf{0}, \quad D(\mathbf{e}) = \sigma^2 \mathbf{V} \quad (4)$$

where  $\mathbf{b}$  ( $\mathbf{b}_0$  or  $\mathbf{b}_1$  or  $\mathbf{b}_2$ ) is the vector of the response variable,  $\mathbf{M}$  is the design matrix of order  $n \times q$ ,  $\boldsymbol{\gamma}$  is the vector of unknown parameters,  $\mathbf{V}$  is symmetric positive definite  $n \times n$  matrix and  $\sigma^2$  is scalar unknown parameter. We can write  $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$  for some non-singular matrix  $\mathbf{L}$ . Applying the transformation  $\mathbf{u} = \mathbf{L}'\mathbf{b}$

$$\mathbf{u} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (5)$$

where  $\mathbf{W} = \mathbf{L}'\mathbf{M}$ , and  $\boldsymbol{\varepsilon} = \mathbf{L}'\mathbf{e}$ , so  $D(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$

The vector of estimates of  $\boldsymbol{\gamma}$  from this model is

$$\hat{\boldsymbol{\gamma}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{u} = (\mathbf{M}'\mathbf{V}^{-1}\mathbf{M})^{-1}\mathbf{M}'\mathbf{V}^{-1}\mathbf{b} \quad (6)$$

where matrix  $\mathbf{U}^{-}$  is the generalized inverse of matrix  $\mathbf{U}$ .

From Searle [18], the sum of squares due to error is

$$\text{SSE} = \mathbf{u}'\mathbf{u} - \hat{\boldsymbol{\gamma}}'\mathbf{W}'\mathbf{u} = \mathbf{b}'\mathbf{V}^{-1}\mathbf{b} - \hat{\boldsymbol{\gamma}}'\mathbf{M}'\mathbf{V}^{-1}\mathbf{b} \quad (7)$$

and its expected value is

$$\begin{aligned} E(\text{SSE}) &= E(\mathbf{u}'\mathbf{u} - \hat{\boldsymbol{\gamma}}'\mathbf{W}'\mathbf{u}) = \text{Rank}(\mathbf{I} - \text{Rank}(\mathbf{W})) \\ &= \text{trace}((\mathbf{I} - \text{Rank}(\mathbf{W}))) = n - \text{Rank}(\mathbf{W}) = (n - p) \sigma^2 \end{aligned} \quad (8)$$

where  $n$  is the number of observations and  $p$  is the number of independent columns of  $\mathbf{M}$  and  $p < q$ .

The sum of square due to model (SSR) is

$$\mathbf{u}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{u} = \mathbf{b}'\mathbf{V}^{-1}\mathbf{M}(\mathbf{M}'\mathbf{V}^{-1}\mathbf{M})^{-1}\mathbf{M}'\mathbf{V}^{-1}\mathbf{b} \quad (9)$$

with degree of freedom equal to  $E(\mathbf{M}) = p$ , and also SSE and SSR are independent of each other.

#### 4. One-way Analysis of Variance

We simplify the above linear model (4) for use in one-way analysis of variance. For one-way analysis of variance, the model (4) can be written as

$$\mathbf{b} = \mu\mathbf{1} + \boldsymbol{\Delta}'\boldsymbol{\tau} + \mathbf{e} \quad (10)$$

where  $\mu$  is overall population mean,  $\mathbf{1}$  is vector of 1's,  $\boldsymbol{\Delta}'$  is the design matrix, and  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_s)'$  is the vector of  $s$  levels (groups) of the factor with  $E(\mathbf{e}) = \mathbf{0}$ ,  $E(\mathbf{e}\mathbf{e}') = \sigma^2\mathbf{V}$ . This model alternatively can be written as

$$b_{ij} = \mu + \tau_i + e_{ij} \tag{11}$$

where  $E(e_{ij}) = 0, \text{Var}(e_{ij}) = \sigma^2 c_{ij}$

$$\text{Cov}(e_{ij}, e_{i'j'}) = 0 \text{ for } i \neq i' \text{ or } j \neq j'$$

$$\mathbf{b} = (b_{11}, \dots, b_{1n_1}, \dots, b_{s1}, \dots, b_{sn_s})', \mathbf{V} = \text{diag} (c_{11}, \dots, c_{1n_1}, \dots, c_{s1}, \dots, c_{sn_s})$$

$\sum n_j = n, \tau_i$  is the effect of  $i^{\text{th}}$  treatment,  $b_{ij}$  is the value of  $b$  for the  $j^{\text{th}}$  experimental unit in the  $i^{\text{th}}$  group,  $c_{ij}$  is the coefficient associated with the  $\sigma^2$  in the variance of  $b_{0ij}$  for the  $j^{\text{th}}$  experimental unit in the  $i^{\text{th}}$  group ;  $i=1, \dots, s$  ;  $j = 1, \dots, n_s$ .

Let  $\mathbf{V}^{-1} = \text{diag} (d_{11}, \dots, d_{1n_1}, \dots, d_{s1}, \dots, d_{sn_s})$ , then we have the following result

*Theorem 1.* Under the heteroscedastic one-way analysis of variance, sum of square due to regression (SSR) =  $\Sigma [(\sum d_{ij} b_{ij})^2 / (\sum d_{ij})]$  and sum of square due to error SSE =  $\Sigma \Sigma d_{ij} b_{ij}^2 - \text{SSR}$ .

*Proof.* From the models (4), (9) and (11), we have

$$\mathbf{M}'\mathbf{V}^{-1}\mathbf{M} = \begin{bmatrix} \sum_i \sum_j d_{ij} & \sum_j d_{1j} & \sum_j d_{2j} & \dots & \sum_j d_{sj} \\ \sum_j d_{1j} & \sum_j d_{1j} & 0 & \dots & 0 \\ \sum_j d_{2j} & 0 & \sum_j d_{2j} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_j d_{sj} & 0 & \dots & 0 & \sum_j d_{sj} \end{bmatrix}$$

On simplifying

$$\text{SSR} = \mathbf{b}'\mathbf{V}^{-1}\mathbf{M}(\mathbf{M}'\mathbf{V}^{-1}\mathbf{M})^{-1}\mathbf{M}'\mathbf{V}^{-1}\mathbf{b} = \Sigma [(\sum d_{ij} b_{ij})^2 / (\sum d_{ij})]$$
 and

$$\text{SSE} = \Sigma \Sigma d_{ij} b_{ij}^2 - \text{SSR}$$

After making an adjustment for mean, the sum of squares due to treatment will be  $\text{SST} = \text{SSR} - \text{CF}$  with  $(s - 1)$  degrees of freedom, where  $\text{CF} = \{(\sum \sum d_{ij} b_{ij})^2 / (\sum d_{ij})\}$ . SST and SSE are independent of each other. The mean square due to treatment (MST) is  $\text{SST}/(s-1)$  and the mean square due to error (MSE) is  $\text{SSE}/(n-s)$ . Under the null hypothesis ( $H_0$ ) that all the treatments have equal effect,  $\text{MST}/\text{MSE}$  is distributed as  $F_{s-1, n-s}$ .

This expression is true for all the estimates  $\mathbf{b}_0, \mathbf{b}_1$  and  $\mathbf{b}_2$ .

### 5. Applications

The difference between our method and the method of Wishart [21] is that Wishart used analysis of variance assuming the equality of variances whereas we have used weighted least square under GLM, in recognition that the variances for each estimated value from each lactation are not equal. It is because the lactation days are not same for each cattle.

We have taken the subset of data of lactation curves of cattle and studied a single factor with multiple levels affecting the location and shape of the lactation curves.

The data pertains to the weekly milk yield for the complete lactation of cows of the same sire, same lactation order for the two consecutive years of calving (because of small number of observations in one year). There were 25 such lactating cows. No significant difference was observed between the years. Thus, the season of calving is considered as a single factor affecting the locations ( $b_0$ ,  $b_1$  and  $b_2$ ) of the shape of lactation curves. Season of calving is classified as Season 1 : Spring 16<sup>th</sup> February to 31<sup>st</sup> March; Season 2 : Summer 1<sup>st</sup> April to 15<sup>th</sup> July; Season 3 : Rainy 16<sup>th</sup> July to 30<sup>th</sup> September; Season 4 : Autumn 1<sup>st</sup> October to 30<sup>th</sup> November; Season 5 : Winter 1<sup>st</sup> December to 15<sup>th</sup> February.

These data were analyzed by the above proposed procedure and by the method given by Wishart and the results are presented in Table 1 for  $b_0$ ,  $b_1$  and  $b_2$ .

**Table 1.** ANOVA for  $b_0$ ,  $b_1$  and  $b_2$  by the two methods

S.V	d.f.	Proposed		Wishart	
		MS	F-value	MS	F-value
<b><math>b_1</math></b>					
Season	4	0.0003767	2.08	0.0004689	1.45
Error	20	0.0001810		0.0003238	
<b><math>b_2</math></b>					
Season	4	0.0011834	3.03*	0.00024355	1.47
Error	20	0.0003908		0.0001653	
<b><math>b_3</math></b>					
Season	4	0.0101678	3.94*	0.000003578	1.54
Error	20	0.0025797		0.000002319	

This table reveals that the F value is higher for all the measures, namely  $b_0$ ,  $b_1$  and  $b_2$  using the proposed method and the effect of season of calving was found to be significant for the locations  $b_1$  and  $b_2$  with the proposed method which could not be established with the method of Wishart. Further,  $R^2$  values were worked out for  $b_0$ ,  $b_1$  and  $b_2$  separately by the proposed method and the method given by Wishart. The respective values for the proposed method were

29.39, 37.72 and 44.08% and for the method given by of Wishart were 22.46, 22.77 and 23.58%. It reveals that  $R^2$  values were higher by the proposed method than that the method given by Wishart. It can, therefore, be concluded that the proposed method gives more precise results than that the method given by Wishart. This method in general can be applied for analyzing the data of incomplete measurements (unequal length) following certain trend.

The difference in the results obtained by the Wishart method and by our method will depend on the amount of variation in the length of lactation curve or the number of individuals at risk i.e. the amount of heterogeneity in the model. If there is no difference between the lactation lengths or there is no variability in  $n_{ij}$  (number of individuals at risk in the  $ij^{\text{th}}$  cell), there will be no difference in the results obtained by the Wishart method and by our method.

#### REFERENCES

- [1] Beale, E.M.L, and Little, R.J.A. (1975). Missing values of the multivariate analysis. *J. Roy. Statist. Soc.*, B37, 129-145.
- [2] Bhoj, D.S. (1973a). Percentage points of the statistics for testing hypotheses on mean vectors of multivariate normal distributions with missing observations. *J. Statist. Comp. and Simul.*, 2, 211-224.
- [3] Bhoj, D.S. (1973b). On the distribution of statistic used for testing multivariate mean vector with incomplete data. *J. Statist. Comp. and Simul.*, 2, 309-346.
- [4] Box, G.E.P. (1950). Problem in the analysis of growth and wear curves. *Biometrics*, 6, 342-369.
- [5] Cerosvsky, H. (1957). The significance of the shape of lactation curve in cows. *Tierzucht*, 11, 185-188.
- [6] Church, A.J. (1966). Analysis of data when the response is a curve. *Technometrics*, 8, 229-246.
- [7] Crepeau, H., Koziol, J., Reid, N. and Yuh, Y.S. (1985). Analysis of incomplete multivariate data from repeated measurement experiments. *Biometrics*, 41, 505-514.
- [8] Danford, M.B., Hages, H.M. and Mcnef, R.C. (1960). On the analysis of repeated measurement experiments. *Biometrics*, 16, 547-665.
- [9] Gerdmann, H. (1964). Investigations on the relationship between milk yield and fertility in cows with particular consideration of shape of lactation curve. *Inaug. Diss. Doct. Med. Vet.*, Hanover.
- [10] Hartley, H.D. and Hocking, R.R. (1971). The analysis of incomplete data (with discussion ). *Biometrics*, 27, 783-817.
- [11] Hocking, R.R. and Mark, D.L. (1979). Estimation with incomplete data; an improved computational method and the analysis of nested data. *Comm. Stat.-Theory Methods*, A8, 1155-1181.

- [12] Ibrahim, M.A.M. (1985). The use of gamma type function in describing the lactation curves of Egyptian buffaloes. *Egyptian J. Anim. Prod.*, **32**(2), 113-123.
- [13] Kumar, D., Sharma, O.P., and Singh, V.P. (1992). A stochastic model to describe the milk yield in buffaloes. *Indian J. Anim. Sci.*, **52**(6), 591-682.
- [14] Maymone, B. and Molossini, F. (1959). The rising phase of the lactation curve in dairy cows. *Ann. Sper. Agr.*, **13**, 1005-1033.
- [15] Nelder, J.A. (1966). Inverse polynomial, a useful group of multi factor response functions. *Biometrics*, **22**, 128-141.
- [16] Sanders, H.G. (1930). The analysis of the lactation curve into maximum yield and persistency. *J. Agric. Sci., Cambridge*, **20**, 145-185.
- [17] Sarkar, S. and Rana, P.S. (2001). Mathematical formulation of lactation curves. *Indian J. Anim. Sci.*, **71**(9), 860-864.
- [18] Searle, S.R. (1971). *Linear Models*. John Wiley, New York.
- [19] Sikka, L.C. (1950). A study of lactation as affected by heredity and environment. *J. Dairy Res.*, **17**, 231-252.
- [20] Snee, R.O. (1972). On the analysis of response curve data. *Technometrics*, **14**, 47-62.
- [21] Wishart, J. (1938). Growth rate determinations in nutritional studies with bacon pig and their analysis. *Biometrics*, **30**, 16-28.
- [22] Wood, P.D.P. (1967). Algebraic method of the lactation curve in cattle. *Nature*, London, **216**, 164-165.