COMPARISON OF TWO METHODS (EPA AND MULTIPLE LINEAR REGRESSION) OF LIFE TIME PRODUCTION ON THE BASIS OF EARLIER LACTATION PERFORMANCE RECORDS IN THARPARKAR CATTLE

By

R.K. Patel and M. Gurnani

N.D.R.I., Karnal (Haryana)

(Received : December, 1981)

Summary

Since the assumptions underlying estimation of Expected producing ability (EPA), of equality of variances of various lactation records and the equality of correlations between various pairs of lactation records on Tharparkar cattle at National Dairy Research Institute, Karnal, were not holding true so it was considered desirable to compare the accuracy of prediction of 'true' producing ability by EPA formula with that on the basis of multiple regression on earlier lactation records. The measure of 'true' producing ability was considered to be life time production (LTP) upto 8 and 10 years separately. It was found that the accuracy of prediction of life time prediction upto 10 years on EPA was slightly higher by multiple regression method whereas the accuracy of prediction of life time production upto 10 years by multiple regression was similar or slightly lower than the regression on EPA. Thus prediction of 'true' producing ability by EPA would be almost as accurate as multiple regression approacch.

1. Introduction

It is of primary interest to genetically improve the milk producing ability of cows by selection. The generally recommended and easily applied criterion of selection of cows on the basis of their own information is by the use of formula for "Expected Producing

Ability" (EPA) or "Most Probable Producing Ability" (MPPA) given by Lush [3]. This formula for EPA is :

$$\text{EPA}=u+\frac{nr}{1+(n-1)r}(\bar{x}+u)$$

Where :

u is the population mean, r is repeatability of lactation milk records and $\bar{x}$ is the average of n lactation records of the cow.

This formula can be derived by taking the regression of a single or average of several furure lactation records on average of earlier lactation records. In derivation of this formula, the implied assumptions are :

(*i*) The variances of various repeated lactation records are same, and

(*ii*) The correlations between the various pairs of lactation records are same and equal to the repeatability of lactation production.

These assumptions were found to be not holding true for the data on Tharpakrar cattle being maintained at this Institute (Annexure Tables A & B). It was, therefore, considered desirable to judge the accuracy of prediction of 'true' producing ability of cows by EPA formula and multiple regression technique on various sequential combinations of availability of earlier lactation milk yield records. The measure of 'true' producing ability was considered to be lifetime milk production upto 8 and 10 years.

2. Material and Methods

Data for the present investigation were taken from the available records of Tharparkar cattle maintained at the National Dairy Research Institute, Karnal (India). The data on lactation milk production upto four lactations of 700 Tharparkar cows which had completed at least two normal lacrations during the period from 1927 to 1980 were considered. The number of cows having milk production records in different orders of lactation were 700 for first, 547 for second, 425 for third and 347 for fourth. The production records upto 8 and 10 years of age were studied. The lactation milk records which had duration of less than 100 days were considered to be abnormal due to improper pre-calving management or due to disease incidence like mastitis and so ignored. The milk yield for lactation

duration between 100 days to 305 days was taken as standard 305-day lactation yield without any correction which is recommended practice in developed countries as well as in organised herds in India. The variance of lactation milk production, separately for each lactation, was calculated. Sanders [4] lactation correction factors as obtained by Gurnani (unpublished) were applied. For testing the homogeneity of variances of first four laetation milk yields, Bartlett's Chi-square test was applied (Amble, [1]).

Simple linear correlation coefficients beween various possible combinations of lactation records were estimated. The Chi-square test was applied for testing the homogeneity of correlation coefficients (Snedecor and Cochran, [5]. Repeatability of lactation milk yield was estimated by the least-square analysis of variance of repeated lactation milk records among cows (Harvey, [2]) as a one-way classification using the model :

$$Y_{ij}=u+c_i+e_{ij},$$

Where Y_{ij} is the milk yield in the jth lactation of ith cow ;

c_i is the effect of ith cow, and

e_{ij} is the random error associated with the jth lactation record of the ith cow and is assumed to be $N(o, \sigma_e^2)$

The simple linear regression of life time production (x) on Expected Producing Ability (EPA) (x) based on first lactation record, average of first two lactation records, average of first three lactation records separately was fitted. The accuracy of prediction was judged by coefficient of determination (R^2) of fitting the corresponding regression model. The data were not corrected for any fixed effects of period and season, under the assumption that such fixed effects will not have any influence on the association between lifetime production and milk yields in various lactations. Such is also the practice in estimation of EPA in organised herds in India.

The prediction of lifetime production upto 8 years (LTP-8) and upto 10 years (LTP-10) was also made on the basis of multiple linear regression on various sequential combinations of first four lactation milk records and corresponding accuracies (R^2) of fitting of each of the models were estimated.

3. Results and Discussion

The accuracy of prediction of lifetime production upto 8 years (LTP-8) and 10 years (LTP-10) on the basis of earlier lactation milk records.

The simple linear regression of life time production (LTP) on Expected producing ability (EPA) based on sequential combinations of first four lacation milk records are given in Table 1 for LTP-8 and in Table 2 for LTP-10. The accuracy $(R)^2$ of prediction of LTP-8, on the basis of EPA, increased consistently when second record was added to first, third record was added to first two and fourth record was added to first three records. The maximum accuracy $(R)^2$ of prediction of LTP-8 on basis of first four lactation records was 43.2 percent (Table 1).

TABLE 1

Prediction of LTP-8 on the basis of EPA for the Tharparkar breed.

Traits		*No. of cows*	*bYx*	*S.E. (bYx)*	*rLTP. EPA Correlation*	*Coefficient of determination R^2 (%)*
Y(kg)	*X(kg) EPA based on*					
8 yrs. lactation milk production	First lactation milk record	228	6.014	0.312	0.483	23.3
8 yrs. lactation milk production	Av. of first two lactation milk records	228	4.858	0.324	0.549	30.1
8 yrs. lactation milk production	Av. of first three lactation milk records	228	3.895	0.288	0.550	30.3
8 yrs. lactation milk production	Av. of first four lactation milk records	228	4.999	0.303	0.657	43.2

bYx : Regression of 8 ycars lactation milk production (*Y*) on *EPA* (*X*).
Repeatability of lactation milk production=0.495

The accuracy of prediction of LTP-10, on the basis of EPA, also increased consistently, when second record was added to first, when third record was added to first two records and fourth record was added to first three records (Table 2). The maximum accuracy of prediction of LTP-10, on the basis of four records, was 45.5 per cent.

TABLE 2

Prediction of Life Time Milk Production (10 years) on the basis of EPA for Tharparkar Breed

Traits		*No. of cows*	*bYx*	*S.E. (bYx)*	*$r_{LTP.EPA}$ Correlation*	*Coefficient of determination R^2(%)*
Y(kg)	*X(kg) EPA based on*					
10 yrs. lactation milk production	First lactation milk record	159	7.192	0.458	0.497	24.8
10 yrs. lactation milk production	Av. of first two lactation milk records	159	5.976	0.424	0.597	35.6
10 yrs. lactation milk production	Av. of first three lactation milk records	159	5.615	0.423	0.618	38.2
10 yrs. lactation milk production	Av. of first four lactation milk records	159	6.217	0.433	0.674	45.4

b_{YX} : Regression of 10 years lactation milk production (Y) on *EPA* (X)
Repeatability of lactation milk production=0.495

Accuracy of prediction of lifetime production by multiple regression on earlier lactation records.

In view of significant differences in variances of different lactation records and significant differences in correlations between pairs of lactation records, the above method of using EPA as a criterion for predicting future yield would not be theoretically appropriate in this data. Therefore, the multiple regressions of life time

TABLE 3

Prediction of LTP-8 on the Basis of Multiple Regression on Sequentiul Combination of First Four Lactation Records in Tharparkar Cattle

Multiple linear regressihn		*No. of cows*	b_0	b_1	@*S.E.* (b_1)	b_2	*S.E.* (b_2)	b_3	*S.E.* (b_3)	b	*S.E.* (b_4)	*Coefficient of determination* R^2 (%)
Dependent trait (kg)	*Independent trait/s (kg)*											
8 years lactation	First lactation production	228	3954.734	2.977	0.312	—	—	—	—	—	—	23.3
8 years lactation	First & secoud lactation production	228	2581.166	2.280	0.352	1.084	0.322	—	—	—	—	32.1
8 years lactation	First, second, and third lactation production	228	1070.461	1.647	0.373	1.647	0.311	0.711	0.278	—	—	42.5
8 years lactation	First, second, third and fourth lactation production	228	904.652	0.652	0.282	0.658	0.322	1.373	0.303	1.342	4.289	44.8

b_0 : Constant in multiple linear regression equation

b_1, b_2, b_3, b_4 : Partial regression coefficients of LTP-8 on first, sedond, third and forrth lactation reeord.

@*S.E.* : Standard Error.

TABLE 4

Prediction of Lifetime milk production (LTP-10) on the basis of multiple linear regression on sequential combination of first four lactation records in Tharparkar Breed

Multiple linear regression		*No. of cows*	b_0	b_1	*S.E.* (b_1)	b_2	*S.E.* (b_2)	b_3	*S.E.* (b_3)	(b_4)	*S.E.* (b_4)	*Coefficient of determination* R^2 (%)
Dependent trait (kg)	*Independent trait/s (kg)*											
10 years lactation production	First lactation record	159	6503.908	3.560	0.458	—	—	—	—	—	—	24.8
10 years lactation production	First and second lactation records	159	5999.439	2.725	0.605	0.951	0.498	—	—	—	—	25.0
10 years lactation production	First, second and third lactation records	159	4099.657	1.827	0.432	0.558	0.432	2.047	0.413	—	—	38.2
10 years lactation production	First, second, third and fourth lactation records	159	3706.232	1.523	0.548	0.313	0449	1.118	0.434	2.165	0.399	41.8

production (as a measure of 'true' producing ability) on various sequential combinatian of first four records were fitted. The coefficients/of various regression equations along with the corresponding R^2 values are given in Table 3 and 4.

The accuracy of prediction $(R)^2$ by linear regression of life time milk production upto 8 years and 10 years of age on first lactation milk production was same, as expected, as that of regression of LTP-8 and LTP-10 on EPA based on first lactation. When second, third and fourth records were added to earlier records for prediction of LTP-8, the R^2 values increased (Table 3). These R^2 values were slightly higher than the corresponding values obtained for regression of LTP-8 on EPA based on first two, first three and first four lactation records (Table 1). Similarly when second, third and fourth records were added to earlier records for prdicetion of LTP-10, the R^2 values increased (Table 4). These R^2 values were generally lower or similar as compared to regression of LTP-10 on EPA based on corresponding number of lactation records. It was thus found that the accuracy of prediction of life time production upto 8 years on EPA was slightly higher than by multiple regression method whereas the accuracy of prediction of life time production upto 10 years by multiple regression was similar or slightly lower than the regression on EPA. It may, therefore, be concluded that for a herd data, wherein the assumptions underlying estimation of producing ability for milk production by EPA, may not hold true, the accuracy of prediction may be as good as multiple regression approach. It is a fortunate result as the computation by EPA formula is much simpler as the development of prediction formulae by multiple regression would require large data size which may not be available under many Indian farm conditions.

REFERENCES

[1] Amble, V.N. (1975) : Statistical Methods in Animal Sciences. *Indian Society of Agricultural Statistics*, New Delhi.

[2] Harvey, W.R. (1966) : Least-squares analysis of data with unequal sub-class numbers. *USDA, ARS*. U.S.A.

[3] Lush, J.L. (1945) : Animal Breeding Plans. Collegiate Press, Inc. *Ames, Iowa*, 2nd Ed.

[4] Sanders, H.G. (1928) : Variation in milk yield caused by seasons of the year, service period, age and dry period and their elimination. *J. Agric. Sci.*, **18** : 46-7.

[5] Snedecor and Cochran (1967) ; Statistical Methods 6th Ed. IndianE d. Oxford and IBH Publ. Co., Calcutta.

ANNEXURE

TABLE A

Test of homogeneity of variances of lactation milk yield records in different lactations in tharparkar cattle

Lactation No.	*I*	*II*	*III*	*IV*	*Bartlett's Chi-square test for homogeneity of variance OBSERVED X^2*
	σ^2x_1	σ^2x_2	σ^2x_3	σ^2x_4	
Variance of lactation milk record (Kg^2) (Uncorrected)	381036	482141	522946	534197	16.84**
Variance of lactation milk records (Kg^2) (Corrected for lactation difference by Sanders' Method)	381036	387935	437824	605900	29.26**

** Significant ($P<.01$)

TABLE B

Phenotypic correlations between various lactation milk records (same for records corrected for lactation differences by sanders ratio factors as well as uncorrected records)

Lactation No.	*Correlation between pairs of lactation milk records*			*x^2 value for test of homogeneity of correlation coefficient*
	II	*III*	*IV*	
I	0.525	0.473	0.321	
II		0.444	0.470	40.005**
III			0.564	

No. of cows for each correlation coefficient=228

** Significant ($P<.01$)