Archivos de zootecnia, vol. 38, núm. 142, 1989, p.249.

ESTIMATE OF (CO)VARIANCE COMPONENTES FOR MILK AND FAT YIELD IN SPANISH HOLSTEIN POPULATION USING REML.

ESTIMACIÓN DE COMPONENTES DE (CO)VARIANZA PARA LA PRODUCCIÓN DE LECHE Y GRASA EN LA POBLACIÓN HOLSTEIN ESPAÑOL UTILIZANDO EL MÉTODO REML.

Cañón, J.*, P.J. Berger**, J.P. Gutiérrez* y A. Muñoz***.

* Departamento de Producción Animal. Facultad de Veterinaria. 28040. Madrid (España).
** Department of Animal Science. ISU. 239 Kildee Hall. Ames, IA 50011 (España).
*** Departamento de Producción Animal. Facultad de Veterinaria. 30071. Murcia (España).

Key words: Variance estimates. Genetic parameters. REML. Milk yield. Fat yield. Holstein.
Palabras clave: Estima de varianzas. Parámetros genéticos.REML. Producción de leche. Producción de grasa. Holstein.

Resumen

Se han estimado las componentes de varianza, así como algunos parámetros genéticos combinaciones de dichas componentes, para la producción de leche y grasa, considerando todas las lactaciones y sólo la primera, mediante el ajuste de dos modelos diferentes a un mismo archivo de datos de vacuno de leche, utilizando para la estimación de componentes de varianza el método REML. Los resultados muestran valores para la heredabilidad de la producciòn de leche moderadamente bajos (0,17-0,21) y valores inferiores para la producción de grasa (0,13-0,15). Las correlaciones genéticas y fenotípicas fuerón elevadas, 0,94 y 0,86 respectivamente. También se apreció heterogeneidad entre niveles de rebaño-año para las componentes de varianza debida a toros, al ambiente permanente, y residual.

Recibido: 28-12-88. Aceptado: 26-9-89.

Summary

Two different models were applied to a dairy cattle data set to estimate variance componentes by REML, and combinations of those variance components were used to estimate genetic parameters for milk and fat yield when all and only first lactation records were included. Results show moderate values for heritability of milk yield (0.17-0.21) and lower values for fat yield (0.13-0.15). Genetic and phenotypic correlations were high, 0.94 and 0.86 respectively. Heterogeneous sire, permanent environment and residual variance by herd level were found.

Introduction

Knowledge of genetic parameters, i.e. heritabilities, genetic and phenotypic correlations, for traits of economic importance are needed to evaluate breeding strategies.

Until now first lactation milk and fat yield have been the most important traits on which selection was based for dairy cattle in Spain. Although, recently more importance has been given to repetitive performance of the animals, and information on all lactations is starting to be included in national dairy evaluation programs (Wiggans, Misztal and Van Vleck, 1988 a; b; Cañón and Chesnais, 1988).

Many different methods and strategies are available to estimate variance components. Among the methods available, REML (Patterson and Thompson, 1971) is probably the best choice in animal breeding applications due to the desirable properties defined by Harville (1977). The properties of REML are: estimates are always within the parameter space; estimates are unique and are not dependent on the prior values used; estimates are not affected by selection or assortative mating if the information on which selection was based is included in the analysis. Although estimates are biased, standard errors of estimation are generally smaller than unbiased procedures. This method may be suitable even when the distribution of the variable to be analyzed is unknown.

Use of EM (Expectation-Maximization) algorithm (Dempster, Laird and Rubin, 1977) is relatively easy to derive and program, and yields non-negative estimates. It is however slow to converge. Other alternatives can be used, i.e., a reparametrization proposed by Thompson and Meyer (1986) and Meyer (1987) or an algorithm presented by Harville (1977) tested by Van Raden (1986).

Linear models used as approximations to explain events are most simple when random variables in the model can be assumed to have homogeneous variances. Milk and fat yield show important scale effects. Residual and sire variances tend to increase as the phenotypic mean of the herd increase (Hill et al., 1983; Van Vleck et al., 1985) and in most cases this is also true for heritability estimates (Powell and Norman, 1984; Van Vleck et al., 1985).

Estimate additive genetic, permanent environmental, and residual variances for milk and fat yield using all lactations was the first objetive of this study. The second objective was to estimate the same variance components for the same two traits at two levels of production.

Material and methods

Data

The original data set consisted of 99,995 305 day lactations of Spanish Holstein cows sired by A.I. (Cañón and Muñoz, 1988. Each cow was required to have a first lactation. Records were coded by herd-year. After elimination of records when sire was not identified or miss-identified, records coming from herd-year's with only one sire represented and records coming from sires with less than two daughters across two herd-year, the distribution of the data is presented in table I.

Data were adjusted to mature cow calving in January by multiplicative factors (Cañón, 1988).

To look at the heterogeneous herd variances records were grouped into two mutually and exclusively independent categories (high and low production level) using herd milk yield mean as criterion.

To study heterogeneus herd variances all lactations were included and model (1) was applied separately to each herd production level set of data.

Model

The model used to estimate variance components when all lactations are included was:

$$Y_{ijklm} = \mu + hy_i + g_j + S_{k(j)} + C_{l(jk)} + e_{ijklm} \qquad (1)$$

where:

Y_{ijklm} represents the mth lactation recorded in the ith herd-year, coming from the 1^{th} daughter of k^{th} sire who is included in the j^{th} genetic group.

Assumptions of the model are:

$$E(Y_{ijklm}) = \mu + hy_i + g_j$$

$$Var\ (S) = A\ \sigma^2_s$$

$$Var\ (C) = I\ \sigma^2_c$$

$$Var\ (E) = I\ \sigma^2_e$$

and S, C, E, A and I are respectively: solution vectors in (1) for S and C, vector of error, genetic relationship matrix, and identity matrix. σ^2_s, σ^2_c and σ^2_e are scalars.

Genetic groups were defined by birth year of the sire. Sire and maternal-grandsire was the pedigree information taked into account in A.

This model could be interpreted as a good aproximation to an animal model after ingnoring relationships among females other than through sires, and non-radom mating (Boldman, personal communication).

Estimation procedure was REML via EM, "k" reparametrization (Meyer, 1987) by Meyer's computer programs.

Heritability and repeatability were estimated by:

$$h^2 = \sigma^2_g\ /\ (\sigma^2_g + \sigma^2_{pe} + \sigma^2_e)$$

$$r = (\sigma^2_g + \sigma^2_{pe})\ /\ (\sigma^2_g + \sigma^2_{pe} + \sigma^2_e)$$

where:

$$\sigma^2_g = 4\sigma^2_s$$

$$\sigma^2_{pe} = \sigma^2_c - 3\sigma^2_s$$

Estimation of variance components for first lactation only was obtained by the following multitrait model:

$$Y_{ijklm} = \mu + hy_{im} + g_{jm} + S_{k(j)m} + e_{ijklm} \qquad (2)$$

where:

Y_{ijklm} is the ijkl observation for the m trait. (m= 1, 2; 1= milk yield, 2= fat yield.) recorded in the i^{th} herd-year for the 1^{th} daughter of sire k in gnetic group j.

hy_{im} is the i^{th} herd-year for the m^{th} trait considered as a fixed effect.

g_{jm} is the j^{th} genetic group fixed effect, j=1 to 5

$S_{k(j)m}$ is the random effect for the m^{th} trait of the sire k in the genetic group j.

e_{ijklm} is the random residual effect for the m trait associated with the $ijkl^{th}$ cow.

In matrix notation model (2) for each of the c traits is:

$$Y_c = Xb_c + Zu_c + e_c$$

where:

c is the trait (c=1 or 2)

Y_c is the vector of observations on trait c,

X is a design matrix connecting b with Y and is the same for both traits,

b_c is a vector of fixed effects for trait c,

Z is a design matrix and is the same for both traits,

u_c is a vector of sire random effects for trait c.

e_c is a vector of random residual effects.

$E(Y_c) = Xb_c$

$Var(u_c) = A*G$, G is the sire component of (co)variance between traits

$Var(e_c) = I*R'$ R is the residual component of (co)variance between traits

* is the symbol for Kroenecker product of matrices

Genetic group definition and relationship information was similar to that used in model (1).

Exact REML solutions were obtained using Van Raden's computing strategy (Van Raden, 1986) which obtain σ^2_i using the expression:

$$\hat{u}'_i A^{-1} \hat{u}i/(n_i - tr(A^{-1}C_{ii})k_i)$$

where:

$\hat{u}'_i A^{-1} \hat{u}i$ is the sum of squares

C_{ii} is a portion of $(Z'MZ + A^{-1}k_i)^{-1}$

A^{-1} is the inver of the numerator relationship matrix

$M = I - X(X'X)^{-1}X'$

n_i is the number of levels of effect i, and k_i is σ^2_e/σ^2_i

(Harville, 1977)

Results and discussion

Variance component estimates and functions of these variances (heritabilities and repeatabilities) when all lactations and only the first one was considered are presented in table II.

Heritability estimates ranged from 0.13, for fat yield when all lactations were considered, to 0.20 for milk yield, when only first lactation was considered. Although these values are lower than those frequently reported in the literature (Neummann, 1969; Cue et al. 1987; Swolve and Van Vleck, 1987) they are similar to other results when herd level was taken into account (de Veer and Van Vleck 1987; Boldman 1988). Gómez and Alenda (1985) reported lower heritability estimates values for first lactation milk yield in the Spanish Holstein population but their model for analysis did not include genetic relationships among sires and they included herd average as a covariable in the model which can account for a partial loss of variance between sires.

Heritability estimates for fat yield were lower than expected and could be caused by a low quality of the measurements for this trait. If milk yield is easy to measure either in a not very well organized national milk recording program, fat measurements are more suitable to a bad management of the samples increasing residual variance and decreasing heritability.

Repeatability estimates were 0.40 and 0.36 for milk and fat yield, respectively. These estimates were lower than those found in the literature (Maijala and Hanna, 1974; Boldman, 1988)

Phenotypic correlation between milk and fat yield was 0.89. This is very similar to most values reported elsewhere (Maijala and Hanna, 1974; Cue et al., 1987). Although a larger number of sires may have been more desirable to estimate the genetic correlation accurately, the estimated genetic correlation between milk and fat yield, 0.94, was in the upper bound of those reviewed in the literature, by O'Connor (1969), higher than estimates reported by the De Jager and Kennedy (1987), and moderately close to the mean weighted genetic correlation of 0.813 reported by Maijala and Hanna (1974).

Herds were divided according to high and low levels of production to study heterogeneous variances. Number of herds, sires and cows whitin each production level is in Table III. Estimated genetic parameters are in Table IV.

The scale effect for milk and fat yield is well known, i.e., phenotypic standard deviation increases with the mean. Phenotypic causal components of variance: genetic, residual and permanent enviromental variance are also related to the mean (Hill et al., 1983; Mirande and Van Vleck, 1985) and general results show that all of these variances increase as production increases, and this relationship is almost linear (de Veer and VanVleck, 1987; Boldman, 1988). In the last two articles referenced heritability estimates for milk yield in low level herds were between 0.12 and 0.19 corresponding to a milk yield average of 5,503 and 6,595 Kg. respectively. These averages were similar to the averages for the low and high herds in this study (Table IV).

As expected, all variance components were higher at a high level than a low level herd average (de Veer, 1987; de Veer and Van Vleck, 1987; Boldman 1988). While heritability estimates also increased with production level of herd in these former studies, heritability was higher at a low production level. In this study was also found to be the case by Robertson et al. (1960) and Mitchell et al. (1961) with three herd production levels.

These results have importan implications for the prediction of genetic merit, e.g. animals (sires and cows) with most of their information in large variance herds tend to be overevaluated. It is of interest to know if it is possible to stabilize genetic variance dividing by the genetic variance within production level (Y_i / σ^2_g) or whether one can use a log transformation (de Veer and Van Vleck, 1987). Certainly some attention must be paid to account for heterogeneous permanent enviromental and residual variances as proposed by Boldman (1988).

Genetic parameters presented in Table II have been used to predict sire and cow genetic merit by an animal model with repeated records in the Spanish Holstein population (Cañón and Chesnais, 1988) and to estimate genetic trends (Cañón and Muñoz, 1988)

Acknowledgements

Computer facilities were provided by the Animal Science Department of Iowa State University.

References

Cañón, J., 1988. Influencia de factores ambientales sistemáticos sobre la producción de leche en la población frisona española. II Cálculo de factores de ajuste aditivos y multiplicativos para la edad y mes de parto. I.T.E.A. (Submitted).

Cañón, J. and Chesnais, J., 1988. Indirect approach to simultaneous sire and cow evaluation using an animal model with repeated records and groups for unknown parents. Z. Tierzüchtg. Züchtgbiol. (In press).

Cañón, J. and Muñoz, A., 1988. Genetic trends for milk production in the Spanish Holstein population. Z. Tierzüchtg. Züchtgsbiol. (Submitted).

Cue, R.I., Monardes, H.G. and Hayes, J.F., 1987. Correlations between production traits in first lactation holstein cows. J. Dairy Sci., 70. 2132-2137.

De Jager, D. and Kennedy, B.W., 1987. Genetic parameters of milk yield and composition and their relationships with alternative breeding goals. J. Dairy Sci., 70: 1258-1266.

De Veer, J.C. and Van Vleck, L.D., 1987. Genetic parameters for first lactation milk yields at three levels of herd production. J. Dairy Sci., 70: 1434-1441.

Gómez, L. and Alenda, R., 1985. Parámetros genéticos y fenotípicos para caracteres productivos en la población española de ganado vacuno frisón en primera lactación. Proc. I Jornadas sobre Producción Animal. Octubre 22-24. Zaragoza. Pgs. 73-76.

Hill, W.G., Edwards, M.R., Akmed, M.K.A. and Thompson, R., 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod., 36: 59-68.

Maijala, K. and Hanna, M., 1974. Reliable phenotypic and genetic parameters in dairy cattle. Proc. Ist World Cong. Genet. Appli. Livestock Prod. Vol 1., 7-11 October. Madrid. Pgs. 541-564.

Mitchell, R.G., Corley, E.L. and Tyler, W.J., 1961. Heritability, phehotypic and genetic correlations between type ratings and milk and fat production in Holstein-Friesian cattle. Ibid., 44: 1501-1510.

Neumann, W., 1969. Ergebnisse der Heritabilitätsschätzung für die Milchmenge, den Fettgehalt und die Fettmenge auf der Grundlage von Nachkommenschaften, die nach Bereichen von Tierzunchtinspektionen geordnet. Arch. Tierzucht, 12: 295-304.

O'Connor, L.K., 1969. Genetic improvement of milk protein in dairy cattle. Proc. 2nd. World Conf. Anim. Prod. Pgs. 498-499.

Robertson, A., O'Connor, L.K. and Edwards, J., 1960. Progeny testing dairy breeds at different management levels. Anim. Prod., 2: 141-152.

Swalve, H. and Van Vleck, L.D., 1987. Estimation of genetic (co) variances for milk yield in first three lactations using an animal model and restricted maximum likelihood. J. Dairy Sci., 70: 842-849.

Wiggans, G.R., Misztal, I. and Van Vleck, L.D., 1988. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. J. Dairy Sci., 71: 54-69. Suppl. 2.

Wiggans, G.R., Misztal, I. and Van Vleck, L.D., 1988. Animal model evaluation of Ayrshire milk yield with all lactations, herd-sire interation, and groups based on unknown parents. J. Dairy Sci. (In press).

Table I. Data structure

	All lactations	First lactation
Sires	244	243
Herds	680	652
Herds-year	1,679	1,623
Cows	7,295	6,991
Records	12,102	6,991

Table II. Variance components, heritability, repeatability, mean and standard deviation estimates for milk and fat yield when all lactations included and when only first lactation is considered.

	All lactations		First lactation	
	Milk Kg	Fat Kg	Milk Kg	Fat Kg
Mean	6,049	200	5,071	160
Standard deviation	1,379	45	1,115	41
Phenotypic variance	1,148,996	1,210	1,035,557	1,105
Sire variance (σ^2_s)	48,634	37.3	51,805	41
Cow variance (σ^2_c)	421,244	395	----	----
Residual variance (σ^2_e)	679,118	777	983,752	1,064
Heritability	0.17	0.13	0.20(.045)	0.15(.039)
Repeatability	0.41	0.36	0.94*	0.89**

standard errors between brackets

* genetic correlation

** phenotypic correlation

Table III. Data structure by herd production level.

	High level	Low level
Sires	197	211
Herds	247	431
Herds-year	710	955
Cows	3,256	3,985
Records	5,142	6,856

Table IV. Variance components, heritability, repeatability, mean and standard desviation estimates for milk and fat yield by level of herd production.

	High level		Low level	
	Milk Kg	Fat Kg	Milk Kg	Fat Kg
Mean	6,661	218	5,231	175
Standard deviation	1,319	43	1,002	36
Phenotypic variance	1,467,519	1,490	699,780	811
Sire variance	62,175	47	36,468	30
Cow variance	560,165	526	240,497	218
Residual variance	845,178	918	422,815	563
Heritability	0.17	0.13	0.21	0.15
Repeatability	0.43	0.39	0.40	0.31