ANIMAL SCIENCE AND GENETICS
Published by the Polish Society of Animal Production
vol. 18 (2022), no 1, 25-31
DOI: 10.5604/01.3001.0015.7774

Research Article ***Open Access***

Assessment of the accuracy of selected models used to estimate the heritability coefficient in a cattle population

Zbigniew Sobek[1#], Jolanta Różańska-Zawieja[1], Idzi Siatkowski[2], Alicja Szabelska-Beręsewicz[2], Joanna Zyprych-Walczak[2]

Department of Genetics and Animal Breeding, Poznan University of Life Sciences, ul. Wołyńska 33, 60-637 Poznań
Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, ul. Wojska Polskiego 28, 60-637 Poznań

SUMMARY

The aim of the study was to examine the use of different effects in models for estimating gestation length in dairy cows in order to determine which effects result in the most accurate assessment. The analysis was conducted on a database of over 2,5 million Polish Holstein-Friesian cows subject to use value assessment in Poland that calved between 2005 and 2010. Nine mixed linear models were analysed. Analysis of variance was performed for each model. The Akaike Information Criterion and Bayesian Information Criterion were calculated for each model. The model which takes into account two random effects, sire and herd, obtained the highest score for the feature of gestation length in the population. Analysis of various models confirms that the model should be adapted to a given population, and that its usefulness must be verified.

KEY WORDS: linear model; dairy cattle; gestation length

INRODUCTION

Farmers and researchers are often faced with the choice of an appropriate mathematical model for data analysis. Existing computational systems make it possible to gather such large datasets that the model can be chosen from a wide range of possibilities. It is essential to choose the best model and a dataset of sufficient size to ensure the accuracy of the results. Authors usually use 'typical' models, even when they possess additional information that could be included (Rönnegård end Lee, 2013). On the other hand, many studies indicate that increasing the number of effects included in models does not necessarily improve the accuracy of the assessment (Donoghue et al., 2004;

[#]Corresponding author e-mail: zbigniew@up.poznan.pl
Received: 30.01.2022 Received in revised form: 05.02.2022
Accepted: 20.02.2022 Published online: 28.02.2022

et al., 2004; Sun et al., 2010).

Gestation length is a very important parameter, particularly in the context of animal welfare and production economics. More accurate prediction of the date of parturition can help dairy producers to meet the needs of pregnant cows and provide better health care during this high-risk stage of life (Norman et al., 2009). It can also help dairy farmers to achieve the targeted dry period length and meet the nutritional needs of cows during late pregnancy and calving, which can minimize the risk of metabolic diseases (Tomasek et al., 2017). Cows treated in a manner appropriate to their physiological state, i.e. pregnancy, are less likely to suffer from postpartum complications, which is subsequently reflected positively in their milk production and subsequent calving.

Gestation length is affected by multiple factors, including the age of the cow, production volume, the number, weight, and sex of the calves, the month of conception, and the season of calving (Tomasek et al., 2017; Hansen et al., 2004; Jenkins et al., 2016; King et al., 1985; Kheirandish et al., 2014; Rezac et al., 2013; Silva et al., 1992) Some studies indicate that gestation length is a highly heritable trait (Hansen et al., 2004; Winkelman et al., 2001) and that it significantly affects milk production and fertility parameters (Jenkins et al., 2016).

Although gestation length is not taken into account in selection programmes, there is no question that it has a major effect on the economics of dairy farm production. For this reason, accurate assessment of this trait for the cattle population in a given country, taking into account the climate zone, is very important.

The main goal of this work is to identify the best model for estimation of a trait, which includes all relevant information that could improve the assessment. The research focuses on the influence of various effects on the accuracy of the assessment. The model that results in the most accurate assessment is the basis for determining the heritability coefficient of gestation length in dairy cows.

MATERIAL AND METHODS

The research was based on data concerning over 2,5 million Polish Black-and-White Holstein-Friesian (HO) dairy cows subject to use value assessment in Poland. The dataset includes cows that gave birth during the years 2005-2010. The feature analysed was gestation length. As the first step of the analysis, data that included missing values were excluded from further analysis. Next, the dataset was filtered to exclude records with gestation length lower than 254 days or higher than 297 days. Finally, only records in which the bulls had sired at least 1000 calves were used for the analysis. The criteria resulted in a dataset of calves (158,783 males and 332,553 females) sired by a large number of bulls.

Linear mixed models were used in the study. Since only some of the bulls were included, the effect of the father was considered random. Similarly, the herd was also considered a random effect. The feature 'HYS' (herd-year-season of calf birth) is considered in two versions: as one effect or as the three separate effects of year, herd and season. The model included the following as fixed effects: year (R), season (SE – summer, i.e. 1 April to 30 September; winter, i.e. 1 October to 30 March), lactation number (L), gender (G) and difference in weight (B).

The following models were analysed:

model_herd_year_season: $Y_{ijklmno} = s_i + L_j + G_k + B_l + o_n + R_m + SE_o$

model_herd_season: $Y_{ijklno} = s_i + L_j + G_k + B_l + o_n + SE_o$

model_herd_year: $Y_{ijklmn} = s_i + L_j + G_k + B_l + o_n + R_m$

model_herd: $Y_{ijkln} = s_i + L_j + G_k + B_l + o_n$

model2: $Y_{ijklp} = s_i + L_j + G_k + B_l + hys_p$
model_year_season: $Y_{ijklmo} = s_i + L_j + G_k + B_l + R_m + SE_o$
model_season: $Y_{ijklo} = s_i + L_j + G_k + B_l + SE_o$
model_year: $Y_{ijklm} = s_i + L_j + G_k + B_l + R_m$
model1: $Y_{ijkl} = s_i + L_j + G_k + B_l,$

where:

Y_{ijkl} is the gestation length;
s_i is the random effect of the i^{th} bull;
L_j is the fixed effect of the j^{th} lactation number, $j = 1, \dots, 11$;
G_k is the fixed effect of the calf's gender, $k = 1,2$;
B_l is the fixed effect of the weight of the l^{th} calf $l = 1, 2$ (for twin calves the average is used);
o_n is the random effect of the n^{th} herd;
R_m is the fixed effect of the m^{th} year, $m = 1, 2, 3, 4, 5$ (2005, …, 2009);
SE_o is the fixed effect of the o^{th} season of calf birth, $o = 1, 2;$
hys_p is the random effect of the p^{th} herd-year-season class.

Analysis of variance (ANOVA) was performed for each model, and then the Akaike Information Criterion (AIC) (Akaike, 1974) and Bayesian Information Criterion (BIC) (Schwarz, 1978) were calculated for each model. The criteria were used to compare the models. The lower the value of the AIC or BIC coefficient, the more suitable the model. The results are ordered according to the values of the AIC and BIC coefficients.

The following phenotype variance (σ_y^2) parameter was derived from the estimated variance components:

$$\sigma_y^2 = \sigma_S^2 + \sigma_o^2 + \sigma_e^2,$$

where σ_S^2 is the variance of the bull effect, σ_o^2 is the variance of the herd effect, and σ_e^2 is the residual variance. Using this parameter, the coefficients of heritability (h_S^2) can be calculated as follows:

$$h_S^2 = {4\sigma_S^2}/{\sigma_y^2}.$$

All calculations were performed in R platform 3.2.0 (R Core Team, 2015) using the packages lme4 (Bates et al., 2014), data.table (Dowle et al., 2014) and dplyr (Wickham and Francois, 2015).

RESULTS AND DISCUSSION

For the 9 models presented in the previous section, the method of maximum likelihood was used to evaluate the AIC and BIC coefficients. In addition, p-values were calculated based on ANOVA for contiguous models from Table 1. For example, for models 7 (Y = s + L + G + B +SE) and 6 (Y = s + L + G + B + R + SE) we obtained a p-value < 0,001. This value indicates a significant difference between model 7 and model 6 directly above it in Table 1.

Table 1 shows that model Y = s + L + G + B + o + R + SE resulted in the lowest AIC and BIC values for the feature of gestation length in the analysed population. The other models yielded higher AIC and BIC values. This analysis confirms the need to verify the model and make a decision based on the fit of the model.

Table 1

AIC and BIC coefficients for assessment of model accuracy

No.	Model	AIC	BIC	p-value
1	Y = s + L + G + B + o + R + SE	3038697	3038830	----
2	Y = s + L + G + B + o + SE	3038791	3038880	< 0,001
3	Y = s + L + G + B + o + R	3040961	3041083	NS
4	Y = s + L + G + B + o	3041047	3041125	< 0,001
5	Y = s + L + G + B + hys	3047988	3048066	NS
6	Y = s + L + G + B + R + SE	3057141	3057286	NS
7	Y = s + L + G + B +SE	3057319	3057419	< 0,001
8	Y = s + L + G + B + R	3059306	3059417	NS
9	Y = s + L + G + B	3059472	3059538	< 0,001

NS – non-significant

Since the BLUP (best linear unbiased prediction) method was introduced for cattle evaluation, many farmers have been uncertain whether the HYS effect should be random or fixed. Henderson (1975) suggested that herd-year-season should be treated as a fixed effect in the populations included in the selection process. Chauhan (1987), however, indicated that the model is more effective when HYS is a random effect. Sobek (1989), to calculate the breeding value of bulls based on the BLUP method, applied two models, with HYS treated as a fixed effect in the first model and as a random effect in the second. Calculated rank correlations obtained for the breeding values indicate that the type of effect influenced the accuracy of the assessment.

Mäntysaari and Mäntysaari (2015) also used the AIC and BIC coefficients to evaluate the accuracy of models with random and fixed effects (mixed model). Norman et al. (2009) assessed the heritability of gestation length using a model in which the fixed effect of herd-year and the fixed effect of the season were used separately.

For the best model of the form Y = s + L + G + B + o + R + SE, the variances of the random effects are as follows:

$\sigma_S^2 = 5{,}364; \sigma_o^2 = 2{,}463$ and $\sigma_e^2 = 27.376$

Based on these calculations, the coefficient of heritability is $h_S^2 = 0{,}152$. The characteristics of the fixed effects are included in Tables 2 and 3. The intercept estimate was calculated for male calves in 2005. The values presented in Table 2 show how the gestation length changes when the variable in increased by one unit (L and B variables) or is changed to different level (G for females, R2006 for 20006 year, R2007 for 20007 year, R2008 for 20008 year, R2009 for 20009).

Table 2

Fixed effects used in the best model

	Estimate	Std. Error		Estimate	Std. Error
(Intercept)	280,087	0,1639	R2007	0,085	0,0315
L	0,109	0,0041	R2008	0,182	0,0355
G	-1,037	0,0173	R2009	0,341	0,0397
B	0,139	0,0021	SE	0,733	0,0154
R2006	-0,016	0,0284			

Table 3

Correlations of fixed effects

	(Intercept)	L	G	B	R2006	R2007	R2008	R2009
L	-0,086							
G	-0,081	0,002						
B	-0,003	-0,107	0,142					
R2006	-0,112	0,009	0,021	-0,038				
R2007	-0,124	0,024	0,012	-0,050	0,642			
R2008	-0,132	0,023	0,007	-0,053	0,562	0,652		
R2009	-0,125	0,022	0,004	-0,060	0,477	0,555	0,705	
SE	-0,049	0,013	0,000	-0,008	0,001	0,005	0,016	0,008

The heritability coefficient of gestation length for the Polish Holstein-Friesian cow population, calculated in the present study according to the best fitted linear model, is low, at 0,152. Nogalski and Piwczyński (2008) adopted a model including the fixed effect of the herd and random effect of the season and obtained a somewhat higher heritability coefficient for estimate gestation length in PHF cows, ranging from 0,201 to 0,210. The same authors obtained much lower values, from 0,054 to 0,073, when their estimation took into account indirect maternal effects. Johanson et al. (2011) obtained gestation length heritability of 0,51 for the Holstein-Friesian breed. Crevs (2006) reported high heritability coefficients for Charolais cattle, ranging from 0,61 to 0,64, estimated according to four different models. Cervantes et al. (2009), in a study on cattle of the meat breed Asturiana de los Valles, obtained gestation length heritability coefficients of 0,325 ± 0,022; 0,331 ± 0,026 and 0,226 ± 0,018; while Oyama et al. (2002) reported a coefficient of 0,4 for Japanese Black cows.

Similar estimates of heritability coefficients of gestation length were obtained in the 1960s by DeFries (1959) for five dairy cattle breeds in the United States, ranging from 0,420 to 0,474; depending on the method.

The value of the heritability coefficient is somewhat different from those reported in the literature, which may be explained in part by the population size and the choice of model best fitted to the data.

CONCLUSIONS

The proposed analysis using AIC and BIC criteria showed significant differences between several of the models considered in the study. The authors concluded that the best model, with significantly lower AIC and BIC values than the others, includes two random effects: bull effect and herd effect.

REFERENCES

1. Akaike H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control. No. 19. pp. 716-723
2. Bates D., Maechler M., Bolker B. and Walker S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-7. URL: http://CRAN.R-project.org/package=lme4
3. Cervantes J., Gutiérrez JP., Fernández I., & Goyache F. (2009). Genetic relationships among calving ease, gestation length, and calf survival to weaning in the Asturiana de los Valles beef cattle breed. J. Anim. Sci.(serial on line).1-22. Availabl from www.http://jas.fass.org accessed Sep 11 2010
4. Chauhan V.P.S. (1987). Dairy sire evaluation fitting some of the herd-year-season effects as random. Livest. Prod. Sci., 16: 117-130
5. Crews DH. (2006). Age of dam and sex of calf adjustments and genetic parameters for gestation length in Charolais cattle. J Anim Sci. 84: 25-31
6. DeFries J.C. (1959). Heritability of the Length of the Gestation Period in Dairy Cattle. J. Dairy Sci. 42, 4, 598-606
7. Donoghue K. A., Rekaya R., K. Bertrand J. (2004). Comparison of methods for handling censored records in beef fertility data: Field data. J. Anim. Sci., 82: 357-361
8. Dowle M., Short T., Lianoglou S., Srinivasan A. with contributions from Saporta R. and Antonyan E. (2014). Data.table: Extension of data.frame. R package version 1.9.4. http://CRAN.R-project.org/package=data.table
9. Hansen M., Lund M. S., Pederson J. and Christensen L. G. (2004). Gestation length in Danish Holsteins has weak genetic associations with stillbirth, calving difficulty, and calf size. Livest. Prod. Sci., 91: 23-33
10. Henderson C.R. (1975): Best linear unbiased estimation and prediction under a selection model. Biometrics, 31: 423-447
11. Jenkins G. M., Amer P., Stachowicz K., Meier S. (2016). Phenotypic associations between gestation length and production, fertility, survival, and calf traits. J. Dairy Sci., 99: 418-426
12. Johanson JM, Berger PJ, Tsuruta S., Misztal I. (2011). Bayesian threshold-linear model evaluation of perinatal mortality, dystocia, birth weight, and gestation length in a Holstein herd. J Dairy Sci., 94: 450-460
13. Kheirandish A., Khalajzadeh S., Yadi J. (2014). Effects of multiple births on gestation length of Iranian Holstein dairy cows. Res. Opin. Anim. Vet. Sci., 4(9): 508-511
14. King KK, Seidel Jr GE, Elsden RP. (1985). Bovine embryo transfer pregnancies. II. Lengths of gestation. J. Anim. Sci., 61: 758-62

15. Mäntysaari P., and Mäntysaari E. A. (2015): Modeling of daily body weights and body weight changes of Nordic Red cows. J. Dairy Sci. 98: 6992-7002
16. Nogalski Z., Piwczyński D. (2012). Association of Length of Pregnancy with Other Reproductive Traits in Dairy Cattle Asian-Australas J Anim Sci., 25(1): 22-27
17. Norman H. D., Wright J. R., Kuhn M. T., Hubbard S. M., Cole J. B., and Vanraden P. M. (2009). Genetic and environmental factors that affect gestation length in dairy cattle. J. Dairy Sci. 92: 2259-2269
18. Oyama K., Katsuta T., Anada K. and Mukai F. (2002). Heritability and Repeatability Estimates for Reproductive Traits of Japanese Black Cows. Asian-Aust. J. Anim. Sci. Vol. 15, No. 12: 1680-1685
19. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
20. Rezac P., Pospisilova D., Slama P., Havlicek Z. (2013). Different effects of month of conception and birth on gestation length in mares. J Anim Vet. Adv., 12: 731-5
21. Rönnegård L., Lee Y. (2013). Exploring the potential of hierarchical generalized linear models in animal breeding and genetics. J. Anim. Breed. Genet., Vol. 130 Issue 6, 415-416
22. Schwarz G. (1978). Estimating the dimension of a model. Ann. Statist. No. 6. pp. 461-464
23. Silva HM., Wilcox CJ., Thatcher WW., Becker RB., Morse D. (1992). Factors affecting days open, gestation length, and calving interval in Florida dairy cattle. J Dairy Sci.,75: 288-293
24. Sobek Z. (1989). Zastosowanie metody BLUP w ocenie wartości hodowlanej buhajów w Wielkopolsce. Roczniki Akademii Rolniczej w Poznaniu Rozprawy Naukowe, Zeszyt 193
25. Sun C., Madsen P., Lund MS., Zhang Y., Nielsen US., Su G. (2010). Improvement in genetic evaluation of female fertility in dairy cattle using multiple-trait models including milk production traits Journal of Animal Science, 88(3): 871-8
26. Tomasek R., Rezac P., Havlicek Z. (2017). Environmental and animal factors associated with gestation length in Holstein cows and heifers in two herds in the Czech Republic Theriogenology, 87: 100-107
27. Wickham H. and Francois R. (2015). dplyr: A Grammar of Data Manipulation. R package version 0.4.1. http://CRAN.R-p Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control. No. 19. pp. 716-723
28. Winkelman, A. M., and R. J. Spelman. (2001).Selection for reduced gestation length in New Zealand dairy cattle. Proc. Assoc. Advmt. Anim. Breed. Genet., 14: 63-66