

Wpływ wielkości grup ojcowskich i stad na optymalną ocenę współczynnika odziedziczalności długości ciąży u bydła

Jolanta Różańska-Zawieja^{1#}, Alicja Szabelska-Beręsewicz²,
Zbigniew Sobek¹, Anna Nienartowicz-Zdrojewska¹,
Joanna Zyprych-Walczak², Idzi Siatkowski²

¹Uniwersytet Przyrodniczy w Poznaniu, Katedra Genetyki i Podstaw Hodowli Zwierząt,
ul. Wołyńska 33, 60-637 Poznań; #e-mail: jolek@up.poznan.pl

²Uniwersytet Przyrodniczy w Poznaniu, Katedra Metod Matematycznych i Statystycznych,
ul. Wojska Polskiego 28, 60-637 Poznań

Estymacja parametrów genetycznych jest kluczowym elementem w procesie doskonalenia populacji. W przypadku zwierząt hodowlanych proces ten jest oparty na próbcie, jako podzbiórce całej populacji. Z tego powodu ważnym jest, aby znać wpływ wielkości grup ojcowskich i wielkości stada na dokładność oszacowania tych parametrów, a w szczególności odziedziczalności. Celem przeprowadzonego badania było wykazanie wpływu wielkości grup ojcowskich i stad na dokładność szacowania współczynnika odziedziczalności (h^2). Materiał do analiz obejmował dane krów rasy holsztyńsko-fryzyjskiej urodzonych w latach 2005-2010 i objętych oceną wartości użytkowej w Polsce. Analizowaną cechą była długość ciąży. Obliczenia z wykorzystaniem liniowego modelu mieszanego przeprowadzono przy użyciu platformy R 3.1.3. Analizy wykazały, że jeśli obliczenia dotyczą córek w grupach ojcowskich, to dokładniejsze są dla tych, które znajdują się w mniejszej liczbie stad, ale stad o większej liczebności. Wzrost liczebności w grupach ojcowskich kosztem ich liczby nie ma takiego wpływu na pogorszenie dokładności oszacowania, jak to jest w przypadku ograniczenia liczby małych stad. Ograniczenie oszacowań tylko do największych stad powoduje pogorszenie dokładności oszacowania.

SŁOWA KLUCZOWE: odziedziczalność / model liniowy / bydło mleczne / długość ciąży

Proces doskonalenia populacji wymaga wiedzy na temat parametrów genetycznych doskonalonych cech, a zwłaszcza ich współczynnika odziedziczalności (h^2). Przy szacowaniu h^2 ważnym jest, aby właściwie dopasować model do struktury analizowanej populacji, poprzez włączenie znaczących efektów mających wpływ na wartości doskonalonych cech [11]. Wartości parametrów genetycznych są określane na podstawie próby będącej podzbiorem całej populacji, dlatego ważne jest, aby wiedzieć, jaki wpływ ma wielkość grup ojcowskich i stad na dokładność szacowania współczynnika odziedziczalności.

Gromadzenie danych, szczególnie w przypadku dużych zwierząt gospodarskich, wymaga długiego okresu czasu (czasami kilka lat). Z tego powodu wskazane jest, aby znać przynajmniej przybliżone wymagania liczebnościowe względem analizowanej populacji, aby mieć pewność, że przeprowadzone oszacowania badanego parametru będą na przewidywalnym poziomie dokładności [3, 9].

Celem badań było wykazanie wpływu wielkości populacji (liczby zwierząt) w grupach ojcowskich i stad na dokładność szacowania h^2 .

Material i metody

Baza danych zawierała informacje o krowach wycielonych w latach 2005-2010. Analizowaną cechą była długość ciąży. Dane obejmowały informacje o krowach rasy polskiej holsztyńsko-fryzyjskiej odmiany czarno-białej (PHF). Plik wyjściowy zawierał rekordy dotyczące 491 336 osobników. W pierwszej fazie, analizując efekt wielkości stada, do obliczeń użyto danych krow, które pochodziły po buhajach będących ojcami przynajmniej 1000 razy. Otrzymany zbiór danych zawierał w rezultacie stada o wielkości między co najmniej 400 a 1700 krow. Z kolei, dla najbardziej efektywnej wielkości stada (przynajmniej 400 krow) została przeprowadzona analiza mająca na celu określenie optymalnej liczby grup ojcowskich, wykorzystująca dane w zakresie od co najmniej 400 do 1700 cieląt po każdym byku. W ostatnim kroku analizowane były jednocześnie wielkość stada oraz grupy ojcowskie. W tym przypadku otrzymane zestawy danych zawierały rekordy, w których wielkość stada liczyła co najmniej 400 do 1100 krow. Jednocześnie liczba cieląt po każdym byku zawierała się między co najmniej 400 a 1500. Dla tak otrzymanych zestawów danych został użyty liniowy model mieszany, w postaci:

$$Y_{ijklmno} = s_i + L_j + G_k + M_l + P_m + o_n + B_o + e_{ijklmno}$$

gdzie:

$Y_{ijklmno}$ – długość ciąży;

s_i – efekt losowy i -tego byka;

L_j – efekt stały związany z j -tym numerem laktacji ($j = 1, \dots, 11$);

G_k – efekt stały płci cielęcia ($k = 1, 2$);

M_l – efekt stały związany z masą ciała l -tego cielęcia;

P_m – efekt stały m -tego roku ($m = 2005, \dots, 2010$);

o_n – efekt losowy n -tego stada;

B_o – efekt stały o -tej pory roku narodzin cielęcia ($o = 1, 2$) (lato, zima);

$e_{ijklmno}$ – błąd losowy.

Parametr wariancji fenotypowej (σ_y^2) został wyznaczony na podstawie wyestymowanych komponentów wariancyjnych, według wzoru:

$$\sigma_y^2 = \sigma_s^2 + \sigma_o^2 + \sigma_e^2$$

gdzie: σ_s^2 oznacza wariancję efektu buhaja (ojca), σ_o^2 to wariancja efektu stada, a σ_e^2 to wariancja reszt. Korzystając z tych parametrów współczynnik odziedziczalności (h_s^2) wyznaczono według wzoru:

$$h_s^2 = \frac{\sigma_s^2}{\sigma_y^2}$$

Jako istotną statystykę podsumowującą model mieszany wykorzystano współczynnik determinacji (R^2). Wykorzystanie R^2 w tego typu modelach jest rzadkie, natomiast często stosowane jest w stałych modelach liniowych (LM) oraz uogólnionych modelach liniowych (GLM). R^2 ma bardzo użyteczną właściwość, dostarczając bezwzględną wartość dopasowania modelu, która nie może być zmierzona za pomocą kryteriów informacyjnych [8]. Dla modeli mieszanych R^2 może być dwojakiego typu: marginalny R^2 reprezentuje wariancję wyjaśnioną przez czynniki stałe oraz warunkowy, który interpretowany jest jako wariancja wyjaśniona przez oba czynniki stałe i losowe (tj. cały model), i jest wyliczana zgodnie ze wzorem:

$$R_c^2 = \frac{\sigma_f^2 + \sum_{l=1}^u (\sigma_l^2)}{\sigma_f^2 + \sum_{l=1}^u (\sigma_l^2) + \sigma_e^2 + \sigma_d^2}$$

gdzie: σ_f^2 to wariancja efektów stałych, σ_l^2 to wariancja l -tego efektu losowego, σ_e^2 to wariancja związana z addytywną dyspersją oraz σ_d^2 to wariancja specyficzna rozkładu (szczegóły Nakagawa i Schielzeth [8], s. 137).

Oszacowania miały na celu zweryfikowanie, jakie wielkości efektów losowych byłyby właściwe, aby uzyskać optymalny model do szacowania współczynnika odziedziczalności, natomiast wybór modelu był dokonywany na podstawie warunkowego R_c^2 .

Wszystkich obliczeń dokonano z wykorzystaniem pakietu statystycznego R 3.1.3 [10], pakietów lme4 [2], data.table [4], MuMIn [1] oraz dplyr [13].

Wyniki i dyskusja

W pierwszej kolejności skupiono się na wpływie wielkości stada na estymację współczynnika R_c^2 oraz h_s^2 . Dla ustalonej wartości powtórzeń buhaja-ojca (przynajmniej 1000 cieląt po każdym byku) sprawdzono, jak wartości tych współczynników zmieniały się dla różnej wielkości stad. Podsumowanie charakterystyk rozważanych zestawów danych przedstawiono w tabeli 1.

Tabela 1 – Table 1

Podsumowanie zestawów danych otrzymanych dla ustalonej liczby powtórzeń byków (>1000) oraz każdej z rozważanych wartości dla wielkości stada (zakres: od >400 do 1700)

Summary of datasets obtained for the fixed number of repetitions for each bull in the dataset (>1,000) and each number of repetitions for the herd in the dataset (range 400 – 1,700)

Wielkość stada Size of the herd	Liczba krów Number of cows	Liczba byków Number of sires	Liczba stad Number of herds
>1700	3584	69	2
>1600	3584	69	2
>1500	3584	69	2
>1400	5064	79	3
>1300	7823	85	5
>1200	7823	85	5
>1100	12 378	97	9
>1000	13 431	102	10
>900	16 287	128	13
>800	20 421	143	18
>700	26 364	161	26
>600	31 709	170	34
>500	40 660	188	50
>400	54 255	208	81

Dla takich zestawów danych przeprowadzono analizy, w celu wyznaczenia współczynnika determinacji R_c^2 oraz wartości współczynnika odziedziczalności h^2 . Otrzymane rezultaty zaprezentowano w tabeli 2.

Tabela 2 – Table 2

Wartości R_c^2 , wariancje efektów losowych oraz współczynnik dla ustalonej liczby powtórzeń byków (>1000) oraz każdej z rozważanych wartości dla wielkości stada (zakres: od >400 do >1700)

R_c^2 values, variances of random effects and heritability coefficient for the number of repetitions for each bull in the dataset (>1,000) and each number of repetitions for the herd in the dataset (range 400 – 1,700)

Wielkość stada Size of the herd	R_c^2	σ_s^2	σ_o^2	σ_e^2	h_s^2
>1700	0,198	0,555	5,362	27,442	0,017
>1600	0,198	0,555	5,362	27,442	0,017
>1500	0,198	0,555	5,362	27,442	0,017
>1400	0,200	0,662	5,015	26,515	0,021
>1300	0,215	0,738	5,199	26,075	0,023
>1200	0,215	0,738	5,199	26,075	0,023
>1100	0,205	0,622	4,597	25,140	0,020
>1000	0,202	0,629	4,398	25,083	0,021
>900	0,199	0,661	4,241	25,463	0,022
>800	0,201	0,989	3,891	25,087	0,033
>700	0,214	1,054	4,295	25,124	0,035
>600	0,211	1,126	4,078	25,120	0,037
>500	0,227	1,180	4,768	25,261	0,038
>400	0,239	1,350	5,161	25,452	0,042

Wartości R_c^2 zawarte w tabeli 2. dowodzą, że im mniejsze są restrykcje odnośnie do wielkości stada, tym lepszy model oraz że wraz ze wzrostem R_c^2 obserwuje się wzrost h_s^2 .

Następnie przeanalizowano wpływ liczby cieląt (potomstwa) po każdym buhaju na estymację współczynnika R_c^2 oraz h_s^2 . Dla ustalonej wielkości stada (przynajmniej 400 krów w stadzie) sprawdzono zmiany wartości tych współczynników dla różnej liczby cieląt po każdym buhaju. Podsumowanie charakterystyk rozważanych zestawów danych przedstawiono w tabeli 3.

Tabela 3 – Table 3

Podsumowanie zestawów danych otrzymanych dla ustalonej wielkości stad (>400) oraz każdej z rozważanych wartości dla liczby cieląt po każdym byku (zakres: od >400 do >1700)

Summary of datasets obtained for the fixed herd size (at least 400) and each number of calves from each sire in the dataset (range 400 – 1,700)

Liczba cieląt Number of calves	Liczba krów Number of cows	Liczba ojców Number of sires	Liczba stad Number of herds
>1700	13659	6	294
>1600	18549	9	318
>1500	21681	11	321
>1400	24576	13	324
>1300	27264	15	328
>1200	34729	21	343
>1100	43776	29	346
>1000	51045	36	346
>900	59480	45	347
>800	70591	58	347
>700	77992	68	347
>600	31709	170	34
>500	40660	188	50
>400	54255	208	81

Dla takich zestawów danych przeprowadzono analizy, aby wyznaczyć współczynnik R_c^2 oraz wartości współczynnika odziedziczalności. Otrzymane wyniki przedstawiono w tabeli 4.

Wartości współczynnika odziedziczalności rosną wraz ze spadkiem liczby cieląt po każdym byku-ojcu. Jednakże, w porównaniu do danych z tabeli 2., wartości tego współczynnika są wyższe, wpadając w przedział między 0,055 a 0,114.

Kolejnym krokiem analiz było wyznaczenie wartości h_s^2 oraz R_c^2 dla każdej kombinacji wielkości stada (w zakresie od >400 do >1100 krów) oraz liczby powtórzeń byka (w zakresie od >400 do >1500 cieląt) w zestawie danych. Wyniki zaprezentowano odpowiednio na rysunkach 1 i 2.

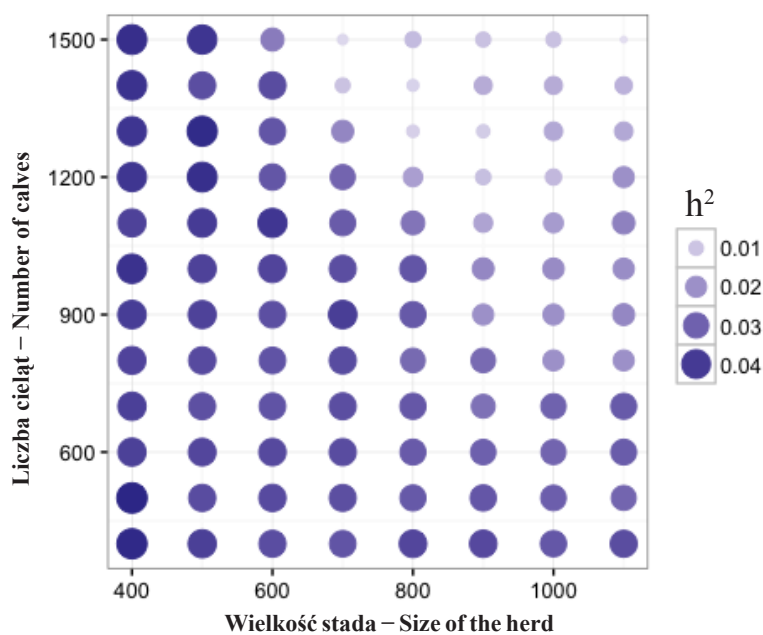
Optymalnym rozwiązaniem byłaby jednoczesna maksymalizacja współczynnika odziedziczalności h_s^2 oraz współczynnika determinacji R_c^2 .

Tabela 4 – Table 4

Wartości R_c^2 , wariancje efektów losowych oraz współczynnik dla ustalonej wielkości stad (>400) oraz każdej z rozważanych wartości liczby powtórzeń byków (zakres: od >400 do >1700)

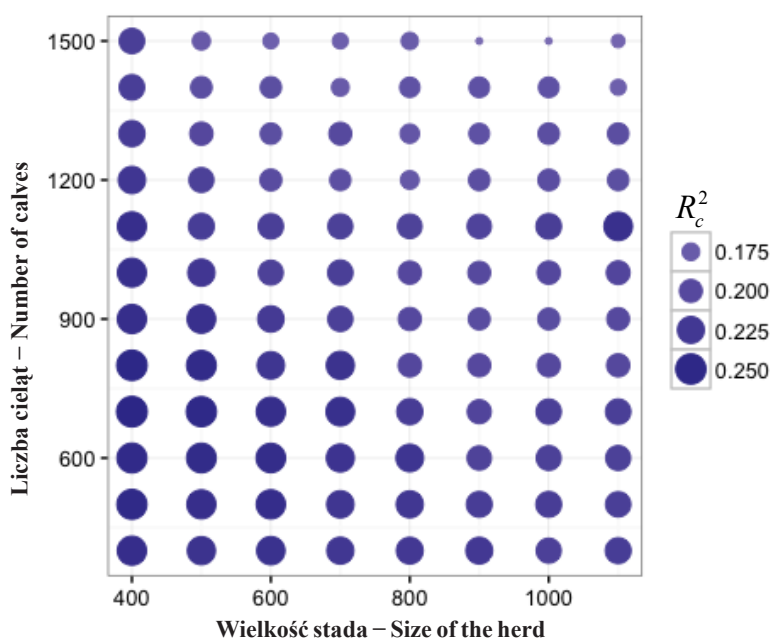
R_c^2 values, variances of random effects and heritability coefficient for the fixed number of repetitions for each herd in the dataset (>400) and each number of repetitions for the bull in the dataset (in range 400 – 1,700)

Liczba cieląt Number of calves	R_c^2	σ_s^2	σ_o^2	σ_e^2	h_s^2
>1700	0,140	1,596	1,409	25,783	0,055
>1600	0,142	1,754	1,377	26,112	0,060
>1500	0,154	2,229	1,351	26,068	0,075
>1400	0,149	1,925	1,436	26,180	0,065
>1300	0,178	3,001	1,439	26,185	0,098
>1200	0,171	2,609	1,448	25,813	0,087
>1100	0,160	2,460	1,261	25,879	0,083
>1000	0,163	2,447	1,260	25,659	0,083
>900	0,171	2,604	1,299	25,617	0,088
>800	0,164	2,390	1,244	25,645	0,082
>700	0,166	2,449	1,250	25,757	0,083
>600	0,173	2,649	1,277	25,582	0,090
>500	0,177	2,764	1,309	25,605	0,093
>400	0,196	3,479	1,338	25,675	0,114



Rys. 1. Współczynnik odziedziczalności dla zestawów danych z wielkością stad w zakresie od >400 do >1100 krów oraz liczbą powtórzeń byka w zakresie od >400 do >1500 cieląt

Fig. 1. Heritability coefficient for the number of repetitions for each herd in the dataset (range 400 – 1,100) and each number of repetitions for the bull in the dataset (range 400 – 1,500)



Rys. 2. Współczynnik R_c^2 dla zestawów danych z wielkością stad w zakresie od >400 do >1100 krów oraz liczbą powtórzeń byka w zakresie od >400 do >1500 cieląt
 Fig. 2. R_c^2 values for the number of repetitions for each herd in the dataset (range 400 – 1,100) and each number of repetitions for the bull in the dataset (in range 400 – 1,500)

Na rysunku 1. można zauważyć, że dla wielkości stad w zakresie od >400 do >600 współczynnik h_s^2 utrzymuje się na podobnym poziomie, natomiast dla większych ograniczeń dotyczących wielkości stad współczynnik ten się zmniejsza.

Z drugiej strony, dane zawarte na rysunku 2. wskazują, że najwyższe wartości współczynnika R_c^2 uzyskano dla zestawów danych z najmniejszą liczbą powtórzeń byków oraz najmniejszym ograniczeniem dla wielkości stada jednocześnie.

W pracy badano wpływ wielkości populacji grup ojcowskich i stad na wartości współczynnika odziedziczalności długości ciąży u bydła. W tym celu wykorzystano liniowy model mieszany, który we wstępnych badaniach autorów uznano za najbardziej dokładny. Najwyższą wartość h^2 uzyskano dla wariantu ponad 400 cieląt po buhaju w stadzie ($h^2=0,114$). Wzrost wielkości stada do ponad 1700 cieląt po ojcu spowodował znaczne zmniejszenie liczby grup ojcowskich, co spowodowało zmniejszenie wartości h^2 do 0,055.

Khan i Khan [7] przedstawili w swoich badaniach wyniki współczynników odziedziczalności dla rodzimej populacji bydła. Wartości te zostały porównane do licznych badań prowadzonych na całym świecie. Wyniki tych oszacowań były zróżnicowane zarówno pod względem wartości, jak i dokładności szacunków. Według Khan i Khan [7] różnice te wynikały zarówno z różnorodności światowej populacji bydła, jak i różnych wielkości

badanych populacji. W prezentowanych badaniach własnych próbowano wykazać wpływ efektu liczby obiektów i ich wielkości w populacji na dokładność oszacowań.

Rzewuska i Strabel [12] wykazali, że w przypadku cech wydajności mleka wpływ efektu kolejnej laktacji na wartość h^2 jest tak duży, że w pełni uzasadnia wprowadzenie efektu kolejnych laktacji do modelu.

W zależności od specyfiki populacji model może być skonstruowany przy użyciu dodatkowych efektów, jak to miało miejsce w badaniach przeprowadzonych przez Ismael i wsp. [5]. W tym przypadku został uwzględniony efekt położenia geograficznego. Takie rozszerzenie modelu było uzasadnione w sytuacji, gdy autorzy posiadali dane dla 811 285 krów pierwiastek z 7458 stad, po 3432 ojcach.

Jenko i wsp. [6] także badali wpływ liczby obserwacji w klasach stado-rok na wartość szacowania h^2 . Analizy przeprowadzili na populacji ponad 42 000 zwierząt. Wyciągnięte wnioski wskazywały niską podatność na dywersyfikację oszacowań, w zależności od przyjętego kryterium wielkości populacji, w klasie stado-rok. Cytowani autorzy wskazali, że wzrost liczebności w grupach stada-rok spowodował minimalny wzrost uzyskanych oszacowań współczynnika odziedziczalności.

Z kolei w badaniach własnych, za pomocą współczynnika determinacji R_c^2 , wykazano, że jeśli obliczenia dotyczą ojcowskich grup buhajów, których córki znajdują się w mniejszej liczbie stad, ale stad o większej liczebności, uzyskane oszacowania są bardziej dokładne i wartości h^2 nieznacznie rosną.

Dla badanej cechy wykazano, że wzrost liczebności w grupach ojcowskich kosztem ich liczby nie ma takiego wpływu na pogorszenie dokładności oszacowania, jak ograniczenie liczby małych stad. Ograniczenie estymacji tylko do najliczniejszych stad powoduje pogorszenie dokładności estymacji.

PIŚMIENNICTWO

1. BARTOŃ K., 2016 – MuMIn: Multi-Model Inference. R package version 1.15.6. <https://CRAN.R-project.org/package=MumIn>.
2. BATES D., MAECHLER M., BOLKER B., 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. BAUSELL R.B., LI Y.-F., 2002 – Power Analysis for Experimental Research. A Practical Guide for the Biological, Medical and Social Science, Cambridge University Press.
4. DOWLE M., SHORT T., LIANOGLU S., SRINIVASAN A. with contributions from SAPORTA R. and ANTONYAN E., 2014 – data.table: Extension of data.frame. R package version 1.9.4. URL: <http://CRAN.R-project.org/package=data.table>.
5. ISMAEL A., STRANDBERG E., BERGLUND B., KARGO M., FOGH A., LØVENDAHL P., 2016 – Genotype by environment interaction for the interval from calving to first insemination with regard to calving month and geographic location in Holstein cows in Denmark and Sweden. *Journal of Dairy Science* 99, 5498-5507.
6. JENKO J., GORJANC G., KOVAČ M., DUCROCQ V., 2013 – Comparison between sire-maternal grandsire and animal models for genetic evaluation of longevity in a dairy cattle population with small herds. *Journal of Dairy Science* 96, 8002-8013.

7. KHAN M.A., KHAN M.S., 2016 – The heritability estimates of linear type traits in sahiwal cows. *Journal of Animal and Plant Sciences* 26 (1), 25-33.
8. NAKAGAWA S., SCHIELZETH H., 2013 – A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2), 133-142.
9. NORMAN H.D., WRIGHT M.T., KUHN S.M., HUBBARD J.B., COLE J.B., VANRADEN P.M., 2009 – Genetic and environmental factors that affect gestation length in dairy cattle. *Journal of Dairy Science* 92, 2259-2269.
10. 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/>.
11. RÖNNEGÅRD L., LEE Y., 2013 – Exploring the potential of hierarchical generalized linear models in animal breeding and genetics. *Journal of Animal Breeding and Genetics* 130 (6), 415-416.
12. RZEWUSKA K., STRABEL T., 2013 – Genetic parameters for milk urea concentration and milk traits in Polish Holstein-Friesian cows. *Journal of Applied Genetics* 54 (4), 473-482.
13. WICKHAM H., FRANCOIS R., 2015 – dplyr: A Grammar of Data Manipulation. R pack-age version 0.4.1. URL: <http://CRAN.R-project.org/package=dplyr>.

Jolanta Róžańska-Zawieja, Alicja Szabelska-Beręsewicz, Zbigniew Sobek,
Anna Nienartowicz-Zdrojewska, Joanna Zyprych-Walczak, Idzi Siatkowski

The effect of population size of paternal groups and herds on optimal estimation of the heritability index for gestation length in cattle

Summary

Estimation of genetic parameters is a crucial element in the process of population improvement. In the case of farm animals this process is based on a sample which is a subset of the whole population. For this reason it is important to know the effect of the size of paternal groups and of the herd on the accuracy of estimations of these parameters, particularly heritability. The aim of this study was to show the effect of the population size of paternal groups and herds on the accuracy of estimation of the heritability index (h^2). The material for the analyses comprised data of Holstein-Friesian cattle born in 2005-2010 and subject to use value assessment in Poland. The trait analysed was gestation length. Calculations using a linear mixed model were performed using the R 3.1.3 platform. The analyses showed that calculations concerning daughters in paternal groups are more accurate when the daughters are in a smaller number of herds, but of greater size. An increase in the size of paternal groups at the expense of their number does not have such a negative effect on the accuracy of the estimate as in the case of a reduced number of small herds. Limiting the estimate to only the largest herds reduces its accuracy.

KEY WORDS: heritability / linear model / dairy cattle / gestation length