Journal of Animal and Feed Sciences, 11, 2002, 485–495

Impact of data transformation on the heritability estimates of reproductive traits in laying hens*

P. Piotrowski and T. Szwaczkowski

August Cieszkowski Agricultural University,
Department of Genetics and Animal Breeding
Wołyńska 33, 60-637 Poznań, Poland

(Received 18 April 2002; accepted 2 August 2002)

ABSTRACT

Two reproductive traits: percentage of fertilized egg (PFE) and percentage of hatched of set eggs (PHC) of four strains of laying hens (15 339 recorded individuals) over nine generations from pedigree farm were studied. The computations for three type of data sets for each trait (untransformed, arcsinsqrt transformed, probit transformed data) were used to estimate direct and maternal genetic variances within strains. Prior to analysis each observation was divided by the average. Error variance estimates and logarithms of likelihood were taken as comparison criteria. Generally, the reproductive traits are low heritable. Negligible differences between direct heritability estimates of PFE have been registered. However, these estimates obtained from untransformed data were larger than the transformed ones. Generally, smaller residual variances have been received from arcsinsqrt transformed data (in a majority of cases, largest error variances were estimated from probit transformed data). It usually corresponds with the second of the employed criteria.

KEY WORDS: data transformation, laying hens, reproductive traits, genetic variance components

INTRODUCTION

Generally, reproductive traits are troublesome in statistical analysis. It is known that one of the main assumption of the classical methods is, among others, normality of residuals. Unfortunately, fertility and hatchability, expressed usually as percentage per dam, do not hold the assumption. Thus, data transformations (see e.g.,

* Supported by the State Committee for Scientific Research, Grant No. 5 PO6D 02019

Foerster, 1993) are recommended. When eggs are treated as units, the threshold model (Gianola and Foulley, 1983) can be used. However, from a computational standpoint the use of a threshold approach in complex animal models is more demanding than the linear model. Therefore, from the practical point of view for "easier" computations (theoretical advantage of the threshold model), the transformation of data sets to normality is still preferable. Moreover some authors (Hagger and Hofer, 1989; Varona et al., 1999) reported similar estimates of genetic parameters obtained via the linear and threshold models.

The last decades have seen an increasing role of reproductive traits in livestock improvement programs. A number of investigations concerning the fertility and hatchability have been conducted in layer populations (Chaudary et al., 1987; Hartmann, 2001). It is known that the development of a chicken embryo depends on the egg environment during incubation. The so-called „egg environment" is determined by both dam genotype and the external environment. Hence, a statistical model including the maternal (indirect) effects is more frequently applied to the genetic study of these traits. For instance, Sevalem et al. (1998) reported that maternal additive genetic variance for laying hen reproduction traits made up a considerable part (0.19) of phenotypic variability.

The objective of this paper is to analyse the adequacy of heritability estimates models for transformed and untransformed data-sets of reproductive laying hens. The Bliss degrees (arcsin square) and probit transformations have been checked.

MATERIAL AND METHODS

Birds

The study is based on information collected on Pedigree Laying Hen Farm of Iwno (West Poland) in the years 1989-97. Four strains have been included into the analysis (more details are given in Table 1). The percentage of fertilized eggs (PFE) and percentage of hatched of set eggs (PHC) were observed. PFE was examined by candling on day 8 of incubation. The number of observations per generation (year) is relatively small because both traits were registered only for dams (chosen as parents). Feeding level and other environmental conditions did not considerably vary. Descriptive statistics of the data files are also shown in Table 1.

Data sets

Prior to analysis the following three data sets for each trait of each strain were formed: DATA-1 (untransformed data), DATA-2 (data transformed as $y=\arcsin \omega_i^{1/2}$ (see e.g. Foerster, 1993), where: is the i-th untransformed observa-

tion (percentage), DATA-3 (the so-called probit transformation). In the case of probit transformation, cumulative frequencies are transformed to a normal probability (probit) scale (see e.g., Lynch and Walsh, 1998).

To compare the results from different data sets (untransformed and transformed data) each observation was divided by the mean.

TABLE 1

Description of the data sets

Strain	n	Average	SD	Skewness	Kurtosis
		PFE			
H77	5422	87.65	(±17.06)	-2.5381	7.1336
N88	3463	89.42	(±17.46)	3.1841	137.0083
R33	2040	87.29	(±17.71)	-2.3286	5.7597
S22	4414	82.94	(±20.04)	-2.0260	4.1124
		PHC			
H77	5386	75.53	(±16.48)	-1.0998	1.4610
N88	3441	78.56	(±15.26)	-1.3716	2.6737
R33	2015	69.4	(±19.73)	-0.8341	0.4238
S22	4333	66.89	(±20.35)	-0.6268	-0.0450

n = number of observations

Genetic model

A single trait animal model has been used to estimate a direct and maternal heritability within strains:

$$\mathbf{y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{m} + \mathbf{e}$$

where: $\mathbf{y}$ is the $nx1$ vector of observations;
$\mathbf{b}_1$ is a $p_1 x1$ vector of fixed generation (year) effects (p_1 =9);
$\mathbf{b}_2$ is a $p_2 x1$ vector of hatch period effects (p_2 =4);
$\mathbf{a}$ is a $qx1$ vector of random direct additive genetic effects;
$\mathbf{m}$ is a $qx1$ vector of random maternal additive genetic effects;
$\mathbf{e}$ is a $nx1$ vector of random errors;
$\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ the nxp_1, nxp_2, nxq and nxq incidence matrices, respectively.
The first and second moments were assumed to be as follows:

$$E\begin{bmatrix}\mathbf{a}\\ \mathbf{m}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix} \text{ and } D\begin{bmatrix}\mathbf{a}\\ \mathbf{m}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & \mathbf{0}\\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where **A** is the *qxq* additive relationship matrix; σ_a^2 is the direct additive genetic variance; σ_e^2 is the error variance; σ_m^2 is the maternal additive genetic variance; σ_{am} is the covariance between direct and maternal additive effects.

Hence, the

$$\mathbf{y} \sim N(\mathbf{X_1 b_1} + \mathbf{X_2 b_2}, \mathbf{Z_1 A Z'}_1\sigma_a^2 + \mathbf{Z_1 A Z'}_2\sigma_{am} + \mathbf{Z_2 A Z'}_1\sigma_{am} + \mathbf{Z_2 A Z'}_2\sigma_m^2 + \mathbf{I}\,\sigma_e^2).$$

The following genetic parameters have been estimated:
- direct heritability ($h_a^2 = \sigma_a^2 / \sigma_p^2$),
- maternal additive heritability ($h_m^2 = \sigma_m^2 / \sigma_p^2$),
- covariance between direct and maternal effects as proportion to phenotypic variance ($d_{am} = \sigma_{am} / \sigma_p^2$),
- total heritability ($h_T^2 = (\sigma_a^2 + 0.5\sigma_m^2 + 1.5\sigma_{am}) / \sigma_p^2$), where σ_p^2 is the phenotypic variance.

Computing algorithm and comparison criteria

The derivative-free restricted maximum likelihood (DFREML) algorithm (Meyer, 1989) under a simplex procedure has been employed. A value of 10^{-8} was used as the convergence criterion on all analysis. The following starting values for each data set were taken: 0.5 for h_a^2, 0.01 for h_m^2 and 0.001 for d_{am}.

The residual variance estimates were used as a criterion of the model's adequacy for the same trait within strain. Moreover, the logarithm likelihood values (log L) were also checked. The computations were performed by the use the DFREML package programs of Meyer (1993).

RESULTS AND DISCUSSION

Estimates of direct and maternal heritabilities and covariance between the effects (as proportion) for the two traits studied are listed in Tables 2 and 3. As expected, reproductive traits have been shown to be low heritable. However, the estimates are higher for PFE than PHC, which corresponds with the number of results reported in the literature (Sewalem, 1998; Szwaczkowski et al., 2000; Hartmann, 2001). It seems that a larger genetic variability of PHC resulted from a relatively long physiological process. On the other hand, opposite relationships of these traits have also been found. Additionally, differences between the strains have been registered. So, many comparisons of estimates from untransformed and two -way transformed data can be performed. As already mentioned two various approaches have been examined. A number of authors (see e.g., Foerster, 1993) have previously employed the Bliss

TABLE 2

Heritability estimates and their standard deviations of PFE for different data sets

Parameter	Data set	Strain							
		H77		N88		R33		S22	
h_a^2	DATA 1	0.093	(±0.023)	0.079	(±0.027)	0.069	(±0.033)	0.111	(±0.028)
	DATA 2	0.087	(±0.023)	0.080	(±0.027)	0.101	(±0.040)	0.124	(±0.030)
	DATA 3	0.029	(±0.013)	0.034	(±0.018)	0.094	(±0.038)	0.112	(±0.028)
h_m^2	DATA 1	0.043	(±0.016)	0.033	(±0.017)	0.019	(±0.017)	0.026	(±0.014)
	DATA 2	0.044	(±0.016)	0.032	(±0.017)	0.037	(±0.024)	0.016	(±0.011)
	DATA 3	0.012	(±0.008)	0.019	(±0.013)	0.012	(±0.014)	0.003	(±0.005)
d_{am}	DATA 1	0.000		0.021	(±0.014)	0.036	(±0.024)	0.001	(±0.003)
	DATA 2	-0.034	(±0.014)	0.001	(±0.003)	0.003	(±0.007)	0.001	(±0.003)
	DATA 3	0.001	(±0.002)	0.002	(±0.004)	0.001	(±0.004)	-0.001	(±0.003)
h_T^2	DATA 1	0.114	(±0.026)	0.128	(±0.034)	0.132	(±0.045)	0.125	(±0.030)
	DATA 2	0.104	(±0.025)	0.097	(±0.030)	0.123	(±0.044)	0.133	(±0.031)
	DATA 3	0.034	(±0.014)	0.046	(±0.021)	0.102	(±0.040)	0.111	(±0.028)

TABLE 3

Heritability estimates and their standard deviations of PHC for different data sets

Parameter	Data set	Strain							
		H77		N88		R33		S22	
h_a^2	DATA 1	0.136	(±0.032)	0.242	(±0.055)	0.307	(±0.108)	0.199	(±0.038)
	DATA 2	0.130	(±0.039)	0.179	(±0.041)	0.202	(±0.056)	0.198	(±0.038)
	DATA 3	0.045	(±0.016)	0.146	(±0.070)	0.180	(±0.020)	0.136	(±0.041)
h_m^2	DATA 1	0.110	(±0.021)	0.057	(±0.073)	0.116	(±0.039)	0.027	(±0.014)
	DATA 2	0.106	(±0.027)	0.033	(±0.017)	0.087	(±0.037)	0.027	(±0.014)
	DATA 3	0.035	(±0.014)	0.036	(±0.014)	0.089	(±-0.000)	0.024	(±0.006)
d_{am}	DATA 1	-0.085	(±0.027)	-0.054	(±-0.130)	-0.082	(±0.068)	-0.001	(±0.003)
	DATA 2	-0.080	(±0.027)	-0.001	(±0.003)	0.002	(±0.006)	-0.001	(±0.003)
	DATA 3	0.000		-0.035	(±0.029)	-0.050	(±0.006)	-0.003	(±0.005)
h_T^2	DATA 1	0.189	(±0.033)	0.189	(±0.042)	0.242	(±0.062)	0.211	(±0.039)
	DATA 2	0.062	(±0.019)	0.193	(±0.042)	0.249	(±0.063)	0.211	(±0.039)
	DATA 3	0.062	(±0.019)	0.112	(±0.032)	0.149	(±0.049)	0.144	(±0.033)

degrees to transform fertility and hatchability observations. On the other hand, some earlier studies conducted by Szwaczkowski and Piotrowski (1998) indicated unsatisfactory approximation of the empirical distribution (PFE and PHC) to normality. Hence, the probit transformation recommended for discrete data has also been checked.

In general, negligible differences between direct heritability estimates of PFE have been observed, although higher estimates (with the exception of strain H77) are obtained from DATA-2. Analogous relationships have been noted for h^2_m and d_{am} estimates. More pronounced differences between DATA-3 and other data sets have been obtained for total heritability estimates in two strains (H77, N88).

In all four strains, heritability estimates of PHC obtained from DATA-1 were larger than the transformed data sets. However, in two strains (H77, S22) the differences between DATA-1 and DATA-2 were small. Similar tendencies have been registered for other estimated functions of (co)variance components including total heritability.

The number of investigations on the effects of non-normality of distribution have been carried out on egg production traits (e.g., Besbes et al., 1993; Szwaczkowski et al., 1994; Koerhuis, 1996). Ibe and Hill (1988) pointed out how transformation of egg production could increase the efficiency of selection through a higher heritability of the transformed data. Unfortunately, the obtained heritabilities do not lead to univocal implications, except some estimates from the probit transformation. Therefore, particular (co)variance components have also been monitored (see Figures 1 and 2). As already mentioned, to compare the magnitudes of some estimates, each observation (within data set) was divided by the average. Higher direct genetic variance components are usually estimated from untransformed data rather than from the transformed ones. However for strain R33 the highest estimates $\hat{\sigma}^2_a$ have been obtained from the probit transformed data. Basically, similar relationships have been obtained for $\hat{\sigma}^2_m$ and $\hat{\sigma}_{am}$. Generally, estimates of maternal genetic variances are more strongly influenced by strains than by transformation approaches. Covariances between direct and maternal effects are also determined by transformation. These covariance estimates for PFE are positive whereas the dependences for PHC are negative.

It seems that transformation can improve the statistical properties of a trait depending on its original distribution. As presented in Table 1 the distributions of two traits over four strains are considerably different. Skewness coefficients of PFE are higher compared to those of PHC. By the way, it should be noted that PHC distributions for R33 and S22 are more symmetrical than the other ones.

What is data best statistical properties? Two criteria of goodness of model (residual variance estimator and magnitude of logarithm likelihood) have been employed (see Table 4). Generally, smaller residual variances (for both traits of all strains) have been obtained from Bliss degree transformed data. It should be stres-

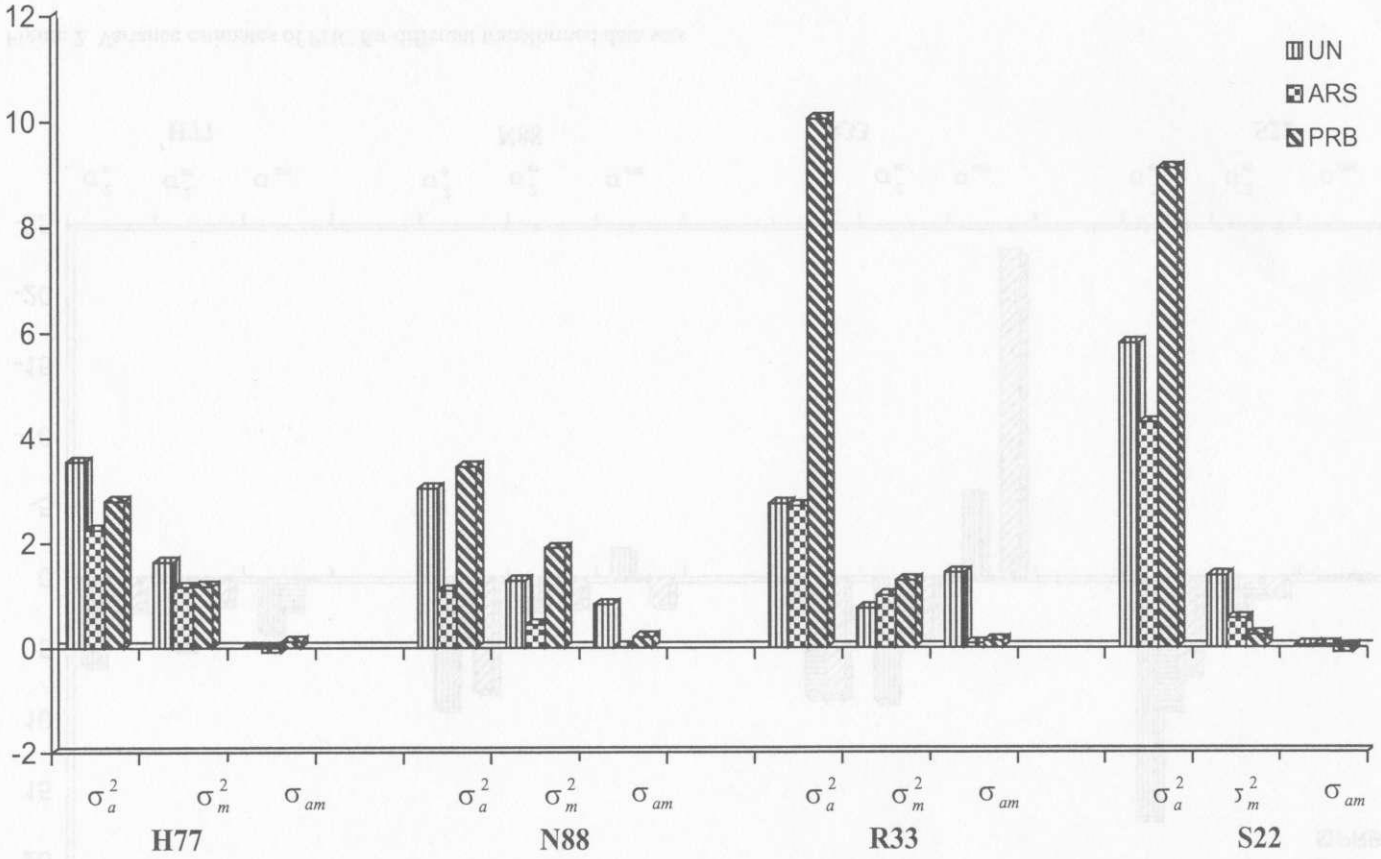

Figure 1. Variance estimates of PFE for different transformed data sets

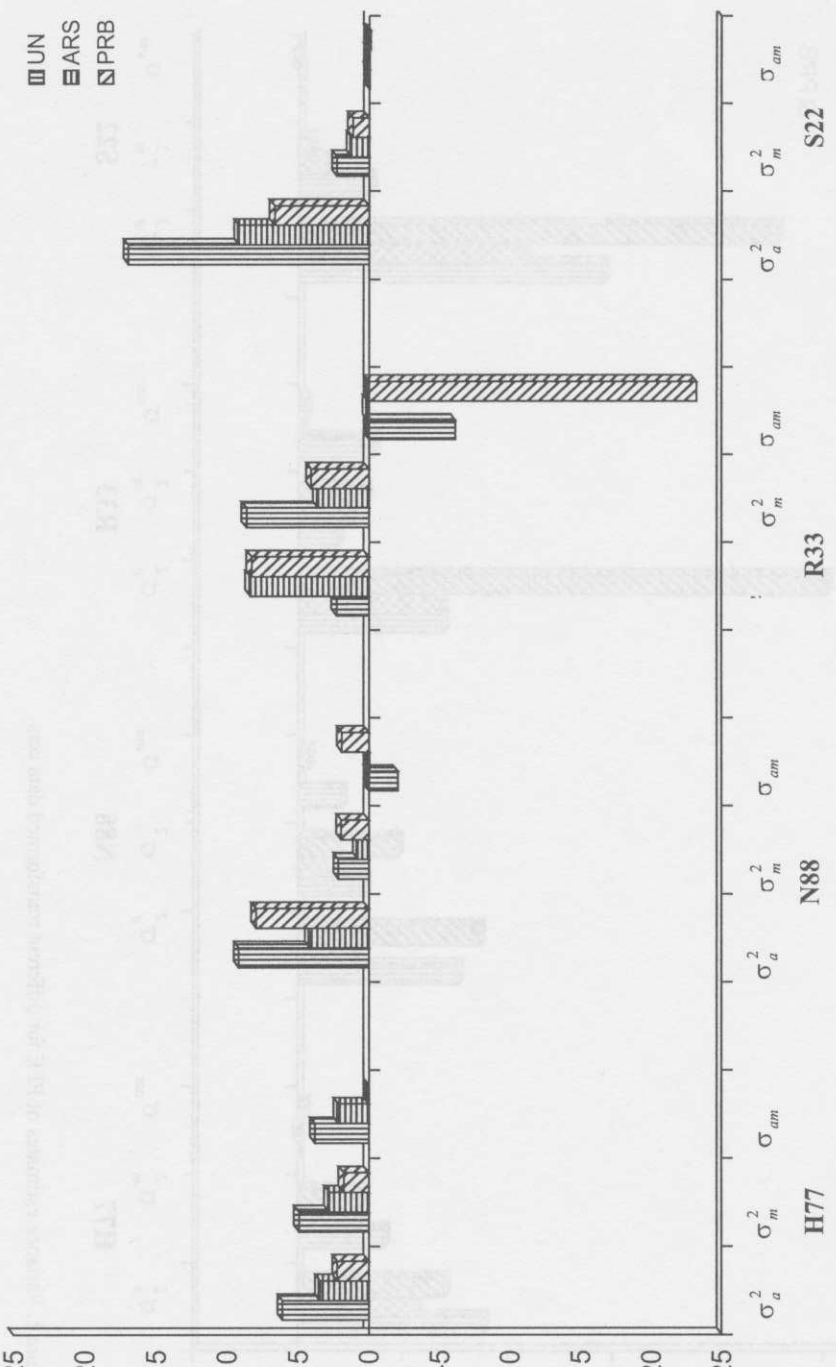

Figure 2. Variance estimates of PHC for different transformed data sets

TABLE 4

Estimates of residual variances ($\hat{\sigma}^2_e$) and logarithm likelihoods (LogL) obtained for different transformed data sets

Trait		Strain							
	Data set	H77		N88		R33		S22	
		$\hat{\sigma}^2_e$	LogL	$\hat{\sigma}^2_e$	LogL	$\hat{\sigma}^2_e$	LogL	$\hat{\sigma}^2_e$	LogL
PFE									
	DATA-1	0.0329	1982.56	0.0333	1230.97	0.0348	414.72	0.0453	327.02
	DATA-2	0.0226	3024.81	0.0121	3008.99	0.0230	814.45	0.0299	1236.45
	DATA-3	0.0961	-574.26	0.0957	-479.76	0.0951	-596.64	0.0724	-657.79
PHC									
	DATA-1	0.0377	1500.00	0.0286	1262.13	0.0489	-188.18	0.0661	-751.27
	DATA-2	0.0224	2910.45	0.0184	2095.28	0.0293	397.45	0.0359	566.17
	DATA	0.0462	1168.98	0.0468	582.76	0.0360	252.33	0.0413	405.16

sed that largest error variances were estimated from DATA-3. For the same data set (especially, PFE of H77 and N88) from DATA-2, were 3-4 times smaller than those for DATA-3. However, for PHC of two strains (R33, S22) error variance estimates from untransformed data are higher compared to the probit transformed ones. According to this criterion it may be concluded that the arcsin transformation leads to the most satisfactory estimates of genetic parameters. Generally, it corresponds with the second of the criteria used (logarithm likelihood). Of all the data sets, the largest logarithms of likelihoods have been obtained for Bliss degree transformed data, whereas for PFE medium values have been received from untransformed data.

Differences between DATA-1 and DATA-3 for PHC are strongly influenced by strains, and in consequences it may also be determined by trait distribution within strain, starting values and convergence criteria which must be specified for each given data file. In spite of all, variability of heritability estimates is not very large.

From the theoretical point of view the arcsin squared transformation can be recommended for the reproductive layer traits. On the other hand, differences between strains in heritability estimates have been registered. Hence, it seems that other transformation approaches may also be considered depending on the trait distribution.

ACKNOWLEDGEMENTS

The authors would like to thank Karin Meyer for the use of her package programs DFREML.

REFERENCES

Besbes B., Ducrocq V., Foulley J.L. Protais M., Tavernier A., Tixier-Boichard M., Beaumont C., 1993. Box-Cox transformation of egg production traits of laying hens to improve genetic parameter estimation and breeding evaluation. Livest. Prod. Sci. 33, 313-326

Chaudary M.L., Sandhu S.J., Brah G.S., 1987. Genetic, phenotypic and environmental relationships of fertility and hatchability with other economic traits in White Leghorns. J. Anim. Breed. Genet. 104, 169-174

Foerster A., 1993. Züchterische Möglichkeiten einer Verbesserung der Schlufrate in reinzuchtlinien eines Zuchtprogrammes für braune Legehybriden. Institute für Tierzucht und Tierhaltung, der Christian-Albrechts Universität, Kiel (Germany)

Gianola D., Foulley J.L., 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15, 201-224

Hagger C., Hofer A., 1989. Correlations between breeding values of dairy sires for frequency of distocia evaluated by a linear and a nonlinear method. J. Anim. Sci. 67, Suppl. 1, 88 (Abstr. 210)

Hartmann C., 2001. Selection for yolk production in laying hens. Acta Univ. Agr. Sueciae, Agraria 298, Swedish University of Agricultural Sciences, Uppsala (Sweden)

Ibe S.N., Hill W.G., 1988. Transformation of poultry egg production data to improve normality, homoscedasticity and linearity of genotypic regression. J. Anim. Breed. Genet. 105, 231-240

Koerhuis A.N.M., 1996. Non-normality of egg production distributions in poultry and the effects of outlier elimination and transformation on size and curvilinearity of heritability. Livest. Prod. Sci. 45, 69-85

Lynch M., Walsh B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc. Publishers Sunderland, Massachusetts (USA)

Meyer K., 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using derivative-free algorithm. Genet. Sel. Evol. 21, 317-340

Meyer K., 1993. Programs to estimate variance components for individual animal models by restricted maximum likelihood (REML). User notes. Institute of Animal Science, Armidale (Australia)

Sewalem A., 1998. Genetic study of reproduction traits and their relationship to production traits in White Leghorn lines. Acta Univ. Agr. Sueciae, Agraria 126, Swedish University of Agricultural Sciences, Uppsala (Sweden)

Szwaczkowski T., Piotrowski P., 1998. On transformation of the hatchability traits in laying fowl. XVIII-th Genetic Days, Ceske Budejovice (Czech Republic). Czech J. Anim. Sci. 43, 428 (Abstr.)

Szwaczkowski T., Wężyk S., Czeluśniak H., 1994. Transformations of performance trait data in laying hens. Roczn. Nauk. Zoot. 21, 25-32

Szwaczkowski T., Wężyk S., Piotrowski P., Cywa-Benko K., 2000. Direct and maternal genetic and environmental effects for fertility and hatchability in laying hens. Arch. Geflugelk. 64, 115-120

Varona L., Misztal I., Bertrand J.K., 1999. Threshold-linear versus linear-linear analysis of birth weight and calving ease using an animal model: II. Comparison of models. J. Anim. Sci. 77, 2003-2007

STRESZCZENIE

Wpływ transformacji danych na wielkość oszacowań odziedziczalności cech reprodukcyjnych kur nieśnych

Badaniami objęto dwie cechy reprodukcyjne (procent zapłodnienia jaj oraz procent wylęgu z jaj nałożonych) 15339 kur nieśnych czterech rodów z fermy zarodowej. Osobniki pochodziły z dziewięciu pokoleń. Obliczenia przeprowadzono w obrębie każdego rodu, dla trzech typów danych: dane nietransformowane, transformowane według stopni Blissa oraz z zastosowaniem transformacji probitowej. Przed przeprowadzeniem obliczeń każdą wartość cechy podzielono przez średnią, co umożliwiło porównanie uzyskanych estymatorów. Oszacowano wariancję genetyczną addytywną bezpośrednią i matczyną oraz kowariancję między efektami genetycznymi bezpośrednimi i matczynymi. Przyjęto dwa kryteria adekwatności modelu: estymator wariancji błędu i logarytm wiarogodności. Zgodnie z oczekiwaniami uzyskano niskie oceny odziedziczalności obydwóch cech. Nie stwierdzono znaczących różnic między estymatorami odziedziczalności bezpośredniej (definiowanej jako iloraz wariancji genetycznej addytywnej bezpośredniej i wariancji fenotypowej) dla procentu jaj zapłodnionych. Oszacowania uzyskane dla danych nietransformowanych były jednak wyższe niż dla transformowanych. Generalnie, najniższe estymatory wariancji błędu otrzymano dla transformacji według stopni Blissa. W większości przypadków najwyższe oceny wariancji błędu stwierdzono dla danych przekształcanych probitowo, co świadczy o jego najgorszym dopasowaniu. Należy podkreślić, że do podobnych wniosków prowadzi także ocena dokonana na podstawie drugiego kryterium.