http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Pakistan Journal of Biological Sciences 10 (9): 1556-1558, 2007
ISSN 1028-8880

Heterogeneity of Variance for Milk Traits at Climitical Regions in Holstein Dairy Cattle in Iran and the Best Method(s) for Data Transformation

Sheyda Varkoohi, Hassan Mehrabani-Yeganeh,
Seyed Reza Miraei-Ashtiani and Navid Ghavi Hossein-Zadeh
Department of Animal Science, Faculty of Agriculture,
University of Tehran, Karaj, Iran

Abstract: In this research, heterogeneity of variance components in cattle populations at climatical regions in the first three lactations was studied. The data set included the following: 161328 records of first, 123369 records of second and 81013 records of third lactations, which was collected by Animal Breeding Center of Iran from 1983 to 2004. Records of three lactations were divided in the base of Domarten method. Bartlett test for heterogeneity of variance components was significant among all subgroups. In order to decrease the heterogeneity of variance components, we used several data transformation methods including Logarithmic, Square root and Arc sin transformations. Logarithmic transformation decreased the heterogeneity of variance components in the three lactations and other methods had not effect for removing the heterogeneity in any group. Genetic parameters and heritability were estimated for three lactations by MATVEC program, using animal model. Results showed that the heritability estimates of milk yield were decreased from the first lactation to the third; also the heritability estimates of transformed data were slightly higher than the original data. Comparison between estimated parameters in single trait and two traits analysis, before and after data transformation, showed that there were not significant differences between derived results.

Key words: Variance components, variance heterogeneity, data transformation, climatical regions, Holstein cattle

INTRODUCTION

The accuracy of genetic evaluation and estimation of (co) variance components depends on how well the assumptions match the data. One of the important assumptions is that of homogeneity of variances of random factor (s) across all possible strata of data. The problem of heterogeneous variances in genetic evaluation of dairy cattle is that, above average animals may be over evaluated in more variable herds and as a result, a greater proportion of animals would be chosen from these herds. Also, heterogeneity of variances can bring about bias in estimations, decrease in accuracy, lowered selection response and changes in ranking.

Heterogeneity of genetic, permanent environmental and residual variance components have been reported by Everett *et al*. (1983), Boldman and Freeman (1990) and Gengler and Wiggans (2001). Nikolaou *et al*. (2004) have introduced output production levels, herd size, birth rate, herd management and climatical condition as the sources of heterogeneity.

To remove the heterogeneity of variance, different data transformation methods have been suggested. If the coefficient of variation is equal in all treatments and there is a relationship with mean, there won't need to transform data. If the ratio of standard deviation to mean is constant, we use the logarithmic transformation and if the ratio of standard deviation to the root of mean is constant, we use the square root transformation and if data are in the form of percentage, we use the Arc sin transformation (Sancristobal *et al*., 1993; Freund and Wilson *et al*., 1993). The objectives of this research were to study the heterogeneity of variance for milk traits at climitical regions and to introduce the best method for removing the heterogeneity of variance components.

MATERIALS AND METHODS

Data classification: In this study, data consisted of 161328 milk production records of first lactation, 123369 milk production records of the second lactation and 81013 milk production records of third lactation, which were collected during 1983-2004 by the Iranian Animal Breeding Center to estimate genetic parameters at three climatical regions. The data were divided based on different climatical regions using advanced Domarten method.

Corresponding Author: Dr. Seyed Reza Miraei-Ashtiani, Department of Animal Science, Faculty of Agriculture, University of Tehran, Karaj, Iran Tel: +98-261-2248082 Fax: +98-261-2246752

(Provinces with the coefficient of dryness of less than 10 placed at dry desert region, provinces with the coefficient of dryness from 10 to 24 placed at Mediterranean region and provinces with the coefficient of dryness of more than 24 placed at humid region).

Bartlett test: Bartlett test performed as a pre-test for studying the heterogeneity of variances between above groups before and after the data transformation by SAS8. 12 program.

Data transformation: For removing the heterogeneity of variance between mentioned groups, we used three methods for data transformation of adjusted milk production (adjusted for 305 day lactation and two times milking per day) including: Logarithmic, Square Root and Arc sin transformations.

Data analysis: Data were analyzed using MATVEC program and animal model.

Univariate analysis: The model equation was:

$$Y_{ijkl} = \mu+HYS_i+Age_j+a_k+e_{ijkl}$$

Which, Y_{ijkl} is the milk yield from cow k in HYS class I, μ is the overall mean, HYS_i is the effect of Herd-Year-Season, Age_j is the effect of the age class j, a_k is the breeding value of animal k and e_{ijkl} is the random residual effect.

Multivariate analysis: Due to a large volume of data in three lactations, three two-trait analyses, i.e. among the three lactations were performed. The model equation was:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} z_1 & 0 \\ 0 & z_2 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Where, y_1 and y_2 are records related to the milk yield, x_1 and x_2 are design matrices of fixed effects, z_1 and z_2 are design matrices of random effects, b_1 and b_2 are vectors of fixed effects, a_1 and a_2 are vectors of random effects and e_1 and e_2 are vectors of residual effects.

RESULTS AND DISCUSSION

Bartlett test before and after data transformation: The result of Bartlett test, pre and post transformation, are shown in Table 1-4. Heterogeneity tests were significant among all groups ($p<0.0001$) before data transformation, indicating the heterogeneity of variance components between climatical regions. Comparison between Chi-Square values before and after data transformation showed that the logarithmic transformation decreased the heterogeneity of variance components at climatical regions (especially in the third lactation), but other methods of data transformation had not only no effect on removing the heterogeneity, rather increased it in some cases.

Table 1: Bartlett test before data transformation

Production level	df	χ^2-value	p-value
1st lactation	2	303.1	0.0001
2nd lactation	2	393.7	0.0001
3rd lactation	2	350.1	0.0001

Table 2: Bartlett test after logarithmic transformation

Production level	df	χ^2-value	p-value
1st lactation	2	112.2	0.0001
2nd lactation	2	29.5	0.0001
3rd lactation	2	3.5	0.1735

Table 3: Bartlett test after Square root transformation

Production level	df	χ^2-value	p-value
1st lactation	2	109.9	0.0001
2nd lactation	2	89.9	0.0001
3rd lactation	2	73.1	0.0001

Table 4: Bartlett test after Arc sin transformation

Production level	df	χ^2-value	p-value
1st lactation	2	422.6	0.0001
2nd lactation	2	554.5	0.0001
3rd lactation	2	368.8	0.0001

Data analysis

Univariate analysis: The results of univariate analysis pre and post Logarithmic transformation are shown in Table 5 and 6.

All variance components in three lactations for dry desert region were the highest and for the humid region were the lowest. Climatical conditions of humid regions are responsible for parasitic diseases and decrease in the average production of animals. There is an environmental variation at dry desert region and additive variance is higher at this region.

Results showed that there were not significant differences between before and after logarithmic transformation estimates of heritabilities of the first three lactation. Heritabilities increased slightly after logarithmic transformation and the same results observed by Hill (1983) and Kominakis and Rogdakis (1998).

Multivariate analysis: Due to lower volume of data at the Mediterranean region, we performed arbitrarily multivariate analysis on the Mediterranean data.

Results showed that, estimates of variance components and heritabilities are more reliable than

Table 5: Estimated variance components and heritabilities before data transformation

	Climatical region	σ_a^2	σ_e^2	σ_p^2	h^2±SE
1st lactation	Mediterranean	288202	870901	1159103	0.249±0.014
	Dry desert	319985	941911	1261896	0.254±0.013
	Humid	170257	598262	768519	0.222±0.011
2nd lactation	Mediterranean	290156	1254372	1544528	0.188±0.015
	Dry desert	318121	1335550	1653671	0.192±0.009
	Humid	175694	945908	1121602	0.157±0.012
3rd lactation	Mediterranean	288242	1643140	1931382	0.149±0.013
	Dry desert	305286	1650409	1955695	0.156±0.015
	Humid	277134	1641931	1919065	0.144±0.017

Table 6: Estimated variance components and heritabilities after logarithmic transformation

	Climatical region	σ_a^2	σ_e^2	σ_p^2	h^2±SE
1st lactation	Mediterranean	0.00741521	0.0221034	0.02951868	0.251±0.012
	Dry desert	0.00715174	0.0208579	0.02800973	0.255±0.014
	Humid	0.00391036	0.0135669	0.01747734	0.224±0.013
2nd lactation	Mediterranean	0.00746550	0.0318359	0.03930147	0.190±0.015
	Dry desert	0.00711007	0.0295748	0.03668493	0.194±0.016
	Humid	0.00403523	0.0214506	0.02548589	0.158±0.013
3rd lactation	Mediterranean	0.00741625	0.0417029	0.04911915	0.151±0.011
	Dry desert	0.00682321	0.0365472	0.0433704	0.157±0.014
	Humid	0.00636505	0.0372346	0.04359965	0.146±0.016

Table 7: Estimated variance components and heritabilities in Mediterranean region

	Climatical region	σ_a^2	σ_e^2	σ_p^2	h^2±SE
1st lactation	Multivariate	319985	941911	1261896	0.253±0.008
	Univariate	321536	940956	1262447	0.255±0.010
2nd lactation	Multivariate	318121	1335550	1653671	0.192±0.012
	Univariate	319590	1320049	1639639	0.195±0.013
3rd lactation	Multivariate	305286	1650409	1955695	0.156±0.011
	Univariate	306172	1646837	1953009	0.157±0.015

univariate analyses due to use more information in the multivariate analyses (Table 7). Heritability estimates decreased from the first to third lactation due to decreases in additive variance. This is consistent with the results of Dahlin *et al*. (1998).

The results showed that only the logarithmic transformation had significant effect on removing heterogeneity of variance components, but did not completely remove heterogeneity. The transformed data resulted in slightly higher values of heritability, which decreased from first to third lactation. Also, not many differences observed between univariate and multivariate analyses.

REFERENCES

Boldman, K.G. and A.E. Freeman, 1990. Adjustment for heterogeneity of variances by production level in dairy cow and sire evaluation. J. Dairy Sci., 73: 503-508.

Dahlin, A., U.N. Khan, A.H. Zafar, M. Saleem, M.A. Chaudhry and J. Philipsson, 1998. Genetic and environmental causes of variation in milk production traits of Sahiwal cattle in Pakistan. Anim. Sci., 66: 307-318.

Everett, R.W., J.F. Keown and J.F. Taylor, 1982. The problem of heterogeneous within herd variances when identifying elite cows. J. Dairy Sci. (Supple. 1): 100 (Abstr).

Freund, R.J. and W.J. Wilson, 1993. Statistical Methods. Academic Press, San Diago, CA.

Gengler, N. and G.R. Wiggans, 2001. Variance of effects of lactation stage within herd by herd yield. J. Dairy. Sci., 84 (Suppl. 1), 216 (Abstr. 869).

Hill, W.G., M.R. Edwards and M.K.A. Ahmed, 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod., 36: 59-67.

Kominakis, A., E. Rogdakis and K. Koutsotolis, 1998. Genetic parameters for milk yield and litter size in Boutsiko dairy sheep. Can. J. Anim. Sci., 78: 525-532.

Nikolaou, M., A.P. Kominakis, E. Rogdakis and S. Zampitis, 2004. Effect of mean and variance heterogeneity on genetic evaluation of Lesbos dairy sheep. Livest. Prod. Sci., 88: 107-115.

Sancristobal, M., S.L. Foulley and E. Manfredi, 1993. In: Ference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. Genet. Sel. Evol., 25: 3-12.