

Estimation of Genetic Variance Components of Body Size Measurements in Hanwoo (Korean Cattle) Using a Multivariate Linear Model

Jungjae Lee and Naesoo Kim*

Department of Animal Science, Chungbuk National University, Korea

ABSTRACT

The objectives of this study were to quantify the combination values of the principal components and factors calculated using body measurements of Hanwoo (Korean Cattle) and estimate their heritabilities. The technique of multivariate analysis was used to reduce a large number of variables to a smaller number of new variables and characterize cattle according to body shape. The analyses were performed using 1,979 cattle at 12 months of age and 936 cattle at 24 months of age. The data for the analyses was obtained from progeny tests performed on Korean Cattle for 6 years from 2003 to 2008. The phenotypic correlations among these traits were estimated to range from 0.32 to 0.90 at 12 months of age and from 0.21 to 0.82 at 24 months of age. The first principal components (PC1s) indicated a weighed average of overall body measurements, accounting for 99.91% of the total variation for both periods of test. The two first PCs had positive coefficients for all body measurements. The major sources of PC, such as chest girth (CG), body length (BL), rump height (RH), and wither height (WH) were similar for both test periods. The heritabilities for PC1, the first factor score (FS1), and the second factor score (FS2) were estimated by multivariate REML method. The estimated heritabilities for PC1, FS1, and FS2 were 0.33, 0.38, and 0.40, respectively, at 12 months of age and 0.26, 0.76, and 0.58 at 24 months of age. Further studies are needed to determine whether the heritabilities of FS1 and FS2 at 24 months of age were overestimated.

(Key words : Hanwoo, Principal component analysis, Factor analysis, Body measurements, multivariate REML)

INTRODUCTION

Korean cattle are called Hanwoo (*Han* means Korean and *woo* means cattle) in Korean. Korean cattle originated from crossbreds of *Bos indicus* and *Bos primigenius*. Traditionally, Koreans raised these cattle as working animals. The rapid growth in the Korean economy after 1970 brought about the mechanization of agricultural methods and led to increasing domestic consumption of beef cattle. As a result, these cattle became more important as sources of beef rather than labor. It is necessary to improve the body measurements of these cattle to enhance their meat production abilities. Traits associated with body size have been used in the selection program in Korea. Therefore, it is very important to estimate the genetic parameters of body measurements and to predict genetic improvement.

Until now, considerable attention has been focused on the genetic parameters for each trait in beef cattle (Veselá et al., 2005) and on the relation between these parameters and meat

production ability (Meyer et al., 1987; Veerkamp and Brotherstone, 1997; Royal et al., 2002), and longevity (Essl, 1998; Setatti et al., 2004). Therefore, the traits of shape and size are considered to be good indirect predictors of production and longevity.

Traditionally, genetic parameters were estimated for each trait directly. However, body size measurements are composed of more than 10 traits that are highly phenotypically correlated. It is not possible to make the assumption that these traits are all independent. Therefore, it is necessary to use multivariate methods (principal component analysis and factor analysis) to remove the collinearity among the variables, so as to produce more efficient and meaningful results than the direct approach in terms of animal breeding.

Multivariate statistical methods have been successfully applied to biology, computer science, sociology, animal breeding and genetics. In animal breeding research, these methods remove redundant information from correlated variables and yield a smaller number of latent variables,

* Corresponding author : Dr. N. S. Kim, Professor Division of Animal Science College of Agriculture, Life & Environment Science Chungbuk National University #12 Gaeshin-dong Cheongju 361-763, Korea. Tel: 82-43-261-2548, Fax: 82-43-273-2240, Email: rvaukim@chungbuk.ac.kr

called ‘principal components’ and ‘factors’, that account for most of the variance in the body measurements. The scores of principal components and factors that are estimated as a linear combination of optimally weighted observed traits can then be used for estimating the genetic parameters. The objective of this study was to identify the combination value (latent variables) of principal components and factors calculated using body measurements of Hanwoo and to estimate their genetic variance components using multivariate REML for the animal model.

MATERIALS AND METHODS

1. Data

Body measurement data of the Hanwoo were obtained from the Livestock Improvement Main Center of NACF (National Agricultural Cooperative Federation). Ten body measurements were recorded for 1,979 yearling cattle (12 months of age) and 936 cattle at the end of the test period (24 months of age) from 2003 to 2008. Each record included information on the age of the yearling animal and at the end of the test period, as well as on the herd and pedigree. The body measurements analyzed were wither height (WH), rump height (RH), body length (BL), chest depth (CD), chest width (CW), rump length (RL), hip width (HW), thurl width (TW), pin bone width (PW), and chest girth (CG).

2. Statistical analyses and modeling

(1) Multivariate method

Principal Component Analysis (PCA) is a statistical technique that linearly transforms a given set of variables into a new set of composite variables. These new variables are orthogonal to each other and capture most of the information in the original variables. A small set of uncorrelated variables is much easier to understand and use in further analysis than a larger set of correlated variables. For these reasons, principal components are often used in exploratory data analysis and data reduction. The concept was originally conceived by Pearson (1991) and independently developed by Hotelling (1993).

The principal components, Z_1, Z_2, \dots, Z_p , are particular linear combinations of the variables, X_1, X_2, \dots, X_p with n measurements. After PCA is performed, Z_1, Z_2, \dots, Z_p will be uncorrelated and can be expressed as:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{22} \\ \vdots \\ a_{p1} \end{bmatrix} Z_1 + \dots + \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{pi} \end{bmatrix} Z_i + \dots + \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{bmatrix} Z_p$$

and $\text{Cov}(Z_i, Z_j) = 0$ for all $i \neq j$ which indicates that all the geometrical variation modes are independent of each other. PCA employs a key idea of matrix algebra: a $p \times p$ symmetrical, nonsingular matrix, such as the variance matrix Σ , may be reduced to a diagonal matrix D through premultiplication and premultiplication a particular orthonormal matrix A such that

$$A \Sigma A' = D$$

The diagonal elements of D , $\lambda_1, \lambda_2, \dots, \lambda_{10}$ are called the eigenvalues of Σ . The columns of A , a_1, a_2, \dots, a_p are called the eigenvectors of Σ . The eigenvalues may be obtained from the following determinant equation

$$|\lambda I - \Sigma| = 0$$

where I is the identity matrix. This produces a p^{th} degree polynomial in λ from which the values $\lambda_1, \lambda_2, \dots, \lambda_{10}$ are obtained.

Therefore:

$$\text{Var}(z) = \begin{bmatrix} \text{Var}(Z_1) & 0 & 0 & 0 \\ 0 & \text{Var}(Z_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \text{Var}(Z_p) \end{bmatrix} = A \Sigma A' = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_p \end{bmatrix}$$

Clearly, $\lambda_i = \text{Var}(Z_i)$, since λ is obtained by letting $\det(\lambda I - \Sigma) = 0$, so $\lambda_1 + \dots + \lambda_p = \text{Var}(X_1) + \dots + \text{Var}(X_p) = \sum_{i=1}^p \sigma_i^2$. It shows clearly that the percentage of total variance of the original variables for each principal component can be explained by its corresponding eigenvalue.

The PRINCOMP and FACTOR procedures of the SAS statistical program package (SAS STAT 9.1 User’s Guide, 2004) were used to estimate the principal components and factors. The main objective of the 2 multivariate approaches (Principal component and factor analyses) was to reduce the number of body measurements traits to estimate heritabilities for both test periods without loss of original information for the 10 body measurements. One of the 10 principal components and 2 of the 10 factors were retained to calculate the scores of principal component and factors for both test periods using the eigenvalue criterion (Cattell, 1978).

Under this criterion, each retained principal component and factors account for most of the total variation in 10 body measurements.

The scoring coefficients of the factors were estimated by the regression method. The principal components and factors were computed from the phenotypic covariance matrices among the 10 body measurements. The original variation in the body measurements can be preserved by using covariance matrices instead of correlation matrices. The factors were rotated using the orthogonal rotation method; to simplify the matrix structure and enhance interpretability, it was assumed that the factors are independent. The principal components and factors retained were used to calculate the scores of principal components (PCs) and factors (FSs) and estimate their genetic parameters.

(2) Statistical Model

To identify the fixed and covariate effects to be included in the model, generalized linear modeling was performed using SAS (SAS STAT 9.1 User’s Guide, 2004). The fixed and covariate effects included were classification date (jointly evaluated animals), age of dam, region of breeding, and quadratic regression of age at evaluation. A summary of the fixed and covariate effects included in the models is shown in Table 1. The classification date was highly significant for both test periods. The non-significance of age of dam was shown for all the variables except the first principal component (PC1) at 12 months of age. The variance components were estimated simultaneously for PC1, factor1 (FS1) and factor2 (FS2) using multivariate REML for an animal model (Meyer, 1986).

$$Y_{ijkl} = \mu + YS_i + Mage_j + RB_k + a \times age^2 + A_l + \varepsilon_{ijkl}$$

where:

Y_{ijkl} = principal component and factors scores

- μ = overall mean
- YS_i = fixed effect of the *i*th classification date (Year-Season)
- $Mage_j$ = fixed effect of the *j*th age of dam (year)
- RB_k = fixed effect of the *k*th region of breeding
- a = quadratic regression coefficients for age of animal (month)
- A_l = random additive genetic effect for the *l*th animal
- ε_{ijkl} = random error

RESULTS

Table 2 lists the overall means, phenotypic standard deviation, and minimum and maximum values of body measurements for both test periods. Table 3 lists the phenotypic correlations among 10 body measurement traits at 12 months of age (above the diagonal) and 24 months of age (below the diagonal). The highest phenotypic correlations were determined between the same traits of WH and RH ($r_p = 0.90$ and 0.82 , respectively) for both test periods. The lowest phenotypic correlations were calculated between the traits of WH and PW ($r_p = 0.32$) at 12 months of age and between WH and CW ($r_p = 0.21$) at the end of the test periods (i.e., 24 months of age). On average, the 10 body measurement traits showed similar and strong positive phenotypic correlations for both test periods; this result was in agreement with that reported by studies from Veselá et al. (2005) and higher than Norris et al. (2008).

The principal components and factors can be interpreted by taking into account the sign and magnitude of their coefficients. Figs. 1-3 (at 12 months of age) and Fig. 4-6 (at 24 months of age) show the linear combination diagram of PC1, FS1, and FS2. The coefficients of the traits were used to calculate the scores of principal components and factors and to describe the new variables (PC1, FS1, and FS2).

The estimated eigenvalues estimated by PCA and the

Table 1. Fixed effects and covariate term included in the analysis of score of principal component and factors at 12 and 24 months of age

Traits	12 months of age				24 months of age			
	YS	Mage	RB	age ²	YS	Mage	RB	age ²
PRIN1	+	+	+	-	+	-	+	+
Factor1	+	-	+	+	+	-	-	-
Factor2	+	-	+	+	+	-	-	-

+ : Included in model; - : Excluded in model
 YS : year-season; Mage : mother’s age ; RB : region of breeding.

Table 2. Simple statistics of body measurements at 12 months and 24 months of age in Hanwoo

Traits (cm)	Abbreviation	12 months of Age (Obs. = 1,979)				24 months of Age (Obs. = 936)			
		Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Wither Height	WH	120.39	3.99	101.00	135.00	136.38	4.41	121.00	154.00
Rump Height	RH	122.50	4.05	102.00	138.00	137.35	4.37	124.00	166.00
Body Length	BL	132.43	6.25	107.00	151.00	155.38	7.53	130.00	178.00
Chest Depth	CD	62.01	2.75	46.00	71.00	76.16	3.18	49.00	90.00
Chest Width	CW	37.97	3.76	20.00	51.00	51.25	3.76	34.00	68.00
Rump Length	RL	44.39	3.11	30.00	55.00	52.04	3.42	35.00	71.00
Hip Width	HW	36.93	2.97	25.00	49.00	49.15	3.61	34.00	61.00
Thurl Width	TW	38.88	2.95	21.00	50.00	47.19	3.49	34.00	61.00
Pinbone Width	PW	21.11	2.44	14.00	29.00	25.86	2.31	16.00	32.00
Chest Girth	CG	165.62	7.01	123.00	188.00	214.76	8.57	187.00	245.00

Table 3. Phenotypic correlations among 10 body measurement traits at 12 months of age (above the diagonal) and 24 months of age (below the diagonal) in Hanwoo

	WH	RH	BL	CD	CW	RL	HW	TW	PW	CG
WH		0.8999	0.5257	0.6041	0.3688	0.4822	0.4379	0.3945	0.3201	0.5000
RH	0.8150		0.5361	0.6012	0.3857	0.5113	0.4695	0.4177	0.3735	0.5107
BL	0.4432	0.4467		0.5209	0.4584	0.5584	0.5108	0.4330	0.3641	0.5307
CD	0.4347	0.4429	0.3683		0.5038	0.5888	0.5728	0.5124	0.3581	0.6291
CW	0.2112	0.2566	0.3093	0.4205		0.4865	0.5554	0.5805	0.4868	0.6306
RL	0.2838	0.3045	0.3391	0.3832	0.3435		0.5916	0.5558	0.4195	0.5254
HW	0.2618	0.3472	0.3482	0.4417	0.5002	0.5179		0.6796	0.4429	0.5258
TW	0.2964	0.3269	0.3166	0.3792	0.4482	0.4452	0.5460		0.5567	0.5721
PW	0.2545	0.2927	0.3520	0.3641	0.4105	0.3424	0.4698	0.4881		0.4950
CG	0.4120	0.3897	0.4715	0.5992	0.5559	0.3424	0.3770	0.3763	0.4211	

proportion of the total variation in the body measurements are shown in Table 4. The two (PC1s) for both test periods are a linear combination of optimally-weighted overall body measurements, accounting for 99.91% of the total variation for both test periods. This result means that 10 body measurements can be represented by PC1s for both test periods. Therefore, PC1s can be a good indicator for evaluating the total body shape characteristic of Hanwoo. The two PC1s had positive coefficients for all body measurements. The major sources of PC, such as CG, BL, RH, and WH, were similar for both test periods. The proportion of PC1s was estimated to be markedly higher than the proportions reported in previous studies. For instance, Vukasinovic et al. (1997) used 18 traits in Brown Swiss cattle and retained 5 principal components, accounting

for 58% of the total phenotypic variance. Roughsedge et al. (2000) used 23 linear type traits in British Holstein cattle and retained 7 principal components, accounting for 57.9% of the total phenotypic variance. Gutiérrez and Goyache (2002) used 10 linear type traits in Asturiana de los Valles beef cattle and retained 2 principal components, accounting for 49.74% of the total phenotypic variance.

Additional factor analysis was performed to investigate the type-characterization of Hanwoo. The FS1 at 12 months of age (Fig. 2) consists of high positive loadings on the traits related to height and length (WH, RH and BL) and high negative loadings on those related to width (CG, CW, TW, and HW). The FS2 at 12 months of age (Fig. 3) consists of high positive loadings on width-related traits (CG, CW, TW, and HW) and high negative loadings on the height- and

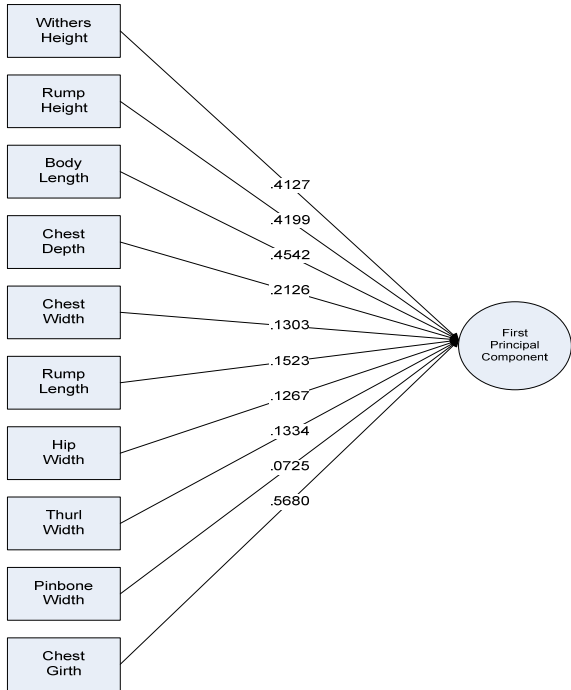


Fig. 1. Linear combination diagram of the first principal component for yearling animals (12 months of age).

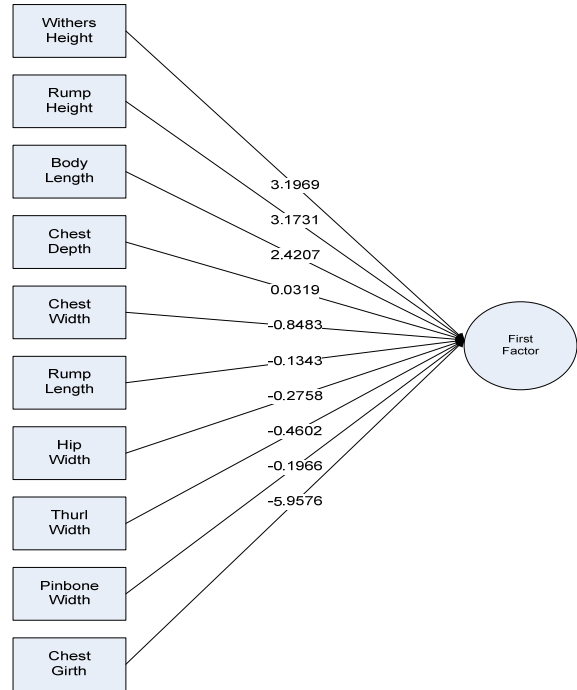


Fig. 2. Linear combination diagram of the first factor score for yearling animals (12 months of age).

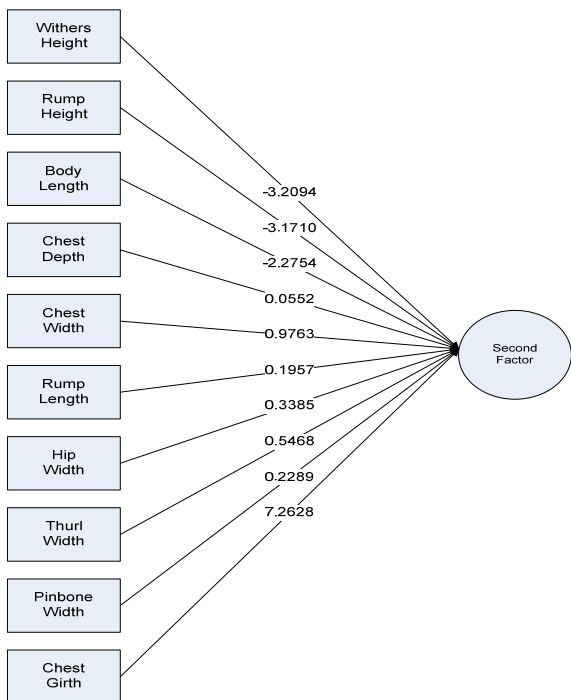


Fig. 3. Linear combination diagram of the second factor score for yearling animals (12 months of age).

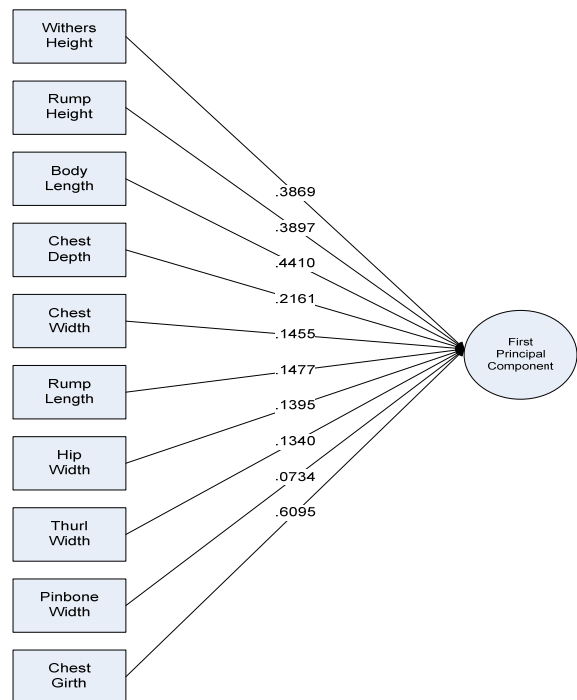


Fig. 4. Linear combination diagram of the first principal component at the end of the test period (24 months of age).

Table 4. Eigenvalues and proportion of the total phenotypic variance explained (in percentage) by principal component at 12 and 24 months of age

Component	12 months			24 months		
	Eigenvalue	Proportion of total variance	Cumulative proportion	Eigenvalue	Proportion of total variance	Cumulative proportion
1	85145.51	99.91	99.91	124293.37	99.91	99.91
2	23.86	0.03	99.94	35.96	0.03	99.94
3	18.88	0.02	99.96	27.75	0.02	99.96
4	11.51	0.01	99.98	19.39	0.02	99.97
5	5.83	0.01	99.98	7.80	0.01	99.98
6	4.44	0.01	99.99	6.43	0.01	99.99
7	3.81	0.00	99.99	5.59	0.00	99.99
8	2.75	0.00	100.00	4.91	0.00	99.99
9	2.52	0.00	100.00	3.39	0.00	100.00
10	1.55	0.00	100.00	3.25	0.00	100.00

length-related traits (WH, RH, and BL). Individual cattle with high positive loadings on FS1 (i.e., long and tall in stature) at 12 months of age would be better than those with high positive loadings on FS2 (large width) at the same period. The results obtained for FS1 and FS2 at 12 months of age

were opposite to each other. Furthermore, the coefficients of traits for FS1 and FS2 at 24 months of age (Figs. 5 and 6) were estimated to be opposite to the coefficients of traits at 12 months of age.

Table 5 provides descriptions of principal components and

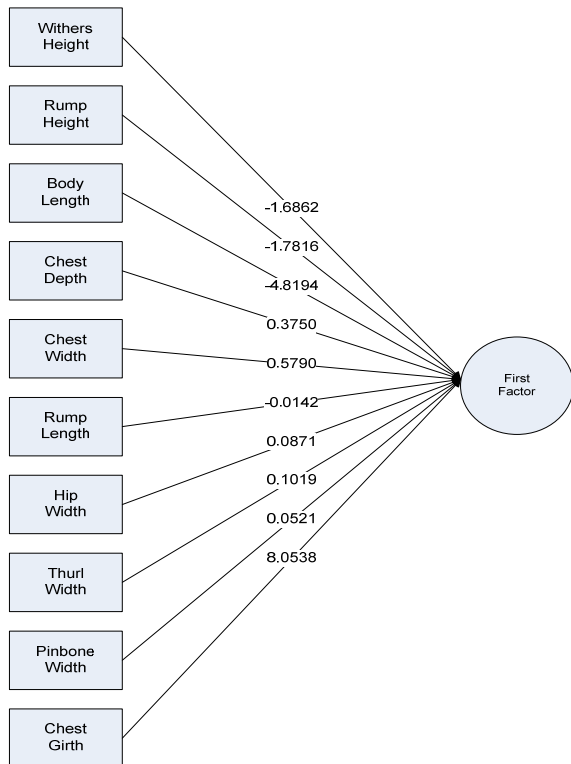


Fig. 5. Linear combination diagram of the first factor score at the end of the test period (24 months of age)

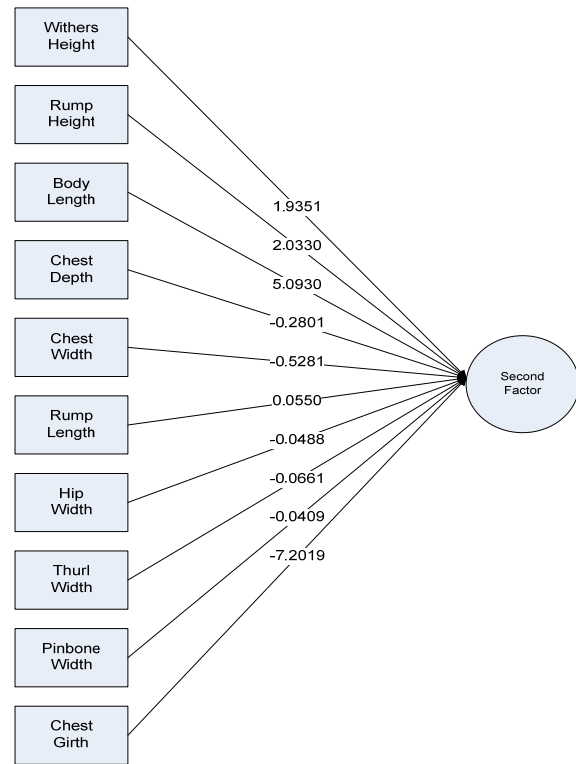


Fig. 6. Linear combination diagram of the second factor score at the end of the test period (24 months of age)

Table 5. Description of principal components and factors of body measurements and heritabilities of their scores at 12 and 24 months of age

Variables	12 months of age		24 months of age	
	Description	h^2	Description	h^2
Prin1	The weighted average of overall body measurements	0.33	The weighted average of overall body measurements	0.26
Factor1	The traits related to height and length	0.38	The traits related to width and wide chest girth	0.76
Factor2	The traits related to width and wide chest girth	0.40	The traits related to height and length	0.58

h^2 : coefficient of heritability.

factors of body measurements and the coefficients of heritability (h^2) of the PC and FS scores at 12 and 24 months of age. The highest h^2 values were calculated for FS1 at 24 months of age ($h^2=0.76$) and FS2 at 24 months of age ($h^2=0.58$). This means that a good selection response is expected from FS1 and FS2 at 24 months of age compared to the other variables, but these variables do not have very suitable statistical models (Table 1). In contrast, FS2 at 12 months of age and PC1 at 24 months of age had the lowest h^2 values ($h^2=0.26$). PC1, FS1, and FS2 at 12 months of age and PC1 at 24 months of age had moderate heritabilities which are in agreement with the results for body traits reported by Veselá et al. (2005) and Gutiérrez and Goyache (2002).

The heritability estimates of FS1 and FS2 at 24 months of age were higher than expected and were in agreement with those reported by Vági (1997) who studied Hungarian beef cattle and obtained estimates (0.70) for body length, (0.63) for chest depth, (0.71) for shoulder muscling and (0.76) for rump muscling. Vági used the same method (factor analysis) as estimate genetic parameters.

DISCUSSION

Body measurements of Hanwoo, especially CG, WH, RH, and BL, were considered as very important traits and the equation of PC1 can be used as an indicator for linear-type scoring rules. When setting up a breeding scheme, it is more useful to employ latent variables (PC1, FS1, and FS2) than employ each of the traits related to body size or shape. In particular, an over-all breeding scheme for the traits related to body size or shape can be set up using only one latent variable (PC1). For this reason, the proportions of PC1s accounted for more than 99% of the variation in the 10

body measurements. The estimated heritabilities of FS1 and FS2 at the end of the test period were considerably higher than expected because (i) the number of animals at the end of the test period was smaller than that at yearling age and (ii) suitable statistical models were not available; fixed effect were only classification date. Therefore, it is necessary to obtain more data of body measurements traits at 24 months of age and to conduct further studies so as to determine whether the heritabilities of FS1 and FS2 at 24 months of age were overestimated here. Further studies of multivariate methods (e.g., structural Equation Model and path analysis) will be needed to identify the causal relationships among the original and latent variables.

REFERENCES

- Cattell, R. B. 1978. *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Plenum Press, New York.
- Essl, A. 1998. Longevity in dairy cattle breeding: a review. *Livestock Prod. Sci.* 57:79-89.
- Gutiérrez, J. P. Goyache, F. 2002. Estimation of genetic parameters of type traits in Asturiana de los Valles beef cattle breed. *J. Anim. Breed. Genet.* 119:93-100.
- Hotelling, H. 1993. Analysis of complex of statistical variable into principal components. *J. Educ. Psychol.* 24(1):498-520.
- Meyer, K. 1986. Restricted maximum likelihood to estimate genetic parameters - in practice. In: *Proc. 3rd World Congr. Genet. Appl. Livest. Prod.*, Lincoln, NE, USA, X: 455-459.
- Meyer, K., Brotherstone, S., Hill, W. G. and Edwards, M. R. 1987. Inheritance of linear type traits in dairy cattle and correlations with milk production. *Anim. Prod.* 44:1-10.
- Norris, D., Selapa, N. W., Banga, C. B. and Ngambi, J. W. 2008. Estimation of (Co) Variance Components for Type Traits in Charolais Cattle. *J. Biol. Sci.* 8(1):229-232.

- Pearson, K. 1991. On line and planes of closet fit to systems of points in space. *Philosophical Magazine*, Sep. 6(2):559-572.
- Roughsedge, T., Brotherstone, S. and Visscher, P. M. 2000. Effect of cow families on type traits in dairy cattle. *Anim. Sci.* 60:391-398.
- Royal, M. D., Pryce, J. E., Woolliams J. A. and Flint, A. P. F. 2002. The genetic relationship between commencement of luteal activity and calving interval, body condition score, production, and linear type traits in Holstein Friesian dairy cattle. *J. Dairy Sci.* 85:3071-3080.
- SAS, 2004. SAS/STAT 9.1 User's Guide Volume 1. SAS Institute Inc., Cary, NC., USA.
- Setatti, M. M., Norris, D., Banga C. B. and Benyl, K. 2004. Relationships between longevity and linear type traits in Holstein cattle population of Southern Africa. *Trop. Anim. Health Prod.* 36:807-814.
- Vági, J. 1997. Utilisation of population genetic multivariate methods in the evaluation of results of cattle type traits judgment. *Acta Biol. Hung.* 48:105-112.
- Veerkamp, R. F. and Brotherstone, S. 1997. Genetic correlations between linear type traits, food intake, live weight and condition score in Holstein Friesian dairy cattle. *Anim. Sci.* 64:385-392.
- Veselá, Z., Pribyl, J., Šafus, P., Vostrý, K. Šeba, K. and Štolc, L., 2005. Breeding value for type traits in beef cattle in the Czech Republic. *Czech J. Anim. Sci.* 50:385-393.
- Vukasinovic, N., Moll, J. and Künzi, N. 1997. Factor analysis for evaluating relationships between herd life and type traits in Swiss Brown cattle. *Livest. Prod. Sci.* 49:227-234.

(Received March 17, 2010; Revised May 31, 2010; Accepted June 1, 2010)