

Effect of Box-Cox Transformation on Genetic Parameter Estimation in Layers*

Yakut ÜNVER, Yavuz AKBAŞ, İsmail OĞUZ

Unit of Biometry and Genetics, Department of Animal Science, Faculty of Agriculture, Ege University, İzmir - TURKEY

Received: 03.07.2001

Abstract: The purpose of the study was to investigate the effects of Box-Cox transformation on estimations of the genetic parameters for egg production traits that do not hold to the assumptions of parametric statistical analysis. Egg production of 1980 animals from 43 sires and 8 dams per sire were used. Five partial record periods were considered: 22-30 weeks of age, 31-34 weeks of age, 22-34 weeks of age, 31-40 weeks of age and 22-40 weeks of age.

The egg production traits showed non-normal distributions in the form of negative skewness and positive kurtosis. The greatest departure from normality was observed in the peak period (31-34 weeks of age). After applying Box-Cox transformation, the distribution of the transformed data became closer to normality. Applying Box-Cox transformation to partial egg production data resulted in an increase in heritabilities and slightly changed genetic and phenotypic correlations. Transformation such as Box-Cox should be preferred for estimating genetic parameters from the data, when the assumptions are not satisfied.

Key Words: Box-Cox transformation, egg production, heritability, correlations

Yumurta Tavuklarında Box-Cox Transformasyon Uygulamasının Genetik Parametre Tahminlerine Etkisi

Özet: Bu çalışma, parametrik istatistik analizlerin önkoşulu olan varsayımları sahip olmayan yumurta verim özelliklerinde genetik parametre tahminleri üzerine Box-Cox transformasyonun etkisini belirlemek amacıyla gerçekleştirilmiştir. Çalışmada baba başına 8 tavuk düşecek şekilde 43 babadan elde edilen toplam 1980 bireyin yumurta verimleri kullanılmıştır. Beş farklı döneme (22-30; 31-34; 22-34; 31-40 ve 22-40 hafta) ait kısmi yumurta verimleri hesaplanmıştır.

Farklı dönemlerde elde edilen kısmi yumurta verimleri negatif çarpıklık ve pozitif basıklık değerleri ile normalden sapmalı bir dağılım göstermişlerdir. Normalden en büyük sapma pik veriminin yer aldığı döneme (31-34 hafta) ait yumurta veriminde saptanmıştır. Box-Cox transformasyon uygulanması sonrasında veriler normale daha fazla yaklaşmışlardır. Transformasyon ile yumurta verimlerine ait kalıtım derecelerinde bir artış, genetik ve fenotipik korelasyonlarda ise hafif bir değişme gözlenmiştir. Varsayımların gerçekleşmediği verilerde genetik parametre tahminlerine yönelik olarak da Box-Cox gibi transformasyon uygulamaları tercih edilmelidir.

Anahtar Sözcükler: Box-Cox transformasyonu, yumurta verimi, kalıtım derecesi, korelasyonlar

Introduction

In biological studies, quantitative variables are assumed to be normally distributed but this is rarely examined. Egg production traits, particularly those based on early records, have a non-normal distribution (1,2) in the form of negative skewness and positive kurtosis (Figure 1). The size of the deviation from normality depends on the trait and the number of individuals in the population. However, body weight and egg size show little or no departure from normality (3).

Environmental and/or genetic factors may cause a departure from normality. Brah and Dev (3) concluded that deviations from normality might be associated with traits of relatively low heritability, extremes of gene frequencies and non-additive genetic effects. An abnormal distribution of reproductive traits in a genetically homogeneous population can be explained in terms of non-random environmental factors (1). Many diseases are well known to have catastrophic effects on egg production. It is postulated that any of these factors and

* This article is a part of the MSc thesis of the first author.

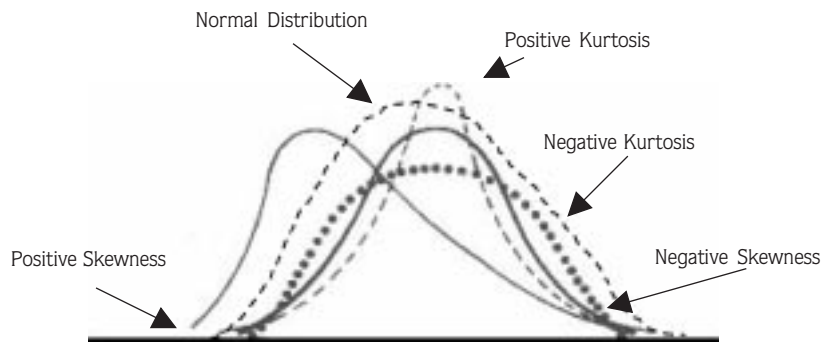


Figure 1. Types of deviations from normality.

possibly many others cause a negative skewness of the distribution.

The common models and methods for estimating phenotypic and genetic parameters have some assumptions that genetic and environmental effects are additive, and that experimental errors are independent in terms of probability and have equal variances and are normally distributed (4). However, these assumptions do not hold for egg production traits, as described by Clayton (1) and Brah and Dev (3).

Non-normality can lead to a loss of efficiency in the estimation of fixed effects with a corresponding loss of power of tests and to the heterogeneity of error variances (2).

To achieve normality, the omission of outliers or the transformation of data into a new scale is suggested to satisfy more closely the above assumptions (5). There are many transformation methods such as logarithmic, square root, arc-sin and Box-Cox. Box-Cox transformation reduced the heterogeneity of error variance and permitted the assumption of equal variance as well as increasing the linearity of genetic regression (2).

The purpose of the present study was to determine the effects of Box-Cox transformation on estimations of genetic parameters for egg production traits.

Materials and Methods

Egg production data from a commercial sire line were used. Hens were artificially inseminated. In total 1980 animals from 43 sires and 8 dams per sire were used. Egg records were collected and recorded daily and individually. The period of data collection was from 22 to

40 weeks of age, known as the early part of the whole egg production record. In egg production, there are some important periods, namely age at sexual maturity, period of peak production and the period after the peak. Therefore, in this study the early egg production period (22-40 weeks of age) is subdivided into 5 subperiods. The first (22-30 weeks of age) is the period from sexual maturity to the peak period. The second (31-34 weeks of age) includes peak production. In the third period (22-34 weeks of age) the first and second periods are combined. The fourth period covers egg production after the peak until 40 weeks of age. The last period (22-40 weeks of age) includes entire early egg production.

The normality of the partial egg production data was examined using skewness, kurtosis and the Shapiro-Wilk test. Skewness measures the asymmetry of the distribution is calculated from sums of the values to the power 3. Kurtosis measures the relative shape of the middle and tails of the distribution. Kurtosis is calculated from sums of the values to the power 4. The Shapiro-Wilk test was also implemented because the sample size was less than 2000.

Egg production for these periods was transformed by Box-Cox transformation as follows (6):

$$Z^t = \frac{y(t)_{ijkl}}{J(t)^{\frac{1}{n}}} = \frac{y^t - 1}{tG_y^{t-1}} \tag{1}$$

where Z^t is the standardized transformed variate, $y(t)_{ijkl}$ is an original untransformed record, $J(t)^{\frac{1}{n}}$ is the n^{th} root of the Jacobian of the transformation to remove the t -dependent scale effect, and G_y is the geometric mean of the n original observations. These variables are then used to calculate the log-likelihood $L_{\text{max}}(t)$ for trial values of t .

$$L_{\max}(t) = \frac{n}{2} \log_e = \left[\frac{S_e(t)}{n} \right] \quad [2]$$

where $S_e(t)$ is the residual sum of squares of the standardized observations. The optimal t value can be chosen to give the maximum $L_{\max}(t)$ and minimum $S_e(t)/n$ (2).

Phenotypic and genetic parameters (heritabilities and genetic correlations) were estimated using 4 different models for each partial egg production:

$$Y_{ijkl} = \mu + f_i + s_j + d_{k(j)} + \beta x + e_{ijk} \quad [3]$$

$$Y_{ijkl} = \mu + f_i + s_j + \beta x + e_{ijk} \quad [4]$$

$$Y_{ijkl} = \mu + f_i + d_{k(j)} + \beta x + e_{ijk} \quad [5]$$

$$Y_{ijkl} = \mu + f_i + a_j + \beta x + e_{ijk} \quad [6]$$

where Y_{ijkl} is partial egg production; μ is the overall mean, f_i is the fixed effect of the i^{th} cage row ($i = 1, \dots, 13$), s_j and $d_{k(j)}$ are random effects of the sire and dam within sire, respectively, a_j is the random effect of the animal ($j=1, \dots, 1980$), β is the regression coefficient, x is body weight at sexual maturity as a covariate, and e_{ijk} is the residual effect.

The first 3 models are called parent models (sire + dam, sire and dam model, respectively) and the last one is known as an animal model. The JMP (7) and DFREML (8) computer programs are used to estimate genetic parameters for the parent and animal models,

respectively. After that, heritabilities and their standard errors were calculated according to Dickerson (9) and Swiger et al. (10) for parent models. Genetic correlations given in the study were estimated from the sire + dam model as described by Becker (11).

Results

Results from analysis of normality for egg production using skewness, kurtosis and probability of the Shapiro-Wilk test are given in Table 1. Distributions of errors from the sire + dam model for different partial egg production periods are shown in Figure 2.

Since the error terms of partial egg production data do not have a normal distribution, Box-Cox transformation is applied. After this transformation, the distribution of the transformed data became closer to normality. However, after the transformation, normality is achieved ($P > 0.05$) only for egg production in the 22-40 weeks of age period, which is the total partial egg production when we consider the Shapiro-Wilk test.

Applying Box-Cox transformation to partial egg production data resulted in an increase in their heritabilities (Table 2).

The genetic correlations are higher than the corresponding phenotypic correlations before and after transformation. Genetic and phenotypic correlations (Table 3) were slightly changed by transformation while there was a clear increase in heritabilities.

Table 1. Statistics related to normality for untransformed and transformed partial egg production.

Egg Production Periods	Untransformed			Transformed		
	Skewness	Kurtosis	P(S-W)*	Skewness	Kurtosis	P(S-W)*
22-30	-0.398	0.814	0.0001	0.111	-0.397	0.00001
31-34	-1.549	5.867	0.0000	-0.188	0.535	0.01300
22-34	-0.134	-0.286	0.0001	-0.170	-0.239	0.00001
31-40	-0.648	3.686	0.0000	0.174	0.808	0.00006
22-40	-0.506	1.430	0.0000	0.082	0.520	0.05340

*P(S-W): probability of Shapiro-Wilk test

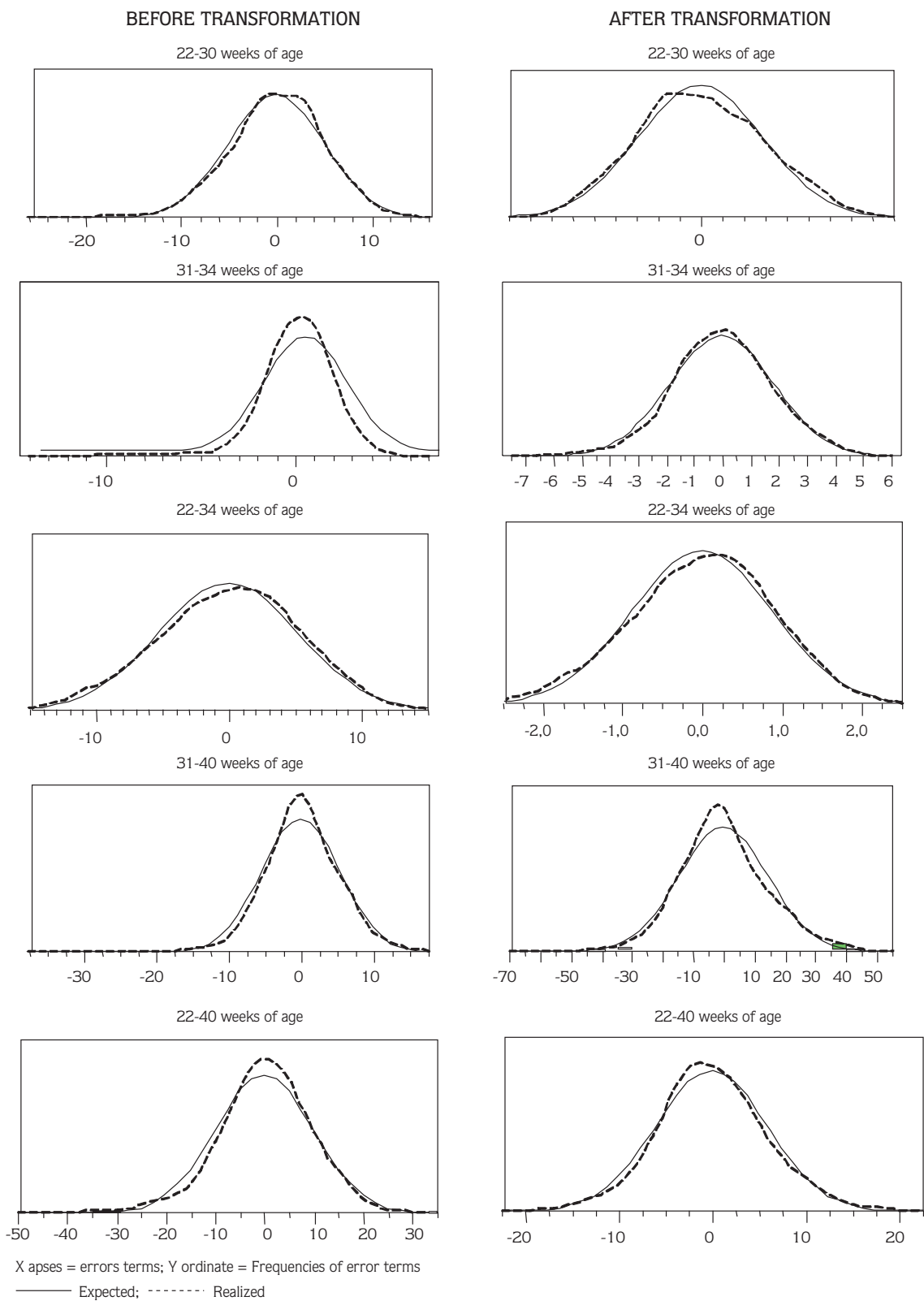


Figure 2. Distributions of errors for different partial egg productions in the sire + dam model.

Table 2. Heritability estimates for untransformed and transformed partial egg production from different periods.

Periods (weeks)	Type*	h_{s+D}^2	h_s^2	h_D^2	h_A^2
22-30	B.T.	0.27 ± 0.11	0.28 ± 0.06	0.26 ± 0.02	0.24 ± 0.01
	A.T.	0.29 ± 0.02	0.27 ± 0.05	0.30 ± 0.02	0.25 ± 0.01
31-34	B.T.	0.18 ± 0.22	0.17 ± 0.10	0.18 ± 0.04	0.18 ± 0.08
	A.T.	0.25 ± 0.30	0.24 ± 0.12	0.26 ± 0.05	0.25 ± 0.00
22-34	B.T.	0.25 ± 0.10	0.28 ± 0.06	0.23 ± 0.02	0.25 ± 0.01
	A.T.	0.26 ± 0.11	0.30 ± 0.08	0.22 ± 0.02	0.22 ± 0.10
31-40	B.T.	0.58 ± 0.14	0.49 ± 0.10	0.66 ± 0.09	0.53 ± 0.01
	A.T.	0.62 ± 0.05	0.54 ± 0.11	0.70 ± 0.04	0.58 ± 0.06
22-40	B.T.	0.44 ± 0.07	0.39 ± 0.08	0.49 ± 0.03	0.38 ± 0.11
	A.T.	0.48 ± 0.12	0.43 ± 0.09	0.53 ± 0.03	0.43 ± 0.11

h_{s+D}^2 , h_s^2 and h_D^2 are heritabilities based on sire + dam, sire and dam components of variance

B.T.: Before transformation, A.T.: After transformation

Table 3. Phenotypic (above diagonal) and genetic (below diagonal) correlations between different partial egg production periods from the sire + dam model.

Periods	BEFORE TRANSFORMATION				
	22-30	31-34	22-34	31-40	22-40
22-30		0.20	0.96	0.004	0.66
31-34	0.29 ± 0.46		0.47	0.49	0.50
22-34	0.95 ± 0.03	0.57 ± 0.33		0.14	0.74
31-40	0.36 ± 0.19	0.80 ± 0.14	0.56 ± 0.15		0.76
22-40	0.74 ± 0.08	0.71 ± 0.15	0.86 ± 0.05	0.89 ± 0.03	
Periods	AFTER TRANSFORMATION				
	22-30	31-34	22-34	31-40	22-40
22-30		0.25	0.94	0.07	0.69
31-34	0.34 ± 0.18		0.47	0.55	0.53
22-34	0.94 ± 0.01	0.56 ± 0.35		0.16	0.72
31-40	0.43 ± 0.04	0.82 ± 0.07	0.56 ± 0.09		0.76
22-40	0.77 ± 0.04	0.72 ± 0.19	0.84 ± 0.07	0.90 ± 0.02	

Discussion

Since the error terms do not have a normal distribution in this study as reported by Clayton (1), the assumptions do not hold for egg production traits, which show markedly non-normal distributions.

Although expected values for skewness and kurtosis for a normal distribution are zero, all partial egg

productions show negative skewness and positive kurtosis, while the third period (22-34 weeks of age) egg production has negative skewness and negative kurtosis (Table 1). The general tendency obtained in this study was the same as the results reported by Clayton (1) and Brah and Dev (3).

Ibe and Hill (2) and Savaş et al (12) reported that Box-Cox transformation produced more nearly normally

distributed egg production compared to untransformed data but it was not completely normal for all instances. Savaş et al. (12) clarified that the egg production that has the closest distribution to normal is total partial egg production, as found in this study.

In our study, the optimum Box-Cox parameter (t) differs from one trait to another. The optimum t values for all periods mentioned are 4.05, 2.70, 0.68, 3.26 and 3.60, respectively. The t values are all positive for these negatively skewed distributions. Ibe and Hill (2) also state that the t values are negative for positively skewed data but positive for negatively skewed data. The greater the departure of the data from normality, the higher the absolute value of the t value required. Beaumont (13) has suggested obtaining the parameter t in an interval between -2 and $+5$.

The biggest deviation from normality is shown in the peak period (31-34 weeks of age). Clayton (1) reported that negative asymmetry and positive kurtosis achieve their greatest values in those periods when egg production is at its peak. This result is supported by other studies (12,14-16).

The heritabilities of partial egg production are increased by Box-Cox transformation. This result is consistent with those of most researchers (2,12,17), while some found lower estimates after transformation (18). Higher heritability from transformed data would give rise to higher expected selection responses than that from untransformed data.

The increase in heritability after transformation is higher for egg production at 31-34 weeks of age because the deviation from normality was also highest in this group. However, the difference in heritability for egg production between before and after transformation is similar in the 22-30 and 22-34 weeks of age groups due to the fact that they cover mainly the same egg production period. In spite of that, this was not the case in the same level for the periods 31-34 and 31-40. The common period between these groups (31-34 and 31-40) is less than that between the groups 22-30 and 22-34. On the other hand, period 31-34 covers the peak period of egg production. For total egg production period the change in heritability between before and after transformation was about 0.04 and this level is nearly 10% of the heritability level before transformation.

Since age at first egg and egg number in the first 8-week period are highly heritable traits (14), heritability for egg production at 22-30 weeks of age is higher than that during the second period but similar to that of the third period of egg production. Although most estimates of heritability for total egg production were between 0.10 and 0.20, heritability estimates tend to be higher for partial egg number than for total egg number (14). Heritability for egg production in the period 31-40 and 22-40 was high and greater than 0.38 in this study.

The levels of heritability estimates for transformed traits were generally higher than those reported in the literature (13,16,19). Besbes et al. (16) examined the effect of transformation on 3 partial egg production records during 19-26, 26-38 and 26-54 weeks of age from lines A and B. After transformation, heritability estimates for egg production during these periods changed from 0.23, 0.07 and 0.12 to 0.26, 0.10 and 0.18, respectively for Line A and from 0.26, 0.09 and 0.11 to 0.27, 0.13 and 0.16 for Line B, respectively.

In spite of clear increases in heritabilities, correlations were slightly changed by transformation. Ibe and Hill (2) and Savaş et al. (12) obtained similar results through Box-Cox transformation. However, Besbes et al. (18) noted that correlations from non-normal data are biased even for non-selected populations

In genetic studies, non-normality affects the results of selection experiments (2,3) not only for genetic parameter estimation but also for breeding value estimation.

In conclusion, the departure from the assumptions for egg production traits is important for genetic parameter estimation since those estimated are biased when the assumptions do not hold. Transforming data to a scale that the assumptions are more closely satisfied by using Box-Cox transformation should be preferred. It can be performed using standard statistical software. Ibe and Hill (2) and Koerhuis (17) emphasize that when normality and homogeneity of variance were not satisfied, selection index and BLUP methods for breeding value estimation became less efficient.

Acknowledgments

This research was supported by the Ege University Research Fund (98-ZRF-050).

References

1. Clayton, G.A.: Normality of Egg Production in Poultry. *Br. Poult. Sci.*, 1975; 16: 431-439.
2. Ibe, S.N., Hill, W.G.: Transformation of Poultry Egg Production Data to Improve Normality, Homoscedascity and Linearity of Genotypic Regression. *J. Anim. Breed. Genet.*, 1988; 105: 231-240.
3. Brah, G.S., Dev, D.S.: Normality of Some Quantitative Traits in Chickens. *Br. Poult. Sci.* 1979; 20: 191-195.
4. Banks, B.D., Mao, I.L., Walter, L.P.: Robustness of the Restricted Maximum Likelihood Estimator Derived under Normality as Applied to Data with Skewed Distributions. *J. Dairy Sci.* 1985; 68: 1785-1792.
5. Cochran, W.G.: Some Consequences When the Assumptions for the Analysis of Variance Are Not Satisfied. *Biometrics*, 1947; 3: 22-38.
6. Box, G.P.E., Cox, D.R.: An Analysis of Transformations. *J.R. Stat. Soc.* 1964; B26: 211-243.
7. SAS Institute: JMP User's guide, Version 3.1, SAS Institute Inc., Cary, NC, 1994.
8. Meyer, K.: DFREML, User Notes, Version 2.1, 1993.
9. Dickerson, G.E.: Techniques for Research in Quantitative Animal Genetics. *Techniques and Procedures in Animal Production Research.* American Soc. An. Prod. Pub. 1960; 56-105.
10. Swiger, L.A., Harvey, W.R., Everson, D.O., Gregory, K.E.: The Variance of Intraclass Correlation Involving Groups with One Observation. *Biometrics*, 1964; 20: 818-826.
11. Becker, W.A.: *Manual of Procedures in Quantitative Genetics.* 2nd ed., Washington State Univ., Pullman, WA, 1967.
12. Savaş, T., Preisinger, R., Röhe, R., Kalm, E.: Yumurta Verimi Devamlılığının Islahında Kısmi Verimlerden Yararlanma Olanakları. YUTAV, Uluslararası Tavukçuluk Fuarı ve Konferansı, 1997; 14-17/05/97, İstanbul.
13. Beaumont, C.: Genetic Parameters of the Duration of Fertility in Hens. *Can. J. Anim. Sci.* 1992; 72: 193-201.
14. Flock, D.K.: Genetic Analysis of Part-Period Production in a Population of White Leghorns under Long-Term RRS. *Tierzüchtg. Züchtgsbiol.*, 1977; 94: 89-103.
15. Boukila, B., Desmarais, M., Pare, J.P., Bolamba, D.: Selection for Increased Egg Production Based on Annual Records in Three Strains of White Leghorns. 1. Comparison of Different Partial Records to Improve Annual Egg Production. *Poult. Sci.*, 1987; 66: 1077-1084.
16. Besbes, B., Ducrocq, V., Foulley, J.L., Protais, M., Tavernier, A., Tixier-Boichard, M., Beaumont, C.: Box-Cox Transformation of Egg-Production Traits of Laying Hens to Improve Genetic Parameter Estimation and Breeding Evaluation. *Livest. Prod. Sci.* 1993; 33: 313-326.
17. Koerhuis, A.N.M.: Non-normality of egg production distribution in poultry and the effects of outlier elimination and transformation on size and curvilinearity of heritability. *Livest. Prod. Sci.* 1996; 45: 69-85.
18. Besbes, B., Ducrocq, V., Tavernier, A., Foulley, J.L., Boichard, M., Beaumont, C., Protais, M.: Application of the Box-Cox Transformation to Egg Production Traits of Laying Hens. 42nd Annual Meeting of the European Association for Animal Production. Berlin, 8-12 September 1991.
19. Crawford, R.D.: *Poultry Breeding and Genetics.* Elsevier Science Publishers B. V. Second Edition, Amsterdam, 1990; 705-759.