Faculty Papers and Publications in Animal Science

Animal Science Department

January 1991

Heritability Estimates of Transformations of Normally Distributed Records

A. L. DeStefano
Cornell University

L. Dale Van Vleck
University of Nebraska-Lincoln, dvan-vleck1@unl.edu

DeStefano, A. L. and Van Vleck, L. Dale, "Heritability Estimates of Transformations of Normally Distributed Records" (1991). *Faculty Papers and Publications in Animal Science*. 141.
https://digitalcommons.unl.edu/animalscifacpub/141

GENETICS AND BREEDING

Heritability Estimates of Transformations of Normally Distributed Records

A. L. DeSTEFANO and L. D. VAN VLECK[1]
Department of Animal Science
Cornell University
Ithaca, NY 14853

ABSTRACT

Milk yields were simulated for identical twin and parent-progeny pairs to determine effect of transformations on genetic variance. Use of identical twin pairs excludes variability due to Mendelian sampling and contribution by the other parent present in parent-progeny records. Phenotypic variances of 616,800; 1,439,200; and 2,261,600 and means from 2267 to 13,603 by increments of 2267 were used for linear scale records. Phenotypic variances of .01000, .01875, and .02750 and means equal to the natural logarithm of the means for linear scale were used for log scale records. Heritability varied from .05 to .95 by increments of .10. For each combination of parameters, 10 replicates of 10,000 pairs of relatives were created. Data normally distributed on a linear scale were log transformed. Log normally distributed data were exponentially transformed. Heritability estimates from correlations among relatives indicated heritability is larger for normally distributed data than for transformed data. For parent-progeny pairs, the difference increased as heritability increased, and for identical twin pairs, the differences were least at extremely large and small heritabilities. For both types of relative pairs, the difference increased as phenotypic variance increased. Absolute differences in heritability ranged from .0000 to .0089, a difference of little practical importance.
(Key words: heritability, logarithmic transformations, log normal distribution)

Abbreviation key: LNORM = normally distributed on log scale, LNORM-EXP = exponentiation transformation of LNORM, NORM = normally distributed records on linear scale, NORM-LN = ln transformation of NORM.

INTRODUCTION

Currently the Northeast sire evaluation utilizes ln transformed records, unlike other sire evaluations. The benefits or disadvantages of transformation for improving the accuracy of evaluation have not been studied. Hill et al. (2) estimated heritability of milk yield at different production levels for untransformed and ln transformed data and concluded that the ln transformation may have resulted in higher heritability estimates due to increased homogeneity of phenotypic variance over herds at different production levels and that the CV of milk production is homogeneous. Brotherstone and Hill (1) later concluded that for yield there was substantial heterogeneity of variance and heterogeneity of CV among herds and that although a ln transformation improves homogeneity of variance, some differences in variability remain. This result suggests that an ln transformation is not appropriate. Other analyses (3, 4, 5) found that ln transformation resulted in unchanged or smaller estimates of heritability.

The object of this study was to determine the change in heritability with ln or exponential transformation when the records being transformed were normally distributed samples from populations with constant mean and variance.

MATERIALS AND METHODS

Records on identical twins were created and the correlation between transformed records calculated. Parent-progeny pairs also were simulated to determine effect of relationship on this method. Data sets were created using

TABLE 1. Parameters to create normally distributed records on linear and natural logarithm (ln) scales.

Normal distribution on linear scale		Normal distribution on natural log scale	
Mean (kg)	Phenotypic variance (kg^2)	Mean (kg)	Phenotypic variance [ln (kg^2)]
2267	616,800	7.73	.01000
4534	1,439,200	8.42	.01875
6801	2,261,600	8.83	.02750
9069		9.11	
11,336		9.34	
13,603		9.52	

Monte Carlo simulation where the single measurement, linear animal model was:

$$P_{xi} = \mu + g_{xi}\sigma_G + e_{xi}\sigma_E$$
$$P_{yi} = \mu + a_{xy}g_{xi}\sigma_G + g_{yi}(1-a_{xy}^2)^{.5}\sigma_G$$
$$+ e_{yi}\sigma_E$$

where P_{xi} is a record on relative X of relative pair i, P_{yi} is a record on relative Y of relative pair i, μ is a constant corresponding to mean production level, g and e are pseudorandom standard normal deviates, σ_G is the SD of additive genetic variance, σ_E is the SD of environmental effects, and a_{xy} is the numerator relationship between relatives X and Y. The genetic term in P_{yi} is composed of two parts; the first, $a_{xy}g_{xi}\sigma_G$, simulates the portion of additive genetic effects in common between relatives X and Y. The second term, $g_{yi}(1-a_{xy}^2)^{.5}\sigma_G$, accounts for Mendelian sampling and contribution of genes to Y from ancestors not related to X.

Two sets of parameters for μ and σ_p^2 (phenotypic variance, $\sigma_G^2 + \sigma_E^2$) were specified to define different data sets. The first set was chosen to simulate populations of normally distributed records on a linear scale (NORM), and the second set was chosen to create normally distributed populations with means and variances that approximate populations of milk records on a log normal scale (LNORM). A wider range of means and variances than observed with actual production records was used to investigate extremes. Table 1 lists the two sets of parameters. Heritability was varied for normally distributed records between .05 and .95 by increments of .10, yielding 180 unique com-

binations of mean, variance, and heritability for each parameter set.

The means in Table 1 for the LNORM were obtained from ln transformation of milk production means on the usually observed scale in increments of 2267 kg. A normal variable, x, transformed to ln(x) has variance, $\sigma_{ln(x)}^2$, approximately equal to

$$\sigma_x^2/\mu^2. \qquad [1]$$

The range of variances for the LNORM was determined by substituting the phenotypic variances for the linear scale into Equation [1] assuming $\mu = 9069$ kg. This range of variances required that mean production level be greater than 6801 kg because a smaller mean combined with some phenotypic variances within the range resulted in the creation of negative milk yield records, for which the logarithm is undefined.

Records were created by applying a specific combination of parameters from one of the two sets of parameters to the model. Each subset consisted of pairs of individual records. The pairing was based on relationship. In separate simulations, two relationships were considered: identical twins with $a_{xy} = 1$ and parent-progeny pairs with $a_{xy} = .5$. When $a_{xy} = 1$, the second genetic term of the record of animal Y becomes 0 so that the only difference between the records of twins is due to environmental differences. The effect of Mendelian sampling and contribution by the other parent on the correlation between the pair is eliminated. When $a_{xy} = .5$, the sampling effect is present in the model, and a parent-progeny pair (e.g., daughter-dam) is simulated. Correlation between environ-

TABLE 2. Description of data sets and transformations.

Distribution of data prior to transformation	Transformation	Distribution of data post transformation
Normal with parameters based on ln(linear scale) (LNORM)	Exponentiation	Log normal distribution (LNORM-EXP)
Normal with parameters based on linear scale (NORM)	Natural logarithm	Nonnormal (NORM-LN)

mental effects for related animals was 0. For each combination of parameters and relationships, 10,000 pairs of records were produced and replicated 10 times, making 100,000 pairs for each combination.

Each record was transformed to create a transformed data set. Distributions of original data, transformations applied, and distributions of transformed data are in Table 2. The first kind of transformation used on the normal distribution, LNORM, was exponentiation. For the new record, $P_{ji}^* = \exp(P_{ji})$, where P_{ji} is the record on relative j (either relative X or relative Y) of relative pair i. The transformed records (LNORM-EXP) have a log normal distribution, i.e., the ln of the LNORM-EXP are normally distributed, which implies the LNORM-EXP have a multiplicative model:

$$P_{ji}^* = (\mu^*)\,(G_i^*)\,(E_i^*)$$

where $\mu^* = e^\mu$, $G_i^* = e^{G_i}$, and $E_i^* = e^{E_i}$. For parent-progeny pairs, the genetic term G_i^*, for P_{yi}^*, a record on relative y, was the product of the two genetic terms in the untransformed model. This transformed data set simulates the distribution of milk yield records based on the assumption that the multiplicative model is the true model and a ln transformation would result in a normal distribution.

The second kind of transformation was to calculate ln for normally distributed records P_{ji} where a transformed record P_{ji}^* was expressed as

$$P_{ji}^* = \ln(P_{ji}).$$

This was a deliberate transformation from NORM, based on an additive model to a distribution that is not normal and to a model that, in

some way, is not additive. This data set (NORM-LN) can be used to assess the effect of a ln transformation on a normally distributed data set.

Heritability estimates for the identical twin data were obtained from correlations between records of twins and for parent-progeny data heritability estimates were twice the correlations between records of parent and progeny. Heritability estimates were averaged over 10 replicate data sets for each combination of heritability, mean, and variance. Heritability estimates from the untransformed data, LNORM, were compared with estimates from the transformed data, LNORM-EXP, whereas estimates from NORM were compared with NORM-LN.

Average estimates of heritability from transformed data were subtracted from average estimates based on normally distributed data, yielding the differences D_h. A three-way ANOVA was performed for each of the data sets with ln or exponentiation transformation and $a_{xy} = 1$ or .5 to determine if mean level, heritability, and phenotypic variance had any effects on D_h. To examine the effect of actual heritability on D_h, the differences were averaged over production level and phenotypic variance and plotted against heritability. Likewise, differences were averaged over production level and heritability and plotted against phenotypic variance to determine effect of phenotypic variance; they were also averaged over heritability and phenotypic variance and plotted against mean to determine effect of production level on differences in estimated heritability.

RESULTS AND DISCUSSION

On average, estimates of heritability from transformed data, NORM-LN or LNORM-EXP, were smaller than estimates from data with a normal distribution, NORM or LNORM. This difference occurred whether data were
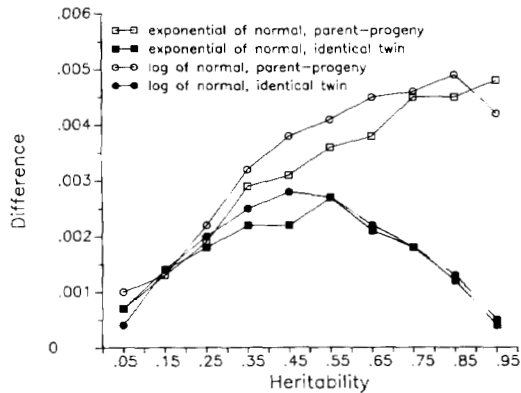
Figure 1. Average heritability estimates from data transformed by exponentiation or natural logarithm as differences from heritability estimates from untransformed, normally distributed records (difference) plotted against heritability for identical twin and parent-progeny pairs.
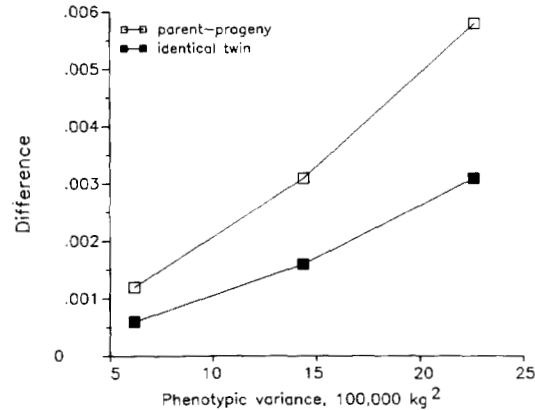

Figure 2. Average heritability estimates from data transformed by natural logarithm as differences from heritability estimates from untransformed, normally distributed records (difference) plotted against phenotypic variance on the actual scale for both identical twin and parent-progeny pairs.

normal on the linear scale or on the ln scale. The differences ranged from $-.0012$ to $.0089$ over all data sets.

Figure 1 shows average D_h over production level and phenotypic variance plotted against heritability for four data sets, ln and exponential transformations with identical twin and parent-progeny pairs. Heritability level was significant ($P \leq .0001$) for all four data sets. For h^2 from .05 to .45, the four data sets gave similar results; D_h increased as heritability increased. For parent-progeny data, this trend continued as heritability increased. Identical twin data resulted in D_h decreasing as heritability became extremely high or low. Because simulation of records for identical twins produces variability solely due to environmental effects, the correlation between these records provides the best measure of what happens to heritability if normal records are transformed. The distribution of genetic values on the transformed scale, however, is unknown.

Plots of D_h averaged over heritability and production level against phenotypic variance on the scale of the untransformed distributions are contained in Figure 2 for the exponential transformation (LNORM-EXP versus LNORM) and Figure 3 for the ln transformation (NORM-LN versus NORM). The figures show that D_h increased as phenotypic variance increased. The effect of phenotypic variance was significant for each data set ($P \leq .0001$).

Differences averaged over heritability and phenotypic variance and plotted against mean level are in Figure 4. Mean level did not have a significant effect on heritability estimates from LNORM compared with those from LNORM-EXP. Production level was significant ($P \leq .0001$) on the difference in heritability estimates of NORM minus NORM-LN, with the difference decreasing as the production level increased.

Although the heritability estimates from transformed data were smaller on average than those from normal data, sample estimates from the normal data for all replications were not always smaller or larger than the parameter h^2.

CONCLUSIONS

This simulation suggests that transformation of normally distributed records results, on average, in smaller heritability estimates compared with estimates from normally distributed records. Although there were definite trends in the magnitude of the differences due to heritability, phenotypic variance, and in some cases mean level, the differences were too small to be of practical importance. The results of this simulation study indicate that for the distributions considered here, ln transformation does not alter genetic variation, as measured by heritability, enough to be meaningful.
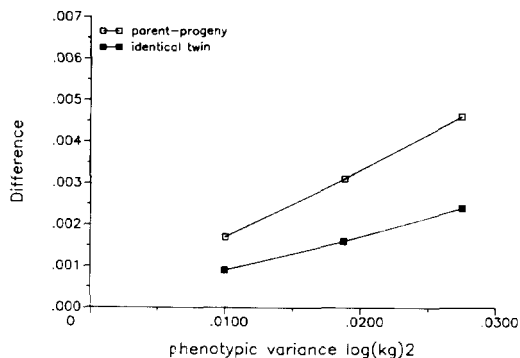
Figure 3. Average heritability estimates from data transformed by exponentiation as differences from heritability estimates from untransformed, normally distributed records (difference) plotted against phenotypic variance on the natural logarithmic scale for both identical twin and parent-progeny pairs.
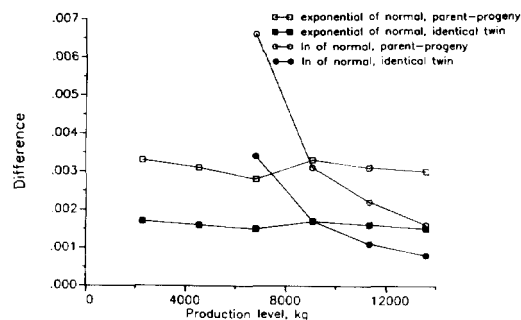

Figure 4. Average heritability estimates from transformed data as differences from heritability estimates from untransformed records (difference) plotted against mean level for identical twin and parent-progeny pairs. Missing data points are due to negative records created by the combination of low mean and high variance. Natural log transformation could not be applied to negative values.

Within a population, there can be heterogeneity of variance among herds due to increases or decreases in either genetic variance or environmental variance. The contribution of these individual components to heterogeneity of variance in this simulation is not clear. Different sources of heterogeneity of variance in a population may cause ln transformation to affect heritability estimates differently. This may explain differences in the results of studies where ln transformation was applied to field data. For the data used by Hill et al. (2), the transformed scale may be more appropriate or more normal because this simulation associates larger heritability estimates with a normal distribution. But for other data, for which transformation resulted in decreased heritability estimates (3, 4, 5), the original scale may be more appropriate. This may not be true with transformations other than logarithmic.

Because heritability estimates in this simulation did not change significantly when ln transformation was applied to a normal distribution, this study suggests that use of ln transformation such as in the Northeast sire evaluation may not affect estimation of genetic variance, even in

the event the transformation is not optimal for the data.

ACKNOWLEDGMENTS

REFERENCES

1 Brotherstone, S., and W. G. Hill. 1986. Heterogeneity of variance amongst herds for milk production. Anim. Prod. 42:297.
2 Hill, W. G., M. R. Edwards, M.K.A. Ahmed, and R. W. Thompson. 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod. 36:59.
3 Mirande, S. L., and L. D. Van Vleck. 1985. Trends in genetic and phenotypic variance for milk production. J. Dairy Sci. 68:2278.
4 Schneider, J. C., and L. D. Van Vleck. 1986. Heritability estimates for first lactation milk yield of registered and nonregistered Holstein cows. J. Dairy Sci. 69: 1652.
5 Van Vleck, L. D., L. R. Cox, and S. L. Mirande. 1985. Heritability estimates of milk production from daughter on dam regression by year and management level. J. Dairy Sci. 68:2964.