Int. J. Agricult. Stat. Sci. Vol. 10, Supplement 1, pp. 243-247, 2014 *ISSN : 0973-1903*

ORIGINAL ARTICLE

ESTIMATION OF HERITABILITY OF MASTITIS DISEASE USING MOMENT ESTIMATORS

Subrat Keshari Behera, A. K. Paul*, S. D. Wahi, M. A. Iquebal, Samarendra Das, Ranjit Paul, Wasi Alam and Anil Kumar

Indian Agricultural Statistics Research Institute, New Delhi - 110 012, India.
E-mail : apaul_66@yahoo.com

Abstract : The present study deals with study of inheritance of mastitis disease in Sahiwal breed of cows. As suggested by Kleinman (1973) two moment estimators $\hat{\rho}_{KEQ}$ and $\hat{\rho}^{*}_{KEQ}$ are used for the estimation of the heritability of mastitis disease from live data through calculation of intra-class correlation coefficient. Bootstrap technique is used for estimation of standard error of heritability as direct formula for the same is not available in literature. The findings are then compared with the results of analysis of variance (ANOVA) method. The method based on ANOVA estimator showed better performance as compared to the moment estimators $\hat{\rho}_{KEQ}$ and $\hat{\rho}^{*}_{KEQ}$ for the estimation of heritability of mastitis disease.

Key words : Heritability, Intra class correlation, Mastitis disease, Moment Estimators.

1. Introduction

Mastitis or inflammation of the mammary gland, is the most common and most expensive disease of dairy cattle throughout most of the world. There are many characters of economic importance in animal and plant breeding, which are polygenic in inheritance, but their phenotypic expressions show discontinuities. The characters are expressed in 'all or none' fashion. Although, lacking a continuous distribution, such characters are known to be multifactorial in their inheritance. The relationship between polygenes and the expression of such characters comes about through the establishment of 'thresholds'. Thus, there are two separate scales for the description of the phenotypic values. The underlying polygenic distribution, which is continuous and the visible phenotypic distribution, which is discontinuous and the two scales are connected by the 'threshold' a point of discontinuity. Heritabilities of these important traits are thus to be obtained by technique other than classical methods employed for continuous trait.

Dempster and Lerner (1950) and Bhatia *et al*. (1992) developed an algorithm for calculating the heritability of such binary traits and further Gianola (1979) generalized it. Van Vleck (1980) used the algorithm in a simulation study of sib and parent offspring analysis of binary trait. Magnusen and Kremer (1995) considered the beta-binomial model for estimating heritability of binary trait in plant breeding using the concept of selection response and realized heritability. Ridout *et al*. (1999) reviewed different estimators of intraclass correlation for binary data and compared them in an extensive simulation study. Carlén *et al*. (2004) estimated heritability of clinical mastitis in Holstein cows as 0.035 using the method of AI-REML. A simulation study for Genetic evaluation of mastitis in dairy cattle using linear models, threshold models and survival analysis was done by Carlén *et al*. (2006). Vazquez *et al*. (2009) reported the heritability estimates for US Holstein as 0.061 for mastitis measured as a binary trait in the probit model and 0.085 for the number of mastitis cases in the ordinal threshold model. Gernand *et al*. (2012) applied threshold methodology for binary distributed health disorders including mastitis and found their genetic parameters. Behera *et al*. (2010) estimated heritability of mastitis disease using ANOVA method through intraclass correlation. Here, in the present investigation live data is used for estimation of heritability of mastitis disease in case of Sahiwal breed of cows through estimation of intraclass correlation coefficient. The intraclass correlation coefficient

Author for correspondence* ***Received* *April 28, 2014* ***Revised*** *August 22, 2014* ***Accepted*** *October 14, 2014*

provides a quantitative measure of similarity between individuals within groups.

2. Materials and Methods

The data for present study has been taken from breeding records of Sahiwal cattle from the Breeding Farm, N.D.R.I., Karnal in 2007-2008. The data is having 86 sires, 1422 progenies and number of progenies affected by mastitis in terms of 0 and 1. Here, 0 indicates resistance to the disease and 1 indicates the appearance of disease. Out of 1422 progenies 474 are affected by mastitis. The methods used for estimation of heritability of mastitis disease are described below.

Moment Estimators

Let us suppose that there are k groups of individuals. The i*th* group having n_i individuals with each having a binary response X_{ij} ($i = 1, \ldots, k; j = 1, \ldots, n_i$). We refer to the two possible values of X_{ij} as success and failure and coded them as one and zero, respectively. Also, let $Y_i = \Sigma_j X_{ij}$ denote the total number of success in the i*th* group.

The probability of success is assumed to be the same for all individual's group; specifically, $\Pr(X_{ij}=1)=\pi$ for all i, j. Furthermore, the responses of individuals from different groups are assumed to be independent. Within each group, the correlation between any pair of responses (X_{ij}, X_{il}) ($j \neq l$) is ρ. In particular, the correlation is assumed not to vary with group size.

Let, $\tilde{\pi}_i = Y_i / n_i$ denote the observed proportion of successes in the i*th* group and define

$$\hat{\pi}_w = \sum_{i=1}^{k} w_i \tilde{\pi}_i$$

and

$$S_w = \sum_{i=1}^{k} w_i \left(\tilde{\pi}_i - \tilde{\pi}_w\right)^2,$$

Where, the w_i are the weights summing to one. By equating $\tilde{\pi}_w$ and S_w to their expected values under the common-correlation model [Kleinman (1973)] derived a class of estimators of the form

$$\hat{\rho} = \frac{S_w - \tilde{\pi}_w\left(1-\tilde{\pi}_w\right)\sum_{i=1}^{k}\frac{w_i\left(1-w_i\right)}{n_i}}{\tilde{\pi}_w\left(1-\tilde{\pi}_w\right)\left[\sum_{i=1}^{k} w_i\left(1-w_i\right) - \sum_{i=1}^{k}\frac{w_i\left(1-w_i\right)}{n_i}\right]}$$

He considered two specific estimators, one with equal weights ($w_i = 1/k$) and label the estimator $\hat{\rho}_{KEQ}$, which is given as

$$\hat{\rho}_{KEQ} = \frac{S_w - \tilde{\pi}_w\left(1-\tilde{\pi}_w\right)\sum_{i=1}^{k}\frac{w_i\left(1-w_i\right)}{n_i}}{\tilde{\pi}_w\left(1-\tilde{\pi}_w\right)\left[\sum_{i=1}^{k} w_i\left(1-w_i\right) - \sum_{i=1}^{k}\frac{w_i\left(1-w_i\right)}{n_i}\right]}$$

Where, $w_i = 1/k$.

Kleinman (1973) also proposed slight variants of the above stated estimator by replacing S_w by $S_w^* = \left(k-1\right)S_w/k$ and label the estimator $\hat{\rho}_{KEQ}^*$, which is given as

$$\hat{\rho}_{KEQ} = \frac{S_w^* - \tilde{\pi}_w\left(1-\tilde{\pi}_w\right)\sum_{i=1}^{k}\frac{w_i\left(1-w_i\right)}{n_i}}{\tilde{\pi}_w\left(1-\tilde{\pi}_w\right)\left[\sum_{i=1}^{k} w_i\left(1-w_i\right) - \sum_{i=1}^{k}\frac{w_i\left(1-w_i\right)}{n_i}\right]}$$

Where, $w_i = 1/k$.

Analysis of Variance (ANOVA) Estimator

Suppose a set of n_i dams selected at random and mated with i*th* sire ($i = 1, 2, \ldots, k$) giving rise to one progeny each. Considering the random linear model

$$X_{ij} = \mu + s_i + e_{ij}$$

Where, X_{ij} is the observation on the progeny of j*th* ($j = 1, 2, \ldots, n_i$) dam mated to the i*th* sire ($i = 1,2,\ldots,k$), μ is the general mean, s_i is the effect of the i*th* sire and e_{ij} is the uncontrolled environmental and genetic deviations attributable to individuals within sire groups. All effects are assumed random with

$$E\left(s_i\right) = E\left(e_{ij}\right) = 0,\ E\left(s_i^2\right) = \sigma_s^2 \text{ and } E\left(e_{ij}^2\right) = \sigma_e^2.$$

The analysis of variance based on this model is shown in Table 1.

The estimator is given by

$$\hat{\rho}_{AOV} = \frac{MS_b - MS_w}{MS_b + \left(\lambda - 1\right)MS_w}$$

Where, MS_b and MS_w are respectively, the between-group and within-group mean squares from a one-way analysis of variance of the binary data X_{ij} and where

$$\lambda = \frac{1}{k-1}\left[N - \sum_{i=1}^{k}\frac{n_i^2}{N}\right] \text{ with } N = \sum_{i=1}^{k} n_i$$

For binary data, explicit formulas for MS_b and MS_w are

$$MS_b = \frac{1}{k-1}\left[\sum_{i=1}^{k}\frac{Y_i^2}{n_i} - \frac{1}{N}\left(\sum_{i=1}^{k} Y_i\right)^2\right]$$

$$MS_w = \frac{1}{N-k}\left[\sum_{i=1}^{k} Y_i - \sum_{i=1}^{k}\frac{Y_i^2}{n_i}\right]$$

Fleiss (1981) used a modification of this estimator, which is denoted by $\hat{\rho}_{AOV}^{*}$, in which the divisor of MS_b is k rather than (k-1). So,

$$\hat{\rho}_{AOV}^{*} = \frac{MS_b - MS_w}{MS_b + (\lambda - 1)MS_w}$$

Where,

$$MS_b = \frac{1}{k}\left[\sum_{i=1}^{k}\frac{Y_i^2}{n_i} - \frac{1}{N}\left(\sum_{i=1}^{k} Y_i\right)^2\right]$$

Estimation of Heritability and its Standard Error

The half-sib heritability estimate is obtained by the formula

$$\hat{h}^2 = 4 \times \hat{\rho}$$

Where, $\hat{\rho}$ is the estimated intra-class correlation coefficient.

According to Falconer (1998), Standard Error for half-sib heritability estimate can be obtained by the formula

$$\text{Standard Error} = 32\frac{\hat{h}^2}{T}$$

Where, $T = nN$ and n is average half sib family size, N is the number of half-sib family. For Analysis of Variance estimators, we have used the above formula for estimation of standard error.

Bootstrap technique is used for estimation of standard error of estimation of heritability by moment estimators. The bootstrap is a computer based technique for estimating the standard errors, biases, confidence interval and other measures of statistical accuracy with just a sample in hand. Here, the objective was to find the standard error of the statistic, just based on the sample values; in other words we are expected to study the sampling distribution of the estimator. As our original data itself is a sample, the regrouping can be statistically termed as "resampling". Here a type of resampling is done in which, we are selecting a group of the same size as that of the mother sample randomly from the mother sample itself with replacement, then the resultant group gave some information about the sample and in turn about the population.

There are number of ways in which the bootstrap samples can be obtained in order to get the bootstrap estimates of the parameters, variance, standard error and confidence intervals. The present sampling procedure is based on the resampling from levels of the sires family directly from the data. In case of half sib family model the sires were resampled based on the resampling directly from the master sample. The bootstrap algorithm used can be described as follows:

1. Divide the whole data material into K classes corresponding to the sire families. Let $\hat{F}$'s stand for multivariate empirical distribution, where each family has mass 1/K.

2. Draw a random sample with replacement of K classes from $\hat{F}$'s.

3. Then select all the progenies in the sire family or choose the progenies randomly with replacement from the progenies in that family fixed size p.

4. Analyze the data according to the model under study.

5. Estimate the parameters h^2 as desired.

Go back to 2 and draw a fresh sample from the data and the same procedure is repeated N times such that we get N estimates of h^2.

We shall discuss the one sample situation in which a random sample of size n is observed from a completely unspecified probability distribution F.

$X_i = X_i$, $X_i \sim$ ind F, $i = 1, 2, \ldots, n$.

Where, F is a distribution on either the real line or plane. Let, $X = (X_1, X_2, \ldots, X_n)$ and $X = (X_1, X_2, \ldots, X_n)$ denote the random sample and its observed realization, respectively.

The problem to be solved is that given a specified random variable R(X, F) with possible dependence on

Table 1 : Form of analysis of half-sib families.

S.V.	d.f.	S.S.	M.S.	E(M.S.)
Between sires	k-1	SS_b	MS_b	$\sigma_e^2 + \lambda\sigma_s^2$
Within sires	N-k	SS_w	MS_w	σ_e^2

Table 2 : ANOVA and Moment Estimators.

Estimators	$\hat{\rho}$ value	$\hat{h}^2$ value	Bootstrap Standard Error		Standard Error by approximate formula
			Sample size=500	Sample size=1000	
ρ_{AOV}	0.0061	0.0244	0.0549	0.0545	0.0005
ρ^*_{AOV}	0.0053	0.0213	0.0537	0.0537	0.0004
ρ_{KEQ}	0.0143	0.0573	0.0251	0.0235	_
ρ^*_{KEQ}	0.0130	0.0523	0.0283	0.0258	_

both X and the unknown distribution F, to estimate the sampling distribution of unknown parameter on the basis of the observed data x.

Let θ (F) be some parameter of interest of F and $\hat{\theta}$ be an estimator of $\theta(F)$ based on the sample. Bootstrap requires us to calculate the sample analogue estimator $\hat{\theta}$ a large number of times (say N), each of which is based on a sample of size 'n' obtained by either sampling with replacement from the original n observations or by generating samples from the population inferred from the data sample.

Let us see Bootstrap procedure

1. Obtain an estimate $\hat{\theta}$ from the original sample values (x_1, ..., x_n).

2. For i = 1, 2, …, N, obtain an estimate $\hat{\theta}^*_i$ from the ith sample of size n obtained by sampling with replacement from the original sample values $(x_1,...,x_n)$.

The expected value of $\hat{\theta}$ will be estimated by

$$\hat{\theta}^*(.) = \frac{1}{N}\sum\hat{\theta}^*_{(i)}$$

The bias correction is estimated by

$$\hat{B}_\theta = \hat{\theta}^*_{(.)} - \hat{\theta}$$

The sample standard deviation of the $\hat{\theta}^*_{(.)}$ is then

$$\hat{SD}^* = \left[\frac{1}{N-1}\sum\left(\hat{\theta}^*_{(i)} - \hat{\theta}^*_{(.)}\right)\right]^{1/2}$$

3. Results and Discussion

The collected dataset was analyzed by SAS 9.3. The analysis of variance estimator method provides the value of $\hat{\rho}^*_{AOV}$ as 0.0061 and its corresponding value of $\hat{h}^2$ is 0.0244 with estimated standard error equals to 0.0^35. By the method given by Fleiss (1981) the $\hat{\rho}^*_{AOV}$ value is found to be 0.0253 and the corresponding value of $\hat{h}^2$ is 0.0213 with its standard error equals to 0.0^34. Nevertheless, bootstrap technique provides the estimated standard error of $\hat{h}^2$ for $\hat{\rho}^*_{AOV}$ as 0.0549 and 0.0545 for the sample sizes 500 and 1000, respectively. Likewise, the bootstrap standard error of $\hat{h}^2$ for $\hat{\rho}^*_{AOV}$ is obtained as 0.0537 and 0.0537 for sample size 500 and 1000, respectively. Following the methods given by Kleinman (1973), $\hat{\rho}_{KEQ}$ value was found to be 0.0143 and the corresponding $\hat{h}^2$ value is 0.0573. Also $\hat{\rho}_{KEQ}$ value is found to be 0.013 and corresponding $\hat{h}^2$ value is 0.0523. Again bootstrap technique provides the estimated standard error of $\hat{h}^2$ for $\hat{\rho}_{KEQ}$ as 0.0251 and 0.0235 for the sample sizes 500 and 1000, respectively. Likewise, the bootstrap standard error of $\hat{h}^2$ for $\hat{\rho}_{KEQ}$ is obtained as 0.0283 and 0.0258 for sample size 500 and 1000, respectively. Thus, the above results affirmed that the standard errors of heritability estimates in case of $\hat{\rho}_{KEQ}$ and $\hat{\rho}^*_{KEQ}$ estimators are slightly decreasing as the sample size increased from 500 to 1000. Weller *et al.* (1997) reported heritability estimates of mastitis disease to be in the range from 0.02 to 0.04. The heritability estimates obtained in case of $\hat{\rho}_{KEQ}$ and $\hat{\rho}^*_{KEQ}$

lies outside this range. Hence, it can be concluded that method based on moment estimators $\hat{\rho}_{KEQ}$ and $\hat{\rho}^{*}_{KEQ}$ are biased for heritability estimation of mastitis disease. Since the direct formula of standard error for estimate of heritability by moment estimators is not available, the bootstrap technique is advantageously used. The standard errors of ANOVA estimators estimated by bootstrap technique are found larger as compared to the $\hat{\rho}_{KEQ}$ and $\hat{\rho}^{*}_{KEQ}$ estimators. From all the results obtained, it can be concluded that ANOVA estimators are better than moment estimators $\hat{\rho}_{KEQ}$ and $\hat{\rho}^{*}_{KEQ}$ for the estimation of heritability of mastitis disease.

References

Bhatia, V. K., P. Narain and P. K. Malhotra (1992). Genetic Parameters of stability of different categories of dairy Animals. *Jour. Ind. Soc. Agril. Statistics*, **44(1)**, 122-137.

Behera, S. K., A. K. Paul and S. D. Wahi (2010). Estimation of heritability of mastitis disease using ANOVA method. *IUP J. Genet. Evol*., **3(1)**, 24-29.

Carlén, E., U. Emanuelson and E. Strandberg (2006). Genetic evaluation of mastitis in dairy cattle using linear models, threshold models and survival analysis : a simulation study. *J. Dairy Sci*., **89(10)**, 4049-57.

Carlén, E., E. Strandberg and A. Roth (2004). Genetic parameters for clinical mastitis, somatic cell score and production in the first three lactations of Swedish holstein cows. *J. Dairy Sci.*, **87(9)**, 3062-70.

Dempster, E. R. and I. M. Lerner (1950). Heritability of threshold characters. *Genetics*, **35**, 212-236.

Efron, B. (1979). Bootstrap Methods : Another look at the Jackknife. *The Annals of Statistics*, **7(1)**, 1-26.

Falconer, D. S. and F. C. Mackay Turdy (1998). *Introduction to Quantitative Genetics*. Longman, England.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*, 2nd edition. New York: Wiley.

Gernand, E., P. Rehbein, U. U. V. Borstel and S. König (2012). Incidences of and genetic parameters for mastitis, claw disorders and common health traits recorded in dairy cattle contract herds. *J. Dairy Sci*., **95(4)**, 2144-56.

Gianola, D. (1979). Heritability of polychotomous characters. *Genetics*, **93(4)**, 1051-1055.

Kleinman, J. C. (1973). Proportions with extraneous variance: Single and independent samples. *Journal of the American Statistical Association*, **68**, 46-54.

Magnussen, S. and A. Kremer (1995). The beta-binomial model for estimating heritability of binary traits. *Theor. Appl. Genet*., **91**, 544-552.

Ridout, S. Martin, G. B. Demetrio Clarice and David Firth (1999). Estimating intra class correlation for binary data. *Biometrics*, **55(1)**, 137-148.

Van Vleck, L. D. (1980). Stability as a categorical trait and by considering other traits. *J. Dairy Sci*., **63**, 1172-1180.

Vazquez, A. I., K. A. Weigel, D. Gianola, D. M. Bates, M. A. Perez-Cabal, G. J. Rosa and Y. M. Chang (2009). Poisson versus threshold models for genetic analysis of clinical mastitis in US Holsteins. *J. Dairy Sci*., **92(10)**, 5239-47.

Weller J. L. and E. Ezra (1997). Genetic analysis of somatic cell score and female fertility of Israeli Holsteins with an individual animal model. *J. Dairy Sci*., **80**, 586-593.