

## ESTIMATION OF VARIANCE COMPONENTS FOR BODYWEIGHT OF GRASSCUTTERS USING EXPECTATION MAXIMIZATION (EM) ALGORITHM OF RESTRICTED MAXIMUM LIKELIHOOD (REML)

I. UDEH

Department of Animal Science, Delta State University, Asaba Campus

Correspondence: [drudeh2005@yahoo.com](mailto:drudeh2005@yahoo.com); 08063867396

### ABSTRACT

The objective of this study was to estimate the variance components and heritability of bodyweight of grass-cutters at 4, 6 and 8 months of age using EM algorithm of REML procedures. The data used for the study were obtained from the bodyweight records of 20 grass-cutters from 4 families at 4, 6 and 8 months of age. The heritability of bodyweight of grasscutters at 4, 6 and 8 months of age were 0.14, 0.10 and 0.12 respectively. This implies that about 10 – 14 % of the phenotypic variability of body weight in this grasscutter population was accounted by additive genetic variance while environmental and gene combination variance made a larger contribution. The implication is that selection of grasscutters in this population should not be based on the information on the animals alone but also information from its relatives.

**Key words:** bodyweight, grasscutters, heritability, REML.

### INTRODUCTION

Grasscutters (*Thryonomys swinderianus*), also known as cane rat, is a fairly large rodent that inhabits sub-saharan Africa, mainly west, central, south and some parts of East Africa (Adenyo *et al*, 2012). Grasscutter is hunted aggressively for its meat because people relish the meat which is a local delicacy in Nigeria (Udeh and Isiorho, 2015). Efforts are being made over the years to domesticate the grasscutters in order to augment the protein need of West African people (Adenyo, 2014). Genetic improvement of grasscutters is needed for economically important traits such as bodyweight, litter size, feed efficiency, etc. Precise estimation of variance components and heritability are important in grasscutter breeding to help in predicting the breeding value of the animals available for selection. For balanced data set, estimation of variance components is implemented by the method of least square or analysis of variance which is similar to REML, although the variance-covariance matrices of the two methods are different (Swallow and Monahan, 1984). However, for unbalanced data, variance components can be estimated by the procedure of maximum likelihood (ML) or restricted maximum likelihood. The two methods are basically the same except that REML takes into account the degree of freedom used for estimating fixed effects when estimating variance components while ML does not (Meyer,

1987). This makes REML estimates of variance components unbiased compared to ML. REML is itself not a method but a goal. The goal of REML is to maximize the likelihood in any possible way (Olvido, 2003). Many methods are used to maximize log likelihood in REML and they include Expectation Maximization, derivative free, Newton Raphson and Average information. This paper deals with the estimation of variance components for bodyweight of grasscutters using EM algorithm of REML.

### MATERIALS AND METHODS

**The study area:** The study was carried out at the grasscutter unit of teaching and research farm, Department of Forestry and Wildlife, Delta State University, Asaba Campus.

**The data:** The pedigree file was arranged chronologically according to animal identity, sire and dam. There were 28 animals in the pedigree. The first 8 animals were parents and did not have records. The remaining 20 animals were progenies of parents and had records. The data file consists of record of bodyweight arranged according to families. Each family has 1 male and 4 female.

**Statistical analysis:** The following mixed linear model was used to fit the data

$$Y = Xb + Za + e$$

Where Y = vector of bodyweight, b = vector of fixed effect associated with family and sex, a =

vector of random additive genetic effect,  $e$  = vector of random errors.  $X$ ,  $Z$  = incidence matrices relating fixed effects and random animal effect respectively to the observation. It is assumed that the expectation of the random variables  $E(Y) = Xb$ ,  $E(a) = E(e) = 0$ . It is also assumed that the residual effects ( $e$ ) which

include random environmental effects and non additive genetic effects are independently distributed with variance  $\sigma_e^2$ , therefore,  $\text{var}(e) = I\sigma_e^2 = R$ ,  $\text{var}(a) = A\sigma_a^2 = G$  and  $\text{cov}(a,e) = \text{cov}(e,a) = 0$ , where  $A$  is the numerator relationship matrix (Mrode,2005).  $\text{Var}(Y) = V = ZGZ' + R$ .

The following mixed model equations were set up and solved to obtain the solution to  $\hat{a}$  and  $\hat{b}$ .

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} X'Y \\ Z'Y \end{pmatrix}$$

Where  $\alpha = 1-h^2/h^2$  or  $\sigma_e^2/\sigma_a^2$ .

REML estimates of variance components using the EM algorithm were obtained by iteration using the following expressions:  $\sigma_a^2 = [\hat{a}'A^{-1}\hat{a} + \text{tr}(A^{-1}C)\sigma_e^2]/q$  and  $\sigma_e^2 = [Y'Y - b'X'Y - \hat{a}'Z'Y]/(N - r(X))$ , where  $N$  is the number of observations,  $q$  is the number of individuals and  $C$  is the part of the inverse of the mixed model equations that corresponds with the random effect ( $C = Z'Z + A^{-1}\alpha$ ). The prior values used for this estimation were  $\sigma_e^2 = 0.4$  and  $\sigma_a^2 = 0.2$ . The prior values were obtained from Mrode (2005). Narrow sense heritability was computed from the expression:  $h^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$ . Estimation of variance components via EM algorithm of REML was implemented using R (version 64. 3:1:3).

## RESULTS AND DISCUSSION

The variance components and heritability of bodyweight of grasscutters at 4, 6 and 8 months of age are shown in Table 1. The number of iterations on convergence varies among the bodyweight groups. It was highest in BWT at 6 months of age (549). Usually EM algorithm takes longer time to converge compared with other REML estimation methods such as Newton Raphson, Fisher's Scoring and Average information algorithm (Misztal, 2008). However, it has the property of always yielding positive estimates as long as prior values are positive (Harville, 1977). The heritability of bodyweight of grasscutters at 4, 6 and 8 months of age were low and ranged from 0.10 to 0.14. This implies that about 10 – 14 % of the phenotypic

variability of bodyweight of grasscutters at 4 – 8 months of age was accounted by additive genetic variance while environmental and non additive genetic variance made a larger contribution. This means that selection of animals in this population should not be based on information on the animals alone but also information on its relatives. It should be noted that heritability is a property of a population and the environmental circumstances the population experienced. In a related study, Annor *et al* (2012) reported heritability estimates of 0.48, 0.66 and 0.84 for bodyweight of grasscutters at 4, 6 and 8 months of age respectively using maternal animal model. The authors concluded that opportunity exists for fast genetic gain in the improvement of growth and bodyweight traits in grasscutters. Similarly, Henry *et al* (2014) reported heritability estimates of 0.56 and 0.22 for litter weights at birth of grasscutters from dam component and sire and dam components of variance.

## CONCLUSION

It was concluded that the estimation of variance components for bodyweight of grasscutters via EM algorithm of REML yielded positive estimates. The heritability of bodyweight computed from the variance components yielded low estimates that range from 0.10 to 0.14.

## REFERENCES

- Adenyo, C; Hayano, A; Inoue, E; Kayang, B. B and Murayama, M. I. (2012). Development of

- microsatellite markers in grasscutters (*Thryonomys swinderianus*, Rodenta) using next generation sequencing technology. Conservation Genetics Resource vol. 4 issue 4: 1011 – 1014.
- Adenyo, C. (2014). Genetic diversity studies of grasscutters (*Thryonomys swinderianus*) in Ghana by microsatellite and mitochondrial markers. Citation: Kyoto University Research Information Repository. URL: <http://hdl.handle.net/2433/188533>.
- Annor, S. Y; Ahunu, B. K; Aboagye, G. S; Boa-Amponsem, K and Cassady, J. P. (2012). Phenotypic and genetic parameter estimates for grasscutter production traits. I. (Co)variance components and heritability. Global Advanced Research Journal of Agricultural Science. Vol. 1(6): 148-155.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Stat. Ass. 72: 320-338.
- Henry, A. J; Ibe, S. N and Asuquo, B. O. (2014). Estimation of genetic parameters for litter trait in grasscutters (*Thryonomys swinderianus*). Anim. Prod. Sci. 54(10): 1762-1764.
- Meyer, K. (1987). Restricted maximum likelihood to estimate variance components for mixed models with two random factors. Genet. Sel. Evol. 19(1): 49-68.
- Misztal, I. (2008). Reliable computing in estimation of variance components. J. Anim. Breed. Genet. 125(6): 363-370.
- Mrode, R. A. (2005). Linear models for the prediction of animal breeding values. 2<sup>nd</sup> Edition. CABI publishing.
- Olivido, A. (2003). MTDFREML workshop notes. Revised. 15 pages.
- R Core Team. (2015). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL: <http://www.R-project.org/>
- Swallow, W. H and Monahan, J. F. (1984). Monte Carlo Comparison of ANOVA, MIQUE, REML and ML estimators of variance components. Technometrics. Vol. 26(1): 47-57.
- Udeh, I and Isiorho, E. (2015). Repeatability of docility in grasscutters (*Thryonomys swinderianus*). Journal of Agriculture and Food Environment. Vol. 2 (1 & 2): 37-40.

**Table 1: REML estimates of variance components and heritability for bodyweight of grasscutters at 4, 6 and 8 months of age.**

| Bodyweight | No of iterations | $\sigma_e^2$ | $\sigma_a^2$ | $h^2$ |
|------------|------------------|--------------|--------------|-------|
| 4 months   | 415              | 0.0588       | 0.0099       | 0.14  |
| 6 months   | 549              | 0.1133       | 0.0131       | 0.10  |
| 8 months   | 480              | 0.0833       | 0.0113       | 0.12  |

Note:  $\sigma_e^2$  = residual variance,  $\sigma_a^2$  = additive genetic variance