Available online at www.equijost.com

Equity Journal of Science and Technology, 2023 10(2): 1 - 4

ISSN 2354-1814; E-ISSN 2683-5961

EQUIJOST *An Official Publication of Kebbi State University of Science and Technology, Aliero, Nigeria*

Comparative analysis of some variance estimation methods on the effect of feeding male and female Albino Rat

*Biodun T. Efuwape, Ahmed O. Olasupo, Wunmi A. Abimbola-Okuneye, Ruth Ale
Olabisi Onabanjo University, Ago Iwoye, Ogun State, Nigeria
*Corresponding Author's email: efuwape.biodun@oouagoiwoye.edu.ng

Abstract
A comparison of three methods for variance components estimation to compute the individual weight in a kilogram of male and female albino rats placed in five different cages after feeding them for a week in the Animal farmhouse of Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria. An analog weighing scale was used to measure the weight, and measurements were taken five times during the week-long experiment. Analysis of variance (ANOVA), Maximum likelihood (ML) and Restricted maximum likelihood (REML) estimations were applied to assess the accuracy and reliability of the weight measurements. However, the comparison shows that estimate of ML and REML are the same, the estimates of the three methods used are not within the same parameter space. It was concluded that for a balanced one-way model, the ANOVA method is better for estimating variance in Animal for both Male and Female.

Keywords: Analysis of variance (ANOVA), Maximum Likelihood (ML), Restricted Maximum Likelihood Estimation (RMLE), Variance components, Parameter space

1. Introduction

The method of estimating variance component was introduced by Anderson et al. [1]. It refers to the processes involved in efficiently calculating the variability within responses or value. It is often used in population genetics and also applied in animal breeding. There have been contradictions on the best method to use in other to estimate the variance component of animals. This problem has led in this research to use the method of Analysis of variance, Maximum likelihood and Restricted maximum likelihood so as to know the relatively best method for estimating the variance in animal for both male and female. The results of Mouhamadou [2] indicated that there is a statistically significant relationship with strong effect size between safety and security index and human development using ANOVA. In their study Ajah et al. [3] also using ANOVA method, indicated that the effect of grazing livestock on cowpea and soybean production significantly varied from one area to another with kwali area council as the most affected while Abaji was the least.

The maximum likelihood estimation did not consider the loss in degrees of freedom that originated from estimation of the fixed effects. Which may result into fixed effect with many levels. These problems can be control by using REML method which was first proposed by Thompson et al. [4]

The main disparity between the Bayesian and ML approaches is the way in which they deal with nuisance parameters [5]

2. Materials and method

2.1 The Study area

The study was conducted at the animal house of Olabisi Onabanjo University, Ago-Iwoye, which served as the primary data collection site. The experiment focused on male and female albino rats and aimed to investigate the effects of feeding on their weight. The animals were housed in five different cages to monitor their individual changes throughout the week. At the commencement of the experiment, all rats were acclimated to the animal house environment for a brief period, the cages were designed to provide ample space for rat to comfortably access food and water. In order to have high level of precision and sensitivity, an analog weighing scale was used and the experiment spanned a week and recorded the weight of each rat five times throughout the week at predetermined intervals in the year 2020.

2.2 Analysis of Variance

Analysis of Variance was used to compare the equality of three or more means. The model of one-way ANOVA can be given as;

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad (1)$$

where, $i = 1,2, \dots k$ and $j = 1,2, \dots n$, ε_{ij} is normally distributed with mean μ and variance σ^2.

2.2 Maximum Likelihood (ML) Estimate

The maximum likelihood (ML) method is a method that is commonly used in statistics in estimating parameters of distribution. These estimators maximize the likelihood of the parameters given the density functions and the data. Also, the following can be obtained to solve the maximum likelihood estimated using the inverse of the co-variance matrix

$$/\varepsilon/ {\sigma^2}_e^{k(n-1)}(\sigma_e^2+n\sigma_\mu^2)^k \qquad (2)$$

$$log_e\ /\varepsilon/\ = k(n-1)\ log_e\ (\sigma_e^2) + klog_e\left(\sigma_e^2+n\sigma_\mu^2\right) \qquad (3)$$

$$(y-lnk\mu)'\varepsilon^{-1}(y-lnk\mu) = \frac{nk(\bar{y}-\mu)}{\sigma_e^2}+\frac{SSE+SSt}{\sigma^2+n\sigma^2} \qquad (4)$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{ns}\left(y_{ij}-\bar{y}\right)^2 \qquad (5)$$

$$SSt = n_s\sum_{i=1}^{k}(\bar{y_i}-\bar{\bar{y}})^2 \qquad (6)$$

Again, using the expressions:

$$-2log_e\left[l\left(\mu,\sigma_e^2,\frac{\sigma_\mu^2}{y}\right)\right] \qquad (7)$$

Equation 2.4 can be written as:

$$e^{(\mu,\sigma_e^2,\frac{\sigma_\mu^2}{y})} = knlog_e(2\bar{x})+k(n-1)log_e(\sigma_e^2) + knlog_e\left(\sigma_e^2+n\sigma_\mu^2\right) + \frac{nk(\bar{y}-\mu)^2+SSE+SSt}{\sigma_e^2+n\sigma_\mu^2} \qquad (8)$$

Likelihood equation is obtained by differentiating $L(\mu,\sigma_e^2,\ \sigma_\mu^2/y)$ with respect to $\mu,\sigma_\mu^2,\sigma_e^2$ and setting the derivatives equal to zero.

$$\frac{\delta l(\mu,\sigma_\varepsilon^2,\frac{\sigma_\mu^2}{y})}{\delta\mu} = \frac{2nk(\bar{y}-\mu)}{\sigma_\varepsilon^2+n\sigma_\mu^2}=0 \qquad (9)$$

$$\frac{\delta l(\mu,\sigma_\varepsilon^2,\frac{\sigma_\mu^2}{y})}{\delta l(\sigma_\varepsilon^2+n\sigma_\mu^2)\overset{2}{\leftarrow}} = \frac{k(n-1)}{(\sigma_\varepsilon^2)\overset{2}{\leftarrow}} + \frac{k-nk(\bar{y}-\mu)\overset{2}{\leftarrow}+SSE+SSR}{\sigma_\varepsilon^2+n\sigma_\mu^2} = 0 \qquad (10)$$

$$\frac{\delta l(\mu,\sigma_\varepsilon^2,\frac{\sigma_\mu^2}{y})}{\delta\mu} = \frac{nk-n^2\,k(\bar{y}-\tilde{\mu})-nSSR}{\sigma_\varepsilon^2+n\sigma_\mu^2} = 0 \qquad (11)$$

The maximum likelihood estimates include:

$$\hat{\mu} = \tilde{\mu} = \bar{y},\ \tilde{\sigma}_\varepsilon^2 = \hat{\sigma}_\varepsilon^2 = \frac{SSE}{k(n-1)} \qquad (12)$$

Therefore, when the estimates of $\sigma^2\mu$ is zero, then the estimate of σ^2_ε is recomputed by pooling SSE and SSt as well as their degrees of freedom to obtain

$$\partial_E^2 = \frac{SSE+SSt}{k(n-1)} \qquad (13)$$

The estimate treatment variance component may be negative. In such a case, it is proposed to either consider zero in case of a negative estimate.

2.3 Restricted Maximum Likelihood (REML)

To obtain the REML estimates, iteration was used. REML estimators were within the parameter space by definition therefore negative estimates were set to zero.

In simple problems where solutions to variance components are closed-form we can remove the bias post hoc by multiplying a correlation factor. However, for complex problems where closed-form solutions do not exist, we need to resort to a more general method to obtain a bias-free estimation for variance components. REML is one of such methods.

Using the one-way random effect model in equation (1), the $-2log_e$ of the likelihood function can be expressed as functions of the sufficient statistic $\bar{y}$, Sum of square error (SSE) and Sum of Square of treatment (SSt) as

$$e^{(\mu,\sigma_\varepsilon^2,\frac{\sigma_\mu^2}{y})} = kn\log_e(2\bar{x}) + k(n-1)\log_e\sigma_\varepsilon^2 + kn\log_e(\sigma_\varepsilon^2+n\sigma_\mu^2)+\frac{nk(\bar{y}-\mu)\overset{2}{\leftarrow}}{\sigma_\varepsilon^2+n\sigma_\mu^2}+\frac{SSE}{\sigma_\varepsilon^2}+\frac{SSt}{\sigma_\varepsilon^2+n\sigma_\mu^2} \qquad (14)$$

$\{[\mathrm{nk}(\bar{y}-\mu)^2/(\sigma_\varepsilon^2+n\sigma_\mu^2) + \log_e(2\bar{x}) + \log_e(\sigma_\varepsilon^2+n\sigma_\mu^2)] + [(\mathrm{kn}-1)\log_e(2\bar{x}) + (n-1)\log_e(\sigma_\varepsilon^2) + (\mathrm{k}-1)\log_e(\sigma_\varepsilon^2+n\sigma_\mu^2) + \frac{SSE}{\sigma_\varepsilon^2}+\frac{SSR}{\sigma_\varepsilon^2+n\sigma_\mu^2}\}$ (15)

$$e^{(\mu,\sigma_\varepsilon^2,\frac{\sigma_\mu^2}{y})} + e^{(\mu,\sigma_\varepsilon^2,SSE,SSt)} \qquad (16)$$

Where;

$e^{(\mu,\sigma_\varepsilon^2,SSE,SSt)} = [(\mathrm{kn}-1)\log_e(2\bar{x}) + (\mathrm{n}-1)\log_e(\sigma_\varepsilon^2) + (\mathrm{k}-1)\log_e(\sigma_\varepsilon^2+n\sigma_\mu^2) + \frac{SSE}{\sigma_\varepsilon^2}+\frac{SSR}{\sigma_\varepsilon^2+n\sigma_\mu^2}\}$ 2.15

The restricted maximum likelihood equations were obtained by differentiating $e^{(\mu,\sigma_\mu^2,\sigma_\varepsilon^2,SSE,SSt)}$ with respect to σ^2_μ and σ^2_ε and then setting the derivatives equal to zero.

$$\frac{\delta l(\mu,\sigma_\varepsilon^2,SSE,SSt)}{\delta\sigma_\mu^2} = \frac{k(n-1)}{\sigma_\varepsilon^2+n\sigma_\mu^2}-\frac{SSE}{\sigma_\varepsilon^2}-\frac{SSt}{\sigma_\varepsilon^2+n\sigma_\mu^2} = 0 \qquad 2.16$$

$$\frac{\delta l(\mu,\sigma_\varepsilon^2,SSE,SSt)}{\delta\sigma_\mu^2} = \frac{n(k-1)}{\sigma_\varepsilon^2+n\sigma_\mu^2}-\frac{nSSt}{\sigma_\varepsilon^2+n\sigma_\mu^2} = 0 \qquad (17)$$

The RMLE solution is

$$\sigma_E^2\frac{SSE}{k(n-1)} = MSE \qquad (18)$$

and

$$\sigma_\mu^2 = \frac{1}{n} = \left[\frac{SSt}{k-1} - \text{MSE}\right]$$

$$= \frac{1}{n}[MSt - MSE] \qquad (19)$$

$$\tilde{\sigma}_E^2 = \hat{\sigma}_E^2 \frac{SSE}{K(n-1)} = MSE \qquad (20)$$

Where, $\hat{\sigma}_\varepsilon^2 = \tilde{\sigma}^2_\mu$ if $\tilde{\sigma}^2_\mu \geq 0 = 0$ if $\tilde{\sigma}^2_\mu < 0$

The estimate of $\hat{\sigma}_\varepsilon^2$ is recomputed by pooling SSE and SSt as well as their degree of freedom to obtain.

$$\hat{\sigma}_E^2 = \frac{SSE + SSt}{kn-1} \qquad (21)$$

3. Results and discussion

Performing the Analysis of Variance (ANOVA) for Male rat

$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \frac{T^2}{N}$

$$135.13^2 + 118.40^2 + \cdots + 125^2 - \frac{(135.13 + 118.40 + \cdots + 125.84)^2}{20}$$

$$SST = 4581.99$$

$$SSt = \frac{1}{n_i}\sum T_i^2 - \frac{T^2}{N}$$

$$\frac{550.79^2}{4} + \frac{478.46^2}{4} + \frac{372.97^2}{4} + \frac{446.8^2}{4} + \frac{498.91^2}{4} - \frac{(135.13 + 118.40 + \cdots + 125.84)^2}{20}$$

$$SSt = 4346.65$$

$$SSE = SST - SSt$$

$$SSE = 235.34$$

To estimate the variance components σ^2_ε and σ^2_μ.

σ^2_ε= Mean square error = 15.69

$$\sigma_\mu^2 = \frac{(MSt - MSE)}{N}$$

$\sigma_\mu^2 = 53.55$

K = 5, n = 2

Also, performing Analysis of Variance(ANOVA) for female rat

$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \frac{T^2}{N}$

$$136.83^2 + 127.18^2 + \cdots + 117.55^2 - \frac{(136.83 + 127.18 + \cdots + 117.55)^2}{20}$$

$$SST = 7486.01$$

$$SSt = \frac{1}{n_i}\sum T_i^2 - \frac{T^2}{N}$$

$$\frac{546.6^2}{4} + \frac{523.44^2}{4} + \frac{391.5^2}{4} + \frac{608.42^2}{4} + \frac{455.24^2}{4} - \frac{(135.13 + 118.40 + \cdots + 125.84)^2}{20}$$

$$SSt = 7210.19$$

$$SSE = SST - SSt$$

$$SSE = 275.83$$

To estimate the variance components σ^2_ε and σ^2_μ.

σ^2_ε= Mean square error = 18.39

$$\sigma_\mu^2 = \frac{(MSt - MSE)}{N} = 89.21$$

Performing maximum likelihood estimator

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{nS}\left(y_{ij} - \bar{y}_i\right)^2$$

$$\widehat{\sigma_\varepsilon^2} = \frac{SSE + SSR}{k(n-1)}$$

Where; SSE = 235.34; SSt = 4346.65

$\bar{Y}$= 117.39

K = 5

n = 20

$$\widehat{\sigma_\varepsilon^2} = 48.23$$

Therefore; $\widehat{\sigma_{MLE}^2} = 48.23$

$$\widehat{\sigma_\mu^2} = \frac{1}{n}\left(\frac{SSt}{K} MSE\right)$$

where;

$$SSt = n_S \sum_{i=1}^{k}(\bar{y}_i - \bar{\bar{y}})^2 = 4346.65$$

$$\widehat{\sigma_\mu^2} = 42.68$$

Performing maximum likelihood estimator:

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{nS}\left(y_{ij} - \bar{y}_i\right)^2$$

$$\widehat{\sigma_{MLE}^2} = \frac{SSE + SSR}{k(n-1)}$$

Where; SSE = 275.83; SSt = 7210.19

$\bar{Y}$= 126.66

K = 5

n = 20

$\hat{\sigma}_\varepsilon^2$ =78.80

Therefore; $\sigma_{MLE}^2 = 78.80$

$$\widehat{\sigma_\mu^2} = \frac{1}{n}\left(\frac{SSR}{K} MSError\right)$$

where;

$$SSR = n_S \sum_{i=1}^{k}(\bar{y}_i - \bar{\bar{y}})^2 = 7210.19$$

$$\widehat{\sigma_\mu^2} = 71.18$$

Performing restricted maximum likelihood estimator

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{nS}\left(y_{ij} - \bar{y}_i\right)^2$$

$$\widehat{\sigma_\varepsilon^2} = \frac{SSE + SSt}{k(n-1)}$$

Where; SSE = 235.34; SSt = 4346.65
$\bar{Y}$= 117.39
K = 5
n = 20
Therefore,

$$\widehat{\sigma_\varepsilon^2} = 48.23$$

$$\sigma_{MLE}^2 = 48.23$$

$$\widehat{\sigma_\mu^2} = \frac{1}{n}\left(\frac{SSR}{K} MSError\right)$$

where;

$$SSR = n_S \sum_{i=1}^{k} (\bar{y_i} - \bar{\bar{y}})^2 = 4346.65$$

$$\widehat{\sigma_\mu^2} = 42.68$$

Performing restricted maximum likelihood estimator:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{nS} \left(y_{ij} - \bar{y_i}\right)^2$$

$$\widehat{\sigma_\varepsilon^2} = \frac{SSE + SSt}{k(n-1)}$$

Where, $SSE = 275.8262; SSt = 7210.187$

$$\widehat{\sigma_\varepsilon^2} = 78.80$$

Therefore; $\widehat{\sigma_{MLE}^2} = 78.80$

$$\widehat{\sigma_\mu^2} = 71.1825$$

4. Conclusions

Considering Male rat, the ANOVA method has 15.69 as its variance for error and 53.55 as its variance for mean. It also shows that the ML and REML has their variances for error to be 48.2314 and their variances for mean to be 42.6820. Also, Female rat shows that the ANOVA method has18.39 as its variance for error and 89.21 as its variance for mean. It also shows that the ML and REML has their variances for error to be78.8001 and their variances for mean to be 71.1825. Thus, the following deductions can be made from the estimates obtained using the ANOVA, ML and REML methods:

i. The estimates of the ML and REML are the same.
ii. The estimates of the three methods of estimating variance component used are not within the parameter space.
iii. The findings shows that ANOVA method seems to give a better estimate than ML and REML because the variances are relatively small compared to that of ML AND REML for a balanced one-way model.

References

1. Anderson RL, Bancroft TA. *Statistical theory in research.* 1st ed. New York: McGraw-Hill; 1952.
2. Mouhamadou TS. Using ANOVA to examine the relationship between safety and Human Development. *Journal of International Business and Economics.* 2014; 2(4):101-106.
3. Ajah J, Oswin CC, Kuntashula E. The effect of grazing livestock on Cowpea (Vigna unguiculata) and Soybean (Glycine max) Production in Nigeria. *International Journal of Agriculture Innovations and Research.* 2015; 4(1): 218-229.
4. Thompson R, Brian C, Alison S, Gilmour AR. A Sparse Implementation of the Average Information Algorithm for Factor Analytic and Reduced rank Variance Model. *Australian and New Zealand Journal of Statistics.* 2003; 45(4): 445-459.
5. Ebru KB, Mehmet ZF. Comparison of Methods of Estimating of Variance Components in Nested Design. *Anadolu University Journal of Science and Technology B- Theoretical Science*. 2015; 4(1): 1-10