Jordan Journal of Mathematics and Statistics (JJMS) 11(1), 2018, pp 69 - 91

EFFICIENCY OF ADAPTIVE METHODS USING SIMULATED ALPHA SKEW NORMAL TWO-STAGE DATA

NESREEN M. AL-OLAIMAT (1) AND LOAI M. A. AL-ZOU'BI (2)

Abstract. Two-stage sampling improves variances for estimators of means and regression coefficients because of intra-class homogeneity. To choose the appropriate way of allowing for clustering using sample data, an adaptive method will be evaluated in this paper based on testing the null hypothesis that the variance component of the random effect is zero. Rejecting the null hypothesis, clustering will be allowed for in variance estimation; otherwise clustering will be ignored. The data will be simulated from alpha-skew normal distribution with different values of the parameter.

1. Introduction

Two-stage sampling designs are used instead of simple random sampling (SRS) when the population is too large or scattered ([15]). Two stage sampling is a sampling technique which is obtained by dividing the population into groups (clusters) called primary sampling units (PSUs), then selecting a random sample from each PSU ([6]). Cluster sampling is a two stage sampling but selecting all PSUs in the first stage ([24]). The advantages of two stage sampling are: obtaining a list of all clusters

2000 *Mathematics Subject Classification.* 40H05, 46A45.

Key words and phrases. Primary Sampling Units, Two-stage Sampling, Linear Mixed Model, Alpha Skew Normal Distribution, Balanced Data, Huber-White estimator.

This work is part of the master thesis written by Nesreen Al-olaimat, supervised by Dr. Loai Al-Zoubi.

Received: Sept. 14, 2017 Accepted: Dec. 7, 2017 .

may be easier and cheaper than obtaining a list of all elements in the population ([36]). It saves time as the data can be collected and summarized quickly. It is more flexible than one stage sampling with respect to the types of information that can be obtained. It gives more accurate results comparing with the other sampling ([6]).

A complication of two-stage sampling is that: values of a variable of interest may tend to be more similar for units from the same PSU than for units from different PSUs. The intraclass correlation (ICC), ρ, is a measure of the association between the observations for members of the same PSU. If the intraclass correlation is non-zero, the clustered nature of the design should be reflected in the analysis procedure. One way of doing this is by fitting a multilevel model (MLM) ([12]).

The intra-class correlation is defined as the ratio of the variance between clusters (σ_b^2) to the total variance ($\sigma_b^2 + \sigma_e^2$) ([38]). It is given by $\rho = \frac{\sigma_b^2}{\sigma_b^2+\sigma_e^2}$. Which can be estimated as: $\hat{\rho} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2+\hat{\sigma}_e^2}$.

In practice the intraclass correlation is often quite small. For example, if units within PSUs are no more homogenous than units over all PSUs, then the intraclass correlation is zero. On the other hand, if units from the same PSU have equal values then the intraclass correlation is 1.

Adaptive methods using LM, LMM and Huber-White for Normal data were used by [3]. These methods were used for Log-normal data ([3]), for Exponential data [2], and for Skew Normal data ([1]). They tested $H_0 : \sigma_b^2 = 0$. If H_0 is rejected they used the LMM or the Huber-White methods to estimate the variance of $\widehat{\beta}$. If H_0 is not rejected, the LM is used to estimate the variance of $\widehat{\beta}$. [3] used another adaptive method based on testing $H_0 : \sigma_b^2 = 0$ and comparing $\widehat{deff}$ to predefined cutoff values (d). If H_0 is rejected and $\widehat{deff} \geq d$ they used the LMM or the Huber-White methods to estimate the variance of $\widehat{\beta}$. Otherwise they used the LM to estimate the variance of $\widehat{\beta}$.

This paper is divided into 9 sections. In Section 2 we have defined the model. The alpha-skew normal distribution is explained in Section 3. Estimation of model parameters using the likelihood theory is discussed in Section 4. In Section 5 the robust Huber-White estimator of $var(\widehat{\beta})$ is discussed. While in Section 6, we have explained the test of $H_0 : \sigma_b^2 = 0$ using the restricted likelihood ratio test. In Section 7 the adaptive methods used throughout this paper and the confidence intervals of the estimated regression parameters are explained. In Section 8, a simulation study is conducted to test the superiority of the adaptive methods. Finally, concluding remarks and future work are included in Section 9.

2. Multilevel analysis

Multilevel models are generalizations of regression models which provide an alternative type of analysis for univariate or multivariate analysis of repeated measures ([14]). Models with fixed effects only, which assume that all observations are independent, are not suitable to analyze some types of correlated data, specifically for clustered data. To analyse multilevel data, cluster effect must be considered and added to the regression model to take into account the correlation of the data. The resulting model is called a linear mixed model ([23]).

A linear mixed model (LMM) is a mix of fixed effects and random effects where the fixed effects called regression coefficients, describe the relationship between the dependent variable and predictor variables for an entire population of units of analysis, the fixed effects are assumed to be unknown fixed quantities in a LMM and we estimate them, but the random effects are random values associated with the levels of a random factor in LMM, the random effects are represented as random variables in a LMM, which is displayed in Equation (2.1) below, containing the random effect b_i and the error e_{ij}. Random effects are not directly estimated but they are estimated according to variance and covariance ([41]).

[12] defined the two-level linear mixed model (LMM) as:

$$y_{ij} = \beta^{'} \mathbf{x_{ij}} + b_i + e_{il}, \quad i = 1, 2, ..., c, \; j = 1, 2, ..., m, \tag{2.1}$$

where y_{ij} is the response of j^{th} unit in PSU i, c is the number of clusters or PSUs in the sample, m is the number of observations selected from PSU i, $\mathbf{x}_{ij}$ is the independent vector of j^{th} unit in PSU i for fixed effects, β is the vector of unknown regression coefficients, b_i is the random effect and e_{ij} is the error term.

A simple special case of (2.1) is the intercept-only model, which assumes that all $x_{ij} = 1$, $\forall\, i, j$, where

$$y_{ij} = \beta + b_i + e_{ij}, \; i = 1, 2, ..., c, \;\; j = 1, 2, ..., m \tag{2.2}$$

([33]).

In this paper, we will assume that b_i has an alpha-skew normal distribution with skewness parameters $a = 0$, ± 1, ± 2 and the error term e_{ij} has an alpha-skew normal distribution with skewness parameter $a = 0$, as well.

3. Alpha Skew Normal distribution

The normal distribution is one of the most important and most widely used distribution in statistics, the density function of the normal probability distribution is symmetric about the mean, the normal probability distribution appears in a wide rang of fields. For example, in Biology, it has been observed that the normal probability distribution fits data on the heights and weights of human and populations ([31]). We know that the probability density function (*pdf*) and cumulative distribution function (*cdf*) of the standard normal random variable are given as:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \;\; \Phi(x) = \int_{-\infty}^{x} \phi(u) du. \tag{3.1}$$

The Skew Normal distribution is an asymmetric distribution which is extended the normal distribution through a shape parameter a; they reduce to the standard normal distribution when $a = 0$ ([4]). The *pdf* of the skew normal distribution is defined as:

$$f_a(x) = 2\phi(x)\Phi(a). \tag{3.2}$$

[28] introduced a new family of distributions, it is called alpha-skew normal (ASN) family, which is a skew-symmetric distribution that is flexible enough to support both unimodal and bimodal shape, this family is denoted as $\{ASN(a) : a \in \mathbb{R}\}$. The *pdf* of the random variable X which follow $ASN(a)$ is defined as:

$$f(x) = \frac{(1 - ax)^2 + 1}{2 + a^2}\phi(x), \quad x \in \mathbb{R},\ a \in \mathbb{R}. \tag{3.3}$$

The ASN distribution has the following properties:

(1) The first and second moments and the variance of the $ASN(a)$ are given by

$$\begin{aligned} \mu = E(X) &= -\frac{2a}{2 + a^2}, \\ E(X^2) &= \frac{2 + 3a^2}{2 + a^2}, \\ and\, \sigma^2 &= \frac{3a^4 + 4a^2 + 4}{(2 + a^2)^2} \end{aligned}$$

(2) There is no close form for the maximum likelihood estimate of the skewness parameter a.

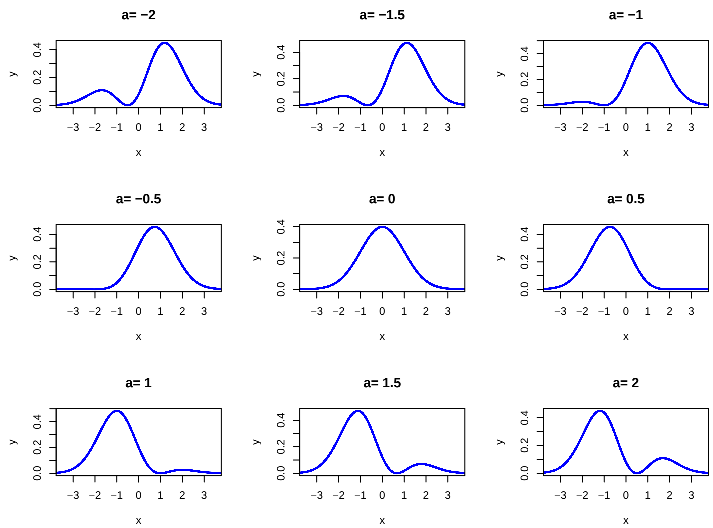
(3) $X \sim N(0, 1)$ when the skewness parameter $a = 0$;

(4) $X \sim BN$ if the skewness parameter $a \Rightarrow \pm\infty$ with $f(x) = x^2\phi(x)$, $x \in \mathbb{R}$, where BN stands for the bivariate normal distribution.

(5) $-X \sim ASN(-a)$.

Figure 1 shows the graph of the alpha-skew normal distribution for different skewness parameters $a = \pm 2, \pm 1.5, \pm 1, \pm 0.5$ and 0. When $a = 0$, it is clear that the distribution is symmetric about 0. This means that the distribution reduces to standard normal

FIGURE 1. Alpha Skew Normal distribution with different skewness parameters



distribution. It is skewed to the left when $a = -2, -1.5, -1,$ and -0.5 and to the right when $a = 0.5, 1, 1.5$ and 2. When $a = \pm 1$, the property of bi-modality starts appear clearly. This property appears much clearer when the absolute value of a increases.

4. Likelihood theory estimation of model parameters

The principle tool for estimating regression coefficient β, variance component σ_b^2 and error term variance, σ_e^2 in multilevel modeling is the maximum likelihood method ([26]). This method is used because it has many advantages as it is generally robust and produces estimates that are asymptotically efficient and consistent, moreover; its estimates has lower variances than other methods. The function used to estimate the variance components σ_b^2 and σ_e^2 is called the restricted maximum likelihood ($REML$) function ([16]). The $REML$ method is basically the same as the maximum likelihood method except for one difference: the $REML$ method takes into account the degrees of freedom used for estimating fixed effects when estimating variance components, while the maximum likelihood method does not use it ([44]).

The REML estimate of $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\beta}}_{ols} = \big(\sum_{i=1}^{c} \mathbf{x}_i'\mathbf{x}_i\big)^{-1} \sum_{i=1}^{c} \big(\mathbf{x}_i'\mathbf{y}_i\big), \tag{4.1}$$

The estimated variance of the regression coefficient $\hat{\boldsymbol{\beta}}$ is defined as:

$$\widehat{var}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{c} \mathbf{x}_i' \widehat{V}_i^{-1} \mathbf{x}_i \right)^{-1} \tag{4.2}$$

where

$$V_i = \sigma_b^2 J_m + \sigma_e^2 I_m \tag{4.3}$$

where J_m is an $m \times m$ matrix with all entries equal to 1, and I_m is an $m \times m$ identity matrix ([33]). Which simplifies when we have balanced data and the intercept-only model to:

$$\widehat{var}(\widehat{\beta}) = \frac{1}{c} \left[\widehat{\sigma}_b^2 + \frac{\widehat{\sigma}_e^2}{m} \right], \tag{4.4}$$

where

$$\begin{aligned} \hat{\sigma_e^2} &= min\bigg(MSE, \, \frac{n-c}{n-1} \, MSE + \frac{c-1}{n-1} \, MSA \bigg); \\ \hat{\sigma_b^2} &= \frac{1}{m} max(MSA - MSE, \, 0); \\ \hat{\beta} &= \bar{y}_{..}. \end{aligned}$$

([34]).

5. Huber-White estimation of $var(\widehat{\beta})$

The ordinary least squares (OLS) is one of the simplest methods for estimating the parameters of multilevel regression models ([16]). This method is applied by minimizing the sum of squares of vertical distance between actual and predicted values. An alternative of ML and $REML$ estimates of $\widehat{var}(\widehat{\beta})$ is the robust variance estimator which was first suggested by [17, 42] then introduced by [25] who applied it to longitudinal data using generalized estimating equations (GEE). This approach is often used for cluster sampling provide that the observations within PSUs might be correlated and the observations in different PSUs are independent. The robust estimator procedure is a general method for estimating the covariance matrix of parameter estimates, which yields asymptotically consistent covariance matrix estimates without making distribution assumptions and even if the assumed model (2.1) is incorrect ([5]).

The estimator $\widehat{var}(\hat{\boldsymbol{\beta}})$ in (4.2) will be approximately unbiased provided that the variance model (4.3) is correct. If this is not the case, $\widehat{var}(\hat{\boldsymbol{\beta}})$ will be biased and inference will be incorrect. This approach can be referred to as robust or Huber-White variance estimation ([17, 42]).

An alternative estimator of V_i is $\widehat{V}_i^{Hub} = \widehat{e}_i\widehat{e}_i'$ where $\widehat{e}_i = y_i - x_i'\widehat{\beta}$, $\widehat{V}_i^{Hub}$ is approximately unbiased for V_i ($i.e$ $E(\widehat{V}_i^{Hub}) \approx V_i$). But

$$var(\hat{\boldsymbol{\beta}}) = var((\sum_{i=1}^{c} \mathbf{x}_i'\widehat{V}_i^{-1}\mathbf{x}_i)^{-1}(\sum_{i=1}^{c} \mathbf{x}_i'\widehat{V}_i^{-1}\mathbf{y}_i)). \tag{5.1}$$

([33]).

Therefore, the Huber-White variance estimator for $\hat{\boldsymbol{\beta}}$ is defined by [25] as:

$$\widehat{var}_{Hub}(\hat{\boldsymbol{\beta}}) = (\sum_{i=1}^{c} \mathbf{x}_i'\widehat{V}_i^{-1}\mathbf{x}_i)^{-1}(\sum_{i=1}^{c} \mathbf{x}_i'\widehat{V}_i^{-1}\widehat{V}_i^{Hub}\widehat{V}_i^{-1}\mathbf{x}_i)(\sum_{i=1}^{c} \mathbf{x}_i'\widehat{V}_i^{-1}\mathbf{x}_i)^{-1} \tag{5.2}$$

where $\widehat{V}_i^{Hub}$ is the robust variance estimator of V_i.

Using the intercept-only with balanced data case, Equation (5.2) reduces to

$$\widehat{var}_{Hub}(\widehat{\beta}) = \frac{1}{c(c-1)}\sum_{i=1}^{c}(\overline{y}_{i.} - \overline{y}_{..})^2 \tag{5.3}$$

6. Testing $H_0 : \sigma_b^2 = 0$ in the linear mixed model

One of the main problems in mixed models is the inclusion or exclusion of the random effect term in the model because of the location of its variance component

(σ_b^2) on the boundary of the parameter space under the null hypothesis H_0. Such tests was first worked by [37] and [39].

The work of [37] focused on deriving the properties of the likelihood by allowing the true parameter value to be on the boundary of its parameter space. Under non-standard conditions, the asymptotic null distribution of the likelihood ratio tests will not have the typical χ^2 distribution. Although, it has been proved that the correct asymptotic distribution is a mixture of χ^2 distributions assuming that response variables are *iid*. [39] applied the results of [37] specifically to test for $H_0 : \sigma_b^2 = 0$ in linear mixed models. They also showed that likelihood ratio test for σ_b^2 has an asymptotic $0.5\chi_0^2 : 0.5\chi_1^2$. This means that under H_0 the χ^2 statistic has mixture distribution of $50 : 50$ χ^2 with zero and one degrees of freedom under the null and alternative hypothesis if the data are *iid*. This is because there is a 50% chance of finding a positive estimate if the null hypothesis is true and a 50% chance of finding a negative estimate. Since the variance can't be negative, usual procedure is to change the negative estimate to zero with a 50% chance of getting zero estimate ([21]). For the important special case of the null hypothesis $H_0 : \sigma_b^2 = 0$; the distribution of the likelihood ratio test (LRT) is the most commonly used test because its desirable theoretical properties and the fact that it is easy to construct, which can be computed using restricted maximum likelihood ($REML$) ([43]). The restricted likelihood ratio test given by the restricted log-likelihood function for the balanced data with the intercept-only model the restricted likelihood ratio test $RLRT$ is defined as:

$$\Lambda = \begin{cases} (n-1)\log(\frac{n-c}{n-1} + \frac{c-1}{n-1}F) - (c-1)\log(F), & \text{if } F > 1 \\ 0 & , \text{ if } F \leq 1 \end{cases} \tag{6.1}$$

where $F = \frac{MSB}{MSE}$, with MSB is the mean square between clusters and MSE is the mean square error within clusters ([40]), where,

$$MSE = \frac{1}{n-c} \sum_{i=1}^{c} \sum_{j=1}^{m} (y_{ij} - \bar{y}_{i.})^2, \text{and } MSB = \frac{m}{c-1} \sum_{i=1}^{c} (\bar{y}_{i.} - \bar{y}_{..})^2$$

.

7. Adaptive strategies

There are number of possible approaches for estimating regression coefficients and their variances when the intraclass correlation (ρ) is thought to be small or has been estimated as a small value. One approach is to fit a linear mixed model regardless. Another is to fit a linear model assuming independent observations, i.e. ρ=0. However, if the sample design is relatively clustered, that is a large number of final units are selected from each PSU, the estimated variances resulting from a linear mixed model can be much larger than those obtained from a linear model assuming independent observations, leading to wider confidence intervals. Moreover, a linear mixed model is more complicated to fit and explain than a simple linear model, so the latter is preferable provided it does not give misleading inference. This paper will explore a third alternative: an adaptive strategy based on testing the null hypothesis that the PSU-level variance component, σ_b^2, is zero. If the null hypothesis is not rejected we use the linear model for estimating the variances of the estimated regression coefficients $\hat{\boldsymbol{\beta}}$. On the other hand, if the null hypothesis is rejected we use the estimated variance for $\hat{\boldsymbol{\beta}}$ either using the standard likelihood theory variance estimator for the LMM or the Huber-White method.

In this paper we will test the methods applied in [3] using non-normal data. Data will be simulated from the alpha-skew normal distribution to see if these methods are still working.

The confidence interval in LMM_s contain confidence intervals for fixed effects and confidence intervals for variance components ([19]). When using the intercept-only model (2.2) in the balanced data case, the confidence interval for β can be defined as

$$(1-\alpha)100\%CI = \widehat{\beta} \pm t_{(df,\, 1-\frac{\alpha}{2})}\widehat{SE}(\widehat{\beta}), \tag{7.1}$$

where $\widehat{SE}(\widehat{\beta}) = \sqrt{\widehat{var}(\widehat{\beta})}$ ([3]). Since the degrees of freedom are unclear, so we need to estimate them from the data, there are several method to estimate the degree of freedom such as Satterthwaite's approximation ([35]), the approximations of ([20]) and [10]. [9] suggested the following form which is based on a scaled t-distribution:

$$(1-\alpha)100\%CI = \widehat{\beta} \pm \delta^{-1} t_{(v,\, 1-\frac{\alpha}{2})} \widehat{SE}(\widehat{\beta}) \tag{7.2}$$

where δ is a scaled factor which is given by

$$\delta = \sqrt{\frac{\nu}{(\nu-2)\widehat{V}(T)}}$$

with ν is the degrees of freedom which is approximated as:

$$\hat{\nu} = \sum_{i=1}^{c} \frac{m}{1+(m-1)\widehat{\rho}} - 1 \tag{7.3}$$

and $\widehat{V}(T)$ is the approximation of the variance of the t-statistic T which is given by:

$$\widehat{V}(T) = 1 + \frac{\widehat{\beta}^2}{4(\widehat{var}(\widehat{\beta}))^3} \widehat{var}(\widehat{var}(\widehat{\beta})) \tag{7.4}$$

where $T = \frac{\widehat{\beta}}{\sqrt{\widehat{var}(\widehat{\beta})}}$.

8. Simulation Study

A simulation study is conducted to evaluate the adaptive methods and to compare them to the non-adaptive methods. Data are generated from Alpha Skew-Normal distribution based on the intercept only model (2.2) assuming each PSU have the same number of observations. The intraclass correlation (ρ), is varied over a range of values of 0, 0.025 and 0.1. The number of PSUs (c) are varied over a range of values of 2, 5 and 25, with number of observations (m) equals to 2, 10 and 50 per PSU. The skewness parameters (a) will have a range of values of 0, ± 1, and ± 2. 1000 samples are generated. The values of σ_b^2 and σ_e^2 are set to be $\frac{\rho}{1-\rho}$ and 1 respectively; to ensure that the intraclass correlation is ρ. We used R statistical programme which is a computer language and software environment for statistical computing and graphics ([30]). The $lmer()$ function which can be found in the $lme4$ packages in R is used to fit the LMM. The lm function is used to fit the linear model as well. These functions are used to estimate the regression coefficients $\widehat{\beta}$ and the estimates of their variances ([41]).

The adaptive approach, which is another approach, relies on the idea of testing $H_0 : \sigma_b^2 = 0$ in model 2.2. If we do not reject H_0 then the linear model is used to estimate $var(\widehat{\beta})$. On the other hand, if we reject H_0, the linear mixed model or the Huber-White robust method of estimation is used to estimate $var(\widehat{\beta})$. 90% confidence intervals was calculated for the LMM method using the method of [10].

Tables 1 - 10 show the number of PSUs, c, and the number of observations per PSU, m. The ratio between the expected value of the estimate of the variance of $\widehat{\beta}$ and the variance of $\widehat{\beta}$ $(E(\widehat{var}(\hat{\beta}))/var(\hat{\beta}))$ using four methods of estimation (ADM, ADH, LMM and $Huber - White$) are also shown. Non-coverage rates for testing $H_0 : \sigma_b^2 = 0$ are also appeared in the tables as well as the lengths of the confidence intervals calculated using the four methods.

Based on these tables, the Huber-White variance ratios are, in general, unbiased regardless the value of ρ and a. The non-coverage rates are close to the nominal rate (10%) as well.

For $\rho = 0$ and regardless of the value of a, the ADM and ADH variance ratios were unbiased in all cases, except when there were 2 sample PSUs with all values of m and when $c = 5$ with values of m of 5, 10 and 50. The LMM method gives bias estimators of $var(\widehat{\beta})$ when $c \leq 5$ for all values of m.

Non-coverage rates, in this case, were close to the nominal rate for adaptive methods in all cases, except when $c = 2$ with $m = 10$, $c = 5$ with $m = 10$ and 50 and when $c = 25$ with $m = 2$ with about 5-20%.

Table 1: Variance ratios, average length and non-coverage of the 90% confidence intervals for β, and power of testing $H_0 : \sigma_b^2 = 0$ using RLRT ρ=0 and all skewness parameters.

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.175	1.175	1.36	0.964	9.9	9.9	13	10.6	10.5	5.071	2.935	5.658	4.961
2	10	1.316	1.316	1.568	1.065	8.4	8.4	9.1	9.9	5.2	0.866	1.062	0.947	2.235
2	50	1.189	1.19	1.465	1.019	10.4	10.4	9.4	8.2	4.5	0.377	0.451	0.423	1.043
5	2	1.078	1.078	1.195	1.013	9.7	9.7	10.4	10.7	10.5	1.189	1.193	1.214	1.282
5	10	1.196	1.196	1.319	1.081	7.9	7.9	8.1	8.3	7.5	0.502	0.506	0.512	0.569
5	50	1.156	1.156	1.279	1.033	8.1	8.1	8.6	9.8	7.2	0.223	0.225	0.229	0.254
25	2	0.949	0.949	0.989	0.924	11.1	11.1	11.6	11.1	10.9	0.478	0.478	0.474	0.478
25	10	1.05	1.05	1.061	0.994	9.1	9.1	9.8	9.7	7.6	0.213	0.213	0.21	0.213
25	50	0.959	0.959	0.984	0.91	10.9	10.9	10.9	11.8	8	0.096	0.095	0.095	0.096

Table 2: ρ=0.1 and skewness parameter $a = -2$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.105	1.105	1.292	0.949	8.3	8.4	12.6	9.2	10	5.353	2.939	5.854	5.095
2	10	1.205	1.205	1.483	1.059	9.8	9.8	10	9.1	5.1	0.851	1.048	0.953	2.346
2	50	1.157	1.158	1.395	1.041	11.9	11.9	11.7	10.1	7.8	0.416	0.548	0.466	1.124
5	2	1.03	1.03	1.134	0.952	11	11	11.2	11	11.5	1.203	1.215	1.225	1.292
5	10	1.091	1.091	1.211	0.999	8.8	8.8	9.9	10.1	7.9	0.507	0.512	0.519	0.577
5	50	1.06	1.06	1.195	1.035	10.1	10.1	10.7	8.2	11	0.233	0.236	0.243	0.276
25	2	0.933	0.933	0.975	0.903	10.2	10.2	11.4	11.5	10	0.48	0.48	0.478	0.479
25	5	1.029	1.029	1.039	0.992	9.9	10	10.4	10.2	12.7	0.307	0.306	0.304	0.307
25	50	0.938	0.938	0.979	0.972	11.8	11.6	11.4	10.1	24.5	0.101	0.101	0.102	0.105

Table 3: Variance ratios, average length and non-coverage of the 90% confidence intervals for β, and power of testing $H_0 : \sigma_b^2 = 0$ using RLRT ρ=0.025 and skewness parameter $a = -1$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.277	1.277	1.506	1.091	8.4	8.4	11.8	10.5	10.2	5.264	2.977	5.743	5.124
2	10	1.157	1.157	1.394	0.959	10.1	10.1	11.2	8.5	5.1	0.858	1.059	0.942	2.271
2	50	1.048	1.048	1.298	1.019	15.5	15.5	13.2	8.6	8.2	0.424	0.564	0.486	1.217
5	2	1.011	1.011	1.113	0.933	10.5	10.4	10.1	11.3	11.8	1.204	1.214	1.224	1.289
5	10	1.141	1.141	1.272	1.077	9.2	9.2	9.8	9.3	10.1	0.514	0.52	0.53	0.591
5	50	1.046	1.047	1.182	1.044	12.8	12.8	11.9	9.7	15.4	0.244	0.249	0.258	0.29
25	2	1.004	1.004	1.049	0.975	9.7	9.6	10.1	9.8	10.9	0.483	0.483	0.482	0.483
25	10	0.98	0.98	0.991	0.966	11	11	11.2	11.3	11.7	0.216	0.216	0.214	0.22
25	50	0.947	0.947	0.99	0.994	11.7	11.7	11.2	10	35.5	0.105	0.105	0.107	0.11

Table 4: ρ=0.1 and skewness parameter $a = -1$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.184	1.184	1.392	1.055	9.6	9.6	12.3	10	11.9	5.525	3.123	6.123	5.348
2	10	1.122	1.122	1.377	1.017	12.1	12.1	11.9	9.4	7.3	0.904	1.177	1.017	2.483
2	50	0.995	0.996	1.177	1.037	23.7	23.7	19.3	10.9	18.7	0.587	0.912	0.667	1.55
5	2	1.144	1.145	1.265	1.065	8.5	8.5	9.1	9.9	12.6	1.219	1.228	1.244	1.304
5	10	1.05	1.05	1.175	1.048	11.1	10.9	10.7	8.5	16.2	0.556	0.564	0.582	0.652
5	50	0.935	0.935	1.01	0.983	16.4	16.1	13.3	10.5	43.4	0.337	0.345	0.356	0.389
25	2	0.922	0.922	0.965	0.906	11.6	11.5	11.2	11.7	14	0.488	0.489	0.488	0.492
25	10	0.946	0.946	0.965	0.989	11.2	11.3	11.8	10.4	31.2	0.235	0.233	0.235	0.243
25	50	0.971	0.971	0.974	0.976	11.1	10.8	11.2	10.4	94.7	0.147	0.147	0.148	0.148

Table 5: Variance ratios, average length and non-coverage of the 90% confidence intervals for β, and power of testing $H_0 : \sigma_b^2 = 0$ using RLRT ρ=0.025 and skewness parameter $a = 0$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.202	1.202	1.414	1.023	8.3	8.3	12.5	10.7	11.1	4.91	2.964	5.351	5.009
2	10	1.125	1.125	1.376	1.009	11.7	11.7	11.8	8.8	7	0.9	1.16	1.009	2.469
2	50	0.95	0.95	1.134	0.993	24.7	24.7	20.3	11.2	18	0.571	0.879	0.651	1.527
5	2	0.984	0.984	1.091	0.919	10.7	10.7	10.8	11.6	11.6	1.207	1.218	1.233	1.301
5	10	0.973	0.973	1.089	0.966	13	12.9	12.7	9.9	15.3	0.549	0.557	0.575	0.643
5	50	0.982	0.982	1.058	1.025	16	15.3	14.5	10.5	43.4	0.332	0.34	0.349	0.381
25	2	1.025	1.025	1.077	1.011	9.8	9.8	10.2	9.5	11.7	0.488	0.489	0.488	0.492
25	10	0.985	0.985	1.006	1.028	11	11.4	10.8	11	27.3	0.231	0.23	0.231	0.24
25	50	0.996	0.996	1.002	1.004	11	10.9	11.1	10.6	91.7	0.143	0.143	0.144	0.144

Table 6: ρ=0.1 and skewness parameter $a = 0$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.091	1.091	1.318	1.015	9.5	9.5	13.9	11	10.3	5.14	3.145	5.825	5.545
2	10	0.887	0.887	1.055	0.9	20.3	20.2	17.7	10	15.2	1.119	1.748	1.266	3.141
2	50	0.895	0.895	0.945	0.92	29.1	28.6	23.9	11.2	45.6	1.281	2.126	1.352	2.622
5	2	1.057	1.057	1.183	1.056	11.1	11	10.6	9.8	16.5	1.293	1.308	1.342	1.434
5	10	0.953	0.953	1.044	1.006	14.6	14.1	13.4	10.2	38.2	0.711	0.729	0.751	0.827
5	50	0.95	0.95	0.96	0.958	13.4	11.9	12.4	10	84.4	0.642	0.631	0.649	0.647
25	2	0.959	0.959	1.009	0.972	10.6	10.4	10.8	10.3	21.7	0.521	0.522	0.525	0.533
25	10	0.974	0.974	0.979	0.989	11.2	11.4	10.8	10	86.5	0.308	0.306	0.309	0.31
25	50	1.054	1.054	1.054	1.054	9	8.7	9	8.7	100	0.244	0.244	0.244	0.244

Table 7: Variance ratios, average length and non-coverage of the 90% confidence intervals for β, and power of testing $H_0 : \sigma_b^2 = 0$ using RLRT ρ=0.025 and skewness parameter $a = 1$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.277	1.277	1.506	1.091	8.4	8.4	11.8	10.5	10.2	5.264	2.977	5.743	5.124
2	10	1.157	1.157	1.394	0.959	10.1	10.1	11.2	8.5	5.1	0.858	1.059	0.942	2.271
2	50	1.187	1.187	1.424	1.089	12.1	12.1	11.4	9	8.9	0.437	0.587	0.488	1.167
5	2	1.011	1.011	1.113	0.933	10.5	10.4	10.1	11.3	11.8	1.204	1.214	1.224	1.289
5	10	1.141	1.141	1.272	1.077	9.2	9.2	9.8	9.3	10.1	0.514	0.52	0.53	0.591
5	50	1.046	1.047	1.182	1.044	12.8	12.8	11.9	9.7	15.4	0.244	0.249	0.258	0.29
25	2	1	1	1.044	0.97	9.8	9.7	10.2	9.9	10.9	0.483	0.483	0.482	0.483
25	10	0.98	0.98	0.991	0.966	11	11	11.2	11.3	11.7	0.216	0.216	0.214	0.22
25	50	0.947	0.947	0.99	0.994	11.7	11.7	11.2	10	35.5	0.105	0.105	0.107	0.11

Table 8: ρ=0.1 and skewness parameter $a = 1$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.184	1.184	1.392	1.055	9.6	9.6	12.3	10	11.9	5.525	3.123	6.123	5.348
2	10	1.122	1.122	1.377	1.017	12.1	12.1	11.9	9.4	7.3	0.904	1.177	1.017	2.483
2	50	0.995	0.996	1.177	1.037	23.7	23.7	19.3	10.9	18.7	0.587	0.912	0.667	1.55
5	2	1.002	1.002	1.11	0.943	10.4	10.4	10.4	10.6	11.2	1.204	1.215	1.23	1.303
5	10	1.05	1.05	1.175	1.048	11.1	10.9	10.7	8.5	16.2	0.556	0.564	0.582	0.652
5	50	1.071	1.071	1.2	0.978	12.6	12.6	12.1	11	6.3	0.222	0.224	0.23	0.256
25	2	0.922	0.922	0.965	0.906	11.6	11.5	11.2	11.7	14	0.488	0.489	0.488	0.492
25	10	0.947	0.947	0.967	0.988	11.2	11.3	11.8	10.6	31.3	0.235	0.233	0.235	0.243
25	50	0.971	0.971	0.974	0.976	11.1	10.8	11.2	10.4	94.7	0.147	0.147	0.148	0.148

Table 9: Variance ratios, average length and non-coverage of the 90% confidence intervals for β, and power of testing $H_0 : \sigma_b^2 = 0$ using RLRT ρ=0.025 and skewness parameter $a = 2$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.063	1.063	1.257	0.869	9.9	9.9	13.6	10.1	7.8	4.466	2.781	4.953	4.871
2	10	1.273	1.273	1.55	1.055	8.6	8.6	9	8.7	4.9	0.849	1.03	0.941	2.258
2	50	1.236	1.236	1.484	1.037	9.5	9.5	9.3	10	5.3	0.384	0.471	0.424	1.03
5	2	1.175	1.175	1.284	1.074	9.6	9.6	9.9	11.8	13.9	1.219	1.231	1.238	1.287
5	10	1.157	1.157	1.284	1.042	8.4	8.4	8.2	9.3	7.5	0.5	0.505	0.512	0.564
5	50	1.034	1.034	1.161	0.966	11	11	10.3	10.1	7.1	0.223	0.225	0.231	0.261
25	2	1.02	1.02	1.064	0.985	9.6	9.6	10.6	10.2	11.1	0.481	0.482	0.479	0.48
25	10	0.994	0.994	1.003	0.941	9.7	9.8	10.7	11.3	7.7	0.213	0.213	0.21	0.213
25	50	0.905	0.905	0.935	0.885	11.3	11.4	12	12.3	11.6	0.097	0.097	0.096	0.098

Table 10: ρ=0.1 and skewness parameter $a = 2$

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non-Coverage of CI for β (%)				Pr(Rej H_0) (%)	Confidence Interval Length			
c	m	VADM	VADH	VLMM	VHub	TADM	TADH	TLMM	THub	RLRT	CADM	CADH	CLMM	CHub
2	2	1.105	1.105	1.292	0.949	8.3	8.4	12.6	9.2	10	5.353	2.939	5.854	5.095
2	10	1.205	1.205	1.483	1.059	9.8	9.8	10	9.1	5.1	0.851	1.048	0.953	2.346
2	50	1.157	1.158	1.395	1.041	11.9	11.9	11.7	10.1	7.8	0.416	0.548	0.466	1.124
5	2	1.03	1.03	1.134	0.952	11	11	11.2	11	11.5	1.203	1.215	1.225	1.292
5	10	1.091	1.091	1.211	0.999	8.8	8.8	9.9	10.1	7.9	0.507	0.512	0.519	0.577
5	50	1.06	1.06	1.195	1.035	10.1	10.1	10.7	8.2	11	0.233	0.236	0.243	0.276
25	2	0.934	0.934	0.976	0.904	10.2	10.2	11.4	11.5	10	0.48	0.48	0.478	0.479
25	10	1.058	1.058	1.071	1.031	8.4	8.4	9.5	8.8	9.2	0.215	0.214	0.213	0.217
25	50	0.938	0.938	0.979	0.972	11.8	11.6	11.4	10.1	24.5	0.101	0.101	0.102	0.105

For $\rho = 0.025$ and $a = 0$, the variance estimates using ADM and ADH methods are unbiased except for $c = 2$ with $m \leq 10$. In this case, the LMM variance estimator is unbiased, except for $c = 2$ with all values of m. For $\rho = 0.1$ and $a = 0$, the ADM and ADH variance estimates are still unbiased except for $c = 2$ with $m = 10$. In this case, the variance estimates using the LMM method were unbiased as well, except for $c = 2$ with $m \leq 10$.

In case of $a = 0$ and $\rho \neq 0$, non-coverage rates increase obviously with high rates when there were 5 or less PSUs with $m \geq 10$ observations per PSU. This rate of increase was less for other values of c.

For $a = 2$ and $\rho = 0.025$, the variance estimates using the ADM and ADH methods of estimation were unbiased, except when $c = 2$ with $m \geq 10$ and when $c = 5$ with $m \leq 10$. The LMM variance estimators were biased for almost all values of $c \leq 5$ with all values of m. The ADM and ADH non-coverage rates were close to the nominal rate (10%), except when $c = 2$ with $m = 10$ and 25, for $c = 5$ with m 10, as well as for $c = 25$ with $m = 50$.

When $\rho = 0.1$, The ADM and ADH variance estimators were unbiased, except for $c = 2$ with m =2 and 50. The LMM methods give biased estimators for $c \leq 5$ with all values of m.

Non-coverage rates using the ADM and ADH, as well as the LMM methods of estimation were in this case, (a=2), away from the nominal rate (10%) with about 5-20%.

In case of $a = 1$ with $\rho = 0.025$, the ADM and ADH variance estimators were unbiased, except when there were 2 PSUs with all numbers of observations each. The LMM variance estimators were biased for all $c \leq 5$ regardless the number of observations each.

The variance ratios using the adaptive methods of estimation were unbiased when $\rho = 0.1$ except when there 2 PSUs with 2 observations each and when there were

5 PSUs with 5 observations each. The LMM method gave biased estimators when $c = 2$ with all m, when $c = 5$ and with $m \leq 10$. In this case, non-coverage rates approximately higher than the nominal rate with about 5-135% for the adaptive methods. For the LMM method, these rates were about 5-90% higher than the nominal rate.

The results for $a = -1$ and $a = -2$ were identical to the results of $a = 1$ and $a = 2$; respectively for all values of ρ.

Confidence intervals get shorter as the number of observations per PSU get higher. Adaptive confidence intervals were, in general shorter than congruent non-adaptive ones with high rate in designs with small number of sample PSUs.

9. Conclusion

(1) The Huber-White variance ratios were approximately unbiased for all values of ρ and a. The non-coverage rates were close to the nominal rate (10%). This is because the degrees of freedom for this method are exact and equal to c-1.

(2) The ADH method perform similar to the Huber-White method except for the extreme designs with $c \leq 5$. The Huber-White method has wider confidence intervals than the ADH method, especially in the extreme designs.

(3) The ADM confidence intervals are noticeably narrower than the LMM for $c \leq 5$. The variance estimates using the ADM method of estimation were approximately unbiased, except for ($c \leq 5$). The non-coverage rates were higher in these designs than other designs.

    Therefore, we recommend avoiding designs with 5 or less PSUs, even if ρ is thought to be low.

(4) $a = \pm 2$ gave the best non-coverage rates, they were close to the nominal rate (10%) with a rate of about 5-20%.

(5) $a = 0$ gave unbiased variance estimators using the adaptive methods of estimation except for the extreme designs with approximately $c = 2$ with $m \leq 10$.

(6) There was no discrimination between the results of positive and negative values of a.

Acknowledgement

We would like to thank the editors and the referees.

References

[1] R. M. Abushrida, L. M. Al-Zoubi, A. I. Al-Omari,(2014). Robustness of adaptive methods for balanced non-normal data: Skewed normal data as an example. Revista Investigation Operacional, 35(2):157-172.

[2] L. M. Al-Zoubi, (2015). Adaptive inference for multi-stage unbalanced exponential survey data based on a simulation from an intercept-only model. Electronic Journal of Applied Statistical Analysis (EJASA), 8(2):136-153.

[3] L. M. Al-Zou'bi, R. G. Clark, D. G. Steel, (2010). Adaptive inference for multi-stage survey data. Communications in Statistics - Simulation and Computation, 39(7):1334-1350.

[4] A. Azzalini, (1985). Aclass of distributions which includes the normal ones. Scandanavian Journal of Statistics, 12(2):171-178.

[5] R. J. Carroll, S. Wang, D. G. Simpson, A. J. Stromberg, and D. Ruppert,(1998). The sandwich (robust covariance matrix) estimator. Technical report.

[6] W. G. Cochran,(1977). Sampling Techniques. John Wiley and Sons, Inc.

[7] D. Commenges, H. Jacqmin, (1994). The intraclass correlation coefficient: Distribution-free definition and test. Biometrics, 50(2):517-526.

[8] L. Duchateau, P. Janssen, G. L. Rowlands, (1998). Linear Mixed Models. An Introduction with Applications in Veterinary Research. International Livestock Research Institute P.O. Box 30709, Nairobi, Kenya and Center for Statistics, Limburgs University Center, Universitaire Campus, B-3590 Diepenbeek, Belgium.

[9] C. Faes, H. Molenberghs, M. Aerts, G. Verbeke, M. G. Kenward, (2004). The effective sample size and a novel small sample degrees of freedom method. Technical report, 0651, Interuniversity Attraction Pole, http://www.stat.ucl.ac.be/IAP.

[10] C. Faes, H. Molenberghs, M. Aerts, G. Verbeke, M. G. Kenward, (2009). The effective sample size and an alternative small-sample degrees-of-freedom method. The American Statistician, 63(4):389-399.

[11] T. B. Fomby, R. C.Hill, S. R. Johnson, (1984). Advanced Econometric Methods. Springer.

[12] H. Goldstein,(2003). Multilevel Statistical Models. Kendall's Library of Statistics 3. Arnold, London, third edition.

[13] M. H. Hansen, W. N. Hurwitz, W. G. Madow, (1953). Sample Survey Methods and Theory, volume 1, 2. John Wiley and Sons, Inc., New York.

[14] R. H. Heck, S. L. Thomas, (2015). An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus. Routledge, third edition.

[15] Hiremath, S. S. (2011). Text Book of Preventive and Community Dentistry. Elsevier India, 2nd edition.

[16] J. Hox, (2002). Multilevel Analysis. Techniques and Applications. Laurence Erlbaum Associates.

[17] P. J. Huber, (1967). The behavior of maximum likelihood estimates under non-standard conditions. Proceedings of the Fifth Berekley Symposium on Mathematical Statistics and Probability, University of California, Berekley, 11:221-233.

[18] P. Kennedy,(2008). A Guide to Econometrics. MIT Press, 6th edition.

[19] J. Jiang, (2007). Linear and Generalized Linear Mixed Models and Their Applications. Springer Series in Statistics.

[20] M. G. Kenward, J. H. Roger, (1997). Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, 53(3):983-997.

[21] D. M. LaHuis, M. W. Ferguson,(2009). The accuracy of significance tests for slope variance components in multilevel random coefficient models. Organizational Research Methods, 12(3):418-435.

[22] L. Kish, (1965). Survey Sampling. John Wiley and Sons, Inc.

[23] Laird, N. M. and Ware, J. H. (1982). Random-effect models for longitudinal data. Biometrics, 38(4):963-974.

[24] P. S. Levy, S. Lemeshow, (2008). Sampling of Populations: Methods and Applications. A John Wiley & Sons, Inc., Publications, fourth edition.

[25] K. Y. Liang, S. L. Zeger, (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1):13-22.

[26] T. M. Shenk, A. B. Franklin, (2001). Modeling in Natural Resource Management: Development, Interpretation and Application. Island Press, 1st edition.

[27] M. O. D. L. Martin, Kelly, R. J. Adams, (1997). Implementation and analysis-primary and middle school years (population 1 and population 2). Technical report, Center for the Study of Testing, Evaluation, and Educational Policy.

[28] D. Olivero, (2010). Alpha skew normal distribution. Proyecciones Journal of Mathematics, 29(3):224-240.

[29] H. D. Patterson, R. Thompson, (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58(3):545-554.

[30] R Development Core Team (2007). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[31] K. M. Ramachandran, C. P. Tsokos, (2009). Have one to sell? Sell it yourself Details about Mathematical Statistics with Applications in R. Elsevier Science Publishing Co Inc, 2nd edition.

[32] P. S. R. S. Rao, (1997). Variance Components Estimation, Mixed Models, Methodologies and Applications. Chapman and Hall.

[33] H. Sahai, M. M. Ojeda, (2004). Analysis of Variance for Random Models, Balanced Data, Theory, Methods, Applications and Data analysis, volume 1. Birkhauser, Boston.

[34] H. Sahai, M. M. Ojeda, (2005). Analysis of Variance for Random Models, Unbalanced Data, Theory, Methods, Applications and Data Analysis, volume 2. Birkhauser, Boston.

[35] F. E. Satterthwaite, (1941). Synthesis of variance. Psychometrika, 6:309-316.

[36] R. L. Scheaffer, W. Mendenhall, R. L. Ott, K. Gerow, (2012). Elementary Survey Sampling. BROOKS/COLE Cengage Learning, 7th edition.

[37] S. G. Self, K. Y. Liang,(1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. journal of the American Statistical Association, 82(398):605-610.

[38] T. A. B. Snijders, R. J. Bosker, (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. SAGE publication.

[39] D. O. Stram, J. W. Lee,(1994). Variance components testing in the longitudinal mixed effects model. Biometrics, 50(4):1171-1177.

[40] P. M. Visscher, (2006). A note on the asymptotic distribution of likelihood ratio tests to test variance components. Twin Research and Human Genetics, 9(4):490-495.

[41] B. T. West, K. B. Welch, A. T. Galecki,(2015). Linear Mixed Model: A Practical Guide Using Statistical Software. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition.

[42] H. White, (1982). Maximum likelihood estimation of misspeciffied models. Econometrica, 50(1):1-25.

[43] D. Zhang, X. Lin, (2008). Random Effect and Latent Variable Model Selection, volume 192 of Lecture Notes in Statistics, chapter 2, pages 19-36. Springer.

[44] K. Meyer. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genetics Selection Evolution, BioMed Central, 1989, 21 (3), pp.317-340. ¡hal-00893806¿

(2) Al al-Bayt University, Department of Mathematics , Mafraq (25113), Jordan

E-mail address: (1) nesrreenolaimat@gmail.com

E-mail address: (2) loai67@aabu.edu.jo