Independent Group Design Analysis of Heteroscedastic Model

Gebriel M. Shamia
Faculty of Science, University of Benghazi
gmshamia@yahoo.com

Intesar N. El-Saeiti
Faculty of Science, University of Benghazi
entesar.el-saeiti@uob.edu.ly

A b s t r a c t

Unbalanced data for independent groups is a common problem in agricultural research on the basis of organized experiments. These data can be easily analysed using linear mixed models. The main problem with the study of unbalanced data that contributes to heterodcedesity models is that many approaches are available and decision-making between them can be difficult. This article deals with the weighted least squares (WLS) method of estimation when cell variances are not equal to the estimators of the mixed model using a form of proportionality condition for the size of the cell sample using a harmonic mean. Special emphasis is given to the independent group design (IGD) on the presumed mixed model and the variability of the data due to unbalanced data in the environmental conditions of the experiments. Estimators of the parameters for this model are calculated to be independent of the weights under this condition. Approximate tests are used when cell variance is known. Initial samples include estimators of cell variances for unknown cell variances, which are used for the analysis of final samples along with correction or bias where necessary. In addition, the procedure is applied to the actual data set used to illustrate the method.
Keywords: Unequal cell variances; Heteroscedasity; Weighted least square (*WLS*); Proportionality condition; Adjustment for bias.

الملخص

تمثل البيانات غير المتوازنة للمجموعات المستقلة مشكلة شائعة في البحوث الزراعية المبنية علي التجارب العملية. يمكن تحليل هذه البيانات بسهولة باستخدام نماذج مختلطة خطية. تكمن المشكلة الرئيسية في دراسة البيانات غير المتوازنة في وجود مشكلة عدم التجانس. تتناول هذه المقالة طريقة تقدير المربعات الصغرى المرجحة (WLS) للتقدير عندما لا تتساوي تباينات الخلية مع مقدري النموذج المختلط وباستخدام شرط التناسب لحجم عينة الخلية استخدم الوسط التوافقي. تم التركيز وبشكل خاص على تصميم المجموعة المستقلة على النموذج المختلط وتباين البيانات بسبب البيانات غير المتوازنة في الظروف البيئية للتجارب. يتم احتساب مقدرات المعلمات لهذا النموذج وتكون مستقلة عن الأوزان تحت هذا الشرط. وتم التطبيق على بيانات فعلية استخدمت لتوضيح الطريقة.

1. Introduction

For the usual analysis of variance balanced models, the error variance is assumed to be constant. But it may sometimes happen that the error variance may be vary from one group of observation to another when model is unbalanced. Talukder (1978) gave a general block design with a heteroscedasitic model of examples of heteroscedasitic errors for cells. Some authors have dealt with this kind of models, for example, Bhuyan (1984) investigate the problem of combined analysis of a group of experiments with the heteroscedasity of error term. Notable among many other researchers are Bhuyan (1984), El-Saeiti (2004), El-Saeiti and Shamia (2006), and Shamia and El-Saeiti (2007). Talukder (1993) gave a general method of analysis of the weighted least squares (WLS) for factorial designs with unequal cell variances. It has been shown that if cell sample sizes are proportional to cell variances, the same proportionality (constant) value is used for all cells, then WLS parameter estimators can be easily obtained as the ordinary least square (OLS) method. This type of proportionality condition may lead to balance data. In addition, Shamia (1991) investigates this method for ordinary factorial designs such as crossed-nested and split-plot designs. El-Saeiti (2004) describes how to deal with an empty cell problem when the cells interact. The design of an independent group is an inter-subject design. The design that allows certain variables to be measured between participants in environmental conditions and others between participants may be the best approach to some research issues. A researcher wishing to compare several types of treatment may choose a mixed design and use a randomized group design in which participants or subjects in each group (or condition) of the experiment are randomly assigned to two or more groups. The participants in one group have absolutely no ties or links to the participants in the other group. See; Kirk, R.E. (1995), Dean and Voss (1999), Anderson (2001), and Ryan (2007).

In this present article, the same way for independent group design (IGD) for unbalanced data using the harmonic mean of cell variances can be considered due to differences to that groups. The case when each group is applied to different numbers of group-experimental units is discussed. In addition, it is assumed that the set of treatments is nested with each group-experimental unit usually occur because of lost data.

For unknown cell variances, sample are drown in two stage. Initial samples provide unbiased estimators of cell variances, which are used in *WLS* estimation and then carried out for analysis of variances of final sample, along with some adjustment for bias wherever necessary.

2. Statistical procedure

Suppose that there is a particular raw material which is available in independent batches in different places or regions that are not large enough to allow all treatment combinations to be run from the same batch. However, if there is a batch contains enough material with $n_{(ij)k}$ replicates using a separate batch of raw material. The model of such crossed-nested design (three-stage independent group design) with unequal cell variance is given by:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(ij)} + \varepsilon_{l(ijk)}, \qquad (1)$$

$$i = 1,2,...,a,\ j = 1,2,...,b,\ k = 1,2,...,c,\ l = 1,2,...,n_{(ij)k}$$

Where y_{ijkl} is the observed number from *l*-th level (replication), μ is the overall mean, α_i is the effect of the *i*-th level of factor-A as an independent group, β_j is the effect of the *j*-th level of factor-B, $(\alpha\beta)_{ij}$ is the effect of the *ij*-th interaction between *A* and *B* as factor-(AB), $\gamma_{k(ij)}$ is the effect of the *k*-th nested within *ij*-th interaction as factor-C(AB), and $\varepsilon_{l(ijk)}$ is the error term (E) which is $NID(0,\sigma^2_{(ij)k})$.

Specifically, we assumed that factor-A is random, factor-B is fixed, factor-(AB) is random, and factor-C(AB) is random too. This is known as the mixed model structure. The model (1) can be written as

$$y_{ijkl} = \mu + \beta_j + \varepsilon^*_{l(ijk)}. \qquad (2)$$

Where, $\{\mu + \beta_j\}$ is the fixed part of the above model with the constrain:

$\sum_{j=1}^{b} n_{(ij)k} w_{(ij)k} \beta_j = 0$, and $\varepsilon^*_{l(ijk)} = \alpha_i + (\alpha\beta)_{ij} + \gamma_{k(ij)} + \varepsilon_{l(ijk)}$

is the random effect part of the model with

$\alpha_i \sim NID(0,\sigma^2_\alpha)$, $(\alpha\beta)_{ij} \sim NID(0,\frac{(b-1)}{b}\sigma^2_{\alpha\beta})$,

$\gamma_{(ij)k} \sim NID(0,\sigma^2_{\gamma(\alpha\beta)})$, and $\varepsilon_{l(ijk)} \sim NID(0,\sigma^2_{(ij)k})$.

Now let us suppose that

$$w_{(ij)k} = \frac{(1/\sigma^2_{(ij)k})}{\sum_i^a \sum_j^b \sum_k^c (1/\sigma^2_{(ij)k})} = \frac{\hbar}{abc}((\sigma^2_{(ij)k})^{-1} = \frac{\hbar}{abc}\left(\frac{\lambda_i}{n_{(ij)k}}\right)$$

Due to Hartmink et al. (2002), where, $\hbar$ is the harmonic mean of the cell variances, and we assume that $n_{(ij)k} \cong \lambda_i \sigma^2_{(ij)k}$ is the proportional condition. For more details, see Shamia and El-Saeiti (2019).

Here, λ_i is the level-specific proportionality constant which varies with the levels of the factor-*A*. This constant produces orthogonality and partial balance. Such proportional condition is equivalent to $n_{(i.).} \equiv \lambda_i \sigma^2_{(i.).}$ For more details of subject, see Searle (1971), Speed *et al*. (1978), Shamia (1991), and El-Saeiti (2004).

By minimizing the *WLS* due to the error term under the usual constraints for model (2), the estimators of parameters are shown as

$\hat{\mu} = \tilde{y}_{....}$ and $\hat{\beta}_j = (\tilde{y}_{.j..} - \tilde{y}_{....})$.

Where

$\bar{y}_{ij..} = \sum_k^c \frac{\bar{y}_{ijk..}}{c}$, $\tilde{y}_{.j..} = \sum_i^a \frac{\lambda_i \bar{y}_{ij..}}{\lambda_o}$, $\tilde{y}_{....} = \sum_i^a \sum_j^b \sum_k^c \frac{\lambda_i \bar{y}_{ijk.}}{bc\lambda_o}$; and $\lambda_o = \sum_i^a \lambda_{i.}$, $\tilde{\lambda}_o = \frac{1}{a}\sum_i^a \lambda_{i.}$.

The sum of squares (SSq) for any effect (main or interaction) can be obtained by the usual method by taking the sum of products from the effect estimators.

$$SS(\mathrm{A}) = \sum_i^a \sum_j^b \sum_k^c n_{ijk} w_{ijk} \hat{\alpha}_i^2 = \frac{\hbar}{a} \sum_i^a \lambda_i (\bar{y}_{i...} - \tilde{y}_{...})^2,$$

$$SS(\mathrm{B}) = \sum_i^a \sum_j^b \sum_k^c n_{ijk} w_{ijk} \hat{\beta}_j^2 = \frac{\hbar}{b} \tilde{\lambda}_0 \sum_j^b (\tilde{y}_{.j..} - \tilde{y}_{...})^2,$$

$$SS(\mathrm{AB}) = \sum_i^a \sum_j^b \sum_k^c n_{ijk} w_{ijk} (\hat{\alpha}\hat{\beta})_{ij}^2 = \frac{\hbar}{ab} \sum_i^a \sum_j^b \lambda_i (\bar{y}_{ij..} - \bar{y}_{i...} - \tilde{y}_{.j..} + \tilde{y}_{...})^2,\ SS\mathrm{C}(\mathrm{AB}) = \sum_i^a \sum_j^b \sum_k^c n_{ijk} w_{ijk} \hat{\gamma}_{k(ij)}^2 = \frac{\hbar}{abc} \sum_i^a \sum_j^b \sum_k^c \lambda_i (\bar{y}_{ijk.} - \bar{y}_{ij..})^2, \text{ and}$$

$$SS(\mathrm{E}) = SS(\mathrm{T}^*) - \Re(\mu) - SS(\mathrm{A}) - SS(\mathrm{B}) - SS(\mathrm{AB}) - SS\mathrm{C}(\mathrm{AB}).$$

Where

$$SS(\mathrm{T}^*) = \frac{\hbar}{abc} \sum_i^a \sum_j^b \sum_k^c \sum_l^{n_{ijk}} (\lambda_i / n_{ijk})\, y_{ijkl}^2, \quad \text{and}$$

$$\Re(\mu) = \frac{\hbar}{abc} \sum_i^a \sum_j^b \sum_k^c \lambda_i \tilde{y}_{....}^2 = \frac{\hbar}{a} \lambda_0 \tilde{y}_{...}^2.$$

In the case of known weights, the sum of the squares is distributed independently as χ^2 with respect to their degrees of freedom (a-1), (b-1), (a-1)(b-1), (c-1) and (N-abc) respectively. The expected mean square EMS's of the three-stage independent group design (IGD) are:

$$EMS(\mathrm{B}) = \frac{ac}{(b-1)} \upsilon_0 \upsilon_3 \sum_{j=1}^b \beta_j^2 + c\, \upsilon_0 \upsilon_3 \sigma^2_{\alpha\beta} + \upsilon_0 \upsilon_3 \sigma^2_{\gamma(\alpha\beta)} + \upsilon_0,$$

$$EMS(\mathrm{AB}) = c\, \upsilon_0 \upsilon_2 \sigma^2_{\alpha\beta} + \upsilon_0 \upsilon_2 \sigma^2_{\gamma(\alpha\beta)} + \upsilon_0,$$

$$EMS\mathrm{C}(\mathrm{AB}) = \upsilon_0 \upsilon_1 \sigma^2_{\gamma(\alpha\beta)} + \upsilon_0,$$

$$EMS(\mathrm{A}) = bc\, \nu_0 \quad \upsilon_2 \sigma_\alpha^2 + \nu_0 \quad \upsilon_2 \sigma^2_{\gamma(\alpha\beta)} + \nu_0 \quad \text{and} \quad EMS(\mathrm{E}) = \upsilon_0.$$

Where

$$\upsilon_0 = \frac{\hbar}{abc}, \quad \upsilon_1 = \tilde{\lambda}_o, \quad \upsilon_2 = \frac{1}{(a-1)} \left\{ [\lambda_o^2 - \sum_{i=1}^a \lambda_i^2] / \lambda_o \right\}, \quad \text{and} \quad \upsilon_3 = \left\{ (\sum_{i=1}^a \lambda_i^2) / \lambda_o \right\}.$$

It is observed from the above results that an approximate F-test due to Satterthwaite (1946) is required to be tested, if sometimes required, is given by

$$F_0^{i} \cong [MS(\mathrm{A}) - \upsilon_0]^{*}\upsilon_1 \,/\, [MS\mathrm{C}(\mathrm{AB}) - \upsilon_0]^{*}\upsilon_2 \sim f_{\alpha,[p_1;p_2]}$$

with

$$p_1 = [MS(\mathrm{A}) - \upsilon_0]^2 \,/\, [MS(\mathrm{A})^2/(a-1) + \upsilon_0^2/(N-abc)], \quad \text{and} \quad p_2 = [MS\mathrm{C}(\mathrm{AB}) - \upsilon_0]^2 \,/\, [MS\mathrm{C}(\mathrm{AB})^2/ab(c-1) + \upsilon_0^2/(N-abc)].$$

The exact F-test with appropriate denominator for testing $H_0^{ii}: \beta_j = 0 \quad \forall \quad j = 1,2.$ is

$$F_0^{ii} \cong [MS(\mathrm{B}) - \upsilon_0]^{*}\upsilon_2 \,/\, [MS(\mathrm{AB}) - \upsilon_0]^{*}\upsilon_3 \sim f_{\alpha,[p_1;p_2]}$$

with

$$p_1 = [MS(\mathrm{B}) - \upsilon_0]^2 \,/\, [MS(\mathrm{B})^2/(b-1) + \upsilon_0^2/(N-abc)], \quad \text{and} \quad p_2 = [MS(\mathrm{AB}) - \upsilon_0]^2 \,/\, [MS(\mathrm{AB})^2/(a-1)(b-1) + \upsilon_0^2/(N-abc)].$$

For testing the interaction between factor-A and factor-B $\Rightarrow \quad H_0^{iii}: \sigma^2_{\alpha\beta} = 0$, we have

$$F_0^{iii} \cong [MS(\mathrm{AB}) - \upsilon_0]^{*}\upsilon_1 \,/\, [MS\mathrm{C}(\mathrm{AB}) - \upsilon_0]^{*}\upsilon_2 \sim f_{\alpha,[p_1;p_2]},$$ with

$$p_1 = [MS(\mathrm{AB}) - \upsilon_0]^2 \,/\, [MS(\mathrm{AB})^2/(a-1)(b-1) + \upsilon_0^2/(N-abc)], \quad \text{and} \quad p_2 = [MSC(AB) - \upsilon_0]^2 \,/\, [MSC(AB)^2/ab(c-1) + \upsilon_0^2/(N-abc)].$$

For testing the nested factor-C within the factor-AB $\Rightarrow \quad H_0^{iv}: \sigma^2_{\gamma(\alpha\beta)} = 0$, we have

$$F_0^{iv} \cong MS\mathrm{C}(\mathrm{AB}) \,/\, MS(\mathrm{E}) \sim f_{\alpha,[ab(c-1);(N-abc)]}.$$

For the random part, the estimator of variances are

$$\hat{\sigma}^2_{\alpha} = \frac{[MS(\mathrm{A}) - \upsilon_0]^{*}\upsilon_1 - [MS\mathrm{C}(\mathrm{AB}) - \upsilon_0]^{*}\upsilon_2}{\{bc\upsilon_o\upsilon_1\upsilon_2\}}, \qquad \hat{\sigma}^2_{\alpha\beta} = \frac{[MS(\mathrm{AB}) - \upsilon_0]^{*}\upsilon_1 - [MS\mathrm{C}(\mathrm{AB}) - \upsilon_0]^{*}\upsilon_2}{\{c\upsilon_o\upsilon_1\upsilon_2\}},$$

and

$$\hat{\sigma}^2_{\gamma(\alpha\beta)} = \frac{MS\mathrm{C}(\mathrm{AB}) - MS(\mathrm{E})}{\{\upsilon_o\upsilon_1\}}.$$

3. Analysis when Cell Variances are Unknown

In practice when cell variances are not known, these are estimated in two stages. At the first stage, an initial sample of constant size, n_0, is selected independently for every cell as follows

$$\hat{\sigma}^2_{(ij)k} = S^2_{(ij)k} = \sum_{l=1}^{n_o} (y_{ijkl} - \tilde{y}_{ijk.})^2 \,/\, (n_o - 1);$$
$$i = 1,2,..., a; \quad j = 1,2,..., b; \quad k = 1,2,..., c.$$

The actual sample sizes, $n_{(ij)k}$, for the cells are then determined by using the estimated variances in place of actual ones in the proportionality condition.

For our design constant λ_i, we have

$$n_{(i.).} \equiv \hat{\lambda}_i S^2_{(i.).}; \quad i = 1,2,..., a.$$

If the final sample is greater than the initial sample size, then the diferences ($n_{(ij)k}$ - n_o) additional observations are to be drawn independently for that cell at the second stage. If not, the

effective final sample is taken at random from initial sample for analysis by the *WLS* estimation. The analysis is then carried out as before using an adjustment as follows.

The estimated design constant $\hat{\lambda}_i$ can be adjusted for bias using the following relation given by Meier (1953).

$$Ave.\{\Re(x_1, x_2,..., x_p)\} = \Re(1,1,...,1) + \sum_{l=1}^{p}(1/m_l)[\partial^2\Re / \partial x_l^2]_{all\ x_l=1} + 0(\sum_{l=1}^{p} m_l^{-2}).$$

Where $\Re$ is a rational function having no singularities for x_l, and m_l. x_l is distributed as a $\chi^2_{m_l}$, l=1,2,…,p.

Thus $\{\Re(x_1, x_2,..., x_p)\} - \sum_{l=1}^{p}(1/m_l)[\partial^2\Re / \partial x_l^2]_{all\ x_l=1}$

is free from bias of order $(1/m_l)$ and thus estimate actual value $\Re(1,1,...,1)$ is more closely. In practice, population variance in $[\partial^2\Re / \partial x_l^2]_{all\ x_l=1}$ are to be replaced by the corresponding sample variances.

In our case, it will be convenient to define quantities x_{ijk} by $(s^2_{ijk}/\sigma^2_{ijk})$, the ratio of estimated variance based on $n_{(ij)k}$ to the actual variance.

Let $\Re(x_{111}, x_{112},..., x_{abc})$ be the rational function defined as

$$\Re(x_{111}, x_{112},..., x_{abc}) = \{\frac{1}{abc}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\lambda_i / x_{ijk})\}$$

and $\Re(1,1,...,1) = \hat{\lambda}_i$.

Applying the above relation, the adjustment of λ_i is given by

$$\hat{\lambda}_i(adj) = \hat{\lambda}_i\left\{1 - \frac{2}{bc}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{1}{(n_{(ij)k}-1)}\right\}.$$

The *ijk*-th estimated weight, $\tilde{w}_{(ij)k} = (\hbar/_{abc})x_{ijk}\sigma^2_{(ij)k}$ is the function of only variable, x_{ijk}. Hence the same adjustment above applies to $\tilde{w}_{ijk}$. The final cell sample sizes are then obtained by using the adjusted $\hat{\lambda}_i$ where $n_{(ij)k}(adj)$ is taken as the nearest integer.

In the case of unknown cell variances, the analysis may be performed for final samples using the approximate design constant along with the bias adjustment, where possible, for the number of squares and other statistics. Since $\hat{\lambda}_i$ has different values for different cells, so the adjustment estimate is to be applied to an average of these values. For extensive details see Shamia (1991), Talukder (1992).

4. Illustrative Example

The data used in this article were corrected for the first lactation period (305 days) of the cow experiment to be discussed below, in which a large amount of the intended data was lost as a result of age groups being mistakenly discarded under three herds (Elhwari, Elmotkamela, Rahba-2) randomly selected from seven herds in the Benghazi region. In fact, the records for each herd were made separately by a number of imported cows, on average, of milk every 15 days two times a day (2×) over the years (1999-2000). Dates of milk production (in 1000 liters) are grouped according to three different age groups, where nested at random, under fixed seasonal effects (Summer, Autumn) El-Tajory (2003).

To analyze these data, first of all, we calculated the estimated cell variances $s^2_{(ij)k} \times 10^2$ for different cell sample sizes. The results are given in Table-1.

Table-1: Summary results of estimated variances.

Summer Season			
Elhwari			
Ages	15-24	24-30	30-44
$n_{(ij)k}$	10	48	26
$s^2_{(ij)k}$	225.344	645.302	373.163
Elmotkamela			
Ages	17-24	24-30	30-54
$n_{(ij)k}$	7	20	13
$s^2_{(ij)k}$	140.812	7 749.855	420.15
Rahba-2			
Ages	22-24	24-30	30-43
$n_{(ij)k}$	5	19	14
$s^2_{(ij)k}$	101.815	6 655.636	5 589.902
Atumun Season			
Elhwari			
Ages	21-24	24-30	30-39
$n_{(ij)k}$	15	98	75
$s^2_{(ij)k}$	508.491	752.532	649.960
Elmotkamela			
Ages	22-24	24-30	30-45
$n_{(ij)k}$	10	71	38

$S^2_{(ij)k}$	7 75.150	1 278.879	841.606
Rahba-2			
Ages	20-24	24-30	30-34
$n_{(ij)k}$	39	78	64
$S^2_{(ij)k}$	6 51.972	8 80.027	748.111

From Table-1, it can be estimated proportionality conditions:

$\hat{\lambda}_1 = 7.00\times10^{-5}$, $\hat{\lambda}_2 = 3.00\times10^{-5}$, and

$\hat{\lambda}_3 = 5.00\times10^{-5} \Rightarrow \hat{\lambda}_0 = 5.00\times10^{-5}$ where

$\tilde{\hbar} = 3.70\times10^{5}$.

Summary results of *ANOVA* can be written in Table-2.

Table- 2: *ANOVA* table.

Source of Variation	d.f.	Sum of Squares	Mean Square	F_0
***Herd* (*H*)**	2	5.02x10^5	2.51x10^5	9.506
***Season* (*S*)**	1	5.55x10^4	5.55x10^4	0.010
H*×*S	2	1.08 x10^4	5.40x10^6	2.22 x10^{-2}
***Age* (*HS*)**	12	5.54x10^4	4.61x10^6	2.40 x10^{-3}
Error	632	.74x10^{10}	1.92x10^7	-

From Table-2 for testing hypothesis the factors; To test the hypothesis $H_0^i : \sigma_\alpha^2 = 0.$

We have,

$F_0^i = 9.506 > f_{0.05(2,4)} = 6.94.$

Then we reject H_0^i at 5% level of significant, that is, *Herds* variability highly differ significantly from zero. This means that, the variant-to-variant purity variation is not null between the three groups of *Herds*. So that

$$\hat{\sigma}_\alpha^2 = \frac{[MS(\mathrm{A}) - \upsilon_0]*\upsilon_1 - [MSC(\mathrm{AB}) - \upsilon_0]*\upsilon_2}{\{bc\upsilon_o\upsilon_1\upsilon_2\}} = 3.53\times10^4$$

Again to test the hypothesis $H_0^{ii} : \beta_j = 0 \quad \forall \quad j = 1,2.$

$F_0^{ii} = 0.010 < 1.000.$

Then we can not reject H_0^{ii}. That is, *Seasons* not differ significantly. Therefore, there is no significant difference between summer and autumn of milk production in cow's projects.

For testing $H_0^{iii}: \sigma^2_{\alpha\beta} = 0.$

$F_{\circ}^{iii} = 221.902 > f_{0.005(2,4)} = 26.275.$

Then we can not accept H_0^{iii} at 0.5% level of significant. That is, the interaction variability between *Herds* and *Seasons* differ significantly from zero. So that

$$\hat{\sigma}^2_{\alpha\beta} = \frac{[MS(\mathrm{AB}) - \upsilon_{\circ}]^{*}\upsilon_{1} - [MSC(\mathrm{AB}) - \upsilon_{\circ}]^{*}\upsilon_{2}}{\{c\upsilon_{\circ}\upsilon_{1}\upsilon_{2}\}} = 1.835\times 10^{6}.$$

Finally for testing $H_0^{iv}: \sigma^2_{\gamma(\alpha\beta)} = 0.$

$F_{\circ}^{iv} = 2.40.\times 10^{-3} < 1.000.$

Then we cannot reject H_0^{vi} at 5% level of significant, that is, the *Age* of cow within the *Herds* and *Seasons* variability not differ significantly from zero.

5. Conclusion

Researchers using an experimental design sometimes classify participants into groups using a median split or extreme group's procedure, but others use analyses that allow them to maintain the continuity of the measured participant variable. In either case, causal inferences may be drawn only about the variables in the design that were experimentally manipulated. In this design participants are divided into entirely separate groups on the basis of random allocation, meaning that each participant has an equal chance of being allocated to either the experimental group. An independent groups design is where the independent variable is operationalized by having different groups of participants take part in the experiment in each condition.

Crossed-nested classified data with unequal replications or unbalanced factors can happen if there are missing data. This missing which leads to a result of unbalanced data are considered as the main reason for the problem of heteroscedastic models. The control is that the final cell sample sizes should be made proportional to the respective cell variances.

Resent development in mixed model theory have been implemented in major studies. Special emphasis is given to the independent group design (IGD) as it contains on random factor as groups or regions.

This study attempts to investigate the theoretical side of such design analysis in the case of unbalanced model in the situation where approximate methods such as unweighted means are inappropriate. It was examined by using proportionality condition on cell sample size using

harmonic mean for variances the cell to get the weighted value to translate the model from unbalanced into the partial balanced form as exact method.

By studying, the independent group design we provide a modification of the analysis of variance method. The results from adjusting (reduction) of sum of squares have been exercised by using example of real data where the problem was obvious. The results proved the capability of this adjustment of analysis as it compared with no adjustment of analysis.

REFERENCES

- **Anderson, N.H. (2001)** Empirical Direction in Design and Analysis. Mahwah, NJ: Lawrence Erlbaum Associates.
- **Bhuyan, K.C. (1984)**. A group of split-plot designs with a heteroscedastic model. Austral. J. Statistic. 26(2), 133-141.
- **Dean, A. and Voss, D. (1999)** Design and Analysis of Experiments. New York: Springer-Verlag.
- **El-Saeiti, I.N. (2004)**. Messy data in heteroscedastic models "case study: mixed-nested design". MSc. thesis, Garyounis (Benghazi) University.
- **El-Saeiti, I.N. and Shamia, G.M. (2006).** Estimating heteroscedastic variance components for four-stage nested design in mixed model. 1st Statistical Conference. Faculty of Science, Al-Fateh (Tripoli) University.
- **El-Saeiti, I.N. and Shamia, G.M. (2007).** Analysis of four-stage nested design with missing data. Journal of Science and its Applications. 1-1, 24-32.
- **El-Tajory, S.S. (2003)**. Comparative study on dairy records from three herds in Benghazi area. MSc. thesis, Garyounis (Benghazi) University.
- **Hartmink, A.J., Gifford, D.K., and Young, R.A. (2002)**. Maximum likelihood estimation of optimal scaling factors for expression array normalization. amink@mit.edu.
- **Kirk, R.E. (1995)** Experimental Design: Procedures for the Behavioural Sciences (3rd edit.). Pacific Grove, CA: Brooks/Cole.
- **Meier, P. (1953)**. Variance of a weighted mean. Biometrics, 9, 59-73.
- Ryan, T.P. (2007) Modern Experimental Design. Hoboken, NJ: Wiley.
- **Satterthwaite, F.E. (1946)**. An approximate distribution of estimates of variance components. Biometrics, 2, 110-112.
- **Searle, S. R. (1971a)**. Linear models, John Wiley and Sons, New York.
- S**hamia, G.M. (1991)**. Analysis of some designs of experiments with heteroscedastic models. M.Sc. thesis, Garyounis (Benghazi) University.
- **Shamia, G.M. and El-Saeiti, I.N. (2007)**. Analysis of split-plot design with a heteroscedastic Models, Journal of Science for 2nd Statistical Coneferance, Al-Fateh (Tripoli) University.
- **Shamia, G.M. and El-Saeiti, I.N. (2019)**. Modifosica-tion of WLS for Analyzing Heteroscedastic modes with Missing Data. Journal of Elixer Statistics, 133, 5373-53576.

- **Snedecor, G.W. and Cochran, W.G. (1967)**. Statistical methods. Seventh Edition. Iowa university press. Ames, Iowa.
- **Speed, F.M., Hocking, R.R., and Hackney, O.P. (1978)**. Methods of analysis of linear models with unbalanced data. Journal of the American Statistical Association, 73, 105-112.
- **Talukder, M.A.H. (1978)**. General block design with a heteroscedastic model. Sankhya, 40B, 217-226.
- **Talukder, M.A.H. (1993)**. A general method of analysis of factorial designs with unequal cell variances. International Statistical Review, 61(3), 409-417.