Journal of the Nigerian Association of Mathematical Physics
***Volume* 22 (November, 2012), pp 199 – 206**
© J. of NAMP

Analysis of Unbalanced Fixed-Effect Non- Interactive Model: An Intra-Factor Approach.

[1*]Ochei L.S, Eze F.C. and [2]Ajibade B.F.

[1]Department of Statistics,
Federal University, Otuoke, Nigeria.
[2]Department of General Studies,
Petroleum Training Institute, Effurun, Warri, Nigeria.

Abstract

This paper examines the analysis of fixed-effect non-interactive unbalanced data by on approach termed "Intra-Factor Design". To derive this design for analysis mathematically, the matrix version of fixed-effect model, $Y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}$ was used. This resulted to the definition and formation of many matrices such as the Information Matrix, $\underline{L}$; Replication Vector, $\underline{r}$, Incidence Matrix, $\underline{Q}$, which is the generalized inverse of the Information Matrix, and other matrices. The Least Squares Method which gave birth to several normal equations was used to estimate for the parameters, $\hat{\tau}$ and $\hat{\beta}$ mathematically. Then, an illustrative example was used to ascertain the workability and application of this Intra-Factor procedure in testing for the main effects under some stated hypothesis for significance. However, before testing for the variance component of the main effects on the illustrative data, it was necessary to first establish that the data was fixed-effect and that interaction was either absent or non-significant.

1.0 Introduction

Over the years many statisticians and scholars in inferential have contrlbuted greatly to the development of basic techniques in statistical analysis especially with balanced data. But a few abound for unbalanced case. This is due to the fact that estimating variance component from unbalanced data is not as straight forward as from balanced data. This is so for two reasons. Firstly, several methods of estimation are available (most of which reduce to the analysis of variance method for balanced data), but no one of them has yet been clearly established as superior to others. Secondly, all the methods involve relatively cumbersome algebra; discussion of unbalanced data can therefore easily deteriorate into a welter of symbols, a situation we do our best to minimize here. However, we are not by this claiming superiority over other methods of estimating variance component of unbalanced data since no comprehensive comparative study of all the methods have yet been conducted.

General Outlay of Unbalanced Data

Balanced data are those in which every one of the subclasses of the model has the same number of observations, that is, equal numbers of observations in all the subclasses. In contrast, unbalanced data are those data wherein the numbers of observations in the subclasses of the model are not all the same, that is, unequal number of observations in the subclasses, including cases where there are no observations at all [1]. Thus, unbalanced data refers not only to situations where all subclasses have some data, namely filled subclasses, but also to cases where some subclasses are empty, with no data in them. The estimation of variance components from unbalanced data is more complicated than from balanced data.

In many areas of research such as this, it is necessary to analyze the variance of data, which is classified into two ways with unequal numbers of observations falling into each cell of the classification. For data of this kind, special methods of analysis are required because of the inequality of the cell members. This we attempt to solve in this paper.

*Corresponding author: ***Ochei L.S,*** E-mail: ocheiluckystephen@yahoo.co.uk, Tel. +234 7035210707

Problem Involved in Random Models

The problem associated with the random-effect models has been the determination of approximate F-test in testing for the main effects say, A and B using F-ratio. In this case, there would be no obvious denominator for testing the hypothesis H_0: σ^2=0, for a level of factor A crossed with the level of factor B in the model such as

$$X_{ijk} = \mu + \alpha_i + \beta_j + \lambda_{ij} + \epsilon_{ijk}$$

where, X_{ijk} is the k^{th} observation (for k = 1,2, .. n_{ij}) in the i^{th} level of factor A and j^{th} level of factor B; i = 1, 2, .., p, j = 1,2 …, q;

μ is the general mean;
σ_i is the effect due to the i^{th} level of factor A;
β_j is the effect due to the j^{th} level of factor B;
λ_{ij} is the interaction between the i^{th} level of factor A and j^{th} level of factor B;
ϵ_{ijk} is the observation error associated with X_{ijk} .

However, Ehiwario [2], in his work on "Common F-test Denominator for Two-way Interactive Balanced Design", discovered and quoted Eze [3] that the non-existence of a clear common denominator for the F-ratio for the different models, was as a result of the presence of interaction in the model/data. To solve this problem of common denominator for F-test, Ehiwario eliminated the interaction effect from the models which consequently led to the violation of homogeneity of variances assumption and making the use of ANOVA test impossible. But he normalized the effects of the violated assumption of homogeneity of variances. The elimination of the interaction effects gave rise to the reduced model which when tested with data collected, yielded same result as the full model. This led to his conclusion that when interaction effects are eliminated from a model, whether fixed, random or mixed the MSE can be used as a common F-test denominator and that the reduced model is more efficient. Hence, Ehiwario [2] provided an alternative method (reduced model) for handling interaction problems without undergoing the rigorous method of calculating the interaction sum of squares, before the conventional analysis of variance. It is on these grounds and findings that this paper has its basis as "non-interactive".

Problem of Mixed-Effect Models

According to Henderson [4], in the customary unbalanced random model for a crossed classification, the nature of the expected value of the sum of squares is such that:

$$E(SSA)\left(= n_{..} - \sum_{i}^{c} n_{i.}^2/n_{..}\right)\sigma_A^2 + \sum\left\{\sum_{i}^{c}\sum_{j}^{b} \frac{n_{ij}^2}{n_i} - \sum \frac{n_{.j}^2}{n_{..}}\ \sigma_B^2\right\} + \sum\sum\left\{\frac{n_{ij}^2}{} - \sum_{i}^{a}\sum_{j}^{b}\frac{n_{ij}^2}{n_{i.}}\ \sigma_{AB}^2 + (\alpha-1)\right\}\sigma_e^2 \ldots\ldots (1.1)$$

This means that with unbalanced data from crossed classification in random model, the expected value of every sum of squares contains every variance component of the model. But in an unbalanced mixed, with α's being fixed and not random effects, the expectation would then be:

$$E(SSA) = \left(\sum_{i}^{a} ni.\propto_i{}^2 - \sum_{i}^{a} ni.\alpha i/n_{..}\right) + \left(\sum_{i}^{a}\sum_{j}^{b}\frac{nij^2}{n_{..}}\sum_{j}^{b}\frac{n.j2}{n_{..}}\right)\sigma_{B+}^2 \left(\sum_{i}^{a}\sum_{j}^{b}\frac{nij^2}{ni.} - \sum_{i}^{a}\sum_{j}^{b}\frac{nij^2}{n_{...}}\right)\sigma_{B\propto}^2 + (\propto -1)\,\bar{\sigma}_e^2 \qquad (1.2)$$

It is observed that the first term of (1.2) which is a function of the fixed effect is different from that in (1.1); and this occurs in expected values of all sum of squares terms (except for SSE). And, more importantly, the function of the fixed effects is not the same from one expected sum of squares term to the next.

For example, with the α's fixed, E(SSE) contains the term $\sum_{j}^{b}\left(\sum_{i}^{a} nij\alpha i\right)^2/n._j - \left(\sum_{i}^{a} nij\alpha i\right)^2/n_{..}$

which differs from the first term in $E\{SSA\}$ of (1.2). Thus $E\{SSA - SSB\}$ does not get rid of the fixed effects even though it does eliminate terms in μ. This is true generally in mixed models, expected values of the sum of squares contain functions of the fixed effects that cannot be eliminated by considering linear combinations of the sum of squares. This means that the equations $E[SS] = P\sigma^2 + \sigma_e^2 f + q$ in the mixed model, where q is a vector of the quadratic functions of the fixed effects in the model. Hence, σ^2 cannot be estimated and the analysis of variance method applied to unbalanced data cannot be used for mixed models as well as for fixed-effect models. It yields biased estimators. Simply put, with unbalanced data, the analysis of variance method for mixed and fixed effect models lead to biased estimators of variance components. Mixed models involve dual estimation problems- estimating both fixed effects and variance components.

As a result of dual estimation problems of the mixed model with unbalanced data which accounted for biased estimators of variance components, Henderson [4], designed a method to correct this deficiency. This he does by his method 2 which uses the data first to estimate fixed effects of the model and then using these estimators to adjusted data. Variance components are estimated from the adjusted data by the analysis of variance method. This whole procedure was designed so that the resulting variance estimators were not biased by the presence of the fixed effects in the model as they were with the analysis of variance estimators derived from the basic data. So far as the criterion of unbiasedness was concerned, this was certainly achieved by this method.

However, the general method of analyzing data adjusted according to some estimator of the fixed effects is open to criticism on other grounds such as: it cannot be uniquely defined, and a simplified form of it, of which Henderson's method 2 is a special case, cannot be used whenever the model includes interactions between the fixed effects and the random effects. This again gives another justification for "Analysis of Unbalanced Fixed-Effect Non-Interaction Model".

2.0 Theoretical Analysis / Experimental Work

This section provides the mathematical derivation of the methodology-Intra-Factor model for unbalanced data analysis, the assumptions of the model, the mathematical derivation of the design matrix of the general unbalanced factor design by the method of least squares as the method of parameter estimation, and the hypothesis to be tested.

Let the unbalanced design be made up by a set of t levels of factor A and b levels of factor B. The model for an Unbalanced Factor Design in which y_{ijk} is the expected value of the response obtained by applying the ith level of factor A in the jth level of factor B on the kth instance could be given by:

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk};\ i=1, 2, \ldots, t,\ j=1, 2, \ldots, b,\ k = 1,2, \ldots, n_{ij} \qquad (2.1)$$

Where,

μ is the parameter for grand means;

τ_i is the parameter measuring the effect of the ith level of factor A;

β_j is the parameter measuring the effect of the jth level of factor B;

ε_{ijk} is the error term associated with y_{ijk};

n_{ij} is the number of times that the ith level of factor A is applied to the jth level of factor B.

Assumptions: $\sum_i \tau_i = \sum_j b_j = 0$

$\varepsilon_{ijk} \sim (0, \sigma_e^2)$

The matrix version of equation (2.1) could be written as [5]

$$\underline{y} = \underline{\tilde{X}}\,\underline{\Theta} + \varepsilon \qquad (2.2)$$

where,

$\underline{y}$ is an $(n x 1)$ vector of responses

$\underline{\tilde{X}}$ is an $(n x p)$ design matrix of zeros and ones

$\underline{\Theta}$ is a $(px1)$ vector of Factor A and Factor B parameters

$\underline{\varepsilon}$ is also an $(nx1)$ vector of error terms, and p = 1+ t + b (while t and b are the numbers of the Factor A and Factor B of the experiment respectively).

The expectation of ε, E[$\underline{\varepsilon}$], is zero while its variance-covariance matrix, V[$\underline{\varepsilon}$], is σ^2I.

We now derive one of the essential matrices involved in our model, called the Information Matrix, $\underline{L}$. In doing this, an unbalanced Factor Design would be illustrated in Figure 2.1 to ease the understanding of the subsequent derivations.

Factor B

Factor A		B_1	B_2	B_3	. . .	B_n
	T_1	XX	X	XXX	. . .	XX
		(y_{11})	(y_{12})	(y_{13})	. . .	(y_{1n})
	T_2	XXX	XXX	--	. . .	X
		(y_{21})	(y_{22})	(y_{23})	. . .	(y_{2n})
	.	–	–	–	. . .	–
	.	–	–	–	. . .	–
	.	–	–	–	. . .	–
	T_n	X	--	XXXX	. . .	XXX
		(y_{n1})	(y_{n2})	(y_{n3})	. . .	(y_{nn})

Figure 2.1: Unbalanced Factor Design (With Response Data Totals y_{ij}'s).

Suppose $\tau_1, \tau_2, \tau_3, \ldots, \tau_n$ are the n factor A parameters for $T_1, T_2, T_3, \ldots$,Tn respectively, while β_1,β_2,β_3, . . . , β_n are the n factor B parameters for B_1, B_2, B_3, . . . ,B_n respectively, and the "X's" indicate number of response data on each cell. Then equation (2.2) could be written as follows:

Factor A Factor B

$$\underline{y} \qquad \underline{\mu} \mid \tau_1\ \tau_2\ \tau_3 \ldots \tau_n \mid \beta_1\ \beta_2\ \beta_3 \ldots \beta_n$$

$$\begin{pmatrix} y_{11}\\ y_{12}\\ .\\ .\\ .\\ y_{nn}\end{pmatrix} = \begin{pmatrix} 1 & \mid & \text{Zeros} & \mid & \text{Zeros} \\ 1 & \mid & & \mid & \\ . & \mid & \text{and} & \mid & \text{and} \\ . & \mid & & \mid & \\ . & \mid & \text{Ones} & \mid & \text{Ones}\\ 1 & \mid & \underbrace{\qquad\qquad}_{\underline{S}_{nxt}} & \mid & \underbrace{\qquad\qquad}_{\underline{U}_{nxb}} \end{pmatrix} \underbrace{\begin{pmatrix}\mu\\ \cdots \\ \tau_1 \\ . \\ \tau_n \\ \cdots \\ \beta_1\\ . \\ \beta_n\end{pmatrix}}_{\underline{\Theta}} + \underbrace{\begin{pmatrix}\varepsilon_1\\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n\end{pmatrix}}_{\underline{\varepsilon}} \qquad (2.3)$$

We notice that the column of $\underline{\tilde{X}}$ can be partitioned as illustrated above into three sub-matrices [6], corresponding to the parameter, μ, the Facto A parameter $\tau_1, \tau_{2, \ldots,} \tau_n$ and $\underline{\Theta}$ can be partitioned as follows:

$$\underline{\tilde{X}} = [\, \underline{I} \mid \underline{S} \mid \underline{U} \,] \qquad (2.4),$$

$$\underline{\Theta} = [\, \mu \mid \tau \mid \beta] \qquad (2.5).$$

where,

$\underline{I}$ is a column vector of ones

$\underline{S}$ is the nxt design matrix for the Factor A and

μ is the nxb design matrix for the Factor B.

We shall therefore rewrite equation (2.2) as

$$\underline{y} = \underline{I}\, \underline{\mu} + \underline{S}\, \underline{\tau} + \underline{U}\, \underline{\beta} + \underline{\varepsilon} \qquad (2.6)$$

In $\underline{S}$, each column is for each factor A, and so, the number of times a Factor A occurs in the design is given by the sum or sum of squares of the entries of corresponding column. In addition, the sum of cross products of any two column, could always be zero due to the fact that any plot or experimental unit accommodates only one Factor A at a time. Therefore, if the number of times each Factor A occurs in the design is r, the r's for all the Factor A's constitute a vector, called the Replication Vector, denoted by $\underline{r}$, then we can write:

$$\underline{S}'\underline{I} = \underline{r},\ \underline{S}'\underline{S} = \underline{r}^{\Theta} \qquad (2.7)$$

$\underline{r}^{\Theta}$ is the diagonal matrix whose diagonal elements are the entries of $\underline{r}$. In the same way, if the number of plots in the j^{th} level of Factor B is denoted by k_j and $\underline{k}$ is a vector of all $k_j^{'s}$, (j = 1,2, …., b), then

$$\underline{U}'\underline{I} = \underline{k},\ \ \underline{U}'\underline{U} = \underline{k}^{\Theta} \qquad (2.8)$$

Furthermore, we define, $\underline{T}$, $\underline{B}$ and G to be Factor A total, Factor B total and grand total, respectively. It is possible to multiply the columns of $\tilde{X}$ with $\underline{y}$ to get:

$$\underline{S}'\underline{y} = \underline{T},\ \ \underline{U}'\underline{y} = \underline{B},\ \ \underline{I}'\underline{y} = G \qquad (2.9)$$

The number of times, n_{ij}, the i^{th} level of Factor A occurs in the j^{th} level of Factor B is given by the sum of cross products of the i^{th} column of $\underline{S}$ and j^{th} column of $\underline{U}$. Define N as the matrix whose $(ij)^{th}$ entry is n_{ij}. This matrix $\underline{N}$ is the same as what is called the Incidence Matrix. Therefore,

$$\underline{S}'\underline{U} = \underline{N} \qquad (2.10)$$

In $\underline{N}$, there always exists a row for each Facto A and a column for each Factor B.

To find the least squares estimator of the vector of parameters, $\underline{\Theta}$, in equation (2.2), we minimize the sum of squares of the error terms, $\underline{\varepsilon}'\underline{\varepsilon}$, with respect to $\underline{\Theta}$. From equation (2.2), this error sum of squares can be expressed as:

$$\underline{\varepsilon}'\underline{\varepsilon} = (y - \underline{\tilde{X}}\, \underline{\Theta})'(y - \underline{\tilde{X}}\, \underline{\Theta}) \qquad (2.11)$$

$$= \underline{y}'\underline{y} - 2\underline{y}'\, \underline{\tilde{X}}\, \underline{\Theta} + \underline{\Theta}'\, X'\underline{\tilde{X}}'\, \underline{\tilde{X}}\, \underline{\Theta} \qquad (2.12)$$

Taking the partial derivatives of equation (2.12 with respect to $\underline{\Theta}$, equating to zero and solving for $\underline{\Theta}$, we obtain the estimate, $\underline{\hat{\Theta}}$, for the parameter Θ via the normal equations.

$$\underline{\tilde{X}}'\underline{\tilde{X}}\ \underline{\hat{\Theta}} = \underline{\tilde{X}}'y \qquad (2.13)$$

Using equations (2.7) to (2.10), it is possible to partition $\underline{\tilde{X}}$ and $\underline{\hat{\Theta}}$ in conformity with equation (2.6) to obtain the following set of normal equations:

$$r\hat{\mu} + \underline{r}^{\Theta}\underline{\hat{\tau}} + \underline{N}\, \underline{\hat{\beta}} = \underline{T} \qquad (2.14)$$

$$r\hat{\mu} + \underline{N}'\, \underline{\hat{\tau}} + \underline{k}^{\Theta}\underline{\hat{\beta}} = \underline{B} \qquad (2.15)$$

$$n\hat{\mu} + \underline{r}'\, \underline{\hat{\tau}} + \underline{k}'\underline{\hat{\beta}} = G \qquad (2.16)$$

No unique solution exists for equations (2.14), (2.15) and (2.16) since they do not have full rank. For instance, it can easily be seen that each of the sum of the t rows of equation (2.14) and b rows of equation (2.15) gives equation (2.16).
Therefore,

$$\text{Rank}\,(\underline{\tilde{X}}'\underline{\tilde{X}}) \leq t + b - 1$$

By equating a set of parameters to zero or any other arbitrary values, for instance, such that the remaining set of equations would be of full rank, some situations of the normal equations (2.13) can be deduced. A typical solution to equation (2.13) will be of the form:

$$\underline{\hat{\Theta}} = \left(\underline{\tilde{X}}'\underline{\tilde{X}}\right)^{-} \underline{\tilde{X}}'\underline{Y}$$

where,

$\left(\underline{\tilde{X}}'\underline{\tilde{X}}\right)^{-}$ is the generalized inverse of $\left(\underline{\tilde{X}}'\underline{\tilde{X}}\right)$ satisfying the following property:
$\left(\underline{\tilde{X}}'\underline{\tilde{X}}\right)\left(\underline{\tilde{X}}'\underline{\tilde{X}}\right)^{-}\left(\underline{\tilde{X}}'\underline{\tilde{X}}\right) = \underline{\tilde{X}}'\underline{\tilde{X}}$.
It is true that most times, the parameters of interest to an experimenter could be those associated to the factor A's. One can therefore, find a simplified solution by dealing with a subset of normal equations involving $\underline{\mu}$ and $\underline{\beta}$. To do this, we pre-multiply equation (2.15) by $\underline{N}k^{-\Theta}$ in the first instance to obtain:

$$r\hat{\mu} + \underline{N}\,\underline{k}^{-\Theta} + \underline{N}'\,\underline{\hat{\tau}} = \underline{N}\,\underline{k}^{-\Theta}\underline{B} \qquad (2.17)$$

and thereafter subtracting the equation (2.17) from equation (2.14), to obtain:

$$\underline{L}\,\underline{\hat{\tau}} = \underline{q} \qquad (2.18)$$

We recall that $\underline{L}$ is the notation used to denote the Information Matrix in the intra factor analysis.
In equation (2.18), $\underline{L}$, is given by

$$\underline{L} = \underline{N}r^{\Theta} - \underline{N}k^{-\Theta}\underline{B} \qquad (2.19)$$

while $\underline{q}$ is given by

$$\underline{q} = \underline{T} - \underline{N}k^{-\Theta}\underline{B} \qquad (2.20)$$

The information matrix, $\underline{L}$, is of order (txt) and has rank less than or equal to (t-1). Of course equality holds if the design is connected. It is always possible to get a solution to the normal equations for $\underline{\hat{\tau}}$ by solving equation (2.18). On the other hand, one can solve for $\hat{\beta}$ in equation (2.15) to obtain

$$\underline{\hat{\beta}} = k^{-\Theta}\underline{B} - k^{-\Theta}\underline{N}'\,\underline{\hat{\tau}} - I\hat{\mu} \qquad (2.21)$$

Equation (2.15) represents a set of equations called the reduced normal equations for the Factor A parameters. At this juncture, it is appropriate to emphasize that equation (2.15) arises by the elimination of the mean and Factor B parameter from the full set of normal equations given in equation (2.13). In equation (2.15), $\underline{q}$ is known as the vector of adjusted Factor A totals. It is so called because the linear combinations of the Factor B totals are subtracted from that of Factor A totals. Indeed $\underline{q}$ involves the liner combination of the responses which can be given as:

$$\underline{q} = \left(\underline{s}' - \underline{N}k^{-\theta}\,\underline{u}'\right)\underline{y} = \underline{s}'\left[\left(1 - \underline{u}\left(\underline{u}\,'\underline{u}\right)^{-1}\underline{u}'\right)\right]\underline{y} \qquad (2.22)$$

It is well known fact from the discussion that the matrix, $\underline{L}$, is not of full rank $\left[rank\left(\underline{L}\right) \leq t-1\right]$ since it can easily be seen that the rows and columns of $\underline{L}$ sum to zero. This characteristic would make it possible to achieve a unique solution for the reduced normal equations.
All solutions to equation (2.18) can be expressed as:

$$\underline{\hat{\tau}} = \underline{Q}\,\underline{q} \qquad (2.23)$$

where, $\underline{Q}$ is the generalized inverse of $\underline{L}$ and is such that $\underline{L}\,\underline{Q}\,\underline{L} = \underline{L}$. In any case, there exist, other versions of the generalized in verse, which can be adopted for analyzing the unbalanced factor designs.
On the whole, from the foregoing, we have identified the design matrix of the general unbalanced factor design with some nice properties to be a matrix of one's and zero's given to give insight into what happens when we have a large area of experimentation.

Test of Hypothesis
Here we assume $\underline{\varepsilon} \sim N(0, I\sigma^2)$.

Sum of squares due to μ, τ and β, $s_R(\mu, \tau, ß) = (\underline{\mu}, \hat{\tau}'\, \beta') \begin{pmatrix} I'y \\ S'y \\ U'y \end{pmatrix}$

$$= \hat{\mu}G + \underline{\hat{\tau}}'\underline{T} + \hat{\beta}'\underline{B} \qquad (2.24)$$

Using equation (2.20) and (2.21), we can write equation (2.24) as:

$$s_R(\underline{\mu}, \underline{\tau}, \underline{\beta}) = \underline{\hat{\tau}}'\,\underline{q} + \underline{\beta}'\underline{k}^{-\Theta}\underline{B} \qquad (2.25)$$

Under Ho: $\underline{\tau} = 0$, the model of the equation (2.2) reduces to:

$\underline{y} = 1\underline{\mu} + U'B + \underline{\varepsilon}$ (2.26)

The normal equations then become

$n\hat{\mu} + \underline{k}'\hat{\beta} = G$ (2.27)

$\underline{k}\hat{\mu} + k^{\Theta}\hat{\underline{\beta}} = \underline{B}$ (2.28)

But with $\hat{\mu} = \bar{y}$, $\hat{\underline{\beta}} = \underline{k}^{-\Theta}\underline{B} - G$ (2.29)

So,

$s_R(\mu, \beta) = \hat{\mu}G + \hat{\underline{\beta}}'\underline{B} = B'\underline{k}^{-\Theta}\underline{B}$ (2.30)

and

$s_R(\tau/\underline{\mu},\underline{\beta}) = s_R(\underline{\mu},\underline{\tau},\underline{\beta}) - s_R(\underline{\mu},\underline{\beta})$

$= \hat{\underline{\tau}}'q$ (2.31)

Similarly,

$s_R(\underline{\beta}/\underline{\mu}) = S_R(\underline{\mu},\underline{\beta}) - S_R(\mu)$

$= \underline{B}'k^{-\Theta}\underline{B} - G^2/n$ (2.32)

By the same token,

$s_R(\underline{\tau}/\underline{\mu}) = \underline{T}r^{-\Theta}\underline{T} - G^2/n$ (2.33)

Since $s_R(\underline{\mu}) = \hat{\underline{\mu}}G = G^2/n$ (2.34)

Recall that when the model of (2.1) is true the Residual Mean Square (say S^2) is an unbiased estimate of σ^2 and has a χ^2 distribution with (n+b+r-1) degrees of freedom (d.f). Similarly, under $H_0 : \hat{\tau} = 0$,

$\hat{\underline{\tau}}'\underline{q}/(r-1)$ has a χ^2 distribution with r-1 degrees of freedom. Therefore,

F = $[\{\hat{\tau}'q(r-1)/S^2\}]$ has an F- distribution with r-1 and n- b – r + 1 degrees of freedom (d.f). The test can be summarized in the following ANOVA Table 2.1.

Table 2.1 Intra-Factor ANOVA

Sources of variation (s.v)	Degrees of freedom (d.f)	Sum of squares (s.s)	Mean squares (m.s)	F- ratio
Factor B (ignoring factor A)	b-1	$\underline{B}'K^{-\theta}\underline{B}-G^2/n$	$(\underline{B}'K^{-\Theta}\underline{B}-G^2/n)/b-1$	$F_B$ms/s^2 (factor B mean squares)
Factor A (eliminating factor B)	r-1	$\hat{\underline{\tau}}'q$	$\hat{\underline{\tau}}'q/(r-1)$	$F_A$ms/s^2 (factor A mean squares)
Residual	n–b–r+1	By subtraction	s^2	
Total corrected for mean	n-1	$y'y-G^2/n$		

In the above procedure, we assume that the Factor A effects are of interest. If we were interested in factor B effects, we will obtain sum of squares for Factor B eliminating Factor A; in other words, we inter- change Factor B for factor A. it is important to note that Adjusted Factor A S.S + Unadjusted Factor B S.S = Adjusted Factor B S.S + Unadjusted Factor A S.S.

Application of the Model

We now apply our model to data in Table 2.2, which shows a two- way crossed classification without interaction between two factors, A, and B of Unbalanced Fixed-Effect model.

Table 2.2 Fixed-Effect Non-Interactive Unbalanced Data

	Factor B				
Factor A	1	2	3	$\underline{T}$	$\bar{X}_{i\cdot}$
1	30,34 (64)	20,24,25 (69)	74 (74)	207	69
2	67 (62)	32,35 (67)	10,29,33 (72)	201	67
$\underline{B}$	126	136	146	408=G	
$\bar{X}_{\cdot j}$	63	68	73		68 = $\bar{X}_{\cdot\cdot}$

Source: Unbalanced version of the two-way crossed classification [1].

Nota: the figures in parenthesis are the cell total; $\bar{X}_{i\cdot}$, $\bar{X}_{\cdot j}$ and $\bar{X}_{\cdot\cdot\cdot}$ are the row, column and overall means respectively.

The Model: The model is $y_{ijk} = \mu_i + \beta_j + \varepsilon_{ijk}$

The Hypothesis:

(i) $H_{01} : \tau_i = 0$ for all i = 1/ 2,3

$H_{02} : \tau_i \neq 0$ for at least one i

(ii) $H_{11} : \beta_j = 0$ for all j = 1,

$H_{12} : \beta_j \neq 0$ for at least one j

Decision rule: Reject H_0, if F calculated > F tabulated at 5% level of significance.

Test for Fixed-Effectness of the Data

To determine the type of data for illustration given in Table 2.2, we shall use a method of calculation of cell by cell population means. For a fixed effect model, $\sum_i \tau_i = \sum_j \beta_j = \sum_{ij} \lambda_{ij} = 0$ and e ~ $N(0, I\sigma_e^2)$. The population means of the data being modeled are expressed as

$$\times_{ijk} + \mu + \tau_i + \beta_j + \lambda_{ij}$$

It is calculated from the least squares estimate of its component parameters.

That is,

$$\left.\begin{aligned} \overline{X}_{...} &= \mu \\ \hat{\tau}_i &= \overline{X}_{i..} - \overline{X}_{...} \\ \hat{\beta}_j &= \overline{X}_{.j.} - \overline{X}_{...} \\ \hat{\lambda}_{ij} &= \overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j..} - \overline{X}_{...} \end{aligned}\right\} \qquad (2.35)$$

Now from Table 2.2 and Equation (2.35) we have the following result summarized in Table 2.3.

Table 2.3 Population Means Without Interaction

Factor B

Factor A		1	2	3
	1	68 + 1 -5 + 0	68 + 1 + 0 + 0	68 + 1 + 5 + 0
	2	68 – 1- 5 + 0	68 – 1 + 0 + 0	68 – 1 + 5 + 0

Here, μ = 68, τ_1=1, τ_2 = -1, ß$_1$ = -5,β_2 = 0,β_3 = 5, $\lambda_{11} = \lambda_{12} = \lambda_{23}$ = 0 from the above we observed that:

$\sum_i \tau_i$ = (1) + (-1) + (1) +(-1) + (1) + (-1) = 1-1 +1-1 +1 -1 = 0

$\sum_j \beta_j = [(-5) + (0) + (5)] + [(-5) + (0) + (5)]$ = -5+ 0 +5 +-5+0 +5 = 0

$\sum_{ij} \lambda_{ij}$ = (0) + (0) + (0) +(0) +(0) + (0) = 0

It follows that $\sum_i \tau_i = \sum_j \beta_j = \sum_i \lambda_{ij}$ (or $\sum_j \lambda_{ij} = \sum_¿ \lambda_{ij}$) = 0 and e_{ijk}~ N(0, I$\sigma_e{}^2$)

From the above findings, we can conclude that the fixed-effect model can be used to model the data in Table 2.2 since the conditions for fixed- effect model, is satisfied adequately.

Test for Absence of Interaction

To test for the absence or non- significance of the interaction effect on the illustrative data of Table 2.2, we shall employ elementary observation of the cell component. In a careful observation of the data, one would notice that the change from one component of Factor A to the other is same for each across Factor B components (ie 69–64 = 67-62; 74-69 = 72-67) and the difference between components of factor A are eual, which is 2 (i.e. 64-62 = 69-67 = 74-72 =2).

Again, from Table 2.3, it is seen that λ_{ij} = 0 for all i and j. Hence, we can conclude that a Two-Factor interaction, which is defined as the change in response due to one factor at different levels of the factor, is absent or non-significant in the fixed-effect data of Table 2.2.

Analysis of Variance for the Main Effects, τ and β

Here, we carry out analysis of variance component on the data of Table 2.2 using our model, test the two hypotheses stated above of the main effects, τ and ß. The result is presented in Table 2.4.

Table 2.4 ANOVA Table for the Model

S.V	d.f	S.S	M.S	f-ratio.
Factor B (Ignoring factor A)	2	100	50	$F_B = {}^{50}/_{2}$ = 25.0
Factor A (Eliminating factor B)	1	2.0	2.0	$F_A\ {}^{2.0}/_{2.0}$ = 1.0
Residual	2	4.0	2.0	
Total	5	106		

Then, $F_{A,tab} = F_{1,2,0.05=18.5}$

$F_{B,tab} = F_{2,2,0.05=19.0}$

Conclusion: since $F_{A,cal}$ (1.0) < $F_{A,tab}$ (18.5), we accept H_0 and conclude that the Adjusted effect due to factor A,τ, is not significant or is the same. Then $F_{B,cal}$ (25.0) > $F_{A,tab}$ (19.0), hence, we conclude that the Unadjusted effect due to Factor B, β, is significant.

However, the test for the Unadjusted effect due to Factor A with Adjusted factor B is summarized in Table 2.5.

Table 2.5 ANOVA Table for Unadjusted Factor A with Adjusted Factor B

S.V	D.F	S.S	M.S	F- RATIO
Factor B (Eliminating factor A)	2	96	48	$F_B = {}^{48}/_{2} = 24.0$
Factor A (Ignoring factor B)	1	6.0	6.0	$F_A = {}^{6.0}/_{2.0} = 3.0$
Residual	2	4.0	2.0	
Total	5	106		

Then $F_{A,tab} = F_{1,2\ 0.05} = 18.5$
$F_{B,tab} = F_{2,2\ 0.05} = 19.0$

3.0 Results / Discussion

From our mathematical derivations of the sum of squares and mean squares using the least squares method, we found out that the Intra-Factor Design or Approach can be used to analyze the variance components of the main effects of fixed but non-interactive unbalanced data.

Secondly, we were able to test for the fixed-effectness of the illustrative data using the method of calculation of cell by cell population means; hence the data was ascertained to be of fixed effect model.

Again, we were able to establish for the absence or non-significance of the interaction effect in the illustrative data by the method of elementary operation of the cell components and by the results from the calculation of cell by cell population means.

Moreso, the analysis of the variance components for the main effect, τ, β were carried out on the illustrative data and the results obtained and summarized were as follows:

1 From the analysis of the variance component for Adjusted Factor A with Unadjusted Factor B effects, it was seen that the Adjusted Factor A effect, τ, was not significant whereas, the Unadjusted Factor B effects, β, was significant.

2 From the analysis of variance component for Unadjusted Factor A with Adjusted Factor B effects, it was observed that the Unadjusted Factor A effect, τ, was also not significant. While on the other hand, the Adjusted effect due to Factor B, β, was significant.

4.0 Conclusion

This study has provided yet another alternative method of analyzing unbalanced data called the Intra-Factor Design to that of Henderson's method 1 & 2 [4]. From the result of the study stated above, the following conclusions were made.

1. When an unbalanced data is non-interactive and is of fixed- effect model, the Intra-Factor Design can be used to test for the variance components of the main effect;
2. The method of calculations of cell by population means can be used to test the type of model of an unbalanced data;
3. The method of elementary operation of the cell components and also by results from same calculation of cell by cell population means can be used to establish the obsence/ non-significance or otherwise of the interaction effect in an unbalanced data;
4. The Intra-Factor Design as an alternative method of analyzing unbalanced data, made provision for the analysis of variance components for both Adjusted and Unadjusted main effects.

References

[1] Eze, F.C. (2002): *Introduction to analysis of variance.* Vol 1, Lano Publishers, Enugu

[2] Ehiwario, J.C. (2008): *Common F-test Denominator for Two-Way Interactive Balanced Design.* M.Sc Thesis, Nnamdi Azikiwe University, Awka.

[3] Eze, F.C. (1998): *On Two Tests for Main Effects in an Unbalanced Interactive Two-W Crossed Model.* M. Sc Project, University of Nigeria, Nsukka.

[4] Henderson, C.R. (1953): Estimation of Variance and Covariance Components*: Biometrics* 9, 226-252.

[5] Chigbu, P.E. (1998): Block *Designs: Efficiency Factors and Optimality Criteria for Comparison.* Limco Press Ltd, Enugu

[6] Onukogu, I.B and Chigbu, P.E. (2002): *Optional Design of Experiments and Mathematical programming.* AP Express Publishers Nsukka