

On Three-Way Unbalance Nested Analysis of Variance

Smaila S. Sanni and Eugene C. Ukaegbu
Department of Statistics, Faculty of Physical Sciences,
University of Nigeria, Nsukka, Nigeria

Abstract: Problem statement: In this study, we give a simple analytically tractable procedure for solving three-way unbalanced nested Analysis of Variance (ANOVA). In many realistic situations, unbalanced design was unavoidable due to natural constraints and missing data. **Approach:** Here, we present a comprehensive approach for addressing solutions of problems arising from unbalanced nested ANOVA. We consider the F-statistics under the different models. **Results:** Special attention was given to the construction of approximate F-test where exact F-test does not exist. Pseudo-degrees of freedoms were derived using the Satterthwaite's type approximation. **Conclusion:** In all derivations, we assume that the effects act independently and that the mean squares are independent. A numerical example is given to illustrate the solution procedure.

Key words: Unbalanced data, mean square, pseudo-degrees of freedom, approximate F-test, nested design, statistical software, least squares method, natural barriers, linear model, model parameters

INTRODUCTION

This study is intended to be a tutorial for those wishing to inform themselves about three-way unbalanced nested ANOVA. It focuses on the background understanding of model parameters estimation, derivation of sum of squares of effects, derivation of expected mean squares of effects, setting up of variance components and construction of approximate F-statistic where exact F-test does not exist. Certainly, these tools are not new, but insight into how they are applied to three-way unbalance nested ANOVA would open new vistas in our way of handling unbalanced hierarchical arrangements. The study is made to be analytically and computationally accessible. Readers need only some prior knowledge of two-way balanced nested design; see, for example, Montgomery (2008) and Dowdy and Chilko (2004) for discussion on balanced Nested Design.

An experiment with three factors A, B and C is said to be three-way unbalanced nested design if one factor say B is nested within another factor say A and factor C is nested within factor B such that each A level has b_1 B levels, each B level has C_{ij} C levels and n_{ijk} observations are drawn from each C level. It is pertinent to mention that this arrangement does not permit interaction between factors. Analysis of variance layout is called unbalanced if it has unequal subclass numbers. In addition, analysis of variance model is said to be unbalanced if the variance

of the difference between any two treatments is not a constant but depends on the treatments.

Many experimental situations could lead to nested arrangement. Thus, this kind of design has found extensive applications in industries, biological sciences, clinical studies. The analysis of nested design is difficult and the problem is complicated when faced with unbalanced nested design. Most statistical software now incorporate commands and guidelines for carrying out computations on unbalanced nested analysis of variance.

Kashiani and Saleh (2010) discussed three methods of estimating variance components for mixed-model. However, there are several methods for estimating variance components when the design is unbalanced. Each method influences the corresponding approximate F-test. Seeger (1970) gave a method for estimating variance components in unbalanced design. He used unweighted mean in his estimates and showed that these estimates are unbiased. Bush and Anderson (1963) considered numerical comparisons between variances of components of variance due to different methods. Tietjen and Moore (1968) developed a fast procedure for computing approximate F-test in unbalanced nested analysis of variance. Sahai and Ojeda (2004) discussed unbalanced nested analysis of variance for random effect model.

This study aims at providing background knowledge on three-way unbalance nested analysis of variance in a simple, straightforward, self-contained

Corresponding Author: Smaila S. Sanni, Department of Statistics, Faculty of Physical Sciences, University of Nigeria, Nsukka, Nigeria

account of the underlying theory. That is, the study exposes the analytical procedures that are hidden when analysis is performed with the aid of statistical software.

Unbalanced nested design could arise from a number of factors which include missing data; lose to follow-up or subjects get sick, limited resources and natural barriers. Missing data can result from overt errors in measurements, patients not showing up for scheduled visits in a clinical trial, loss of samples, (Bolton and Bon, 2009).

MATERIALS AND METHODS

Statistical model: The linear model for three-way unbalanced nested arrangement is given by Sahai and Ojeda (2004) as Eq. 1:

$$y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \lambda_{k(ij)} + e_{ijkl} \quad (1)$$

$$\begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b_i \\ k = 1, 2, \dots, c_{ij} \\ l = 1, 2, \dots, n_{ijk} \end{cases}$$

Where:

y_{ijkl} = The l^{th} observation within the k th level of factor C within j th level of factor B within i th level of factor A

μ = The overall mean

α_i = The effect due to the i th level of factor A

$\beta_{j(i)}$ = The effect due to the j th level of factor B nested within the i^{th} level of factor A

$\lambda_{k(ij)}$ = The effect due to the k th level of factor C nested within the j th level of factor B nested within the i th level of factor A

e_{ijkl} = The residual error of the observation y_{ijkl}

The following restrictions are imposed on Eq. 2:

$$\sum_i N_i \alpha_i = 0, \sum_{ij} n_{ij} \beta_{j(i)} = 0, \text{ and } \sum_{ijk} n_{ijk} \lambda_{k(ij)} = 0 \quad (2)$$

Note that:

$$\sum_{jkl} C = N_i C, \sum_{kl} C = n_{ij} C, \sum_l C = n_{ijk} C,$$

$$\sum_{ikl} C = N_j C, \sum_{ijl} C = N_k C \text{ and } \sum_{ijkl} C = N C$$

where, C is any constant.

Restriction (2) allows for estimating the model parameters using the least squares method.

Estimation of model parameters and sums of squares: The Least squares method has been widely used in estimating model parameters, see, for example Naisipour *et al.* (2008); Kavitha and Duraiswamy (2011) and Rencher and Bruce (2008). The least squares method and restriction (2) are used to estimate the model parameters. The estimates are as follows:

$$\hat{\mu} = \bar{y}_{\dots}, \hat{\alpha}_i = \bar{y}_{i\dots} - \bar{y}_{\dots}, \hat{\beta}_{j(i)} = \bar{y}_{ij\dots} - \bar{y}_{i\dots},$$

$$\hat{\lambda}_{k(ij)} = \bar{y}_{ik\dots} - \bar{y}_{ij\dots} \text{ and } e_{ijkl} = y_{ijkl} - \bar{y}_{ijk}$$

Where:

$$\bar{y}_{i\dots} = \frac{\sum_{ijkl} y_{ijkl}}{N} = \frac{T_{i\dots}}{N}, \bar{y}_{i\dots} = \frac{\sum_{jkl} y_{ijkl}}{N_i} = \frac{T_{i\dots}}{N_i}, \bar{y}_{ij\dots} = \frac{\sum_{kl} y_{ijkl}}{n_{ij}} = \frac{T_{ij\dots}}{n_{ij}}$$

Note that dots on subscripts denote summing or averaging over the subscripts.

See appendix I on how these estimates are derived. The sums of squares are given below:

$$SS_A = \sum_{ijkl} (\bar{y}_{i\dots} - \bar{y}_{\dots})^2 = \sum_i N_i (\bar{y}_{i\dots} - \bar{y}_{\dots})^2$$

$$= \frac{\sum_i T_{i\dots}^2}{N_i} - \frac{T^2}{N}$$

$$SS_B = \sum_{ijkl} (\bar{y}_{ij\dots} - \bar{y}_{i\dots})^2 = \sum_{ij} n_{ij} (\bar{y}_{ij\dots} - \bar{y}_{i\dots})^2 = \frac{\sum_{ij} T_{ij\dots}^2}{n_{ij}} - \frac{\sum_i T_{i\dots}^2}{N_i}$$

$$SS_C = \sum_{ijkl} (\bar{y}_{ijk} - \bar{y}_{ij\dots})^2 = \sum_{ijk} n_{ijk} (\bar{y}_{ijk} - \bar{y}_{ij\dots})^2$$

$$= \frac{\sum_{ijk} T_{ijk}^2}{n_{ijk}} - \frac{\sum_{ij} T_{ij\dots}^2}{n_{ij}}$$

$$SS_E = \sum_{ijkl} (y_{ijkl} - \bar{y}_{ijk})^2 = \sum_{ijkl} y_{ijkl}^2 - \frac{\sum_{ijk} T_{ijk}^2}{n_{ijk}}$$

$$= SS_{Total} = \sum_{ijkl} y_{ijkl}^2 - \frac{T^2}{N}$$

Expected mean squares: We give the expected mean squares for fixed effect model (model I), random effect model (model II) and mixed effect model (model III). We refer the reader to appendix II for the derivation of the expected mean squares under the different models.

Fixed effect model (Model I): Factor A is fixed, factor B is fixed and factor C is fixed.

Assumptions of the model: The assumptions of the model are given in Eq. 3:

$$\sum_i \alpha_i = 0, \sum_{ij} \beta_{j(i)} = 0, \sum_{ijk} \lambda_{k(ij)} = 0, e_{ijkl} \sim N(0, \sigma_e^2) \quad (3)$$

The expected mean squares are:

$$E(MS_A/I) = \frac{\sum_i N_i \alpha_i^2}{a-1} + \sigma_e^2, E(MS_B/I) = \frac{\sum_{ij} n_{ij} \beta_{j(i)}^2}{f_B} + \sigma_e^2$$

$$E(MS_C/I) = \frac{\sum_{ijk} n_{ijk} \lambda_{k(ij)}^2}{f_C} + \sigma_e^2, E(MS_E/I) = \sigma_e^2$$

Where:

$$f_B = a(b_i - 1) = \sum_i b_i - a, f_C = ab_i(c_{ij} - 1) = \sum_{ij} c_{ij} - \sum_i b_i$$

$$f_E = ab_i c_{ij} (n_{ijk} - 1) = N - \sum_{ij} c_{ij}$$

and E(MS) stands for expected mean square due to a given factor under a particular Model.

Random effect model (Model II): Factor A is random, factor B is random and factor C is random.

Assumptions of the model Eq. 4:

$$\alpha_i \sim N(0, \sigma_\alpha^2), \beta_{j(i)} \sim N(0, \sigma_\beta^2), \quad (4)$$

$$\lambda_{k(ij)} \sim N(0, \sigma_\lambda^2) \text{ and } e_{ijkl} \sim N(0, \sigma_e^2)$$

The expected mean squares are:

$$E(MS_A/II) = k_\alpha \sigma_\alpha^2 + k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2,$$

$$E(MS_B/II) = k_1 \sigma_\beta^2 + k_2 \sigma_\lambda^2 + \sigma_e^2$$

$$E(MS_C/II) = k_3 \sigma_\lambda^2 + \sigma_e^2$$

Where:

$$k_\alpha = \left(N - N^{-1} \sum_i N_i^2 \right) / a - 1,$$

$$k_\beta = \left(\sum_i N_i^{-1} \sum_j n_{ij}^2 - N^{-1} \sum_{ij} n_{ij}^2 \right) / a - 1$$

$$k_\lambda = \left(\sum_i N_i^{-1} \sum_{jk} n_{ijk}^2 - N^{-1} \sum_{ijk} n_{ijk}^2 \right) / a - 1$$

Mixed effect model (Model III_A): Factor A is fixed, factor B is random and factor C is random.

Assumptions of the model Eq. 5:

$$\sum_i \alpha_i = 0, \beta_{j(i)} \sim N(0, \sigma_\beta^2), \quad (5)$$

$$\lambda_{k(ij)} \sim N(0, \sigma_\lambda^2) \text{ and } e_{ijkl} \sim N(0, \sigma_e^2)$$

The expected mean squares for the model are:

$$E(MS_A/III_A) = \frac{\sum_i N_i \alpha_i^2}{a-1} + k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2,$$

$$E(MS_B/III_A) = k_1 \sigma_\beta^2 + k_2 \sigma_\lambda^2 + \sigma_e^2$$

$$E(MS_C/III_A) = k_3 \sigma_\lambda^2 + \sigma_e^2 \text{ and } E(MS_E/III_A) = \sigma_e^2$$

Mixed effect model (Model III_B): Factor A is random, factor B is fixed and factor C is random.

Assumptions of the model Eq. 6:

$$\alpha_i \sim N(0, \sigma_\alpha^2), \sum_j \beta_{j(i)} = 0 \quad (6)$$

$$\lambda_{k(ij)} \sim N(0, \sigma_\lambda^2) \text{ and } e_{ijkl} \sim N(0, \sigma_e^2)$$

The expected mean squares are given below:

$$E(MS_A/III_B) = k_\alpha \sigma_\alpha^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2,$$

$$E(MS_B/III_B) = \frac{\sum_{ij} n_{ij} \beta_{j(i)}^2}{f_B} + k_2 \sigma_\lambda^2 + \sigma_e^2,$$

$$E(MS_C/III_B) = k_3 \sigma_\lambda^2 + \sigma_e^2 \text{ and } E(MS_E/III_B) = \sigma_e^2$$

Mixed effect model (Model III_C): Factor A is random, factor B is random and factor C is fixed.

Assumptions of the model Eq. 7:

$$\alpha_i \sim N(0, \sigma_\alpha^2), \beta_{j(i)} \sim N(0, \sigma_\beta^2) \quad (7)$$

$$\sum_{ijk} \lambda_{k(ij)} = 0 \text{ and } e_{ijkl} \sim N(0, \sigma_e^2)$$

The expected mean squares for the model are:

$$E(MS_A/III_C) = k_\alpha \sigma_\alpha^2 + k_\beta \sigma_\beta^2 + \sigma_e^2,$$

$$E(MS_B/III_C) = k_1 \sigma_\beta^2 + \sigma_e^2$$

$$E(MS_C/III_C) = \frac{\sum_{ijk} n_{ijk} \lambda_{k(ij)}^2}{f_C} + \sigma_e^2$$

And:

$$E(MS_E/III_C) = \sigma_e^2$$

Mixed effect model (Model III_{AC}): Factor A is fixed, factor B is random and factor C is fixed.

Assumptions of the model Eq. 8:

$$\sum_i \alpha_i = 0, \beta_{j(i)} \sim N(0, \sigma_\beta^2),$$

$$\sum_{ijk} \lambda_{k(ij)} = 0 \text{ and } e_{ijkl} \sim N(0, \sigma_e^2) \tag{8}$$

The expected mean squares for the model are:

$$E(MS_A/III_{AC}) = \frac{\sum_i N_i \alpha_i^2}{a-1} + k_\beta \sigma_\beta^2 + \sigma_e^2,$$

$$E(MS_B/III_{AC}) = k_1 \sigma_\beta^2 + \sigma_e^2,$$

$$E(MS_C/III_{AC}) = \frac{\sum_{ijk} n_{ijk} \lambda_{k(ij)}^2}{f_c} + \sigma_e^2 \text{ and } E(MS_E/III_{AC}) = \sigma_e^2$$

Estimation of the variance components: First, we consider the random effect model (Model II) and give procedure for estimating the variance components as follows:

$$E(MS_E/II) = \sigma_e^2 \text{ and set } \hat{\sigma}_e^2 = MS_E \text{ since}$$

$$E(MS_E/II) \hat{=} MS_E;$$

where, " $\hat{=}$ " reads "estimated by".

This procedure is similar to Henderson's Method I. Sahai and Ojeda (2004) discussed Henderson's methods for estimating variance components for unbalanced data:

$$E(MS_C/II) = k_3 \sigma_\lambda^2 + \sigma_e^2 \text{ and } MS_C = k_3 \hat{\sigma}_\lambda^2 + MS_E$$

since $E(MS_C/II) \hat{=} MS_C$

$$\hat{\sigma}_\lambda^2 = \frac{MS_C - MS_E}{k_3}$$

$$E(MS_B/II) = k_1 \sigma_\beta^2 + k_2 \sigma_\lambda^2 + \sigma_e^2 \text{ and}$$

$$MS_B = k_1 \hat{\sigma}_\beta^2 + k_2 \hat{\sigma}_\lambda^2 + \hat{\sigma}_e^2 \text{ since } E(MS_B/II) \hat{=} MS_B$$

$$MS_C = k_1 \hat{\sigma}_\beta^2 + k_2 \left(\frac{MS_C - MS_E}{k_3} \right) + MS_E$$

$$\hat{\sigma}_\beta^2 = \frac{MS_B - \frac{k_2}{k_3} MS_C - \left(1 - \frac{k_2}{k_3}\right) MS_E}{k_1},$$

$$\hat{\sigma}_\beta^2 = \frac{MS_B - \gamma MS_C - (1 - \gamma) MS_E}{k_1},$$

Where:

$$\gamma = \frac{k_2}{k_3}.$$

$$E(MS_A/II) = k_\alpha \sigma_\alpha^2 + k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2 \text{ and set}$$

$$MS_A = k_\alpha \hat{\sigma}_\alpha^2 + k_\beta \hat{\sigma}_\beta^2 + k_\lambda \hat{\sigma}_\lambda^2 + \hat{\sigma}_e^2$$

$$k_\alpha \hat{\sigma}_\alpha^2 = MS_A - k_\beta \left[\frac{MS_B - \frac{k_2}{k_3} MS_C - \left(1 - \frac{k_2}{k_3}\right) MS_E}{k_1} \right]$$

$$- k_\lambda \left(\frac{MS_C - MS_E}{k_3} \right) - MS_E$$

$$= MS_A - \frac{k_\beta}{k_1} MS_B - (k_\lambda/k_3 - k_\beta k_2/k_1 k_3) MS_C$$

$$- \left[1 - k_\beta/k_1 - (k_\lambda/k_3 - k_\beta k_2/k_1 k_3) \right] MS_E$$

$$\hat{\sigma}_\alpha^2 = \frac{MS_A - \theta_1 MS_B - \theta_2 MS_C - (1 - \theta_1 - \theta_2) MS_E}{k_\alpha}$$

Where:

$$\theta_1 = k_\beta/k_1$$

And:

$$\theta_2 = (k_\lambda/k_3 - k_\beta k_2/k_1 k_3)$$

The same procedure is used to obtain the variance components for the different models and the results are given below.

Mixed effect model (Model III_A): Factor A is fixed; factor B and factor C are random.

The variance components are:

$$\hat{\sigma}_e^2 = MS_E, \sigma_\lambda^2 = \frac{MS_C - MS_E}{k_3}$$

$$\hat{\sigma}_\beta^2 = \frac{MS_B - \gamma MS_C - (1 - \gamma) MS_E}{k_1}$$

Where:

$$\gamma = k_2/k_3$$

Mixed effect model (Model III_B): Factor B is fixed, factor A and factor C are random.

The variance components are:

$$\hat{\sigma}_e^2 = MS_E, \hat{\sigma}_\lambda^2 = \frac{MS_C - MS_E}{k_3} \text{ and}$$

$$\hat{\sigma}_\alpha^2 = \frac{MS_A - \upsilon MS_C - (1 - \upsilon) MS_E}{k_\alpha}$$

where, $\nu = k_\lambda/k_3$.

Mixed effect model (Model III_C): Factor C is fixed, factor A and factor B are random.

The variance components are:

$$\hat{\sigma}_e^2 = MS_E, \hat{\sigma}_\beta^2 = \frac{MS_B - MS_E}{k_1} \text{ and}$$

$$\hat{\sigma}_\alpha^2 = \frac{MS_A - \theta_1 MS_C - (1 - \theta_1) MS_E}{k_\alpha}$$

Mixed effect model (Model III_{AC}): Factor B is random, factor A and factor C are fixed.

The variance components are:

$$\hat{\sigma}_e^2 = MS_E \text{ and } \hat{\sigma}_\beta^2 = \frac{MS_B - MS_E}{k_1}$$

Test procedures for the models: The expected mean squares help in the determination of appropriate test statistics for testing hypotheses about the effects. Expected mean squares determine which hypotheses are tested by each mean square. An F-statistic can only be formed when, under appropriate hypothesis, two expected mean squares have the same value (Mason *et al.*, 2003). In this study, we shall consider the test statistics under the different models.

Model I (Fixed-effect model): The hypothesis $H:\alpha_i=0$ is tested by MS_A/MS_E , $H:\beta_{j(i)}=0$ is tested by MS_B/MS_E and $H:\lambda_{k(ij)}=0$ is tested by MS_C/MS_E . The ANOVA estimate of σ_e^2 is MS_E .

It is observed that under null hypothesis, $E(MS_A/I) = \sigma_e^2$, $E(MS_B/I) = \sigma_e^2$ and $E(MS_C/I) = \sigma_e^2$.

Model II (Random effect model): The hypotheses to be tested are:

- $H_{0A} : \sigma_\alpha = 0$
- $H_{0B} : \sigma_\beta = 0$
- $H_{0C} : \sigma_\lambda = 0$

We observe that $E(MS_A/II)$, under the null hypothesis $H_{0A} : \sigma_\alpha = 0$, does not have the same value as $E(MS_B/II)$ because $k_\beta \neq k_1$ and $k_\lambda \neq k_2$. Hence, there is no exact F-test for testing $H_{0A} : \sigma_\alpha = 0$. Therefore, we construct a new mean square, $M\hat{S}_\theta$ which is independent of MS_A . We employ the approach proposed.

Let $MS_\theta = k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2$. This is the expected mean square due to factor A under the null hypothesis $H_{0A} : \sigma_\alpha = 0$.

The variance estimates are:

$$\hat{\sigma}_\beta^2 = \frac{MS_B - \frac{k_2}{k_3} MS_C - \left(1 - \frac{k_2}{k_3}\right) MS_E}{k_1},$$

$$\hat{\sigma}_\lambda^2 = \frac{MS_C - MS_E}{k_3} \text{ and } \hat{\sigma}_e^2 = MS_E$$

Therefore:

$$M\hat{S}_\theta = k_\beta \left[\frac{MS_B - \frac{k_2}{k_3} MS_C - \left(1 - \frac{k_2}{k_3}\right) MS_E}{k_1} \right]$$

$$+ k_\lambda \left(\frac{MS_C - MS_E}{k_3} \right) + MS_E$$

$$= \theta_1 MS_B + \theta_2 MS_C + (1 - \theta_1 - \theta_2) MS_E$$

Where:

$$\theta_1 = k_\beta/k_1$$

$$\theta_2 = (k_\lambda/k_3 - k_\beta k_2/k_1 k_3)$$

We now show that:

$$E(M\hat{S}_\theta) = k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2$$

$$E(M\hat{S}_\theta) = \theta_1 E(MS_B) + \theta_2 E(MS_C) + (1 - \theta_1 - \theta_2) E(MS_E)$$

$$= \theta_1 k_1 \sigma_\beta^2 + \theta_1 k_2 \sigma_\lambda^2 + \theta_2 k_3 \sigma_\lambda^2 + \sigma_e^2$$

$$= k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2$$

The approximate F-statistics: The hypothesis $H_{0A} : \sigma_\alpha = 0$ is tested by:

$$F_A = \frac{MS_A}{M\hat{S}_\theta}$$

where, F_A is approximately F-distributed with a-1 and f_θ degrees of freedom.

Where:

$$M\hat{S}_\theta = \theta_1 MS_B + \theta_2 MS_C + (1 - \theta_1 - \theta_2) MS_E$$

And:

$$f_{\theta} = \left(M\hat{S}_{\theta} \right)^2 \left[\frac{\theta_1^2 (MS_B)^2}{f_B} + \frac{\theta_2^2 (MS_C)^2}{f_C} + \frac{(1-\theta_1-\theta_2)^2 (MS_E)^2}{f_E} \right]^{-1}$$

Derivation of pseudo-degrees of freedom, f_{θ} : We construct the degrees of freedom f_{θ} by applying an approximation due to Satterthwaite (1946). A Satterthwaite approximation is based on assuming that a variance estimator has a chi-square distribution and solving for the implied degrees of freedom, using the method of moments (Valliant and Rust, 2010):

$$M\hat{S}_{\theta} = \theta_1 MS_B + \theta_2 MS_C + (1-\theta_1-\theta_2) MS_E$$

If we assume that MS_B and MS_C are independent, then Eq. 9 and 10:

$$\begin{aligned} \text{Var}(M\hat{S}_{\theta}) &= \theta_1^2 \text{Var}(MS_B) \\ &+ \theta_2^2 \text{Var}(MS_C) + (1-\theta_1-\theta_2)^2 \text{Var}(MS_E) \end{aligned} \quad (9)$$

Recall that:

$$\begin{aligned} MS_E &= SS_E/f_E \text{ and that } SS/\sigma_e^2 \sim \chi_e^2 : \\ \text{Var}(MS_E) &= \frac{1}{f_E^2} \text{Var}(SS_E) \end{aligned} \quad (10)$$

$\text{Var}(SS_E/\sigma_e^2) \approx \text{Var}(\chi_{f_E}^2) \Rightarrow \frac{1}{\sigma_e^4} \text{Var}(SS_E) \approx 2f_E$, since the variance of chi-square distribution is $2f_E$.

Substituting the result, $\text{Var}(SS_E) \approx 2\sigma_e^4 f_E$ in (10), we have Eq. 11:

$$\text{Var}(MS_E) \approx 2(\sigma_e^2)^2 f_E^{-1} \quad (11)$$

By extending our idea of (11) in (9), we get:

$$\begin{aligned} 2(M\hat{S}_{\theta})^2 f_{\theta}^{-1} &= 2\theta_1^2 (MS_B)^2 f_B^{-1} \\ &+ 2\theta_2^2 (MS_C)^2 + 2(1-\theta_1-\theta_2)^2 (MS_E)^2 f_E^{-1} \\ f_{\theta} &= \left(M\hat{S}_{\theta} \right)^2 \left[\frac{\theta_1^2 (MS_B)^2}{f_B} + \frac{\theta_2^2 (MS_C)^2}{f_C} + \frac{(1-\theta_1-\theta_2)^2 (MS_E)^2}{f_E} \right]^{-1} \end{aligned}$$

Next, we consider testing $H_{0B}:\sigma_{\beta} = 0$. There is no exact F-test for testing this hypothesis because $E(MS_B/II)$, under $H_{0B}:\sigma_{\beta} = 0$, does not have the same value as $E(MS_C/II)$ since $k_2 \neq k_3$. We derive the approximate F-test using the same procedure as presented above and the result is as follows:

The hypothesis $H_{0B}:\sigma_{\beta} = 0$ is tested by $F_B = \frac{MS_B}{M\hat{S}_{\gamma}}$.

F_B is approximately F-distributed with f_B and f_{γ} degrees of freedom, where:

$$M\hat{S}_{\gamma} = \gamma MS_C + (1-\gamma) MS_E$$

And:

$$f_{\gamma} = \left(M\hat{S}_{\gamma} \right)^2 \left[\frac{\gamma^2 (MS_C)^2}{f_C} + \frac{(1-\gamma)^2 (MS_E)^2}{f_E} \right]^{-1}$$

Finally, the hypothesis $H_{0C}:\sigma_{\lambda} = 0$ is tested by $\frac{MS_C}{MS_E}$.

Model III_A (Mixed effect model; where factor A is fixed): The hypotheses to be tested are:

- $H_{0A} : \alpha_i = 0$
- $H_{0B} : \sigma_{\beta} = 0$
- $H_{0C} : \sigma_{\lambda} = 0$

The hypothesis $H_{0A}:\sigma_{\alpha}=0$ is tested by:

$$F_A = \frac{MS_A}{M\hat{S}_{\theta}}$$

where, F_A is approximately F-distributed with $a-1$ and f_{θ} degrees of freedom, where:

$$M\hat{S}_{\theta} = \theta_1 MS_B + \theta_2 MS_C + (1-\theta_1-\theta_2) MS_E$$

And:

$$f_{\theta} = \left(M\hat{S}_{\theta} \right)^2 \left[\frac{\theta_1^2 (MS_B)^2}{f_B} + \frac{\theta_2^2 (MS_C)^2}{f_C} + \frac{(1-\theta_1-\theta_2)^2 (MS_E)^2}{f_E} \right]^{-1}$$

The hypothesis $H_{0B}:\sigma_\beta=0$ is tested by $F_B = \frac{MS_B}{MS_\gamma}$; F_B is approximately F-distributed with $f_\theta f_B$ and $f_\theta f_\gamma$ degrees of freedom.

Where:

$$MS_{\hat{\gamma}} = \gamma MS_C + (1-\gamma)MS_E \text{ and}$$

$$f_\gamma = (MS_{\hat{\gamma}})^2 \left[\frac{\gamma^2 (MS_C)^2}{f_C} + \frac{(1-\gamma)^2 (MS_E)^2}{f_E} \right]^{-1}$$

The hypothesis, $H_{0C}:\sigma_\lambda=0$ is tested by $\frac{MS_C}{MS_E}$.

Model III_B (Mixed effect model; where factor B is fixed): The hypotheses to be tested are:

- $H_{0A} : \sigma_\alpha = 0$
- $H_{0B} : \beta_{j(i)} = 0$
- $H_{0C} : \sigma_\lambda = 0$

The hypothesis $H_{0A}:\sigma_\alpha=0$ is tested by:

$$F_A = \frac{MS_A}{MS_{\hat{v}_0}}$$

where, F_A is approximately F-distributed with a-1 and f_v degrees of freedom, where:

$$MS_{\hat{v}_0} = v MS_C + (1-v)MS_E \text{ and}$$

$$f_v = (MS_{\hat{v}_0})^2 \left[\frac{v^2 (MS_C)^2}{f_C} + \frac{(1-v)^2 (MS_E)^2}{f_E} \right]^{-1}$$

The hypothesis $H_{0B}:\beta_{j(i)}=0$ is tested by:

$$F_B = \frac{MS_B}{MS_{\hat{\gamma}}}$$

where, F_B is approximately F-distributed with f_B and f_γ degrees of freedom.

The hypothesis $H_{0C}:\sigma_\lambda=0$ is tested by $\frac{MS_C}{MS_E}$.

Model III_C (Mixed effect model; where factor C is fixed): The hypotheses to be tested are:

- $H_{0A} : \sigma_\alpha = 0$
- $H_{0B} : \sigma_\beta = 0$
- $H_{0C} : \lambda_{k(ij)} = 0$

The hypothesis $H_{0A}:\sigma_\alpha=0$ is tested by:

$$F_A = \frac{MS_A}{MS_{\hat{\theta}_1}}$$

where, F_A is approximately F-distributed with a-1 and F_θ degrees of freedom, where:

$$MS_{\hat{\theta}_1} = \theta_1 MS_C + (1-\theta_1)MS_E \text{ and}$$

$$f_{\theta_1} = (MS_{\hat{\theta}_1})^2 \left[\frac{\theta_1^2 (MS_B)^2}{f_B} + \frac{(1-\theta_1)^2 (MS_E)^2}{f_E} \right]^{-1}$$

The hypothesis $H_{0B}:\beta_{j(i)}=0$ is tested by:

$$F_B = \frac{MS_B}{MS_E}$$

and the hypothesis $H_{0C}:\lambda_{k(ij)}=0$ is tested by:

$$F_C = \frac{MS_C}{MS_E}$$

Model III_{AC} (Mixed effect model; where factor A and factor C are fixed): The hypotheses to be tested are:

- $H_{0A} : \alpha_i = 0$
- $H_{0B} : \sigma_\beta = 0$
- $H_{0C} : \lambda_{k(ij)} = 0$

The hypothesis $H_{0A}:\sigma_\alpha=0$ is tested by:

$$F_A = \frac{MS_A}{MS_{\hat{\theta}_1}}$$

where, F_A is approximately F-distributed with a-1 and f_{θ_1} degrees of freedom, where:

$$MS_{\hat{\theta}_1} = \theta_1 MS_C + (1-\theta_1)MS_E \text{ and}$$

$$f_{\theta_1} = (MS_{\hat{\theta}_1})^2 \left[\frac{\theta_1^2 (MS_B)^2}{f_B} + \frac{(1-\theta_1)^2 (MS_E)^2}{f_E} \right]^{-1}$$

The hypothesis $H_{0B} : \sigma_{\beta} = 0$ is by $F_B = \frac{MS_B}{MS_E}$ and the hypothesis $H_{0C} : \lambda_{k(ij)} = 0$ is tested by $\frac{MS_C}{MS_E}$.

RESULTS AND DISCUSSION

Numerical example: We give illustrative example with hypothetical data on the hardness (crushing strength) of a particular tablet produced by a pharmaceutical company. The company has two production sites within a particular region. Two of the machines for producing the tablet are randomly selected from site one and three machines from site two. Based on the production capacity of each machine, two batches of the produced tablets are randomly selected from each of the machines at site one. At site two, two batches are selected from machine one, three batches from machine two and one batch from machine three. The measures of the crushing strength of the tablets randomly selected from each of the batches are recorded in Table 1 below. The company wants to investigate if the batch to batch variability within machines, machine to machine variability within sites and site effect are significant sources of variation in the crushing strength of the tablet. This is a case where the site is fixed, machines and batches are random.

We now proceed to construct the n_{ijk} -table (Table 3) to make the computation easier. The n_{ijk} -table is table of counts of the number of observations for the i^{th} Site, j^{th} Machine and k^{th} batch.

Computation of sum of squares:

Let:

$$R = T^2 / N = (369.14)^2 / 78 = 1746.978713,$$

$$R_i = \sum_i T_i^2 / N_i = \frac{(152.77)^2}{32} + \frac{(216.37)^2}{46} = 1747.072156$$

$$R_{ij} = \sum_{ij} T_{ij}^2 / n_{ij} = (85.27)^2 / 18 + (67.50)^2 / 14 + (113.66)^2 / 10 + (58.30)^2 / 12 = 1748.129827$$

$$R_{ijk} = \sum_{ijk} T_{ijk}^2 / n_{ijk} = (36.82)^2 / 8 + \dots + (58.30)^2 / 12 = 1751.519083$$

$$R_{ijkl} = \sum_{ijkl} y_{ijkl}^2 = (4.08)^2 + \dots + (3.90)^2 = 1836.7302$$

Therefore, the sums of squares are:

$$SS_{\text{site}} = R_i - R = 0.093443, SS_{\text{machine}} = R_{ij} - R_i = 1.057671,$$

$$SS_{\text{Batch}} = R_{ijk} - R_{ij} = 3.389256, SS_{\text{error}} = R_{ijkl} - R_{ijk} = 85.211117,$$

$$SS_{\text{Total}} = R_{ijkl} - R = 89.751487$$

The results of these computations are summarized in Table 3 below. The values marked by asterisk (*) in

column five are obtained by the approximate F-test developed in this study. No exact F-test exists for such tests.

How the degrees of freedom are calculated:

$$f_{\beta} = \sum_i b_i - a = b_1 + b_2 - a = 2 + 3 - 2 = 3$$

$$f_c = \sum_{ij} C_{ij} - \sum_i b_i = (C_{11} + C_{12} + C_{21} + C_{22} + C_{23}) - (b_1 + b_2)$$

$$(2 + 2 + 3 + 2 + 1) - (2 + 3) = 5$$

$$f_e = N - \sum_{ij} C_{ij} = 78 - 10 = 68$$

Calculation of the weighting factors: The values of the n_{ijk} 's in Table 2 are used for computing the weighting factors as follows:

$$k_{\alpha} = \left(N - N^{-1} \sum_i N_i^2 \right) / a - 1$$

$$= \left[78 - \frac{1}{78} (32^2 + 46^2) \right] / 39 = 37.74358974$$

$$k_1 = \left(N - \sum_i N_i^{-1} \sum_{ij} n_{ij}^2 \right) / f_{\beta}$$

$$= \left[78 - \left(\frac{1}{32} (18^2 + 14^2) + \frac{1}{46} (24^2 + 10^2 + 12^2) \right) \right] / 3$$

$$= 14.64130435$$

$$k_2 = \left(\sum_{ij} n_{ij}^{-1} \sum_k n_{ijk}^2 - \sum_i N_i^{-1} \sum_{jk} n_{ijk}^2 \right) / f_{\beta}$$

$$= \left[\left(\frac{1}{18} (8^2 + 10^2) + \frac{1}{14} (6^2 + 8^2) + \frac{1}{24} (10^2 + 6^2 + 8^2) \right) + \left(\frac{1}{10} (6^2 + 4^2) + \frac{1}{12} (12^2) \right) - \left(\frac{1}{32} (8^2 + 10^2 + 6^2 + 8^2) + \frac{1}{46} (10^2 + 6^2 + 8^2 + 6^2 + 4^2 + 12^2) \right) \right] / 3$$

$$= 8.309535312.$$

$$k_3 = \left[78 - \frac{13163}{315} \right] / 5 = 7.242539683.$$

$$k_{\beta} = \left[\frac{3135}{92} - \frac{670}{39} \right] = 16.89659978.$$

$$k_{\lambda} = \left[\frac{1551}{92} - \frac{110}{13} \right] = 8.397157191.$$

$$\theta_1 = 1.15403651, \theta_2 = -0.1646314679.$$

$$\gamma = 1.147323408, \nu = 1.159421634.$$

$$M\hat{S}_{\theta} = 0.3085446029, M\hat{S}_{\gamma} = 0.5931028995.$$

Table 1: Drug hardness

Site	I				II						
	1		2		1			2		3	
Machine	1		2		1			2		3	
Batch	B ₁	B ₂	B ₁	B ₂	B ₁	B ₂	B ₃	B ₁	B ₂	B ₁	
Hardness (kg)	4.08	5.69	3.55	4.67	4.79	3.84	3.13	3.34	5.01	5.71	
	5.11	6.24	4.37	3.82	5.11	5.73	7.21	4.69	5.07	4.23	
	4.56	5.42	6.31	5.03	5.20	4.42	5.89	5.38	3.14	4.77	
	5.74	4.13	4.64	6.90	3.68	5.65	4.45	3.54	6.24	6.22	
	4.80	5.22	6.17	4.30	6.41	4.50		3.14	4.30	5.17	
	3.83	5.47	6.24	3.01	3.93	4.82	5.03	3.70		6.10	
	5.08	4.11		3.71	2.25			4.56		5.41	
	3.62	5.05		4.78	6.86	5.29				4.05	
											3.22
											3.76
T _{ijk.}	36.82	48.45	31.28	36.22	46.00	28.96	38.70	24.95	19.46	58.30	
T _{ij.}	85.27			67.50		113.66			44.41	58.30	
T _{i...}		152.77							216.37		
T _{....}					369.14						

Table 2: n_{ijk}- table

n ₁₁₁ =8	N ₁₂₁ =6	n ₂₁₁ =10	n ₂₂₁ =6	n ₂₃₁ =12
n ₁₁₂ =10	n ₁₂₂ =8	n ₂₁₂ =6	n ₂₂₂ =4	
		n ₂₁₃ =4		
n ₁₁ =18	n ₁₂ =14	n ₂₁ =24	n ₂₂ =10	n ₂₃ =12
	N ₁ =32		N ₂ =46	
		N=78		

Table 3: ANOVA Table for drug hardness

Source of variation	Degree of freedom	Sum of squares	Mean square	Expected mean square	F-test
Site	1	0.093443	0.093443	$\sum_i N_i \alpha_i^2 / a - 1 + 16.9\sigma_\beta^2 + 8.4\sigma_\lambda^2 + \sigma_c^2$	0.3028509*
Machine	3	1.057671	0.352557	$14.6\sigma_\beta^2 + 8.3\sigma_\lambda^2 + \sigma_c^2$	0.5944281*
Batch	5	3.389256	0.6778512	$7.2\sigma_\lambda^2 + \sigma_c^2$	0.5409374
Error	68	85.211117	1.2531047	σ_c^2	
Total	77	89.751487			

The values of the approximate F-tests for factors A and B are obtained as follows:

$$F_A = MS_A / MS_{\hat{\theta}} = 0.3028509$$

$$F_B = MS_B / MS_{\hat{\gamma}} = 0.5944281$$

On the basis of the results shown in Table 3, we conclude that there is no difference in Sites, there is no significant difference in machine-to-machine variability within the Sites and there is no difference in batch-to-batch variability within the machines at one percent and five percent significant levels, respectively.

Appendix I: Procedure for estimating model parameters: We provide here procedure for estimating the parameters of (1). The least squares method and restriction (2) are used to estimate the model parameters as follows Eq. 12 and 13:

$$\sum e_{ijkl}^2 = \sum (y_{ijkl} - \mu - \alpha_i - \beta_{j(i)} - \lambda_{k(ij)})^2 \tag{12}$$

Differentiating (12) with respect to μ and equating to zero, we have:

$$\sum (y_{ijkl} - \mu - \alpha_i - \beta_{j(i)} - \lambda_{k(ij)}) = 0 \tag{13}$$

Solve for μ by summing on i, j, k and l in (13), we get:

$$\sum_{ijkl} y_{ijkl} - N\hat{\mu} = 0$$

by Eq.2. Therefore:

$$\hat{\mu} = \frac{\sum_{ijkl} y_{ijkl}}{N} = \frac{T_{\dots}}{N} = \bar{y}_{\dots}$$

Solve for α_i by summing on j, k and l in (4):

$$\sum_{jkl} y_{ijkl} - N_i \hat{\mu} - N_i \alpha_i = 0$$

Since:

$$\sum_j c_{ij} \beta_{j(i)} = \sum_{jk} n_{ijk} \lambda_{k(ij)} = 0 \text{ by (2):}$$

$$\hat{\alpha}_i = \frac{\sum_{jkl} y_{ijkl}}{N_i} - \hat{\mu} = \frac{T_{i...}}{N_i} - \hat{\mu} = \bar{y}_{i...} - \bar{y}_{...}$$

Solve for $\beta_{j(i)}$ by summing on k and l in (13):

$$\sum_{kl} y_{ijkl} - n_{ij} \hat{\mu} - n_{ij} \hat{\alpha}_i - n_{ij} \hat{\beta}_{j(i)} = 0 ;$$

Since:

$$\sum_{kl} \lambda_{k(ij)} = 0$$

by(2):

$$\hat{\beta}_{j(i)} = \frac{\sum_{kl} y_{ijkl}}{n_{ij}} - \hat{\mu} - \hat{\alpha}_i = \frac{T_{ij..}}{n_{ij}} - \hat{\mu} - \hat{\alpha}_i = \bar{y}_{ij..} - \bar{y}_{i...}$$

Solve for $\lambda_{k(ij)}$ by summing on l in (13):

$$\begin{aligned} \hat{\lambda}_{k(ij)} &= \frac{\sum_{l} y_{ijkl}}{n_{ijk}} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_{j(i)} - \hat{\lambda}_{k(ij)} \\ &= \frac{T_{ijk.}}{n_{ijk}} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_{j(i)} - \hat{\lambda}_{k(ij)} = \bar{y}_{ik..} - \bar{y}_{ij..} \end{aligned}$$

Solve for e_{ijkl} by substituting $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_{j(i)}$ and $\lambda_{k(ij)}$ in (1), we get:

$$e_{ijkl} = y_{ijkl} - \bar{y}_{ijkl}$$

Appendix II. Derivation of expected mean squares:

We derive the expected mean squares for fixed effect model (model I), random effect model (model II) and mixed effect model (model III).

Fixed effect model (Model I):

$$E(MS_A/I) = E \left\{ \frac{\sum_i N_i (\bar{y}_{i...} - \bar{y}_{...})^2}{a-1} \right\}$$

where, $E(MS_A/I)$ stands for expected mean square due to factor A under Model I.

We obtain the equivalents of $\bar{y}_{i...}$ and $\bar{y}_{...}$ using the assumptions of the model and (1) as:

$$\bar{y}_{i...} = \mu + \alpha_i + \bar{e}_{i...}$$

And:

$$\bar{y}_{...} = \mu + \bar{e}_{...} ;$$

Where:

$$\bar{e}_{i...} = \frac{\sum_{jkl} e_{ijkl}}{N_i}$$

And:

$$\bar{e}_{...} = \frac{\sum_{ijkl} e_{ijkl}}{N}$$

Therefore:

$$E(MS_A/I) = E \left\{ \frac{\sum_i N_i [\alpha_i + (\bar{e}_{i...} - \bar{e}_{...})]^2}{a-1} \right\} = \frac{\sum_i N_i \alpha_i^2}{a-1} + \sigma_e^2$$

$$E(MS_B/I) = E \left\{ \frac{\sum_{ij} n_{ij} (\bar{y}_{ij..} - \bar{y}_{i...})^2}{f_B} \right\}$$

We obtain the equivalents for $\bar{y}_{i...}$ and $\bar{y}_{ij..}$ using (1) and (3) as:

$$\bar{y}_{i...} = \mu + \alpha_i + \bar{e}_{i...}$$

And:

$$\bar{y}_{ij..} = \mu + \alpha_i + \beta_{j(i)} + \bar{e}_{ij..}$$

Therefore:

$$E(MS_B/I) = E \left\{ \frac{\sum_{ij} n_{ij} [\beta_{j(i)} + (\bar{e}_{ij..} - \bar{e}_{i...})]^2}{f_B} \right\} = \frac{\sum_{ij} n_{ij} \beta_{j(i)}^2}{f_B} + \sigma_e^2$$

$$E(MS_C/I) = E \left\{ \frac{\sum_{ijk} n_{ijk} (\bar{y}_{ijk.} - \bar{y}_{ij..})^2}{f_C} \right\}$$

We obtain the equivalents for $\bar{y}_{i...}$ and $\bar{y}_{ij..}$ using (1) and (3) as:

$$\bar{y}_{ij..} = \mu + \alpha_i + \beta_{j(i)} + \bar{e}_{ij..}$$

And:

$$\bar{y}_{ijk.} = \mu + \alpha_i + \beta_{j(i)} + \lambda_{k(ij)} + \bar{e}_{ijk.}$$

Therefore:

$$E(MS_C/I) = E \left\{ \frac{\sum_{ijk} n_{ijk} [\lambda_{k(ij)} + (\bar{e}_{ijk.} - \bar{e}_{ij..})]^2}{f_C} \right\} = \frac{\sum_{ijk} n_{ijk} \lambda_{k(ij)}^2}{f_C} + \sigma_e^2$$

$E(MS_E/I)$

$$= E \left\{ \frac{\sum_{ijkl} (y_{ijkl} - \bar{y}_{ijk.})^2}{f_E} \right\} = E \left\{ \frac{\sum_{ijkl} (e_{ijkl} - \bar{e}_{ijk.})^2}{f_E} \right\} = \sigma_e^2$$

Random effect model (Model II):

$$E(MS_A/II) = E \left\{ \frac{\sum_i N_i (\bar{y}_{i...} - \bar{y}_{....})^2}{a-1} \right\}$$

We obtained the equivalents of $\bar{y}_{i...}$ and $\bar{y}_{....}$ using (1) and (4) as:

$$\bar{y}_{i...} = \mu + \alpha_i + \bar{\beta}_{(i)} + \bar{\lambda}_{(i)} + \bar{e}_{i...}$$

And:

$$\bar{y}_{....} = \mu + \bar{\alpha} + \bar{\beta}_{(.)} + \bar{\lambda}_{(.)} + \bar{e}_{....}$$

Where:

$$\bar{\alpha} = \frac{\sum_i N_i \alpha_i}{N}, \bar{\beta}_{(i)} = \frac{\sum_j n_{ij} \beta_{j(i)}}{N_i}, \bar{\lambda}_{(i)} = \frac{\sum_{jk} n_{ijk} \lambda_{k(ij)}}{N_i}$$

$$, \bar{\beta}_{(.)} = \frac{\sum_{ij} n_{ij} \beta_{j(i)}}{N} \text{ and } \bar{\lambda}_{(.)} = \frac{\sum_{ijk} n_{ijk} \lambda_{k(ij)}}{N}$$

Therefore:

$$E(MS_A/II) = E \left\{ \frac{\sum_i N_i [(\alpha_i - \bar{\alpha}) + (\bar{\beta}_{(i)} - \bar{\beta}_{(.)}) + (\bar{\lambda}_{(i)} - \bar{\lambda}_{(.)}) + (\bar{e}_{i...} - \bar{e}_{....})]^2}{a-1} \right\}$$

$$= E \left\{ \frac{\sum_i N_i \alpha_i^2 - N^{-1} \sum_i N_i^2 \alpha_i^2}{a-1} \right\}$$

$$+ E \left\{ \frac{\sum_i N_i^{-1} \sum_j n_{ij}^2 \beta_{j(i)}^2 - N^{-1} \sum_{ij} n_{ij}^2 \beta_{j(i)}^2}{a-1} \right\}$$

$$+ E \left\{ \frac{\sum_i N_i^{-1} \sum_{jk} n_{ijk}^2 \lambda_{k(ij)}^2 - N^{-1} \sum_{ijk} n_{ijk}^2 \lambda_{k(ij)}^2}{a-1} \right\}$$

$$+ E \left\{ \frac{\sum_i N_i (\bar{e}_{i...} - \bar{e}_{....})^2}{a-1} \right\}$$

$$= k_\alpha \sigma_\alpha^2 + k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2$$

Where:

$$k_\alpha = \left(N - N^{-1} \sum_i N_i^2 \right) / a - 1$$

$$, k_\beta = \left(\sum_i N_i^{-1} \sum_j n_{ij}^2 - N^{-1} \sum_{ij} n_{ij}^2 \right) / a - 1$$

And:

$$k_\lambda = \left(\sum_i N_i^{-1} \sum_{jk} n_{ijk}^2 - N^{-1} \sum_{ijk} n_{ijk}^2 \right) / a - 1.$$

$E(MS_B/II)$

$$= E \left\{ \frac{\sum_{ij} n_{ij} (\bar{y}_{ij..} - \bar{y}_{i...})^2}{f_B} \right\}$$

We obtained the equivalents for $\bar{y}_{i...}$ and $\bar{y}_{ij..}$ using (1) and (4) as:

$$\bar{y}_{i...} = \mu + \alpha_i + \bar{\beta}_{(i)} + \bar{\lambda}_{(i)} + \bar{e}_{i...}$$

And:

$$\bar{y}_{ij..} = \mu + \alpha_i + \beta_{j(i)} + \bar{\lambda}_{.(ij)} + \bar{e}_{ij..}$$

Where:

$$\bar{\lambda}_{.(ij)} = \sum_k n_{ijk} \lambda_{k(ij)} / n_{ij}$$

Therefore:

$$\begin{aligned} E(\text{MS}_B/\text{II}) &= E \left\{ \frac{\sum_{ij} n_{ij} [(\beta_{(i)} - \bar{\beta}_{(i)}) + (\bar{\lambda}_{.(ij)} - \bar{\lambda}_{.(i)}) + (\bar{e}_{ij..} - \bar{e}_{i...})]^2}{f_B} \right\} \\ &= E \left\{ \frac{\sum_{ij} n_{ij} \beta_{j(i)}^2 - \sum_i N_i^{-1} \sum_j n_{ij}^2 \beta_{j(i)}^2}{f_B} \right\} \\ &+ E \left\{ \frac{\sum_{ij} n_{ij}^{-1} \sum_k n_{ijk}^2 \lambda_{k(ij)}^2 - \sum_i N_i^{-1} \sum_{jk} n_{ijk}^2 \lambda_{k(ij)}^2}{f_B} \right\} \\ &+ E \left\{ \frac{\sum_{ij} n_{ij} (\bar{e}_{ij..} - \bar{e}_{i...})^2}{f_B} \right\} \\ &= k_1 \sigma_\beta^2 + k_2 \sigma_\lambda^2 + \sigma_e^2 \end{aligned}$$

Where:

$$k_1 = \left(N - \sum_i N_i^{-1} \sum_j n_{ij}^2 \right) / f_B$$

And:

$$k_2 = \left(\sum_{ij} n_{ij}^{-1} \sum_k n_{ijk}^2 - \sum_i N_i^{-1} \sum_{jk} n_{ijk}^2 \right) / f_B$$

$$E(\text{MS}_C/\text{II}) = E \left\{ \frac{\sum_{ijk} n_{ijk} (\bar{y}_{ijk.} - \bar{y}_{ij..})^2}{f_C} \right\}$$

We obtained the equivalents of $\bar{y}_{ijk.}$ and $\bar{y}_{ij..}$ using (1) and (4) as:

$$\bar{y}_{ijk.} = \mu + \alpha_i + \beta_{j(i)} + \lambda_{k(ij)} + \bar{e}_{ijk.}$$

And:

$$\bar{y}_{ij..} = \mu + \alpha_i + \beta_{j(i)} + \bar{\lambda}_{.(ij)} + \bar{e}_{ij..}$$

Therefore:

$$\begin{aligned} E(\text{MS}_C/\text{II}) &= E \left\{ \frac{\sum_{ijk} n_{ijk} [(\lambda_{k(ij)} - \bar{\lambda}_{.(ij)}) + (\bar{e}_{ijk.} - \bar{e}_{ij..})]^2}{f_C} \right\} \\ &= E \left\{ \frac{\sum_{ijk} n_{ijk} \lambda_{k(ij)}^2 - \sum_{ij} n_{ij}^{-1} \sum_k n_{ijk}^2 \lambda_{k(ij)}^2}{f_C} \right\} \\ &+ E \left\{ \frac{\sum_{ijk} n_{ijk} (\bar{e}_{ijk.} - \bar{e}_{ij..})^2}{f_C} \right\} = k_3 \sigma_\lambda^2 + \sigma_e^2 \end{aligned}$$

Where:

$$k_3 = \left(N - \sum_{ij} n_{ij}^{-1} \sum_k n_{ijk}^2 \right) / f_C$$

Mixed effect model (Model III_A):

$$E(\text{MS}_A/\text{III}_A) = E \left\{ \frac{\sum_i N_i (\bar{y}_{i...} - \bar{y}_{....})^2}{a-1} \right\}$$

We obtained the equivalents of $\bar{y}_{i...}$ and $\bar{y}_{....}$ using (1) and (5) as:

$$\bar{y}_{i...} = \mu + \alpha_i + \bar{\beta}_{.(i)} + \bar{\lambda}_{.(i)} + \bar{e}_{i...}$$

And:

$$\bar{y}_{....} = \mu + \bar{\beta}_{.(.)} + \bar{\lambda}_{.(.)} + \bar{e}_{....}$$

Therefore:

$$E(\text{MS}_A/\text{III}_A) = E \left\{ \frac{\sum_i N_i [\alpha_i + (\bar{\beta}_{.(i)} - \bar{\beta}_{.(.)}) + (\bar{\lambda}_{.(i)} - \bar{\lambda}_{.(.)}) + (\bar{e}_{i...} - \bar{e}_{....})]^2}{a-1} \right\}$$

$$f_\theta = 1.6506917, f_\gamma = 2.895969529.$$

$$= \frac{\sum_i N_i \alpha_i^2}{a-1} + k_\beta \sigma_\beta^2 + k_\lambda \sigma_\lambda^2 + \sigma_e^2$$

$$E(MS_B/III_A) = E \left\{ \frac{\sum_{ij} n_{ij} (\bar{y}_{ij.} - \bar{y}_{i...})^2}{f_B} \right\}$$

We obtained the equivalents of $\bar{y}_{i...}$ and $\bar{y}_{ij.}$ using (1) and (5) as:

$$\bar{y}_{i...} = \mu + \alpha_i + \bar{\beta}_{(i)} + \bar{\lambda}_{(i)} + \bar{e}_{i...}$$

And:

$$\bar{y}_{ij.} = \mu + \alpha_i + \beta_{j(i)} + \bar{\lambda}_{(ij)} + \bar{e}_{ij.}$$

Therefore:

$$\begin{aligned} E(MS_B/III_A) &= E \left\{ \frac{\sum_{ij} n_{ij} [(\beta_{j(i)} - \bar{\beta}_{(i)}) + (\bar{\lambda}_{(ij)} - \bar{\lambda}_{(i)}) + (\bar{e}_{ij.} - \bar{e}_{i...})]^2}{f_B} \right\} \\ &= k_1 \sigma_\beta^2 + k_2 \sigma_\lambda^2 + \sigma_e^2 \end{aligned}$$

$$E(MS_C/III_A) = E \left\{ \frac{\sum_{ijk} n_{ijk} (\bar{y}_{ijk.} - \bar{y}_{ij.})^2}{f_C} \right\}$$

We obtained the equivalents of $\bar{y}_{ijk.}$ and $\bar{y}_{ij.}$ using (1) and (5) as:

$$\bar{y}_{ijk.} = \mu + \alpha_i + \beta_{j(i)} + \lambda_{k(ij)} + \bar{e}_{ijk.}$$

And:

$$\bar{y}_{ij.} = \mu + \alpha_i + \beta_{j(i)} + \bar{\lambda}_{(ij)} + \bar{e}_{ij.}$$

$$\begin{aligned} E(MS_C/III_A) &= E \left\{ \frac{\sum_{ijk} n_{ijk} [(\lambda_{k(ij)} - \bar{\lambda}_{(ij)}) + (\bar{e}_{ijk.} - \bar{e}_{ij.})]^2}{f_C} \right\} = k_3 \sigma_\lambda^2 + \sigma_e^2 \end{aligned}$$

$$E(MS_E/III_A) = E \left\{ \frac{\sum_{ijkl} (y_{ijkl} - \bar{y}_{ijk.})^2}{f_E} \right\} = E \left\{ \frac{\sum_{ijkl} (e_{ijkl} - \bar{e}_{ijk.})^2}{f_E} \right\} = \sigma_e^2$$

Using the same procedure as described above, the expected mean squares for other models can be easily derived.

CONCLUSION

We have attempted to provide solutions to problems arising from three-way unbalanced nested ANOVA resulting from a combination of missing data and natural restriction on the design. A primary goal was to construct pseudo-F-test where exact F-test does not exact. The approximate degree of freedom derived in this study often result in non-integer value which in turn leads to interpolation in the table of Percentage points of the F-distribution. The layout for the illustrative example has large degrees of freedom for the estimation of the error component which compensates for the high sum of squares due to error (that is experimental error).

The procedure provided here can be easily implemented and extended to m-way unbalanced nested analysis of variance.

REFERENCES

- Bolton, S. and C. Bon, 2009. *Pharmaceutical Statistics: Practical and Clinical Applications*. 5th Edn., Informa Healthcare USA., New York, ISBN: 1420074229, pp: 656.
- Bush, N. and R.L. Anderson, 1963. A comparison of three different procedures for estimating variance components. *Technometrics*, 5: 421-440.
- Dowdy, S.M. and D. Chilko, 2004. *Statistics for Research*. 3rd Edn., Wiley-Interscience, Chichester, New Jersey, ISBN: 047126735X, pp: 627.
- Kashiani, P. and G. Saleh, 2010. Estimation of genetic correlations on sweet corn inbred lines using SAS mixed model. *Am. J. Agric. Biol. Sci.*, 5: 309-314. DOI: 10.3844/ajabssp.2010.309.314
- Kavitha, S. and K. Duraiswamy, 2011. Adaptive neuro-fuzzy inference system approach for the automatic screening of diabetic retinopathy in fundus images. *J. Comput. Sci.*, 7: 1020-1026. DOI: 10.3844/jcsp.2011.1020.1026
- Mason, R.L., R.F. Gunst and J.L. Hess, 2003. *Statistical Design and Analysis of Experiments: With Applications to Engineering and Science*. 2nd Edn., John Wiley and Sons, New Jersey, ISBN: 0471372161, pp: 728.
- Montgomery, D.C., 2008. *Design and Analysis of Experiments*. 7th Edn. John Wiley and Sons, USA., ISBN: 0470128666, pp: 656.
- Naisipour, M., M.H. Afshar, B. Hassani and A.R. Firoozjaee, 2008. Collocation Discrete Least Square (CDLS) method for elasticity problems and grid irregularity effect assessment. *Am. J. Applied Sci.*, 5: 1595-1601. DOI: 10.3844/ajassp.2008.1595.1601

- Rencher, A.C. and G.B. Bruce, 2008. *Linear Models in Statistics*. 2nd Edn. John Wiley and Sons, New Jersey, ISBN: 0471754986, pp: 672.
- Sahai, H. and M.M. Ojeda, 2004. *Analysis of Variance for Random Models: Unbalanced Data*. 1st Edn., Birkhauser, Boston, ISBN: 0817632298, pp: 480.
- Satterthwaite, F.E., 1946. An approximate distribution of estimates of variance components. *Biometrics*, 2: 110-114. PMID: 20287815
- Seeger, P., 1970. A method of estimating variance components in unbalanced designs. *Technometrics*, 12: 207-218.
- Tietjen, G.L. and R.H. Moore, 1968. On testing significance of components of variance in the unbalanced nested analysis of variance. *Biometrics*, 24: 423-429. PMID: 5683880
- Valliant, R. and K.F. Rust, 2010. Degrees of freedom approximations and rules-of-thumb. *J. Official Stat.*, 26: 585-602.