One-sided tolerance interval in a two-way balanced nested model with mixed effects

A. Silva*,†, F. Carvalho**,†, J. Mexia† and M. Fonseca†

*Departamento de Ciência e Tecnologia, Universidade de Cabo Verde
†Centro de Matemática e Aplicações, Faculdade de Ciência e Tecnologia, Universidade nova de Lisboa
**Unidade Departamental de Matemática e Física, Instituto Politécnico de Tomar

Abstract. In this paper we approach the construction of the both upper and lower tolerance limit in a two-way nested model with mixed effects in balanced data. In order to do so we proceed as Fonseca et al [1] did in order to derive the upper tolerance limit in a two-way nested model with mixed effects in unbalanced data, by using the generalized confidence interval idea earlier used by Krishnamoorthy and Mathew [2] to perform the construction of the upper tolerance limit in a one-way nested model with mixed or random effects model in balanced and unbalanced data. The underlying idea goes through the construction of an approximation for the quantile of the general pivotal quantity for a convinient parametric function.

Keywords: Mixed model; Balanced data; Upper tolerance limit; Lower tolerance limit; Confidence limit.

PACS: 02.50.Tt

1. INTRODUCTION

In many research areas (such as public health, environmental contamination, and others) one deals with the necessity of using data to infer whether some proportion (%) of a population of interest is (or one wants it to be) below and/or over some threshold, through the computation of tolerance interval. The idea is, once a threshold is given, one computes the tolerance interval or limit (which might be one or two - sided bounded) and then to check if it satisfies the given threshold.

Krishnamoorthy and Mathew [2] performed the computation of upper tolerance limit in balanced and unbalanced one-way random effects models, whereas Fonseca et al [1] performed it based in a similar ideas but in a two-way nested mixed or random effects model. In case of random effects model, Fonseca et al [1] performed the computation of such interval only for the balanced data, whereas in the mixed effects case they dit it only for the unbalanced data. For the computation of two-sided tolerance interval in models with mixed and/or random effects we recomend, for instance, Sharma and Mathew [3].

The purpose of this paper is the computation of upper and lower tolerance interval in a two-way nested mixed effects models in balanced data. For the case of unbalanced data, as mentioned above, Fonseca et al [1] have already computed upper tolerance interval. Hence, using the notions persented in Fonseca et al [1] and Krishnamoorthy and Mathew [2], we present some results on the construction of one-sided tolerance interval for the balanced case. Thus, in order to do so at first instance we perform the construction for the upper case, and then the construction for the lower case.

2. NESTED MODEL DESIGN - BASICS NOTIONS

A statistical model is said to have a mixed effects if it consists of a mixture of fixed and random effects factors. Such a model is said to be a two-way nested one, if it consists of two factors, say A and B, where levels of factor B are nested within levels of the factor A. There is many publications concerning nested models. For complete notion concerning such models we suggest Montegomery [4] and/or Dowdy and Chilko [5].

The effects associated with any factor in a nested model are the effects which the levels have on the interest response variable.
Let suppose that the factor \( B \) has \( b_i \), \( i = 1, \ldots, a \), levels nested within the \( i \)th level of the factor \( A \). Thus, the two-way nested mixed effects model is given by

\[
Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{(ij)}, \quad k = 1, \ldots, n_{ij}, \quad j = 1, \ldots, b_i, \tag{1}
\]

where \( \mu \) is the general mean, \( \tau_i \) is considered to be the fixed effect term due to the \( i \)th, \( i = 1, \ldots, a \), level of the factor \( A \), \( \beta_{j(i)} \), \( j = 1, \ldots, b_i \), the random effect term due to the \( j \)th level of the factor \( B \) nested within the \( i \)th level of the factor \( A \), and \( \epsilon_{(ij)} \) the error term associated to the observed value \( Y_{ijk} \). It is assumed that \( \beta_{j(i)} \sim N(0, \sigma^2_B) \), and \( \epsilon_{(ij)} \sim N(0, \sigma^2) \) are independent from each other.

Recalling that our model considers the data to be balanced, that is, \( b_i = b, \ i = 1, \ldots, a \), and \( n_{ij} = n, \ i = 1, \ldots, a, \ j = 1, \ldots, b \), and letting \( \bar{Y}_{i\cdot} = \frac{1}{b} \sum_{j=1}^{b} Y_{i\cdot j} \), the sums of squares are given by

\[
SS_{\beta} = \sum_{i=1}^{a} \sum_{j=1}^{b} \left( \bar{Y}_{i\cdot} - \frac{1}{b} \sum_{j=1}^{b} \bar{Y}_{i\cdot j} \right)^2 \quad \text{and} \quad SS_{\epsilon} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{i\cdot j})^2.
\]

Now \( \bar{Y}_{i\cdot}, SS_{\beta} \) and \( SS_{\epsilon} \) are independent distributed variables.) we define the independent variables

\[
U_{e} = \frac{SS_{\epsilon}}{\sigma^2_{\epsilon}} \quad \text{and} \quad U_{\beta} = \frac{SS_{\beta}}{\sigma^2_{\beta} + \frac{1}{n} \sigma^2_{\epsilon}}.
\]

Thus (a generalization of the Theorem 5.3.1 of Casella and Berger [6] together with Fonseca et al [1] may be very useful here!)

\[
U_{\beta} \sim \chi^2_{a(b-1)} \quad \text{and} \quad U_{e} \sim \chi^2_{a(b-1)},
\]

with \( \chi^2 \) the chi-square distribution with \( r \) degrees of freedom.

**In this paper we approach the two following problems:**

1. The construction of both **upper and lower tolerance limit** for the observable random variable \( Y \), where \( Y \sim N(\mu, \sigma^2_{\beta} + \sigma^2_{\epsilon}) \), with \( \mu_i = \mu + \tau_i \);
2. The construction of both **upper and lower tolerance limit** for the unobserved “true effect” \( Y^* = \mu_i + \tau_i + \beta_{j(i)} \sim N(\mu, \sigma^2_{\beta}) \).

### 3. UPPER TOLERANCE LIMIT

Let \( Y = \{Y_{ij}, \ldots, Y_{ijk}\} \) be a sample of the random variable \( Y \sim N(\mu, \sigma^2_{\beta} + \sigma^2_{\epsilon}) \). A statistic \( C \) is a \((p, \gamma)\)-upper tolerance limit for \( Y \), if the equation

\[
P_Y[P_Y(Y \leq C|Y) \geq p] = \gamma \iff P_Y[q_p \leq C] = \gamma, \tag{2}
\]

holds for \( 0 < \gamma < 1 \), and \( 0 < p < 1 \), where \( q_p \) is the \( p \)th quantile of \( N(\mu, \sigma^2_{\beta} + \sigma^2_{\epsilon}) \). Thus, clearly, \( C \) is also a \( \gamma \)-upper confidence limit for the quantile \( q_p \). More over, \( C \) is a \( \gamma \)-upper confidence limit for the parametric function

\[
\mu_i + z_p \left( \sqrt{\sigma^2_{\beta} + \sigma^2_{\epsilon}} \right),
\]

where \( z_p \) denotes the \( p \)th quantile of the \( N(0, 1) \) (the standard normal distribution), as we are about to show:

\[
P_Y[P_Y(Y \leq C|Y) \geq p] = \gamma \iff P_Y \left[ \mu_i + z_p \left( \sqrt{\sigma^2_{\beta} + \sigma^2_{\epsilon}} \right) \leq C \right] = \gamma,
\]

where \( z_p \) is the standard normal distribution \( i \)th quantile, since \( \frac{\gamma - \mu_i}{\sqrt{\sigma^2_{\beta} + \sigma^2_{\epsilon}}} \sim N(0, 1) \).

Now, with out lost of generality, let \( Y = \{Y_{ij}, \ldots, Y_{ijk}\} \) be a sample of the random variable \( Y^* \sim N(\mu, \sigma^2_{\beta}) \). Then, proceeding identically to the case of \( Y \sim N(\mu, \sigma^2_{\beta} + \sigma^2_{\epsilon}) \), since the unobserved “true effect” variable \( Y^* \sim N(\mu, \sigma^2_{\beta}) \), the

\[
(p, \gamma)\)-upper tolerance limit for \( Y^* \)
is simply the 
\(\gamma\)-upper confidence limit for the parametric function \(\mu_i + z_{\alpha} \sigma_\beta\).

Here, for the observable random variable \(Y \sim N(\mu_i, \sigma_\beta^2 + \sigma^2_e)\) the upper tolerance limit is performed and then the one for the unobserved “true effects” variable \(Y^* \sim N(\mu_i, \sigma_\beta^2)\) follows.

4. LOWER TOLERANCE LIMIT

Let \(Y = \{Y_{ij1}, \ldots, Y_{ijk}\}\) be a sample of the random variable \(Y \sim N(\mu_i, \sigma^2_\beta + \sigma^2_e)\). A statistic \(D\) is a \((p, \gamma)\)-lower tolerance limit for \(Y\), if

\[
P_Y[Y \geq D \mid Y] \geq p \iff P_Y[D \leq q_{1-p}] = \gamma, \quad (3)
\]

holds for \(0 < \gamma < 1\), and \(0 < p < 1\), where \(q_{1-p}\) is the \((1-p)\)th quantile of \(N(\mu_i, \sigma^2_\beta + \sigma^2_e)\). Like at the section 3, it is readily shown that \(D\) is a \(\gamma\)-lower confidence limit for the parametric function \(\mu_i + z_{1-p} \sqrt{\sigma^2_\beta + \sigma^2_e}\), where \(z_{1-p}\) denotes the \((1-p)\)th quantile of the \(N(0,1)\), as we are about to show:

\[
P_Y[Y \geq D \mid Y] \geq p \iff P_Y[D \leq \mu_i + z_{1-p} \sqrt{\sigma^2_\beta + \sigma^2_e}] = \gamma,
\]

where \(z_{1-p}\) is the standard normal distribution, since \(\frac{Y - \mu_i}{\sqrt{\sigma^2_\beta + \sigma^2_e}} \sim N(0,1)\).

Now, again with out lost of generality, let \(Y = \{Y_{ij1}, \ldots, Y_{ijk}\}\) be a sample of the random variable \(Y^* \sim N(\mu_i, \sigma^2_\beta)\). Then, proceeding in a similar way as at section 3, one concludes that the \((p, \gamma)\)-lower tolerance limit for the “true effects” variable \(Y^* = \mu_i + \tau + \beta_{j(i)} \sim N(\mu_i, \sigma^2_\beta)\) is simply the \(\gamma\)-lower confidence limit for the parametric function \(\mu_i + z_{1-p} \sigma_\beta\).

Here, for the observable random variable \(Y \sim N(\mu_i, \sigma^2_\beta + \sigma^2_e)\) the lower tolerance limit is performed and then the one for the unobserved “true effects” variable \(Y^* \sim N(\mu_i, \sigma^2_\beta)\) follows.

ACKNOWLEDGMENTS

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through PEst-OE/MAT/UI0297/2011 (CMA). It was also partially supported by the Núcleo de Investigação em Matemática e Aplicações of Universidade de Cabo Verde.

REFERENCES