A Parametric Bootstrap Approach for Two-way ANOVA under Unequal Variances with Unbalanced Data

Guoyi Zhang*

Abstract
This research is to provide a solution of two-way ANOVA without using transformation when variances are heteroscedastic and group sizes are unequal. Parametric bootstrap tests for main effects and interaction terms have been shown to be competitive with some other methods. We extend the parametric bootstrap algorithm to multiple comparison procedures. Simulation results show that the parametric bootstrap approach works well for two-way ANOVA.

Key Words: ANOVA, Parametric bootstrap, Multiple comparison, Simulations, Unequal variance

1. Introduction

Consider the ANOVA problem of ab normal populations with unequal population variances σ_{ij}^2, $i = 1, 2, \cdots, a$ and $j = 1, 2, \cdots, b$ and let $Y_{ij1}, Y_{ij2}, \cdots, Y_{ij,n_{ij}}$ be a random sample from $N(u_{ij}, \sigma_{ij}^2)$. The two-way ANOVA full model is as follows,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \tag{1}$$

and additive model is as follows

$$Y_{ijk} = u + \alpha_i + \beta_j + \epsilon_{ijk}, \tag{2}$$

where $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_{ij}^2), i = 1, 2, \cdots, a, j = 1, 2, \cdots, b$, $k = 1, 2, \cdots, n_{ij}$, α_i's, β_j's and γ_{ij}'s are subject to the constraints (first introduced by Scheffé (1959) to solve the model identifiability problem)

$$\sum_{i=1}^{a} w_i\alpha_i = 0, \sum_{j=1}^{b} v_j\beta_j = 0, \sum_{i=1}^{a} w_i\gamma_{ij} = 0, \sum_{j=1}^{b} v_j\gamma_{ij} = 0, \tag{3}$$

$w_1, \cdots, w_a$ and $v_1, \cdots, v_b$ are nonnegative weights. In full model (1), we are interested in testing the interaction effect, i.e.

$$H_{0AB} : \gamma_{ij} = 0 \text{ for } i = 1, \cdots, a, j = 1, \cdots, b \text{ vs } H_\alpha : H_0 \text{ not true }. \tag{4}$$

In additive model (2), we are interested in testing the main effects A and B,

$$H_{0A} : \alpha_i = 0, i = 1, \cdots, a, \text{ vs } H_\alpha : H_0 \text{ not true} \tag{5}$$

$$H_{0B} : \beta_j = 0, j = 1, \cdots, b, \text{ vs } H_\alpha : H_0 \text{ not true.} \tag{6}$$

This research intends to provide a solution of two-way ANOVA: testing main effects, interaction effects Xu et al. (2013) and all pairwise comparisons (research in this article) under the assumption of heteroscedastic variances and unequal sizes.

*Guoyi Zhang, Assistant Professor, Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, 87131-0001

The application of the classical F-test for testing main effects and interaction effects may lead to misleading conclusions Wilcox (1989); Krutchkoff (1989); Ananda and Weerahandi (1997) under heteroscedastic variances. The size of the F-test can be much larger or much smaller than the nominal size when the cell variance homogeneity assumption is violated Zhang (2012). Furthermore, F-tests suffer from serious lack of power even under moderate heteroscedasticity. Some alternative tests have been developed due to these issues. James (1954) proposed a first order approximation solution. Johansen (1980) discussed approximate distribution of the residual sum of squares in a linear model and proposed an approximate degrees of freedom (ADF) test. Wilcox (1989) proposed a modified Wilcox test and a James-type test. Krutchkoff (1989) investigated a simulation-based approximate test. Ananda and Weerahandi (1997) suggested a generalized F-test based on the concept of generalized p-value Wang and Chow (1994). Zhang (2012) proposed a simple and accurate ADF test for heteroscedastic two-way ANOVA, and showed that ADF test is generally more powerful than the classical F-test. Xu et al. (2013) extended the work of Krishnamoorthy et al. (2007) and proposed a Parametric bootstrap (PB) test for two-way ANOVA. Xu et al. (2013) compared the PB test with the generalized F (GF) test and showed that the PB test performs better than the GF test and also performs very satisfactorily even for small samples.

Another problem in ANOVA is multiple comparisons (all pairwise simultaneous comparisons). Scheff$\acute{e}$'s method, the Bonferroni inequality-based method, and Tukey-Kramer method are widely used for pairwise comparisons among the group means when variances of sample means are equal. Research of multiple comparisons under the assumption of heteroscedasticity is limited. Hochberg (1976) generalized the Spjϕtvoll and Stoline's procedure 1973 to heterogeneous variance cases. Games and Howell (1976) presented a method for constructing simultaneous confidence intervals based on the Behrens-Fisher statistic with Welch's 1948 approximate t solution for degrees of freedom. Kaiser and Bowden (1983) discussed simultaneous confidence intervals for all linear contrasts in a one-way ANOVA with unequal variances. The above multiple comparison prcedures (MCPs) either involve Studentized range statistics Einot and Gabriel (1975) or alternatively as Student's t statistics Games (1971). Factors such as the degree of variance and sample size heterogeneity, the shape of the population etc,. can affect the rates of Type I error and power characteristics. Therefore most of the MCP tests are relatively data pertinent. Zhang (2013) proposed PB approach multiple comparisons for one-way ANOVA under unequal variances with unbalanced data.

In this research, we extend Zhang (2013)'s PB test for multiple comparisons in one-way ANOVA to two-way ANOVA. This paper is organized as follows. In Section 2, we review PB test for main effects and interaction effects from Xu et al. (2013). In Section 3, we propose PB algorithm of multiple comparisons for two-way ANOVA. In Section 4, we present simulation studies. Section 5 gives conclusions.

2. The Parametric Bootstrap Test for Main Effects and Interaction Effects in Two-way ANOVA

In this section, we review the PB tests of two-way ANOVA Xu et al. (2013) for testing main effects and interaction effects.

2.1 Tests for the Interaction Effects

Assume σ_{ij}^2's are known, a natural test statistic for (4) is the standardized interaction sum of squares $\tilde{S}_I(\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab}; \sigma_{11}^2, \cdots, \sigma_{ab}^2) = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}(\bar{Y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2/\sigma_{ij}^2$.

It can be shown that

$$\tilde{S}_I(\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab}; \sigma^2_{11}, \cdots, \sigma^2_{ab}) = \bar{\mathbf{Y}}'\mathbf{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{\Sigma}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{\Sigma}^{-1/2})\mathbf{\Sigma}^{-1/2}\bar{\mathbf{Y}}, \tag{7}$$

where $\bar{Y}_{ij} = \sum_{k=1}^{n_{ij}} Y_{ijk}/n_{ij}$, $\bar{\mathbf{Y}} = (\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab})'$, $\mathbf{\Sigma} = \text{diag}(\sigma^2_{11}/n_{11}, \sigma^2_{12}/n_{12}, \cdots, \sigma^2_{ab}/n_{ab})$, $\mathbf{1}_k$ denotes the $k \times 1$ vector of ones, $\mathbf{I}_k$ denotes the identity matrix with rank k, and $\mathbf{X} = (\mathbf{1}_{ab}, \mathbf{I}_a \otimes \mathbf{1}_b, \mathbf{1}_a \otimes \mathbf{I}_b)$ with $\otimes$ representing Kronecker product. Let $\mathbf{y} = (y_{11}, y_{12}, \cdots, y_{ab})$ be the observed value of $\mathbf{Y}$. We reject H_{0AB} in (4) when $\tilde{S}_I(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab}; \sigma^2_{11}, \cdots, \sigma^2_{ab})$ $> \chi^2_{(a-1)(b-1),\alpha}$, where $\chi^2_{(a-1)(b-1),\alpha}$ is the upper αth quantile of a chi-square distribution with $(a-1)(b-1)$ degrees of freedom.

In general, σ^2_{ij}'s are unknown. In this case, test statistic in (7) can be obtained by replacing σ^2_{ij} by S^2_{ij}, $i = 1, 2, \cdots, a$, $j = 1, 2, \cdots, b$ and is given by

$$\tilde{S}_I(\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab}; S^2_{11}, \cdots, S^2_{ab}) = \bar{\mathbf{Y}}'\mathbf{S}^{-1/2}(\mathbf{I}-\mathbf{S}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{S}^{-1/2})\mathbf{S}^{-1/2}\bar{\mathbf{Y}}, \tag{8}$$

where $S^2_{ij} = \sum_{k=1}^{n_{ij}}(Y_{ijk}-\bar{Y}_{ij})^2/(n_{ij}-1)$ and $\mathbf{S} = \text{diag}(S^2_{11}/n_{11}, S^2_{12}/n_{12}, \cdots, S^2_{ab}/n_{ab})$,

The parametric bootstrap involves sampling from the estimated models. Under H_{0AB}, the vector $\mathbf{Y}$ has the mean $\mathbf{X}\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\mu, \alpha_1, \cdots, \alpha_a, \beta_1, \cdots, \beta_b)$. The test statistic in (7) is location invariant. Without loss of generality, we take $\mathbf{X}\boldsymbol{\theta} = \mathbf{0}$. Let $s^2_{ij} = \sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij})^2/(n_{ij}-1)$ be the observed S^2_{ij}, the parametric bootstrap pivot variable can be developed as follows. For a given $(y_{11}, y_{12}, \cdots, y_{ab}; s^2_{11}, s^2_{12}, \cdots, s^2_{ab})$, $\bar{Y}_{Bij} \sim N(0, s^2_{ij}/n_{ij})$, $S^2_{Bij} \sim s^2_{ij}\chi^2_{n_{ij}-1}/(n_{ij}-1), i = 1, \cdots, a, j = 1, \cdots, b$. The PB pivot variable based on test statistic (7) is given by

$$\tilde{S}_{IB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S^2_{B11}, \cdots, S^2_{Bab}) = \bar{\mathbf{Y}}'_B\mathbf{S}_B^{-1/2}(\mathbf{I}-\mathbf{S}_B^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{S}_B^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{S}_B^{-1/2})\mathbf{S}_B^{-1/2}\bar{\mathbf{Y}}_B \tag{9}$$

where $\bar{\mathbf{Y}}_B = (\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab})'$, $\mathbf{S}_B = \text{diag}(S^2_{B11}/n_{11}, S^2_{B12}/n_{12}, \cdots, S^2_{Bab}/n_{ab})$. For a given level α, the PB test rejects H_{0AB} in (4) when

$$P\left(\tilde{S}_{IB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S^2_{B11}, S^2_{B12}, \cdots, S^2_{Bab}) > \tilde{s}_I\right) < \alpha, \tag{10}$$

where $\tilde{s}_I = \tilde{S}_I(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab}; s^2_{11}, s^2_{12}, \cdots, s^2_{ab})$ is an observed value of $\tilde{S}_I(\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab}; S^2_{11}, S^2_{12}, \cdots, S^2_{ab})$ in (8). The probability in (10) can be estimated by Algorithm 1.

Algorithm 1: For a given $(n_{11}, n_{12}, \cdots, n_{ab})$, $(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab})$, and $(s^2_{11}, s^2_{12}, \cdots, s^2_{ab})$:
compute $\tilde{S}_I(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab}; s^2_{11}, s^2_{12}, \cdots, s^2_{ab})$ in (8) and call it $\tilde{s}_I$
For $k = 1, \cdots, m$
generate $\bar{Y}_{Bij} \sim N(0, s^2_{ij}/n_{ij})$ and $S^2_{Bij} \sim s^2_{ij}\chi^2_{n_{ij}-1}/(n_{ij}-1)$, $i = 1, \cdots, a, j = 1, \cdots, b$
compute $S_{IB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S^2_{B11}, S^2_{B12}, \cdots, S^2_{Bab})$ using (9)
if $S_{IB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S^2_{B11}, S^2_{B12}, \cdots, S^2_{Bab}) > \tilde{s}_I$, set $Q_k = 1$
(end loop)
$(1/m)\sum_{k=1}^{m} Q_k$ is a Monte Carlo estimate of the p-value in (10).

2.2 Tests for the Main Effects

For the additive model (2), the hypotheses of interests are the main effect A as described in (5) and main effect B as described in (6). A natural test statistic for (5) is the standardized sum of squares due to factor A. It can be shown that

$$\begin{aligned} \tilde{S}_A(\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab}; \sigma_{11}^2, \cdots, \sigma_{ab}^2) &= \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{n_{ij}}{\sigma_{ij}^2}(\bar{Y}_{ij} - \hat{\mu} - \hat{\beta}_j)^2 \qquad (11) \\ &= \bar{\mathbf{Y}}'\mathbf{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{\Sigma}^{-1/2}\mathbf{X}_A(\mathbf{X}_A'\mathbf{\Sigma}^{-1}\mathbf{X}_A)^{-}\mathbf{X}_A'\mathbf{\Sigma}^{-1/2})\mathbf{\Sigma}^{-1/2}\bar{\mathbf{Y}}, \end{aligned}$$

where $\mathbf{X}_A = (\mathbf{1}_{ab}, \mathbf{1}_a \otimes \mathbf{I}_b)$. We reject H_{0A} in (5) when $\tilde{S}_A(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab}; \sigma_{11}^2, \cdots, \sigma_{ab}^2)$ $> \chi^2_{(a-1)b,\alpha}$. The PB pivot variable can be derived as

$$\tilde{S}_{AB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S_{B11}^2, \cdots, S_{Bab}^2) = \bar{\mathbf{Y}}_B'\mathbf{S}_B^{-1/2}(\mathbf{I} - \mathbf{S}_B^{-1/2}\mathbf{X}_A(\mathbf{X}_A'\mathbf{S}_B^{-1}\mathbf{X}_A))^{-}\mathbf{X}_A'\mathbf{S}_B^{-1/2})\mathbf{S}_B^{-1/2}\bar{\mathbf{Y}}_B \qquad (12)$$

For a given level α, the PB test rejects H_{0A} in (5) when

$$P\left(\tilde{S}_{AB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S_{B11}^2, S_{B12}^2, \cdots, S_{Bab}^2) > \tilde{s}_A\right) < \alpha, \qquad (13)$$

where $\tilde{s}_A = \tilde{S}_A(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \cdots, s_{ab}^2)$ is an observed value of $\tilde{S}_A(\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab}; S_{11}^2, S_{12}^2, \cdots, S_{ab}^2)$ as in (11). The probability in (13) can be estimated by Algorithm 2.

Algorithm 2: For a given $(n_{11}, n_{12}, \cdots, n_{ab})$, $(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab})$, and $(s_{11}^2, s_{12}^2, \cdots, s_{ab}^2)$:
compute $\tilde{S}_A(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab}; s_{11}^2, s_{12}^2, \cdots, s_{ab}^2)$ in (11) and call it $\tilde{s}_A$
For $k = 1, \cdots, m$
generate $\bar{Y}_{Bij} \sim N(0, s_{ij}^2/n_{ij})$ and $S_{Bij}^2 \sim s_{ij}^2\chi^2_{n_{ij}-1}/(n_{ij}-1)$, $i = 1, \cdots, a, j = 1, \cdots, b$
compute $S_{AB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S_{B11}^2, S_{B12}^2, \cdots, S_{Bab}^2)$ using (12)
if $S_{AB}(\bar{Y}_{B11}, \bar{Y}_{B12}, \cdots, \bar{Y}_{Bab}; S_{B11}^2, S_{B12}^2, \cdots, S_{Bab}^2) > \tilde{s}_A$, set $Q_k = 1$
(end loop)
$(1/m)\sum_{k=1}^{m} Q_k$ is a Monte Carlo estimate of the p-value in (13).

A natural test statistic for testing main effect B as in (6) is the standardized sum of squares due to factor B. It can be shown that

$$\begin{aligned} \tilde{S}_B(\bar{Y}_{11}, \bar{Y}_{12}, \cdots, \bar{Y}_{ab}; \sigma_{11}^2, \cdots, \sigma_{ab}^2) &= \sum_{j=1}^{b}\sum_{i=1}^{a} \frac{n_{ij}}{\sigma_{ij}^2}(\bar{Y}_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 \qquad (14) \\ &= \bar{\mathbf{Y}}'\mathbf{\Sigma}^{-1/2}(\mathbf{I} - \mathbf{\Sigma}^{-1/2}\mathbf{X}_B(\mathbf{X}_B'\mathbf{\Sigma}^{-1}\mathbf{X}_B)^{-}\mathbf{X}_B'\mathbf{\Sigma}^{-1/2})\mathbf{\Sigma}^{-1/2}\bar{\mathbf{Y}}, \end{aligned}$$

where $\mathbf{X}_B = (\mathbf{1}_{ab}, \mathbf{1}_b \otimes \mathbf{I}_a)$. We reject H_{0B} in (6) when $\tilde{S}_B(\bar{y}_{11}, \bar{y}_{12}, \cdots, \bar{y}_{ab}; \sigma_{11}^2, \cdots, \sigma_{ab}^2)$ $> \chi^2_{(b-1)a,\alpha}$. PB test and PB algorithm can be derived similarly as testing main effect A.

3. The Parametric Bootstrap Method for Multiple Comparisons with Heteroscedastic Variances and Unequal Sizes

The multiple comparison procedure applies when the family of interest is the set of all pairwise comparisons of factor level means. Pairwise comparisons of the factor level means $\mu_{i\cdot}$ can be described as

$$H_0 : \mu_{i\cdot} - \mu_{i'\cdot} = 0 \text{ v.s. } H_\alpha : \mu_{i\cdot} - \mu_{i'\cdot} \neq 0. \qquad (15)$$

Pairwise comparisons of the factor level means $\mu_{\cdot j}$ can be described as

$$H_0 : \mu_{\cdot j} - \mu_{\cdot j'} = 0 \text{ v.s. } H_\alpha : \mu_{\cdot j} - \mu_{\cdot j'} \neq 0. \qquad (16)$$

When all the σ_{ij}^2's are equal and cell sizes are equal, the Tukey's multiple comparison confidence limits for all pairwise comparisons $\mu_{i\cdot} - \mu_{i'\cdot}$ with family confidence coefficient of at least $1-\alpha$ are $\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot} \pm q(\alpha)\sqrt{2MSE/bn}$, where n is the cell size and $q(\alpha)$ is the upper αth quantile of the studentized range distribution. Similarly, the Tukey's multiple comparison confidence limits for all pairwise comparisons $\mu_{\cdot j} - \mu_{\cdot j'}$ with family confidence coefficient of at least $1-\alpha$ are $\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot j'\cdot} \pm q(\alpha)\sqrt{2MSE/an}$. However, Tukey's method fails to work when the population variances are unequal. In this section, we propose PB algorithms of multiple comparison procedures for two-way ANOVA under heteroscedastic variances and unequal sizes.

3.1 Multiple Comparison Procedure for Factor A Level Means

Under heteroscedastic variances and unequal sizes, the estimate of factor level means would be a weighted average of the corresponding cell means. We propose the estimator of factor level means $u_{i\cdot}$ as

$$\begin{aligned}
\bar{Y}_{i\cdot\cdot} &= \frac{\sum_j v_j \bar{Y}_{ij}}{\sum_j v_j} \\
&= \frac{\sum_j v_j(\mu + \alpha_i + \beta_j + \gamma_{ij} + \bar{\epsilon}_{ij})}{\sum_j v_j} \\
&= \mu + \alpha_i + \frac{\sum_j v_j \beta_j}{\sum_j v_j} + \frac{\sum_j v_j \gamma_{ij}}{\sum_j v_j} + \frac{\sum_j v_j \bar{\epsilon}_{ij}}{\sum_j v_j} \\
&= \mu + \alpha_i + \frac{\sum_j v_j \bar{\epsilon}_{ij}}{\sum_j v_j},
\end{aligned}$$

where v_j's are the weights reflected in (3) and $\bar{Y}_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \bar{\epsilon}_{ij}$ are derived from (1) with $\bar{\epsilon}_{ij} \sim N(0, \sigma_{ij}^2/n_{ij})$. Let N be the sample size with $N = \sum_i \sum_j n_{ij}$. We propose the following two sets of weights v_i and use them in simulation studies.

$$v_1 = v_2 = \cdots = v_b = 1/b, \tag{17}$$

$$v_1 = \frac{n_{\cdot 1}}{N}, v_2 = \frac{n_{\cdot 2}}{N}, \cdots, v_b = \frac{n_{\cdot b}}{N}, \tag{18}$$

where $n_{\cdot j} = \sum_i n_{ij}$. Variance of $\bar{Y}_{i\cdot\cdot}$ can be derived as

$$v(\bar{Y}_{i\cdot\cdot}) = \frac{1}{(\sum_j v_j)^2}(\sum_j \frac{\sigma_{ij}^2}{n_{ij}} v_j^2) \text{ and } \hat{v}(\bar{Y}_{i\cdot\cdot}) = \frac{1}{(\sum_j v_j)^2}(\sum_j \frac{s_{ij}^2}{n_{ij}} v_j^2) \tag{19}$$

The PB variables $\bar{Y}_{Bij} \sim N(0, s_{ij}^2/n_{ij})$, $S_{Bij}^2 \sim s_{ij}^2 \chi^2_{(n_{ij}-1)}/(n_{ij}-1), i = 1, \cdots, a, j = 1, \cdots, b$. Hence $\bar{Y}_{Bij} \overset{d}{=} Z_i(s_{ij}/\sqrt{n_{ij}})$ and $S_{B_{ij}}^2 \overset{d}{=} s_{ij}^2 \chi^2_{n_{ij}-1}/(n_{ij}-1)$, where $\overset{d}{=}$ means the same distribution. Let $\bar{Y}_{Bi\cdot\cdot}$ be the PB estimator of factor A level mean $\mu_{i\cdot}$, we have the following variance estimate

$$v(\bar{Y}_{Bi\cdot\cdot}) = \frac{1}{(\sum_j v_j)^2}(\sum_j \frac{s_{ij}^2}{n_{ij}} v_j^2) \tag{20}$$

The test statistic $q_{ii'}^A$ can be written as the following

$$q_{ii'}^A = \frac{|\bar{Y}_{Bi\cdot\cdot} - \bar{Y}_{Bi'\cdot\cdot}|}{\sqrt{v(\bar{Y}_{Bi\cdot\cdot}) + v(\bar{Y}_{Bi'\cdot\cdot})}}, \tag{21}$$

which has the same distribution as

$$q_{ii'}^{A} \stackrel{d}{=} \frac{\left| \frac{\sum_j v_j Z_i(\frac{s_{ij}}{\sqrt{n_{ij}}})}{\sum_j v_j} - \frac{\sum_j v_j Z_{i'}(\frac{s_{i'j}}{\sqrt{n_{i'j}}})}{\sum_j v_j} \right|}{\sqrt{\frac{1}{(\sum_j v_j)^2} \sum_j \frac{s_{ij}^2 \chi^2_{n_{ij}-1}}{n_{ij}(n_{ij}-1)} v_j^2 + \frac{1}{(\sum_j v_j)^2} \sum \frac{s_{i'j}^2 \chi^2_{n_{i'j}-1}}{n_{i'j}(n_{i'j}-1)} v_j^2}}, \tag{22}$$

for $i < i'$, for $i = 1, \cdots, a-1, i' = i+1, \cdots, a$.

For a given $(n_{11}, n_{12} \cdots, n_{ab})$, $(\bar{y}_{11}, \cdots, \bar{y}_{ab})$ and $(s_{11}^2, \cdots, s_{ab}^2)$, let

$$q_{ii'}^{A0} = |\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}| / \sqrt{v(\bar{Y}_{Bi\cdot\cdot}) + v(\bar{Y}_{Bi'\cdot\cdot})} \text{ for } i = 1, \cdots, a-1, i' = i+1, \cdots, a. \tag{23}$$

Given a significance level α, the multiple comparison confidence limits for simultaneous comparisons $\mu_{i\cdot} - \mu_{i'\cdot}$ with family confidence coefficient at least $1-\alpha$ are $\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot} \pm q_\alpha \sqrt{v(\bar{Y}_{Bi\cdot\cdot}) + v(\bar{Y}_{Bi'\cdot\cdot})}$, where q_α can be estimated using parametric bootstrap method given in Algorithm 3.

Algorithm 3.

For a given $(n_{11}, n_{12} \cdots, n_{ab})$, $(\bar{y}_{11}, \cdots, \bar{y}_{ab})$ and $(s_{11}^2, \cdots, s_{ab}^2)$:

For $l = 1, \cdots, L$

Generate $Z_i \sim N(0,1)$ and $\chi^2_{n_{ii'}-1}, i = 1, \cdots, a$

Compute $q_{ii'}^{A}$ using (22) for $i = 1, \cdots, a-1$, $i' = i+1, \cdots, a$

Find $q_l = \max(q_{ii'})$

(end loop)

q_α is the $1-\alpha$ percentile of the simulated distribution of q.

3.2 Multiple Comparison Procedure for Factor B Level Means

PB multiple comparison procedure for factor B level means can be derived similarly. We propose the following estimators for factor level means $u_{\cdot j\cdot}$, $j = 1, 2, \cdots, b$,

$$\bar{Y}_{\cdot j\cdot} = \mu + \beta_j + \frac{\sum_i w_i \bar{\epsilon}_{ij}}{\sum_i w_i}, \tag{24}$$

where w_i's are the weights reflected in (3). We propose two sets of weights w_i and use them in simulation studies,

$$w_1 = w_2 = \cdots = w_a = 1/a, \tag{25}$$

$$w_1 = \frac{n_{1.}}{N}, w_2 = \frac{n_{2.}}{N}, \cdots, w_a = \frac{n_{a.}}{N}, \tag{26}$$

where $n_{i.} = \sum_j n_{ij}$. Let $\bar{Y}_{B\cdot j\cdot}$ be the PB estimator of factor B level mean $\mu_{\cdot j\cdot}$. The variance of $\bar{Y}_{B\cdot j\cdot}$ can be found as follows

$$v(\bar{Y}_{B\cdot j\cdot}) = \frac{1}{(\sum_i w_i)^2} (\sum_i \frac{s_{ij}^2}{n_{ij}} w_i^2). \tag{27}$$

The test statistic $q_{jj'}^{B}$ can be written as the following

$$q_{jj'}^{B} = \frac{|\bar{Y}_{B\cdot j\cdot} - \bar{Y}_{B\cdot j'\cdot}|}{\sqrt{v(\bar{Y}_{B\cdot j\cdot}) + v(\bar{Y}_{B\cdot j'\cdot})}}, \tag{28}$$

which has the same distribution as

$$q_{jj'}^{B} \stackrel{d}{=} \frac{\left| \frac{\sum_i w_i Z_j(\frac{s_{ij}}{\sqrt{n_{ij}}})}{\sum_i w_i} - \frac{\sum_i w_i Z_{j'}(\frac{s_{ij'}}{\sqrt{n_{ij'}}})}{\sum_i w_i} \right|}{\sqrt{\frac{1}{(\sum_i w_i)^2} \sum_i \frac{s_{ij}^2 \chi^2_{n_{ij}-1}}{n_{ij}(n_{ij}-1)} w_i^2 + \frac{1}{(\sum_i w_i)^2} \sum \frac{s_{ij'}^2 \chi^2_{n_{ij'}-1}}{n_{ij'}(n_{ij'}-1)} v_{j'}^2}}, \tag{29}$$

for $j < j', j = 1, \cdots, b-1, j' = j+1, \cdots, b$.

For a given $(n_{11}, n_{12} \cdots, n_{ab})$, $(\bar{y}_{11}, \cdots, \bar{y}_{ab})$ and $(s_{11}^2, \cdots, s_{ab}^2)$, let

$$q_{jj'}^{B0} = |\bar{y}_{\cdot j \cdot} - \bar{y}_{\cdot j' \cdot}| / \sqrt{v(\bar{Y}_{B \cdot j \cdot}) + v(\bar{Y}_{B \cdot j' \cdot})} \text{ for } j = 1, \cdots, b-1, j' = j+1, \cdots, b. \tag{30}$$

Given a significance level α, the multiple comparison confidence limits for simultaneous comparisons $\mu_{i\cdot} - \mu_{i'\cdot}$ with family confidence coefficient at least $1-\alpha$ are $\bar{y}_{\cdot j \cdot} - \bar{y}_{\cdot j' \cdot} \pm q_\alpha \sqrt{v(\bar{Y}_{B \cdot j \cdot}) + v(\bar{Y}_{B \cdot j' \cdot})}$, where q_α can be estimated using parametric bootstrap method given in Algorithm 4.

Algorithm 4.
For a given $(n_{11}, n_{12} \cdots, n_{ab})$, $(\bar{y}_{11}, \cdots, \bar{y}_{ab})$ and $(s_{11}^2, \cdots, s_{ab}^2)$:
For $l = 1, \cdots, L$
Generate $Z_i \sim N(0,1)$ and $\chi^2_{n_{jj'}-1}, j = 1, \cdots, b$
Compute $q_{jj'}^B$ using (29) for $j = 1, \cdots, b-1$, $j' = j+1, \cdots, b$
Find $q_l = \max(q_{jj'})$
(end loop)
q_α is the $1-\alpha$ percentile of the simulated distribution of q.

4. Simulations

In this section, we use simulation to study the main effects tests, interaction effects test and multiple comparisons of two-way ANOVA using PB approach under the assumption of heteroscedastic variances and unequal sizes. The simulation settings follow from Krishnamoorthy et al. (2007) and Xu et al. (2013).

The tests we consider are location-scale invariant. Consider $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$, $i = 1, \cdots, a, j = 1, \cdots, b$ and $\epsilon_{ijk} \sim N(0, \sigma_{ij})$. Without loss of generality, we take $\mu_{ij} = 0, \sigma_{ij}^2 = 1$ in simulation studies. The sample statistics $\bar{y}_{ij}$ and s_{ij}^2 are generated independently as $\bar{y}_{ij} \sim N(0, \sigma_{ij}^2/n_{ij}), s_{ij}^2 \sim \sigma_{ij}^2 \chi^2_{n_{ij}-1}/(n_{ij}-1)$, with $0 < \sigma_{ij}^2 < 1$.

The simulation study was performed with factors: (1) number of factor levels: $a = 2$ and $b = 3$ to form 6 combinations; (2) population standard deviation $\boldsymbol{\sigma}_i = (\sigma_{11}, \cdots, \sigma_{23})$:
$\boldsymbol{\sigma}_1^2 = (1,1,1,1,1,1), \boldsymbol{\sigma}_2^2 = (0.1,0.1,0.1,0.5,0.5,0.5), \boldsymbol{\sigma}_3^2 = (1,1,1,0.5,0.5,0.5)$,
$\boldsymbol{\sigma}_4^2 = (0.1,0.2,0.3,0.4,0.5,1.0), \boldsymbol{\sigma}_5^2 = (0.3,0.9,0.4,0.7,0.5,1), \boldsymbol{\sigma}_6^2 = (0.01,0.1,0.1,0.1,0.1,1)$;
(3) Significance level α: .05 and .1; (4) group sizes $\mathbf{n}_i = (n_{11}, \cdots, n_{23}) : \mathbf{n}_1 = (5,5,5,5,5,5), \mathbf{n}_2 = (10,10,10,10,10,10), \mathbf{n}_3 = (3,3,4,5,6,6), \mathbf{n}_4 = (4,6,8,12,16,20)$; (5) weight variable w_i and v_j: two sets for w_i and two sets for v_j. For a given sample size and parameter configuration, we generated 2500 observed vectors $(\bar{y}_{11}, \cdots, \bar{y}_{ab}, s_{11}^2, \cdots, s_{ab}^2)$ and used 5000 runs to estimate the p-value. The following is used to derive p-value of simultaneous tests (15): (a) calculate $q_m^0 = \max(q_{ii'}^{A0})$ using (23), use Algorithm 3 to find q_α, the $1-\alpha$ percentile of the simulated distribution of q; (b) repeat step (a) for 2500 times, p-value is the proportion of the 2500 simulations when $q_m^0 > q_\alpha$. p-value for simultaneous tests (16) can be derived similarly. Table 1 gives the results of main effects tests and interaction test. Table 2 gives the simulation results of multiple comparisons. From Table 1 and Table 2, we can see that the actual levels of main effects test, interaction effects test and multiple comparisons are close to the nominal levels.

Table 1: Simulation results of main effects and interaction effects tests: "Inter" means interaction test; "Factor A" and "Factor B" mean the main effect tests. Numbers in the Table are simulated p-values. We consider four different sizes, $\mathbf{n}_1 = (5,5,5,5,5,5), \mathbf{n}_2 = (10,10,10,10,10,10), \mathbf{n}_3 = (3,3,4,5,6,6), \mathbf{n}_4 = (4,6,8,12,16,20)$. $\boldsymbol{\sigma}_i^2$ is a vector of unequal variances, we consider, $\boldsymbol{\sigma}_1^2 = (1,1,1,1,1,1), \boldsymbol{\sigma}_2^2 = (0.1,0.1,0.1,0.5,0.5,0.5), \boldsymbol{\sigma}_3^2 = (1,1,1,0.5,0.5,0.5), \boldsymbol{\sigma}_4^2 = (0.1,0.2,0.3,0.4,0.5,1.0), \boldsymbol{\sigma}_5^2 = (0.3,0.9,0.4,0.7,0.5,1), \boldsymbol{\sigma}_6^2 = (0.01,0.1,0.1,0.1,0.1,1)$.

		$\alpha = 0.05$			$\alpha = 0.1$		
	$\boldsymbol{\sigma}_i^2$	Inter	Factor A	Factor B	Inter	Factor A	Factor B
$\mathbf{n}_1$	$\boldsymbol{\sigma}_1^2$	0.0415	0.0405	0.0470	0.0975	0.0895	0.0780
	$\boldsymbol{\sigma}_2^2$	0.0440	0.0445	0.0405	0.1040	0.092	0.0955
	$\boldsymbol{\sigma}_3^2$	0.0430	0.0390	0.043	0.0870	0.086	0.0985
	$\boldsymbol{\sigma}_4^2$	0.0490	0.0490	0.0510	0.0860	0.088	0.0900
	$\boldsymbol{\sigma}_5^2$	0.0510	0.0375	0.044	0.0895	0.0965	0.0905
	$\boldsymbol{\sigma}_6^2$	0.0600	0.0535	0.0600	0.0975	0.1095	0.0920
$\mathbf{n}_2$	$\boldsymbol{\sigma}_1^2$	0.0485	0.0455	0.0440	0.1060	0.0995	0.0945
	$\boldsymbol{\sigma}_2^2$	0.0505	0.0410	0.0505	0.0995	0.0905	0.0960
	$\boldsymbol{\sigma}_3^2$	0.0555	0.0490	0.0425	0.0985	0.1010	0.0965
	$\boldsymbol{\sigma}_4^2$	0.0475	0.0505	0.0510	0.0935	0.1015	0.1070
	$\boldsymbol{\sigma}_5^2$	0.0535	0.0485	0.0485	0.0945	0.097	0.1005
	$\boldsymbol{\sigma}_6^2$	0.0525	0.0450	0.0505	0.0985	0.0985	0.1010
$\mathbf{n}_3$	$\boldsymbol{\sigma}_1^2$	0.0465	0.0500	0.0375	0.0760	0.0985	0.0860
	$\boldsymbol{\sigma}_2^2$	0.0435	0.0460	0.0415	0.0815	0.1030	0.0930
	$\boldsymbol{\sigma}_3^2$	0.0495	0.0520	0.0390	0.0955	0.1050	0.0755
	$\boldsymbol{\sigma}_4^2$	0.0430	0.0450	0.0420	0.1015	0.0805	0.0900
	$\boldsymbol{\sigma}_5^2$	0.0510	0.0625	0.0355	0.0955	0.0845	0.0930
	$\boldsymbol{\sigma}_6^2$	0.0525	0.0525	0.0435	0.1055	0.0925	0.0945
$\mathbf{n}_4$	$\boldsymbol{\sigma}_1^2$	0.0485	0.0455	0.046	0.1135	0.104	0.1075
	$\boldsymbol{\sigma}_2^2$	0.0380	0.0610	0.05	0.0950	0.106	0.1015
	$\boldsymbol{\sigma}_3^2$	0.0535	0.0625	0.0535	0.1045	0.1145	0.1000
	$\boldsymbol{\sigma}_4^2$	0.0520	0.0530	0.0355	0.0945	0.108	0.0930
	$\boldsymbol{\sigma}_5^2$	0.0425	0.0505	0.0435	0.0940	0.111	0.1015
	$\boldsymbol{\sigma}_6^2$	0.0450	0.0505	0.0515	0.0960	0.0945	0.0925

Table 2: Simulation results of multiple comparisons: "MCP" means multiple comparisons; "MCPA1" is the multiple comparison of factor A using weights described in (17); "MCPA2" is the multiple comparison of factor A with weights described in (18); "MCPB1" is the multiple comparison of factor B with weights described in (25); "MCPB2" is the multiple comparison of factor B with weights described in (26). Numbers in Table are simulated p-values. We consider four different sizes, $\mathbf{n}_1 = (5,5,5,5,5,5), \mathbf{n}_2 = (10,10,10,10,10,10), \mathbf{n}_3 = (3,3,4,5,6,6), \mathbf{n}_4 = (4,6,8,12,16,20)$. $\boldsymbol{\sigma}_i^2$ is a vector of unequal variances, we consider, $\boldsymbol{\sigma}_1^2 = (1,1,1,1,1,1), \boldsymbol{\sigma}_2^2 = (0.1,0.1,0.1,0.5,0.5,0.5), \boldsymbol{\sigma}_3^2 = (1,1,1,0.5,0.5,0.5), \boldsymbol{\sigma}_4^2 = (0.1,0.2,0.3,0.4,0.5,1.0), \boldsymbol{\sigma}_5^2 = (0.3,0.9,0.4,0.7,0.5,1), \boldsymbol{\sigma}_6^2 = (0.01,0.1,0.1,0.1,0.1,1)$.

		$\alpha = 0.05$				$\alpha = 0.1$			
	$\boldsymbol{\sigma}_i^2$	MCPA1	MCPA2	MCPB1	MCPB2	MCPA1	MCPA2	MCPB1	MCPB2
$\mathbf{n}_1$	σ_1^2	0.0375	0.0515	0.0495	0.0460	0.0850	0.1015	0.0940	0.0825
	$\boldsymbol{\sigma}_2^2$	0.0465	0.0530	0.0515	0.0565	0.0895	0.0995	0.0865	0.0945
	$\boldsymbol{\sigma}_3^2$	0.0400	0.0460	0.0380	0.0420	0.0890	0.0885	0.0885	0.0800
	$\boldsymbol{\sigma}_4^2$	0.0470	0.0395	0.0450	0.0510	0.0965	0.1010	0.0880	0.1000
	$\boldsymbol{\sigma}_5^2$	0.0515	0.0415	0.0515	0.0415	0.0915	0.0990	0.0925	0.0915
	$\boldsymbol{\sigma}_6^2$	0.0520	0.064	0.0490	0.0505	0.1110	0.1110	0.099	0.0970
$\mathbf{n}_2$	$\boldsymbol{\sigma}_1^2$	0.0395	0.0520	0.0440	0.0585	0.0945	0.0940	0.1050	0.1010
	$\boldsymbol{\sigma}_2^2$	0.0505	0.0480	0.0430	0.0480	0.0920	0.1095	0.0995	0.0985
	$\boldsymbol{\sigma}_3^2$	0.0415	0.0455	0.0510	0.0465	0.1020	0.1020	0.0955	0.1000
	$\boldsymbol{\sigma}_4^2$	0.0520	0.0495	0.0475	0.0380	0.1020	0.0925	0.1010	0.0885
	$\boldsymbol{\sigma}_5^2$	0.0585	0.0550	0.0470	0.0500	0.0895	0.1030	0.1000	0.1040
	$\boldsymbol{\sigma}_6^2$	0.0500	0.0550	0.051	0.0460	0.0990	0.1025	0.0850	0.1075
$\mathbf{n}_3$	$\boldsymbol{\sigma}_1^2$	0.0480	0.0520	0.0445	0.0390	0.0930	0.1035	0.1015	0.0905
	$\boldsymbol{\sigma}_2^2$	0.0470	0.0395	0.0430	0.0500	0.0940	0.0860	0.0835	0.0815
	$\boldsymbol{\sigma}_3^2$	0.0620	0.0545	0.0465	0.0415	0.0945	0.0915	0.0985	0.0835
	$\boldsymbol{\sigma}_4^2$	0.0440	0.0405	0.0450	0.0435	0.0870	0.1125	0.0890	0.0835
	$\boldsymbol{\sigma}_5^2$	0.0430	0.0435	0.0460	0.0375	0.0930	0.1035	0.092	0.0965
	$\boldsymbol{\sigma}_6^2$	0.0545	0.0540	0.0490	0.0540	0.0910	0.1060	0.1065	0.0980
$\mathbf{n}_4$	$\boldsymbol{\sigma}_1^2$	0.0530	0.0485	0.0520	0.0495	0.0980	0.1045	0.1025	0.0965
	$\boldsymbol{\sigma}_2^2$	0.0525	0.0485	0.0505	0.0570	0.0890	0.0940	0.0980	0.1045
	$\boldsymbol{\sigma}_3^2$	0.0555	0.0465	0.0525	0.0500	0.1030	0.0890	0.1005	0.1050
	$\boldsymbol{\sigma}_4^2$	0.0595	0.0545	0.0460	0.0490	0.1005	0.1075	0.1120	0.0835
	$\boldsymbol{\sigma}_5^2$	0.0405	0.0555	0.0625	0.0495	0.0895	0.0905	0.0940	0.0960
	$\boldsymbol{\sigma}_6^2$	0.0485	0.0515	0.0415	0.0480	0.1170	0.1015	0.0940	0.0965

5. Conclusions

Parametric bootstrap approach Krishnamoorthy et al. (2007) for testing the equality of several means under the assumption of heteroscedastic variances has been extended to a multiple comparison procedure in two-way ANOVA. Together with the main effects and interaction effects tests from Xu et al. (2013), a complete study of two-way ANOVA under heteroscedastic variances and unequal sizes from parametric bootstrap approach without using transformations is derived. Simulation studies show that the Type I error of main effects test, interaction effects test and multiple comparisons are close to the nominal level.

References

Ananda, M. and Weerahandi, S. (1997), "Two-way ANOVA with unequal cell frequencies and unequal variances," *Statistics Sinica*, 7, 631–646.

Einot, I. and Gabriel, K. (1975), "A Study of the Powers of Several Methods of Multiple Comparisons," *Journal of the American Statistical Association*, 70, 574–583.

Games, P. (1971), "Multiple Comparisons of Means," *American Educational Research Journal*, 8, 531–565.

Games, P. and Howell, J. (1976), "Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study," *J.Educ.Statist.*, 1, 113–125.

Hochberg, Y. (1976), "A Modification of the T-Method of Multiple Compaiisons for a One-Way Layout with Unequal Variances," *Journal of the American Statistical Association*, 71, 200–203.

Johansen, S. (1980), "The WelchJames approximation to the distribution of the residual sum of squares in a weighted linear regression," *Biometrika*, 67, 85–95.

Kaiser, L. and Bowden, D. (1983), "Simultaneous confidence intervals for all linear contrasts of means with heterogenous variances," *Comm. Statist. Theory Meth.*, 12, 73–88.

Krishnamoorthy, K., Lu, F., and Mathew, T. (2007), "A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models," *Computational statistics and data analysis*, 51, 5731–5742.

Krutchkoff, R. G. (1989), "Two-way fixed effects analysis of variance when the error variances may be unequal," *Journal of Statistical Computation and Simulation*, 32, 177–183.

Scheffé, H. (1959), *The Analysis of Variance*, New York: Wiley.

Spjøvoll, E. and Stoline, M. R. (1973), "An Extension of the T-Method of Multiple Comparison to Include the Cases with Unequal Sample Sizes," *Journal of the American Statistical Association*, 69, 975–978.

Wang, S. and Chow, S. C. (1994), *Advanced Linear Models*, New York: Marcel Dekker.

Welch, B. (1948), "Further Note on Mrs. Aspin's Tables and on Certain Approximations to the Tabled Functions," *Biometrics*, 56, 293–296.

Wilcox, R. R. (1989), "Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA model," *Journal of Statistics Education*, 14, 269–278.

Xu, L., Yang, F., Abula, A., and Qin, S. (2013), "A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances," *Journal of Multivariate Analysis*, 115, 172–180.

Zhang, G. (2013), "A Parametric Bootstrap Approach for One-way ANOVA under Unequal Variances with Unbalanced Data," *Communications in Statistics Simulation and Computation*, to appear.

Zhang, J. (2012), "An approximate degrees of freedom test for heteroscedastic two-way ANOVA," *Journal of Statistical Planning and Inference*, 142, 336–346.