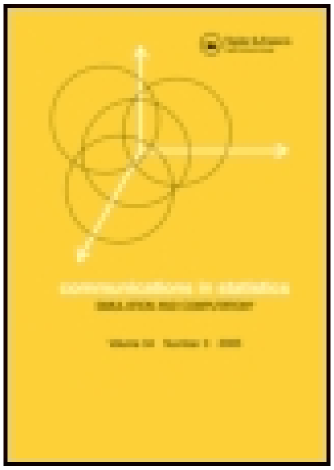

Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

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

Guoyi Zhang [a]

[a] Department of Mathematics and Statistics , University of New Mexico , Albuquerque , NM , 87131-0001 , U.S.A Phone: 505-277-4744

Accepted author version posted online: 07 Mar 2014.

To cite this article: Guoyi Zhang (2014): A Parametric Bootstrap Approach for One-way ANOVA under Unequal Variances with Unbalanced Data, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2013.794288

To link to this article: http://dx.doi.org/10.1080/03610918.2013.794288

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

Guoyi Zhang

Department of Mathematics and Statistics

University of New Mexico

Albuquerque, NM, 87131-0001

U.S.A

505-277-4744

gzhang123@gmail.com

**Abstract**

This research is to provide a solution of one-way ANOVA without using transformation when variances are heteroscedastic and group sizes are unequal. Parametric boothstrap test (Krishnamoorthy, Lu, & Mathew, 2007) has been shown to be competitive with many other methods when testing the equality of group means. We extend the parametric bootstrap algorithm to a multiple comparison procedure. Simulation results show that the parametric bootstrap approach works well for one-way ANOVA.

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



# 1 Introduction

Consider the ANOVA problem of $r$ normal populations with unequal population variances $\sigma_i^2$, $i = 1, 2, \cdots, r$ and let $Y_{i1}, Y_{i2}, \cdots, Y_{i,n_i}$ be a random sample from $N(u_i, \sigma_i^2)$. The one-way ANOVA model is as follows,

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

where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_i^2), i = 1, 2, \cdots, r, j = 1, 2, \cdots, n_i$. This research intends to provide a solution of one-way ANOVA: testing equality of the factor level means and all pairwise comparisons under the assumption of heteroscedastic variances and unequal sizes.

When the population variances are unequal, the classical $F$ test fails to reject the null hypothesis of equal factor level means even for large samples. Many alternative methods were developed due to this issue. Parametric bootstrap test (Krishnamoorthy et al., 2007) is one of such tests. Yiğit and Gokpinar (2010) carried out a simulation study to compare the size performance of the $F$, W (Welch, 1951), SS (Scott & Smith, 1971), BF (Brown & Forsythe, 1974), Chen-Chen's One Stage (OS) (Chen & Chen, 1998), Chen-Chen's One Stage Range (OSR) (Chen, 2001), Weerahandi's Generalized F (GF) (Weerahandi, 1995), Xu-Wang's Generalized F (XW) (Xu & Wang, 2007(a), 2007(b)) and Parametric bootstrap (PB) (Krishnamoorthy et al., 2007) tests when population variances are unequal for one-way ANOVA problem. The Type I error rates and powers of the tests are compared using various sample sizes under various parameter combinations. PB test is shown to be one of the best for testing the equality of factor level means under the assumption of heteroscedastic variances.

Another problem in ANOVA is multiple comparisons (all pairwise simultaneous comparisons). Scheffé'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. However, research of multiple comparisons under the assumption of heteroscedasticity is limited. Hochberg (1976) generalized the Spj$\phi$tvoll 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 & 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 and no uniformly preferable choice has been reached yet.

In this research, we propose a parametric bootstrap test of multiple comparison for use in one-way ANOVA under the assumption of heteroscedastic variances and unequal sizes. Research in this paper together with PB test (Krishnamoorthy et al., 2007) provide a complete solution of the one-way ANOVA. This paper is organized as follows. In Section 2, we review PB test. In Section 3, we propose parametric bootstrap algorithm of multiple comparison for one-way ANOVA. In Section 4, we present simulation studies. Section 5 gives conclusions.

## 2 The Parametric Bootstrap Test for Population Means

In applied statistics an experimenter wants to compare two or more populations, i.e.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0 \text{ v.s. } H_\alpha : \text{ at least one } \alpha_i \neq 0. \tag{2}$$

The classical $F$ test fails to reject the null hypothesis even for large samples when the population variances are unequal. Many alternative methods are developed due to this issue. In this section, we review the PB test suggested by Krishnamoorthy et al. (2007).

Assume $\sigma_i^2$ are unknown, a natural test statistic is the standardized between group sum of squares

$$T_N(S_1^2, \cdots, S_r^2) = \sum_{i=1}^{r} \frac{n_i}{S_i^2} \bar{Y}_i^2 - \frac{(\sum_{i=1}^{r} n_i \bar{Y}_i / S_i^2)^2}{\sum_{i=1}^{r} n_i / S_i^2}, \tag{3}$$

where $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$ and $S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2/(n_i - 1)$. The test rejects $H_0$ in (2) when $T_N(\sigma_1^2, \cdots, \sigma_r^2) > \chi^2_{r-1,\alpha}$ with $\chi^2_{r-1,\alpha}$ the upper $\alpha$th quantile of a chi-square distribution with degrees of freedom $r - 1$. The PB pivot variable can be obtained by replacing $\bar{Y}_i, S_i^2$ in (3) by $\bar{Y}_{B_i}, S^2_{B_i}$ with $\bar{Y}_{B_i} \sim N(0, S_i^2/n_i)$ and $S^2_{B_i} \sim S_i^2 \chi^2_{n_i-1}/(n_i - 1)$, $i = 1, \cdots, r$. Let $Z_i$ be a standard normal random variable. The $PB$ pivot variable has the same distribution as

$$T_{NB}(S_1^2, \cdots, S_r^2) = \sum_{i=1}^{r} \frac{Z_i^2 (n_i - 1)}{\chi^2_{n_i-1}} - \frac{(\sum_{i=1}^{r} \sqrt{n_i} Z_i (n_i - 1)/S_i \chi^2_{n_i-1})^2}{\sum_{i=1}^{r} n_i (n_i - 1)/S_i^2 \chi^2_{n_i-1}}. \tag{4}$$

Krishnamoorthy et al. (2007) suggested generating a simulated distribution of $T_N$ using (4) to estimate the $p$-value of $T_N^*$ (test statistic derived using (3)). The following is the procedure:

**Algorithm 1**

For a given $(n_1, \cdots, n_r)$, $(\bar{y}_1, \cdots, \bar{y}_r)$ of $(\bar{Y}_1, \cdots, \bar{Y}_r)$ and $(s_1^2, \cdots, s_r^2)$ of $(S_1^2, \cdots, S_r^2)$:

compute $T_N(s_1^2, \cdots, s_r^2)$ using (3), label it as $T_N^*$

For $l = 1, \cdots, L$

generate $Z_i \sim N(0,1)$ and $\chi^2_{n_i-1}, i = 1, \cdots, r$

compute $T_{NB}(s_1^2, \cdots, s_r^2)$ using (4)

if $T_{NB}(s_1^2, \cdots, s_r^2) > T_N^*$, set $Q_l = 1$

(end loop)

$\sum_{l=1}^{L} Q_l / L$ is a Monte Carlo estimate of the p-value of $T_N^*$.

# 3 The Parametric Bootstrap Method for Multiple Comparison with Heteroscedastic Variances

The multiple comparison procedure applies when the family of interest is the set of all pairwise comparisons of factor level means; in other words, the family consists of estimates of all tests of the form

$$H_0 : \mu_i - \mu_j = 0 \text{ v.s. } H_\alpha : \mu_i - \mu_j \neq 0. \tag{5}$$

When all $\sigma_i^2$'s are equal, the Tukey's multiple comparison confidence limits for all pairwise comparisons $D = \mu_i - \mu_j$ with family confidence coefficient of at least $1 - \alpha$ are $\hat{D} \pm q(\alpha) S(\hat{D})$, where $q(\alpha)$ is the upper $\alpha$th quantile of the studentized range distribution, $S^2(\hat{D}) = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / \left[ (n-r)(1/n_i + 1/n_j) \right]$ and $n = \sum_{i=1}^{r} n_i$. However, Tukey's method fails to work when the population variances are unequal. In the following, we propose a parametric bootstrap method for multiple comparison procedure for use under heteroscedastic variances.

Recall that in Section 2, $\bar{Y}_{B_i} \sim N(0, S_i^2/n_i)$ and $S_{B_i}^2 \sim S_i^2 \chi^2_{n_i-1}/(n_i - 1), i = 1, \cdots, r$. Hence $\bar{Y}_{B_i} \overset{d}{=} Z_i (S_i / \sqrt{n_i})$ and $S_{B_i}^2 \overset{d}{=} \dfrac{S_i^2}{n_i - 1} \chi^2_{n_i-1}$, where $\overset{d}{=}$ means the same distribution. Let $q_{ij} = |\bar{Y}_{B_i} - \bar{Y}_{B_j}| / \sqrt{S_{B_i}^2/n_i + S_{B_j}^2/n_j}$. $q_{ij}$ has the same distribution as

$$q_{ij} \overset{d}{=} \frac{|Z_i(\frac{S_i}{\sqrt{n_i}}) - Z_j(\frac{S_j}{\sqrt{n_j}})|}{\sqrt{\frac{S_i^2}{n_i(n_i-1)} \chi^2_{n_i-1} + \frac{S_j^2}{n_j(n_j-1)} \chi^2_{n_i-1}}}, \text{ for } i < j, i, j = 1, 2, \cdots, r. \tag{6}$$

For a given $(n_1, \cdots, n_r)$, $(\bar{y}_1, \cdots, \bar{y}_r)$ and $s_1^2, \cdots, s_r^2$, let

$$q_{ij}^0 = |\bar{y}_i - \bar{y}_j| / \sqrt{s_i^2/n_i + s_j^2/n_j} \text{ for } i = 1, \cdots, r-1, j = i+1, \cdots, r. \quad (7)$$

Given a significance level $\alpha$, the multiple comparison confidence limits for simultaneous comparisons $\mu_i - \mu_j$ with family confidence coefficient at least $1 - \alpha$ are $\bar{y}_i - \bar{y}_j \pm q_\alpha \sqrt{s_i^2/n_i + s_j^2/n_j}$, where $q_\alpha$ can be estimated using parametric bootstrap method given in Algorithm 2.

**Algorithm 2**.

For a given $(n_1, \cdots, n_r)$, $(\bar{y}_1, \cdots, \bar{y}_r)$ and $(s_1^2, \cdots, s_r^2)$:

For $l = 1, \cdots, L$

Generate $Z_i \sim N(0, 1)$ and $\chi^2_{n_i-1}, i = 1, \cdots, r$

Compute $q_{ij}$ using (6) for $i = 1, \cdots, r-1,\ j = i+1, \cdots, r$

Find $q_l = max(q_{ij})$

(end loop)

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

# 4 Simulations

In this section, we use simulation to study the overall test and multiple comparisons of one-way ANOVA model using parametric bootstrap under the assumption of heteroscedastic variances and unequal sizes. The simulation settings follow from Krishnamoorthy et al. (2007) and Yiğit and Gokpinar (2010).

The tests we consider are location-scale invariant. Without loss of generality, we take $\mu_1 = \cdots = \mu_r = 0, \sigma_1^2 = 1$ and $0 < \sigma_i^2 < 1$, for $i = 2, \cdots, r$ in our simulation studies. The sample statistics $\bar{y}_i$ and $s_i^2$ are generated independently as $\bar{y}_i \sim N(0, \sigma_i^2/n_i)$ and $s_i^2 \sim \sigma^2\chi^2_{n_i-1}/(n_i - 1)$, with $0 < \sigma_i^2 < 1, i = 2, \cdots, r$.

The simulation study was performed with factors: (1) population standard deviation $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_r)$: various combinations; (2) number of levels $r$: $r = 3$ and $r = 10$; (3) Significance level $\alpha$: .01, .05 and .1; (4) group sizes $\mathbf{n} = (n_1, \cdots, n_r)$: various combinations. For a given sample size and parameter configuration, we generated 2500 observed vectors $(\bar{y}_1, \cdots, \bar{y}_r, s_1^2, \cdots, s_r^2)$ and used 5000 runs to estimate the p-value. Algorithm 1 is used to estimate the p-value of overall test (2). The following is used to derive p-value of simultaneous tests (5): (a) calculate $q_m^0 = max(q_{ij}^0)$ using (7), use Algorithm 2 to find $q_\alpha$, the $1 - \alpha$ percentile of the simulated distribution of $q_m$; (b) repeat step (a) for 2500 times, p-value is the proportion of the 2500 simulations when $q_m^0 > q_\alpha$. Table 1 and Table 2 give the simulation results. From Table 1 and Table 2, we can see that the actual levels of overall test and multiple comparison procedure are close to the nominal levels.

## 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. Therefore, a complete study of one-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 overall test and multiple comparison procedure are close to the nominal level.

## References

Brown, M., & Forsythe, A. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, *16*, 129–132.

Chen, S. (2001). One stage and two stage statistical inference under heteroscedasticity. *Communications in statistics simulation*, *30*, 991–1009.

Chen, S., & Chen, J. (1998). Single-stage analysis of variance under heterosecedastictiy. *Communications in statistics simulation*, *27*, 641–666.

Einot, I., & 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., & 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.

Kaiser, L., & 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., & Mathew, T. (2007). A parametric bootstrap approach for anova with unequal variances: Fixed and random models. *Computational statistics and data analysis*, *51*, 5731–5742.

Scott, A., & Smith, T. (1971). Interval estimates for linear combinations of means. *Applied Statistics*, *20*, 276–285.

SpjØvoll, E., & 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.

Weerahandi, S. (1995). ANOVA under unequal error variances. *Biometrics*, *51*, 589–599.

Welch, B. (1948). Further note on mrs. aspin's tables and on certain approximations to the tabled functions. *Biometrics*, *56*, 293–296.

Welch, B. (1951). On the comparison of several mean values: An alternative approach. *Biometrica*, *38*, 330–336.

Xu, L., & Wang, S. (2007(a)). A new generalized p-value and for anova under heteroscedasticity. *Statistics and Probability Letters*, *78*, 963–969.

Xu, L., & Wang, S. (2007(b)). A new generalized p-value and its upper bound for anova under unequal errors variances. *Communications in Statistics Theory and Methods*, *37*, 1002–1010.

Yiğit, E., & Gokpinar, F. (2010). A simulation study on tests for one-way anova under the unequal variance assumption. *Commun.Fac.Sci.Univ.Series A*, *59*(2), 15–34.



ACCEPTED MANUSCRIPT

Table 1: Simulation result 1: **n** is a vector of unequal group sizes; $\boldsymbol{\sigma}$ is a vector of unequal variances; "Overall" means PB test (Krishnamoorthy et al., 2007) for equality of group means; "MCP" is parametric bootstrap multiple comparison procedure (proposed method); Numbers in Table are simulated p-values.

| | | $\alpha$ = .01 | | $\alpha$ = .05 | | $\alpha$ = .1 | |
|---|---|---|---|---|---|---|---|
| **n** | $\boldsymbol{\sigma}$ | Overall | MCP | Overall | MCP | Overall | MCP |
| (3,5,7) | (1,1,1) | .0155 | .0125 | .0595 | .063 | .0970 | .1005 |
| | (4,4,4) | .0135 | .0110 | .0485 | .0525 | .1045 | .1145 |
| | (1,2,4) | .0075 | .0070 | .0400 | .0515 | .0915 | .095 |
| | (1,4,9) | .0120 | .0100 | .0500 | .0545 | .109 | .0965 |
| | (4,2,1) | .0300 | .0285 | .0715 | .0685 | .1150 | .1095 |
| | (9,4,1) | .0275 | .0270 | .0685 | .063 | .1200 | .1155 |
| (7,10,13) | (1,1,1) | .0065 | .0120 | .0455 | .0445 | .1005 | .0920 |
| | (4,4,4) | .0115 | .0075 | .0515 | .0520 | .0975 | .0960 |
| | (1,2,4) | .007 | .0085 | .0515 | .0485 | .090 | .1040 |
| | (1,4,9) | .0075 | .0095 | .046 | .0535 | .0975 | .1000 |
| | (4,2,1) | .012 | .0145 | .0515 | .0555 | .1125 | .0985 |
| | (9,4,1) | .011 | .0095 | .0505 | .0490 | .1025 | .1080 |

Table 2: Simulation result 2: **n** is a vector of unequal group sizes; $\boldsymbol{\sigma}$ is a vector of unequal variances; "Overall" means PB test (Krishnamoorthy et al., 2007) for equality of group means; "MCP" is parametric bootstrap multiple comparison procedure (proposed method); Numbers in Table are simulated p-values.

| **n** = (3, 3, 3, 4, 4, 4, 5, 5, 5, 5) | $\alpha$ = .01 | | $\alpha$ = .05 | | $\alpha$ = .1 | |
|---|---|---|---|---|---|---|
| $\boldsymbol{\sigma}$ | Overall | MCP | Overall | MCP | Overall | MCP |
| $(1, 1, \cdots, 1)$ | .0035 | .004 | .038 | .0285 | .083 | .0725 |
| $(4, 4, \cdots, 4)$ | .0045 | .005 | .044 | .0395 | .0855 | .084 |
| (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) | .0055 | .0085 | .039 | .0385 | .087 | .0775 |
| (1,1,1,4,4,4,9,9,9,9) | .007 | .007 | .035 | .0355 | .089 | .086 |
| (9,9,9,4,4,4,1,1,1,1) | .015 | .009 | .066 | .054 | .0875 | .097 |
| (10,9,8,7,6,5,4,3,2,1) | .014 | .014 | .069 | .056 | .106 | .105 |