Statistics and Probability Letters 174 (2021) 109095

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Parametric boostrap and objective Bayesian testing for heteroscedastic one-way ANOVA

Guoyi Zhang [a,*], Ronald Christensen [a], John Pesko [b]

[a] Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, 87131-0001, United States of America
[b] AbbVie Inc., IL, United States of America

ARTICLE INFO

Article history:
Received 15 December 2019
Received in revised form 2 March 2021
Accepted 4 March 2021
Available online 18 March 2021

Keywords:
Heteroscedastic ANOVA
Parametric bootstrap
Objective Bayesian
Overall mean tests
Simulations

ABSTRACT

This article explores relationship between parametric bootstrap (PB) and objective Bayesian (OB) approaches to test equality of the factor level means (overall mean test) of heteroscedastic one-way analysis of variance (ANOVA) problem. We compared overall mean tests based on PB, OB, and OB using posterior predictive distribution approaches by simulation studies. We also proved that the PB and OB overall mean tests for one-way heteroscedastic ANOVA are asymptotically equivalent.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Consider an ANOVA problem from r normal populations with unequal population variances σ_i^2, $i = 1, 2, \ldots, r$. Let n_i be the group size of factor level i, and let n be the sample size with $n = \sum_{i=1}^r n_i$. Let $y_{i1}, y_{i2}, \ldots, y_{in_i}$ be a random sample from $N(u_i, \sigma_i^2)$. The one-way heteroscedastic ANOVA (heteANOVA) model with r factor levels is:

$$y_{ij} = \mu_i + e_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n_i, \quad e_{ij} \sim N(0, \sigma_i^2). \tag{1}$$

This research intends to investigate some methods of the overall mean test for one-way heteANOVA

$$\mathrm{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_r \text{ versus } H_\alpha : \text{not all } \mu_i \text{ are equal.} \tag{2}$$

When population variances are unequal, Weerahandi (1995) showed that the p-value given by the classical F-test is much larger than the p-value obtained under unequal variance assumption. Many alternative methods were developed for tests in heteANOVA. Yiğit and Gokpinar (2010) compared nine tests in heteANOVA, such as the F, W (Welch, 1951), BF (Brown and Forsythe, 1974), Weerahandi's Generalized F (GF) (Weerahandi, 1995), and PB (Krishnamoorthy et al., 2007) tests, etc. PB test was shown to be one of the best in regarding the type-I error rates and powers for overall mean test. PB approach was also shown to be competitive for multiple comparison tests in heteANOVA (Zhang, 2015a,b).

Another approach for one-way heteANOVA problem is Objective Bayes (OB). OB uses Bayes' Theorem to obtain posterior distributions of parameters based on some "objective" (also called non-informative or flat) prior distribution and likelihood function. In most cases, objective priors have little effect on the posterior analysis and give answers that resemble frequentist solutions, so that data can speak for themselves as much as possible. The pioneers of Bayesianism,

* Corresponding author.
E-mail addresses: gzhang123@gmail.com (G. Zhang), fletcher@stat.unm.edu (R. Christensen), jpesko@unm.edu (J. Pesko).

https://doi.org/10.1016/j.spl.2021.109095
0167-7152/© 2021 Elsevier B.V. All rights reserved.

Thomas Bayes and Pierre–Simon Laplace, employed flat priors on the unobserved parameters in the context of their "inverse probability" approach (Bayes and Price, 1683–1775). But Jeffreys (1946) is widely considered as the originator of the OB methodology.

In this research, we are interested in the relationship between PB and OB approaches to overall mean test for heteANOVA problem (thereafter called PB or OB test). We want to show the asymptotic equivalence of the different tests, and to compare the type I error rates and powers of the tests under various settings. This paper is organized as follows: Section 2 includes a literature review to help establish some of the background topics related to the research; Section 3 proves that the PB and OB tests are asymptotically equivalent; Section 4 compares tests based on PB and OB by simulation studies; Section 5 gives a real example to illustrate the usage of the PB and OB tests; and Section 6 gives conclusions and possible future research topics.

2. Background

In this section, we review PB, OB, and OB using posterior predictive distribution (OBpred) approaches to test H_0 in Eq. (2) for one way heteANOVA model (1). The Wald-type weighted test statistic is given by:

$$T = \sum_{i=1}^{r} \frac{n_i}{\sigma_i^2} \bar{y}_{i\cdot}{}^2 - \frac{\left(\sum_{i=1}^{r} \frac{n_i}{\sigma_i^2} \bar{y}_{i\cdot}\right)^2}{\sum_{i=1}^{r} \frac{n_i}{\sigma_i^2}},$$

where $\bar{y}_{i.}$ is the group i sample mean, i.e., $\sum_{j=1}^{n_i} y_{ij}/n_i$. The observed test statistic is obtained by replacing the σ_i's with their respective sample variances s_i^2 as follows,

$$T_{Obs} = \sum_{i=1}^{r} \frac{n_i}{s_i^2} \bar{y}_{i\cdot}{}^2 - \frac{\left(\sum_{i=1}^{r} \frac{n_i}{s_i^2} \bar{y}_{i\cdot}\right)^2}{\sum_{i=1}^{r} \frac{n_i}{s_i^2}}, \tag{3}$$

where $s_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2/(n_i - 1)$. The distribution of T_{Obs} is obtained through simulations.

2.1. Parametric bootstrap (PB) approach

For the goal of simulating T_{Obs} under H_0, the PB method generates the sufficient statistics $\bar{y}_i$ and s_i^2 as follows:

$$\bar{y}_{iB} \sim N\left(\bar{y}_{\cdot\cdot}, \frac{s_i^2}{n_i}\right) \quad \text{or} \quad \bar{y}_{iB} \sim N\left(0, s_i^2/n_i\right) \quad \text{and} \quad s_{iB}^2 \sim \frac{s_i^2 \chi^2_{n_i-1}}{n_i - 1}, \tag{4}$$

where $\bar{y}_{\cdot\cdot} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} y_{ij}/n$, and $\chi^2_{n_i-1}$ is a chi-square random variable with $n_i - 1$ degrees of freedom. The tests we consider are location-scale invariant, so without loss of generality, we can generate $\bar{y}_{iB} \sim N\left(0, s_i^2/n_i\right)$.

For each value of $B = 1, \ldots, M$, we generate $\bar{y}_{iB}$, s_{iB}^2 , for $i = 1, \ldots, r$ and compute

$$T_{PB} = \sum_{i=1}^{r} \frac{n_i}{s_{iB}^2} \bar{y}_{iB}^2 - \frac{\left(\sum_{i=1}^{r} \frac{n_i}{s_{iB}^2} \bar{y}_{iB}\right)^2}{\sum_{i=1}^{r} \frac{n_i}{s_{iB}^2}}. \tag{5}$$

With a sufficiently large number of draws M, we can flesh out an estimate of the sampling distribution of T_{Obs} under H_0, and estimate the generalized p-value $Pr(T_{PB} > T_{Obs})$ with $\mathcal{P}_{PB} = \sum_{B=1}^{M} \mathrm{I}(T_{PB} > T_{Obs})/M$, where I() is an indicator variable such that

$$\mathrm{I}(T_{PB} > T_{Obs}) = \begin{cases} 1 & \text{if } T_{PB} > T_{Obs}, \\ 0 & otherwise. \end{cases}$$

2.2. Objective Bayesian (OB) approach

Empirically, the OB approach to the significance test of $H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$ is similar to the PB approach. A Bayesian significance test looks at how far away T_{Obs} is from the posterior distribution of:

$$\tilde{T} = \sum_{i=1}^{r} \frac{n_i}{s_i^2} (\bar{y}_{i.} - \mu_i)^2 - \frac{\left(\sum_{i=1}^{r} \frac{n_i}{s_i^2} (\bar{y}_{i.} - \mu_i)\right)^2}{\sum_{i=1}^{r} \frac{n_i}{s_i^2}}.$$

Note that when H_0 is true, $\tilde{T} = T_{Obs}$. With flat priors on the μ_i, we have

$$\frac{\bar{y}_{i.} - \mu_i}{s_i/\sqrt{n_i}}|\boldsymbol{Y} \sim t_{n_i-1} \quad \text{therefore } \bar{y}_{i.} - \mu_i|\boldsymbol{Y} \sim \frac{s_i}{\sqrt{n_i}} t_{n_i-1}. \tag{6}$$

We repeatedly generate and compute

$$\begin{aligned} T_{OB} &= \sum_{i=1}^{r} \frac{n_i}{s_i^2}\left(\frac{s_i}{\sqrt{n_i}} t_{n_i-1}\right)^2 - \frac{\left(\sum_{i=1}^{r} \frac{n_i}{s_i^2}\frac{s_i}{\sqrt{n_i}} t_{n_i-1}\right)^2}{\sum_{i=1}^{r}\frac{n_i}{s_i^2}} \\ &= \sum_{i=1}^{r} t_{n_i-1}^2 - \frac{\left(\sum_{i=1}^{r} \frac{\sqrt{n_i}}{s_i} t_{n_i-1}\right)^2}{\sum_{i=1}^{r}\frac{n_i}{s_i^2}} \end{aligned} \tag{7}$$

Similar to the PB test, we generate T_{OB} M times, and estimate the generalized p-value with $\mathcal{P}_{OB} = \sum_{B=1}^{M} \mathrm{I}(T_{OB} > T_{Obs})/M$.

2.3. Objective Bayesian with posterior predictive distribution (OBpred) approach

The objective posterior predictive distribution is the distribution of unobserved future responses given the observed data with non-informative prior. The predictive distribution can be dated back to Aitchison (1975).

The test using T_{Obs} for $H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$ is a one-sided test, so performance is primarily related to the behavior of the right-tail of the sampling distribution of T_{Obs} under H_0. Hence, even if the predictive approach gives a better estimate of this sampling distribution overall, it may not lead to a better test. For one-way heteANOVA, under H_0, new data observations come from a t distribution such as

$$\frac{y_{n_i+1} - \bar{y}_{i.}}{s_i\sqrt{1+\frac{1}{n_i}}}|\boldsymbol{Y} \sim t_{n_i-1}.$$

For a new sample of size n_i: $y_{n_i+1}, y_{n_i+2}, \ldots, y_{n_i+n_i}$, we have

$$\bar{y}_{i.}^* = \frac{\sum_{j=1}^{n_i} y_{n_i+j}}{n_i} = \bar{y}_{i.} + s_i\sqrt{1+\frac{1}{n_i}}\frac{\sum_{j=1}^{n_i} t_j}{n_i},$$

where t_j is the t distribution with degree of freedom $n_i - 1$, $\mathrm{E}[t_i] = 0$ and $\mathrm{Var}(t_i) = (n_i-1)/(n_i-3)$, so that $\mathrm{E}[\bar{y}_{i.}^*|\boldsymbol{Y}] = \bar{y}_{i.}$, and $\mathrm{Var}(\bar{y}_{i.}^*|\boldsymbol{Y}) = s_i^2\left(1+\frac{1}{n_i}\right)\frac{n_i-1}{n_i(n_i-3)}$. Implementing a normal approximation, we have

$$\bar{y}_{i.}^*|\boldsymbol{Y} \sim N\left(\bar{y}_{i.}, \frac{(n_i-1)(n_i+1)}{n_i^2(n_i-3)}s_i^2\right), \quad s_i^{2*}|\boldsymbol{Y} \sim \frac{\chi^2_{n_i-1}}{n_i-1}\frac{(n_i-1)(n_i+1)}{n_i(n_i-3)}s_i^2. \tag{8}$$

Under H_0, and let z_i be a standard normal random variable, we can compute draws of T_{pred} as follows:

$$T_{pred} = \sum_{i=1}^{r}\frac{n_i}{s_i^{*2}}(\bar{y}_{i.}^* - \bar{y}_{i.})^2 - \frac{\left(\sum_{i=1}^{r}\frac{n_i}{s_i^{*2}}(\bar{y}_{i.}^* - \bar{y}_{i.})\right)^2}{\sum_{i=1}^{r}\frac{n_i}{s_i^{*2}}} \tag{9}$$

$$\begin{aligned} &= \sum_{i=1}^{r}\frac{n_i}{s_i^2\left(\frac{\chi^2_{n_i-1}}{n_i-1}\frac{(n_i-1)(n_i+1)}{n_i(n_i-3)}\right)}\left(z_i\sqrt{\frac{s_i^2(n_i-1)(n_i+1)}{n_i^2(n_i-3)}}\right)^2 \\ &- \frac{\left(\sum_{i=1}^{r}\frac{n_i}{s_i^2\left(\frac{\chi^2_{n_i-1}}{n_i-1}\frac{(n_i-1)(n_i+1)}{n_i(n_i-3)}\right)}\left(z_i\sqrt{\frac{s_i^2(n_i-1)(n_i+1)}{n_i^2(n_i-3)}}\right)\right)^2}{\sum_{i=1}^{r}\frac{n_i}{s_i^2\left(\frac{\chi^2_{n_i-1}}{n_i-1}\frac{(n_i-1)(n_i+1)}{n_i(n_i-3)}\right)}}. \end{aligned} \tag{10}$$

To approximate the generalized p-value $\Pr(T_{Pred} > T_{Obs})$, we take M draws of T_{Pred}, and calculate $\mathcal{P}_{pred} = \sum_{B=1}^{M} \mathrm{I}(T_{Pred} > T_{Obs})/M$.

3. PB and OB relationship for one-way heteroscedastic ANOVA

From a philosophical standpoint, the PB and OB tests are fundamentally coming from different places. The PB considers data as random and parameters as fixed, while OB considers data as fixed and parameters as random. Empirically, the PB and OB approaches tend to give similar results. Bayarri and Berger (2004) considered OB to be perhaps "the most promising route to the unification of Bayesian and frequentist statistics". Efron (2013) discussed relationship between the PB and OB, specifically demonstrating their near-equivalency for the problem of estimating the correlation parameter of a bivariate normal distribution. Efron (2012) showed the existence of a "Bayes/bootstrap" conversion factor for multidimensional exponential families. In this section, we establish the asymptotic equivalence between the PB and OB approaches of the overall mean test for one-way heteANOVA.

Theorem 1. *For the overall mean test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$ of the one way ANOVA model (1), the PB test statistic (5), OB test statistic (7), and the OBpred test statistic (9) are asymptotically equivalent to each other, i.e., the limiting distributions of the three statistics are the same as $n_i \to \infty$ for $i = 1, 2, \ldots, r$.*

Proof. Noting the relationship between the standard normal, chi-squared, and t distributions such as $z_i/\sqrt{\chi^2_{n_i-1}/(n_i-1)} \overset{d}{=} t_{n_i-1}$, we can rewrite T_{PB} in terms of draws from t_{n_i-1} distribution as follows:

$$T_{PB} = \sum_{i=1}^{r} t^2_{n_i-1} - \frac{\left(\sum_{i=1}^{r} \frac{\sqrt{n_i}}{s_i} t_{n_i-1} \cdot \sqrt{\frac{n_i-1}{\chi^2_{n_i-1}}}\right)^2}{\sum_{i=1}^{r} \frac{n_i(n_i-1)}{s_i^2 \chi^2_{n_i-1}}}.$$

Compare to T_{OB} in Eq. (7), we first want to show that $\chi^2_{n_i-1}/(n_i-1) \overset{a.s.}{\to} 1$.

By the strong law of large numbers, as $n_i \to \infty$ for $i = 1, 2, \ldots, r$,

$$\frac{\chi^2_{n_i-1}}{n_i-1} \overset{d}{=} \frac{\sum_{j=1}^{n_i-1} z_j^2}{n_i-1} \overset{a.s.}{\to} 1.$$

By Slutsky's theorem, T_{PB} and T_{OB} have the same limiting distribution.

Now consider OBpred test statistics (9). When $n_i \to \infty$ for $i = 1, 2, \ldots, r$,

$$g_i = \frac{n_i^2 - 1}{n_i^2 - 3n_i} = \frac{(n_i+1)(n_i-1)}{n_i(n_i-3)} \to 1.$$

A similar argument can be applied to prove that T_{pred} has the same limiting distribution as T_{PB}. Therefore, T_{PB}, T_{OB}, T_{pred} have the same limiting distributions. □

4. Simulation study

Theorem 1 shows that the PB and OB approaches for overall mean test for heteANOVA problem are asymptotically equivalent. In this section, we use a small simulation study to investigate the performance of the PB and OB methods in regarding type I error rates and power of the tests.

The simulation study was performed with factors: (1) number of levels $r = 3$; (2) population standard deviation $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$: various combinations; (3) significance level α: .01, .05 and .1; (4) group sizes $\mathbf{n} = (n_1, \ldots, n_3)$: small size with $\mathbf{n} = (3, 5, 7)$, medium size with $\mathbf{n} = (7, 10, 13)$, and large size with $\mathbf{n} = (21, 30, 39)$. Since the tests we consider are location-scale invariant, without loss of generality, we take H_0 as $\mu_1 = \mu_2 = \mu_3 = 0$, and take two alternatives as $\mu_1 = 0, \mu_2 = 1, \mu_3 = 2$, and $\mu_1 = 0, \mu_2 = 3, \mu_3 = 6$ for power calculation. The following gives detailed steps, and Tables 1 and 2 give the simulation results.

(1) For a given group size $\mathbf{n}$ and parameter configuration, generate $(\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \bar{y}_{3\cdot}, s_1^2, s_2^2, s_3^2)$ according to $\bar{y}_{i\cdot} \sim N(0, \sigma_i^2/n_i)$ and $s_i^2 \sim \sigma_i^2 \chi^2_{n_i-1}/(n_i-1)$, $i = 1, 2, 3$;

(2) For a given (n_1, n_2, n_3), $(\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \bar{y}_{3\cdot})$ and (s_1^2, s_2^2, s_3^2), compute T_{Obs} by (3);

(3) For $B = 1, \ldots, 5000$

(a) PB: generate $\bar{y}_{iB}$ and s^2_{iB} by (4), compute T_{PB} using (5)
if $T_{PB} > T_{Obs}$, set $I_{PB} = I(T_{PB} > T_{Obs}) = 1$, (end loop)
$\mathcal{P}_{PB} = \sum_{B=1}^{5000} I_{PB}/5000$ is a Monte Carlo (MC) estimate of the p-value of T_{PB}.

Table 1
Simulation result: **n** is a vector of unequal group sizes; $\boldsymbol{\sigma}$ is a vector of unequal variances; Numbers in Table are empirical significance levels.

n	σ	$\alpha = .01$			$\alpha = .05$			$\alpha = .1$		
		PB	OB	OBpred	PB	OB	OBpred	PB	OB	OBpred
(3,5,7)	(1,1,1)	.0155	.0025	.0000	0.0595	.0210	.0135	.0970	.0575	.0490
	(4,4,4)	.0135	.0015	.0000	.0485	.0220	.0135	.1045	.0585	.0495
	(1,2,4)	.0075	.0020	.0000	.0400	.0190	.0009	.0915	.0505	.0395
	(1,4,9)	.0120	.0015	.0000	.0500	.0250	.0085	.1090	.0770	.0425
	(4,2,1)	.0300	.0040	.0015	.0715	.0345	.0460	.1150	0.0910	.0825
	(9,4,1)	.0275	.0125	.0035	.0685	.0410	.0475	.1200	.0895	.1070
(7,10,13)	(1,1,1)	.0065	.0040	.0090	.0455	.0385	.0530	.1005	.0835	.1055
	(4,4,4)	.0115	.0040	.0120	.0515	.0435	.0415	.0975	.0725	.1085
	(1,2,4)	.0070	.0045	.0085	.0515	.0390	.0480	.0900	.0985	.0995
	(1,4,9)	.0075	.0075	.0105	.0460	.0340	.0500	.0975	.0975	.1055
	(4,2,1)	.0120	.0085	.0080	.0515	.0530	.0515	.1125	.0905	.0965
	(9,4,1)	.0110	.0115	.0145	.0505	.0505	.0530	.1025	.0985	.1135
(21,30,39)	(1,1,1)	.0085	.0115	.0090	.0500	.0465	.0430	.0975	.0820	.0985
	(4,4,4)	.0165	.0085	.0070	.0460	.0535	.0495	.0990	.0940	.0970
	(1,2,4)	.0095	.0090	.0065	.0540	.0510	.0510	.1135	.0980	.0980
	(1,4,9)	.0100	.0070	.0085	.0485	.0385	.0525	.0915	.0960	.0950
	(4,2,1)	.0110	.0070	.0140	.0575	.0500	.0590	.0950	.0920	.1045
	(9,4,1)	.0090	.0115	.0070	.0475	.0455	.0590	.1065	.1040	.0875

Table 2
Simulation result: significance level for the tests is .05; Numbers in Table are empirical statistical powers under the two alternatives $\mu_1 = 0, \mu_2 = 1, \mu_3 = 2$, and $\mu_1 = 0, \mu_2 = 3, \mu_3 = 6$.

n	σ	$\mu_1 = 0, \mu_2 = 1, \mu_3 = 2$			$\mu_1 = 0, \mu_2 = 3, \mu_3 = 6$		
		PB	OB	OBpred	PB	OB	OBpred
(3,5,7)	(1,1,1)	.5325	.3225	.2845	1.0000	.9990	1.0000
	(4,4,4)	.0820	.0310	.0300	.3345	.1725	.1390
	(1,2,4)	.1525	.0852	.0420	.8785	.7175	.5560
	(1,4,9)	.0705	.0480	.0255	.3560	.2695	.1295
	(4,2,1)	.1290	.0760	.0670	.6130	.5205	.5355
	(9,4,1)	.0855	.0615	.0605	.2120	.1430	.1415
(7,10,13)	(1,1,1)	.9480	.9300	.9335	1.0000	1.0000	1.0000
	(4,4,4)	.1325	.1005	.1240	.7415	.7045	.7525
	(1,2,4)	.3585	.3225	.3775	.9990	.9995	.9990
	(1,4,9)	.1235	.1170	.1210	.7415	.7375	.7535
	(4,2,1)	.2795	.2615	.2900	.9970	.9955	.9955
	(9,4,1)	.1050	.1060	.1215	.5835	.5815	.5880
(21,30,39)	(1,1,1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(4,4,4)	.3565	.3310	.3580	1.0000	1.0000	.9995
	(1,2,4)	.8825	.8750	.8740	1.0000	1.0000	1.0000
	(1,4,9)	.3530	.3490	.3500	.9985	.9990	.9995
	(4,2,1)	.8190	.8150	.8005	1.0000	1.0000	1.0000
	(9,4,1)	.2760	.2680	.2915	.9935	.9935	.9925

(b) OB: generate $\bar{y}_{i.} - \mu_i$ by (6), compute T_{OB} using (7)
if $T_{OB} > T_{Obs}$, set $I_{OB} = I(T_{OB} > T_{Obs}) = 1$, (end loop)
$\mathcal{P}_{OB} = \sum_{B=1}^{5000} I_{OB}/5000$ is a MC estimate of the p-value of T_{OB}.

(c) OBpred approach: generate $\bar{y}^*_{i.}|\boldsymbol{Y}$ and $s_i^{2*}|\boldsymbol{Y}$ according to Eq. (8), compute T_{pred} using (9)
if $T_{pred} > T_{Obs}$, set $I_{pred} = I(T_{pred} > T_{Obs}) = 1$, (end loop)
$\mathcal{P}_{pred} = \sum_{B=1}^{5000} I_{pred}/5000$ is a MC estimate of the p-value of T_{pred}.

(4) Repeat step (1) to step (3) 2000 times, calculate the proportion of rejections for the three cases when $\mathcal{P}_{PB}$, $\mathcal{P}_{OB}$ and $\mathcal{P}_{pred}$ are less than significance level α.

Table 1 reported empirical significance levels of the tests. It shows that when the group sizes are small such as $\mathbf{n} = (3, 5, 7)$, the empirical level from PB test is closer to the nominal level than the other two OB tests. When group sizes are medium such as $\mathbf{n} = (7, 10, 13)$ or large such as $\mathbf{n} = (21, 30, 39)$, all the three tests perform comparably and control the type I error reasonably well.

Table 2 reported power of the three tests under different settings given $\alpha = 0.05$. We can see that for fixed standard deviation, power increases when group size increases. For example, when fixing $\boldsymbol{\sigma} = (1, 1, 1)$, power for the PB test

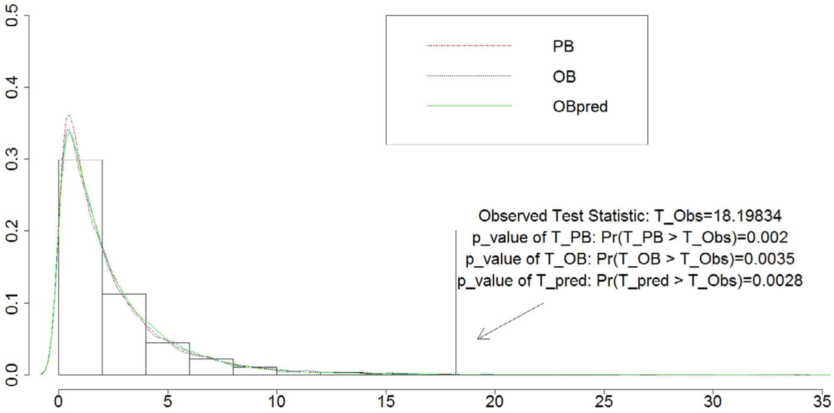

Fig. 1. PB, OB and OBpred Simulated Test Statistics Distribution.

increases from .5325 (small group) to .948 (medium group), and to 1.000 (large group). On the other hand, with fixed group size, power increases when standard deviation decreases. For example, the smallest standard deviation combination $\boldsymbol{\sigma} = (1, 1, 1)$ setting always achieves the highest power. For another example, consider standard deviation $\boldsymbol{\sigma} = (4, 4, 4)$ with group size of $\mathbf{n} = (3, 5, 7)$, and with alternative $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 2$, power of the three tests are as small as .082, .031 and .03. This is because $\boldsymbol{\sigma} = (4, 4, 4)$ are large compared to group means $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 2$, and group size is small, which reduce the statistical power dramatically. With the same group size and standard deviation combination, power under alternative $\mu_1 = 0$, $\mu_2 = 3$, $\mu_3 = 6$ is always greater than power under alternative $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 2$, because $\mu_1 = 0$, $\mu_2 = 3$, $\mu_3 = 6$ is further apart from the null hypothesis. In general, power from PB test is higher than the other two OB tests for small groups, and all the three tests perform comparably with similar statistical power for medium or large groups.

5. Example

To illustrate the testing procedure in a more practical light, we demonstrate the PB and OB approaches for a real one-way heteroscedastic ANOVA problem. The experiment consists of measuring insulin levels in rats a certain length of time (0, 30 and 60 min) after a fixed dose of insulin was injected into their portal vein (Erhardt et al., 2016). The experiment features groups of various sizes: a 0-min group (control group) with $n_1 = 12$, a 30-min group with $n_2 = 10$, and a 60-min group with $n_3 = 12$. A boxplot suggests that heteroscedasticity and outlier are both present. A Bartlett's test confirms the heteroscedasticity with a p-value of .011.

To conduct the PB and OB tests, first calculate the group means $(\bar{y}_1., \bar{y}_2., \bar{y}_3.) = (81.92, 172.90, 128.50)$, and sample standard errors $(s_1, s_2, s_3) = (27.74, 76.12, 49.72)$. Next, calculate the observed test statistic T_{Obs} using Eq. (3) as 18.198. Follow simulation step (3), we generated 5000 draws for T_{PB}, T_{OB} and T_{pred}, and obtained the generalized p-values $\mathcal{P}_{PB} = .0020$, $\mathcal{P}_{OB} = .0035$, and $\mathcal{P}_{pred} = .0028$. For example, $\mathcal{P}_{PB} = P[T_{PB} > T_{Obs}] = 10/5000 = .0020$. Hence, we reject H_0 and conclude that time effect is significant, i.e., the mean insulin levels in rats after a fixed dose of insulin was injected into their portal vein are significantly different for at least two time groups. Fig. 1 depicts the estimated distribution of T_{PB}, T_{OB} and T_{OBpred} under H_0, along with a black vertical line indicating the observed test statistic. We can see from Fig. 1 that the three density curves are very close to each other. For comparison purpose, a regular ANOVA is performed after a log transformation which is used to stabilize the unequal variance and to remove the outliers. The p-value of .0007459 shows a rejection of the equality of the group means as PB and OB tests do. One advantage of the PB and OB tests is that they are simple to use. Another advantage is that they are easy to interpret since there is no transformation needed even for unbalanced heteANOVA problem.

6. Conclusions

For one-way heteroscedastic ANOVA, the PB, OB, and OBpred approaches all control the type I error of the overall mean tests well and have reasonable statistical powers when group sizes are not small. We proved that the PB, OB and OBpred tests are asymptotically equivalent. Future work may investigate the robustness of the PB and OB tests to outliers, and extend current research to multi-way ANOVA problem, and to problems under special designs such as the randomized complete block design and split plot design.

CRediT authorship contribution statement

Guoyi Zhang: Conceptualization, Methodology, Software, Writing - review & editing. **Ronald Christensen:** Conceptualization, Methodology, Supervision. **John Pesko:** Software, Writing - original draft.

Acknowledgments

The authors thank the referees for helpful comments and constructive suggestions to improve the manuscript.

References

Aitchison, J., 1975. Goodness of prediction fit. Biometrika 62 (3), 547–554.
Bayarri, M.J., Berger, J.O., 2004. The interplay of Bayesian and frequentist analysis. Statist. Sci. 19 (1), 58–80.
Bayes, M., Price, M., 1683–1775. An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. Phil. Trans. 53 (1763), 370–418.
Brown, M., Forsythe, A., 1974. The small sample behavior of some statistics which test the equality of several means. Technometrics 16, 129–132.
Efron, B., 2012. Bayesian inference and the parametric bootstrap. Ann. Appl. Stat. 6 (4), 1971.
Efron, B., 2013. A 250-year argument: belief, behavior, and the bootstrap. Bull. Amer. Math. Soc. 50 (1), 129–146.
Erhardt, E., Bedrick, E., Schrader, R., 2016. Lecture Notes for Advanced Data Analysis 2 (ADA2) Stat 428/528. University of New Mexico, Albuquerque, NM.
Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Proc. R. Soc. A 186 (1007), 453–461.
Krishnamoorthy, K., Lu, F., Mathew, T., 2007. A parametric bootstrap approach for anova with unequal variances: Fixed and random models. Compu. Stat. Data Anal. 51, 5731–5742.
Weerahandi, S., 1995. ANOVA under unequal error variances. Biometrics 51, 589–599.
Welch, B., 1951. On the comparison of several mean values: An alternative approach. Biometrica 38, 330–336.
Yiğit, E., Gokpinar, F., 2010. A simulation study on tests for one-way anova under the unequal variance assumption. Commun. Fac. Sci. Univ. A 59 (2), 15–34.
Zhang, G., 2015a. A parametric bootstrap approach for one-way ANOVA under unequal variances with unbalanced data. Comm. Statist. Simulation Comput. 44 (4), 827–832.
Zhang, G., 2015b. Simultaneous confidence intervals for pairwise multiple comparisons in a two-way unbalanced design with unequal variances. J. Stat. Comput. Simul. 85 (13), 2727–2735.