Efficiency Comparison of Statistic for Testing Three Population Means in case of Homogeneity and Heterogeneity of Variance

Jirapa Khomduean[1*] and Autcha Araveeporn[2]

[1]Department of Appiled Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung, Ladkrabang, Bangkok 10520, Thailand. E-mail: jirapadoublea@gmail.com

[2]Department of Appiled Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung, Ladkrabang, Bangkok 10520, Thailand. E-mail: kaautcha@hotmail.com

Abstract

The objective of this research is to compare the efficiency comparison of test statistic for testing three population means in case of homogeneity and heterogeneity of variance. In analysis of variance using F-test (F), Brown–Forsythe's test (BF), modified Brown–Forsythe's test (MBF), and Welch's test (W) are test statistic for computing probability of type I error and power of a test. The Bradley's is a criterion to control the probability of type I error at 0.01 and 0.05 significance level. The data of this research is simulated by using the Monte Carlo technique and each case is replicated 1,000 times from normal and gamma distributions based on equal and unequal sample sizes. For homogeneity of variance, most of F, BF, MBF are test statistical efficiency with respect to probability of type I error and power of the test in normal distribution. For gamma distribution, F, and BF perform very satisfactorily in almost all cases at significance level 0.01, while F and W exhibit a good power of a test at significance level 0.05. In case of heterogeneity of variance, W performs better than F, BF, and MBF at significance levels 0.01. but F, BF, and MBF are reasonable working as good as W at significance levels 0.05.

Keywords : Heterogeneity of variance, Homogeneity of variance, Power of a test, Type I error

1. Introduction

General research is to study the population parameter but it is very difficult to collect entire target population because of population size, a lot of cost, and waste of time. To overcome this problem, sampling technique is a process of selecting a sample from a target population and obtains the estimator which referred to the population parameter. The estimation and hypothesis test is a part of statistical inference that concerned the estimator and population parameter. The strong knowledge about the hypothesis test, sampling techniques, and the test statistic makes the output of research which are reliability, accuracy, and precision. However, the researcher should to know the assumption of each test statistic and characteristic of data before the process of data analysis. If an inappropriate assumption is used, it is possible to produce misleading conclusion.

In Analysis of Variance (ANOVA), is the F test used for testing equal means when the populations are more than two groups. In the conventional ANOVA is the assumptions that based on the random sample is drawn from normal distribution, all populations have a common variances, and the sample data is independent. If the observation of sample data may not be according to the assumption, ANOVA does not give accurate results. So there are the another test statistic under the violation of assumption of ANOVA that studied extensively such as Brown–Forsythe's test (BF), modified Brown–Forsythe's test (MBF), and Welch's test (W) etc.

Brown and Forsythe [2] proposed the BF test for testing the equal means among more than two independent populations in case of heterogeneity of variance or called Brown – Forsythe's Test. Mehrotra [3] adapted Brown–Forsythe's test (BF) to new test statistic or called Modified Brown–Forsythe's test (MBF) which weighted proportion values in term of degree of freedom. Welch [4] modified the original Cochran test by weighting values in test statistic for testing difference of means especially the unequal sample sizes in normal distribution. So the BF, MBF, and Welch are widely used methods for testing hypothesis by the weighted F, and approximated new degree of freedom.

At this time, the researchers use computer to program for data analysis such as SPSS, Minitab, and SAS etc. The popular test statistic is considered for testing means among more than two populations in SPSS program such as analysis of variance using F-test (F), Brown–Forsythe's test (BF), and Welch's test (W).

For this reason, we interested to compare four test statistics : analysis of variance using F-test (F), Brown–Forsythe's test (BF), modified Brown–Forsythe's test (MBF), and Welch's test (W) for homogeneity and heterogeneity of variance. Normality is one of the important assumptions of ANOVA. In this research, we applied to compare the normal distribution and gamma distribution (non-normal) of equal and unequal sample sizes. The R Program is used for the simulation and data analysis.

The objective of this research is to consider the probability of type I error depended on the criteria of Bradley, and to compare a power of a test of four test statistics. The maximum of power of a test is presented the performance of these test statistics.

2. Research Methodology

In this research, we focus to test the different means of k populations and given observations x_{ij} $\{x_{ij} : i = 1,2,..,k; j = 1,2,...,n_i\}$ from the i^{th} population with mean μ_i and variance σ_i^2. The null hypothesis is $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ and the alternative hypothesis is $H_1 : \mu_i \neq \mu_j$ for at least one $i \neq j$.

2.1 Description of statistical testing

2.1.1 Analysis of Variance (ANOVA) : F

Analysis of variance (ANOVA) [1] is a collection of statistical models used to analyze the differences among group means. By having assumptions as follows:

Each group sample is drawn from a normally distribution population.

All populations have a common variances.

All samples are drawn independently of each other.

Within each sample, the observations are sampled randomly and independently of each other.

Test statistic is

$$F_{cal} = \frac{\sum_{i=1}^{k} n_i (x_{i\bullet} - \bar{x})^2 \Big/ (k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 \Big/ (n-k)},$$

where $n = \sum_{i=1}^{k} n_i$ and $x_{i\bullet} = \sum_{j=1}^{n_i} x_{ij}$.

The null hypothesis H_0 is rejected at $F_{cal} > F_{\alpha,k-1,N-k}$, where $F_{\alpha,k-1,N-k}$ is an upper critical value of the F-distribution with degree of freedom $k-1$ and $N-k$ at a significance levels of α .

2.1.2 Brown – Forsythe's Test : BF

Brown and Forsythe (1974) [2] was a test statistic for testing the difference of means among more than two independent populations in case of heterogeneity of variance.

Test statistic is

$$BF = \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^{k}\left(1-\frac{n_i}{N}\right)S_i^2},$$

where $\bar{x}_i = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{x_{ij}}{n_i}$, $\bar{x} = \sum_{i=1}^{k}\frac{n_i\bar{x}_i}{N}$, and $S_i^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{\left(x_{ij}-\bar{x}_i\right)^2}{\left(n_i-1\right)}$.

The null hypothesis H_0 is rejected at $BF > F_{\alpha;k-1,f^*}$, where $F_{\alpha;k-1,f^*}$ is an upper critical value of the F-distribution with degree of freedom $k-1$ and f^* at a significance levels of α. The definitions of f^* and v_i are written as

$$f^* = \left[\sum_{i=1}^{k}\frac{v_i^2}{(n_i-1)}\right]^{-1},$$

and

$$v_i = \left(1-\frac{n_i}{N}\right)S_i^2 \Big/ \left[\sum_{i=1}^{k}\left(1-\frac{n_i}{N}\right)S_i^2\right].$$

2.1.3 Modified Brown – Forsythe's Test : MBF

Mehrotra (1997) [3] modified the original Brown-Forsythe [1] by weighting proportion values in term of degree of freedom as follows :

Test statistic is

$$MBF = \frac{\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^{k}\left(1-\frac{n_i}{N}\right)S_i^2}.$$

The null hypothesis H_0 is rejected at $MBF > F_{\alpha;v_2^*,v_1^*}$, where $F_{\alpha;v_2^*,v_1^*}$ is an upper critical value of the F- distribution with degree of freedom v_1^* and v_2^* at a significance levels of α. The definitions of v_1^* and v_2^* are written as

$$v_1^* = \frac{\left[\sum_{i=1}^{k}\left(\frac{1-n_i}{N}\right)S_i^2\right]^2}{\sum_{i=1}^{k}\frac{\left(\frac{1-n_i}{N}\right)^2 S_i^4}{n_i-1}},$$

and

$$v_2^* = \frac{\left[\sum_{i=1}^{k}\left(\frac{1-n_i}{N}\right)S_i^2\right]^2}{\sum_{i=1}^{k}S_i^4 + \left[\frac{\sum_{i=1}^{k}n_iS_i^2}{N}\right]^2 - 2\frac{\sum_{i=1}^{k}n_iS_i^4}{N}}.$$

The BF statistic is the same as the MBF test, but the MBF test has $F_{\alpha;v_2^*,v_1^*}$ distribution with v_1^* and v_2^* degree of freedom.

2.1.4 Welch's Test : W

Welch (1947, 1951) [4] modified the original Cochran test by weighting values in test statistic to test the difference of means of k populations. So it made consistent in Welch's test when the observations are unequal in normal distribution.

Test statistic is

$$W_0 = \frac{\sum_{i=1}^{k} w_i\left[(\bar{x}_i - \tilde{x})^2/(k-1)\right]}{1+\frac{2(k-2)}{(k^2-1)}\left[\sum_{i=1}^{k}\left(1-\frac{w_i}{u}\right)^2 \Big/ n_i - 1\right]},$$

where $w_i = n_i / S_i^2$, $u = \sum_{i=1}^{k} w_i$, and $\tilde{x} = \sum_{i=1}^{k} w_i \bar{x}_i / u$.

The null hypothesis H_0 is rejected at $W_0 > F_{\alpha;k-1,f}$, where $F_{\alpha;k-1,f}$ is an upper critical value of the F-distribution with degree of freedom $k-1$ and f at a significance levels of α, and

$$f = \left[\left(3\Big/(k^2-1)\sum_{i=1}^{k}\left(1-\frac{w_i}{u}\right)^2\right)\times\frac{1}{n_i-1}\right]^{-1}.$$

2.2 Research design

The scope of research as follows :

(1) Let's consider 3 populations.

(2) Let sample sizes in two cases, contain equal sample sizes and unequal sample sizes. Details of

sample sizes in this study is shown in Table 1.

Table 1 The 3 population and sample size patterns.

	Sample sizes	
	Equal	**Unequal**
Small	10:10:10	5:10:15
Medium	30:30:30	25:30:35
Large	50:50:50	45:50:55

(3) Let two distributions as follows :

- Normal Distribution

The probability density of the normal distribution is

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0,$$

where the mean is $E(X) = \mu$, and variance is $Var(X) = \sigma^2$.

- Gamma Distribution

The probability density of the gamma distribution is

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} & ,x>0,\ \alpha>0, \beta>0 \\ 0 & ,x \text{ otherwise} \end{cases}$$

where the mean is $E(X) = \alpha\beta$, and variance is $Var(X) = \alpha\beta^2$.

(4) In case of homogeneity of variance, let the variances are set as 2, 4, 8, and 16 respectively.

(5) In case of heterogeneity of variance, let the different population variance ratios by non-centrality parameter (ϕ) [5], this ϕ is a measure of the different population variance. The formula is defined by

$$\phi = \frac{\sqrt{\sum_{i=1}^{k}(\sigma_i^2 - \bar{\sigma}^2)^2 / k}}{\sigma_1^2},$$

where

σ_1^2 is the population variance with the lowest,

σ_i^2 is the population variance with i^{th} group; $i = 1,2,...,k$,

$\bar{\sigma}^2$ is the mean of population variance with k group.

The details are given in Table 2 as follows :

Table 2 The different population variance ratios by non-centrality parameter (ϕ)

Levels	Ratios	ϕ
Slightly ($0<\phi<1.5$)	1.5:3:4.5	1
Moderately ($1.5\leq\phi<3.0$)	1.5:3.6:7.2	1.9
Highly ($\phi\geq 3.0$)	1.5:12:18	5.6

(6) Let the significance levels (α) are defined in two level as 0.01 and 0.05.

(7) The R program version 3.2.2 is used to simulated data and data analysis in 1,000 replications.

(8) The probability of type I error is computed from the proportion of rejection of null hypothesis when the null hypothesis is truth by requiring the same mean of each distribution. The probability of type I error is compared under the criterion of Bradley [6] according to the significance level. At the significance level α = 0.01 and 0.05, the range of the criterion is [0.005,0.015], and [0.025,0.075].

(9) Power of the test is computed from the proportion of rejection of null hypothesis when the null hypothesis is false by requiring the different mean of each distribution. The mean of distribution is determined the ratio of 1 : 2 : 3.

(10) The maximum power of a test is called the best test statistic when the test statistics can control the probability of type I error in a range under criterion of Bradley.

3. Research Results and Discussions

The results of this research are divided into two parts : homogeneity of variance and heterogeneity of variance. The probability of type I error and power of a test are computed from the test statistic such as analysis of variance using F-test (F), Brown–Forsythe's test (BF), modified Brown–Forsythe's test (MBF), and Welch's test (W).

3.1 The homogeneity of variance

3.1.1 Probability of type I error

Table 3-4 show the probability of type I error from normal and gamma distribution at the significance level at 0.01.

Table 3 The probability of type I error of test statistic from normal distribution at significance level 0.01.

Sample sizes (n_1,n_2,n_3)	Test statistics	Variances (σ^2)			
		6	12	18	36
(10,10,10)	F	0.010*	0.004	0.013*	0.011*
	BF	0.009*	0.004	0.013*	0.009*
	MBF	0.009*	0.004	0.013*	0.009*
	W	0.007*	0.008*	0.014*	0.007*
(30,30,30)	F	0.011*	0.016	0.013*	0.006*
	BF	0.009*	0.016	0.013*	0.006*
	MBF	0.009*	0.016	0.013*	0.006*
	W	0.009*	0.016	0.014*	0.007*
(50,50,50)	F	0.013*	0.011*	0.016	0.017
	BF	0.013*	0.011*	0.016	0.017
	MBF	0.013*	0.011*	0.016	0.017
	W	0.010*	0.010*	0.014*	0.016
(5,10,15)	F	0.016	0.006*	0.011*	0.012*
	BF	0.014*	0.005*	0.012*	0.009*
	MBF	0.012*	0.005*	0.011*	0.009*
	W	0.019	0.006*	0.005*	0.008*
(25,30,35)	F	0.014*	0.004	0.006*	0.012*
	BF	0.014*	0.005*	0.009*	0.012*
	MBF	0.014*	0.005*	0.009*	0.012*
	W	0.015*	0.005*	0.007*	0.013*
(45,50,55)	F	0.012*	0.009*	0.011*	0.006*
	BF	0.012*	0.009*	0.011*	0.006*
	MBF	0.012*	0.009*	0.011*	0.006*
	W	0.006*	0.008*	0.013*	0.004

* Test statistic is shown the probability of type I error between (0.005,0.015).

From table 3, the results appear following :

For $\sigma^2 = 6$, BF and MBF can control the probability of type I error in all cases. The F and W can control the probability of type I error in almost all cases except sample sizes as (5,10,15).

For $\sigma^2 = 12$, W can control the probability of type I error in almost all cases except sample sizes as (30,30,30). The BF and MBF can control the probability of type I error in almost all cases except sample sizes as (10,10,10), and (30,30,30). The F can control the probability of type I error in almost all cases except sample sizes as (50,50,50), (5,10,15), and (45,50,55).

For $\sigma^2 = 18$, W can control the probability of type I error in all cases. The F, BF, and MBF can control the probability of type I error in almost all cases except sample sizes as (50,50,50).

For $\sigma^2 = 36$, F, BF, MBF, and W can control the probability of type I error in almost all cases except sample sizes as (50,50,50).

Table 4 The probability of type I error of test statistic from gamma distribution at significance level 0.01.

Sample sizes (n_1,n_2,n_3)	Test statistics	Variances (σ^2)			
		6	12	18	36
(10,10,10)	F	0.011*	0.012*	0.007*	0.005*
	BF	0.009*	0.008*	0.005*	0.002
	MBF	0.009*	0.008*	0.004	0.002
	W	0.009*	0.008*	0.005*	0.005*
(30,30,30)	F	0.008*	0.015*	0.005*	0.013*
	BF	0.008*	0.015*	0.004	0.012*
	MBF	0.008*	0.015*	0.004	0.012*
	W	0.007*	0.016	0.008*	0.012*
(50,50,50)	F	0.010*	0.015*	0.010*	0.005*
	BF	0.010*	0.014*	0.010*	0.004
	MBF	0.010*	0.014*	0.010*	0.004
	W	0.010*	0.014*	0.009*	0.007*
(5,10,15)	F	0.012*	0.006*	0.011*	0.011*
	BF	0.009*	0.007*	0.008*	0.008*
	MBF	0.008*	0.005*	0.007*	0.003
	W	0.016	0.018	0.013*	0.020
(25,30,35)	F	0.011*	0.003	0.006*	0.008*
	BF	0.013*	0.003	0.006*	0.009*
	MBF	0.013*	0.003	0.005*	0.008*
	W	0.013*	0.006*	0.007*	0.010*
(45,50,55)	F	0.012*	0.007*	0.005*	0.010*
	BF	0.011*	0.007*	0.005*	0.008*
	MBF	0.011*	0.007*	0.005*	0.008*
	W	0.011*	0.009*	0.010*	0.013*

* Test statistic is shown the probability of type I error between (0.005,0.015).

From table 4, the results appear following :

For $\sigma^2=6$, F, BF, and MBF can control the probability of type I error in all cases. The W can control the probability of type I error in almost all cases except sample sizes as (5,10,15).

For $\sigma^2=12$, F, BF, and MBF can control the probability of type I error in almost all cases except sample sizes as (25,30,35). The W can control the probability of type I error in almost all cases except sample sizes as (30,30,30), and (5,10,15).

For $\sigma^2=18$, F and W can control the probability of type I error in all cases. The BF can control the probability of type I error in almost all cases except sample sizes as (30,30,30). The MBF can control the probability of type I error in almost all cases except sample sizes as (10,10,10), and (30,30,30).

For $\sigma^2=36$, F and W can control the probability of type I error in all cases. The BF and MBF can control the probability of type I error in almost all cases except sample sizes as (30,30,30), (25,30,35), and (45,50,55).

Table 5-6 show the probability of type I error from normal and gamma distribution at the significance level at 0.05.

Table 5 The probability of type I error of test statistic from normal distribution at significance level 0.05.

Sample sizes (n_1,n_2,n_3)	Test statistics	Variances (σ^2)			
		6	12	18	36
(10,10,10)	F	0.057*	0.050*	0.045*	0.046*
	BF	0.056*	0.049*	0.044*	0.045*
	MBF	0.056*	0.049*	0.044*	0.045*
	W	0.054*	0.046*	0.049*	0.046*
(30,30,30)	F	0.045*	0.048*	0.048*	0.047*
	BF	0.044*	0.048*	0.048*	0.046*
	MBF	0.044*	0.048*	0.048*	0.046*
	W	0.046*	0.056*	0.046*	0.045*
(50,50,50)	F	0.048*	0.046*	0.059*	0.053*
	BF	0.048*	0.046*	0.059*	0.053*
	MBF	0.048*	0.046*	0.059*	0.053*
	W	0.045*	0.043*	0.061*	0.060*
(5,10,15)	F	0.063*	0.045*	0.052*	0.051*
	BF	0.064*	0.047*	0.047*	0.048*
	MBF	0.063*	0.046*	0.046*	0.048*
	W	0.061*	0.045*	0.053*	0.052*
(25,30,35)	F	0.048*	0.052*	0.040*	0.045*
	BF	0.047*	0.052*	0.043*	0.044*
	MBF	0.047*	0.052*	0.043*	0.044*
	W	0.045*	0.058*	0.046*	0.042*
(45,50,55)	F	0.039*	0.050*	0.061*	0.039*
	BF	0.037*	0.051*	0.061*	0.038*
	MBF	0.037*	0.051*	0.061*	0.038*
	W	0.040*	0.051*	0.059*	0.030*

*** Test statistic is shown the probability of type I error between (0.025,0.075) .**

Table 6 The probability of type I error of test statistic from gamma distribution at significance level 0.05.

Sample sizes (n_1,n_2,n_3)	Test statistics	Variances (σ^2)			
		6	12	18	36
(10,10,10)	F	0.047*	0.044*	0.044*	0.036*
	BF	0.045*	0.039*	0.040*	0.026*
	MBF	0.045*	0.036*	0.038*	0.026*
	W	0.041*	0.053*	0.040*	0.038*
(30,30,30)	F	0.044*	0.053*	0.047*	0.053*
	BF	0.044*	0.053*	0.047*	0.052*
	MBF	0.044*	0.053*	0.047*	0.050*
	W	0.048*	0.055*	0.056*	0.061*
(50,50,50)	F	0.042*	0.048*	0.052*	0.049*
	BF	0.041*	0.048*	0.052*	0.049*
	MBF	0.041*	0.048*	0.052*	0.049*
	W	0.041*	0.046*	0.057*	0.060*
(5,10,15)	F	0.053*	0.042*	0.052*	0.052*
	BF	0.041*	0.047*	0.055*	0.058*
	MBF	0.040*	0.042*	0.048*	0.047*
	W	0.060*	0.066*	0.058*	0.075*
(25,30,35)	F	0.048*	0.036*	0.043*	0.047*
	BF	0.045*	0.035*	0.044*	0.049*
	MBF	0.045*	0.035*	0.044*	0.048*
	W	0.044*	0.047*	0.041*	0.066*
(45,50,55)	F	0.049*	0.041*	0.052*	0.045*
	BF	0.050*	0.041*	0.052*	0.046*
	MBF	0.050*	0.041*	0.052*	0.046*
	W	0.049*	0.048*	0.053*	0.050*

* Test statistic is shown the probability of type I error between (0.025,0.075) .

From table 5 and 6, the results appear that all test statistics can control the probability of type I error in all cases of normal distribution and gamma distribution.

3.1.2 Power of the test

Table 7-8 show the power of the test from normal and gamma distribution at the significance level 0.01.

Table 7 The highest power of a test of test statistic at significance level 0.01.

Distribution	Variances (σ^2)	Sample sizes (n_1,n_2,n_3)					
		(10,10,10)	(30,30,30)	(50,50,50)	(5,10,15)	(25,30,35)	(45,50,55)
Normal	6	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	BF,MBF	F,BF, MBF,W	F,BF, MBF,W
	12	W	-	F,BF, MBF,W	F	BF,MBF, W	F,BF, MBF,W
	18	F,BF, MBF	F,BF, MBF,W	W	F	F,BF, MBF,W	F,BF, MBF,W
	36	F	F,BF, MBF,W	-	F	F,BF, MBF,W	F,BF, MBF
Gamma	6	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF	F,BF, MBF,W	F,BF, MBF,W
	12	F,BF	F,BF, MBF	F,BF, MBF,W	F	W	F,BF, MBF,W
	18	F	F,W	F,BF, MBF,W	F	F,BF, MBF,W	F,BF, MBF,W
	36	F	F,BF, MBF	F,W	F	F,BF, MBF,W	F,BF, MBF,W

From table 7, the results from normal distribution appear following :

- F shows the highest power of a test in almost all cases except on $\sigma^2=6$ and sample sizes as (5,10,15), on $\sigma^2=12$ and sample sizes as (10,10,10), (30,30,30), and (25,30,35), on $\sigma^2=$ 18 and 36, and sample sizes as (50,50,50).
- BF and MBF show the highest power of a test in almost all cases except on $\sigma^2=12$ and sample sizes as (10,10,10), (30,30,30), and (25,30,35), on $\sigma^2=18$ and sample sizes as (10,10,10), (50,50,50), and (5,10,15), on $\sigma^2=36$ and sample sizes as (50,50,50), and (5,10,15).
- W shows the highest power of a test in almost all cases except on $\sigma^2=6$ and sample sizes as (5,10,15), on $\sigma^2=12$ and sample sizes as (30,30,30), and (5,10,15), on $\sigma^2=18$ and sample sizes as (10,10,10), and (5,10,15), on $\sigma^2=36$ and sample sizes as (10,10,10), (50,50,50), (5,10,15), and (45,50,55).

The results from gamma distribution appear following :

- F shows the highest power of a test in almost all cases except on $\sigma^2=12$ and sample sizes as (25,30,35).
- BF shows the highest power of a test in almost all cases except on $\sigma^2=12$ and sample sizes as (5,10,15), and (25,30,35), on $\sigma^2=$ 18 and sample sizes as (10,10,10), (30,30,30), and (5,10,15), on $\sigma^2=36$ and sample sizes as (10,10,10), (50,50,50), and (5,10,15).
- MBF shows the highest power of a test in almost all cases except on $\sigma^2=12$ and sample sizes as (10,10,10), (5,10,15), and (25,30,35), on $\sigma^2=18$ and sample sizes as (10,10,10), (30,30,30), and (5,10,15), on $\sigma^2=36$ and sample sizes as (10,10,10), (50,50,50), and (5,10,15).
- W shows the highest power of a test in almost all cases except on $\sigma^2=6$ and sample sizes as (5,10,15), on $\sigma^2=12$ and sample sizes as (10,10,10), (30,30,30), and (5,10,15), on $\sigma^2=18$ and sample sizes as (10,10,10), and (5,10,15), on $\sigma^2=36$ and sample sizes as (10,10,10), (30,30,30), and (5,10,15).

Table 8 The highest power of a test of test statistic at significance level 0.05.

Distribution	Variances (σ^2)	Sample sizes (n_1,n_2,n_3) (10,10,10)	(30,30,30)	(50,50,50)	(5,10,15)	(25,30,35)	(45,50,55)
normal	6	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W
	12	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,W	F,BF, MBF,W	F,BF, MBF,W
	18	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F	F,BF, MBF,W	F,BF, MBF,W
	36	F	F,BF, MBF,W	F,BF, MBF,W	F	F,BF, MBF,W	F,BF, MBF,W
gamma	6	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W
	12	W	F,BF, MBF,W	F,BF, MBF,W	F,W	F,BF, MBF,W	F,BF, MBF,W
	18	F	F,BF, MBF,W	F,BF, MBF,W	F	F,BF, MBF,W	F,BF, MBF,W
	36	F	F,BF, MBF,W	F,BF, MBF,W	F	F,BF, MBF,W	F,BF, MBF,W

From table 8, the results from normal distribution appear following :

- F shows the highest power of a test in all cases.
- W shows the highest power of a test in almost all cases except on $\sigma^2 = 18$ and sample sizes as (5,10,15), on $\sigma^2 = 36$ and sample sizes as (10,10,10), and (5,10,15).
- BF and MBF show the highest power of a test in almost all cases except on $\sigma^2 = 12$ and 18 and sample sizes as (5,10,15), on $\sigma^2 = 36$ and sample sizes as (10,10,10), and (5,10,15).

The results from gamma distribution appear following :

- F shows the highest power of a test in all cases.
- W shows the highest power of a test in almost all cases except on $\sigma^2 = 18$ and 36 and sample sizes as (10,10,10), and (5,10,15).
- BF and MBF show the highest power of a test in almost all cases except on $\sigma^2 =$ 12, 18 and 36 and sample sizes as (10,10,10), and (5,10,15).

3.2 The Heterogeneity of Variance

3.2.1 Probability of Type I Error

Table 9-10 show the probability of type I error from normal and gamma distribution at the significance level at 0.01.

Table 9 The probability of type I error of test statistic from normal distribution at significance level 0.01.

Sample sizes (n_1,n_2,n_3)	Test statistics	Non-Centrality Parameters (ϕ)		
		1	1.9	5.6
(10,10,10)	F	0.011*	0.016	0.018
	BF	0.011*	0.014*	0.016
	MBF	0.011*	0.014*	0.014*
	W	0.010*	0.012*	0.009*
(30,30,30)	F	0.018	0.018	0.016
	BF	0.016	0.018	0.015*
	MBF	0.016	0.018	0.014*
	W	0.012*	0.014*	0.008*
(50,50,50)	F	0.011*	0.020	0.017
	BF	0.011*	0.019	0.016
	MBF	0.011*	0.019	0.016
	W	0.010*	0.012*	0.013*
(5,10,15)	F	0.004	0.004	0.004
	BF	0.009*	0.019	0.013*
	MBF	0.009*	0.014*	0.006*
	W	0.005*	0.007*	0.008*
(25,30,35)	F	0.004	0.003	0.013*
	BF	0.008*	0.006*	0.017
	MBF	0.008*	0.006*	0.017
	W	0.006*	0.007*	0.017
(45,50,55)	F	0.007*	0.014*	0.015*
	BF	0.011*	0.020	0.016
	MBF	0.011*	0.020	0.016
	W	0.008*	0.010*	0.011*

* Test statistic is shown the probability of type I error between (0.005,0.015).

From table 9, the results appear following :

For $\phi = 1$, W can control the probability of type I error in all cases. The BF and MBF can control the probability of type I error in almost all cases except sample sizes as (30,30,30). The F can control the probability of type I error in almost all cases when sample sizes are (10,10,10), (50,50,50), and (45,50,55).

For $\phi = 1.9$, W can control the probability of type I error in all cases. The MBF can control the probability of type I error in almost all cases when sample sizes are (10,10,10), (5,10,15), and (25,30,35). The BF can control the probability of type I error in almost all cases especially sample sizes as (10,10,10), and (25,30,35). The F can control the probability of type I error in almost all cases at sample sizes as (45,50,55).

For $\phi = 5.6$, W can control the probability of type I error in almost all cases except sample sizes as (25,30,35). The MBF can control the probability of type I error in almost all cases when sample sizes are (10,10,10), (30,30,30), and (5,10,15). The BF can control the probability of type I error in almost case especially sample sizes as (30,30,30), and (5,10,15). The F can control the probability of type I error in almost all cases especially sample sizes are (25,30,35), and (45,50,55).

Table 10 The probability of type I error of test statistic from gamma distribution at significance level 0.01.

Sample sizes (n_1,n_2,n_3)	Test statistics	Non-Centrality Parameters (ϕ)		
		1	1.9	5.6
(10,10,10)	F	0.019	0.021	0.014*
	BF	0.018	0.019	0.012*
	MBF	0.017	0.018	0.012*
	W	0.015*	0.015*	0.024
(30,30,30)	F	0.010*	0.015*	0.020
	BF	0.009*	0.014*	0.019
	MBF	0.009*	0.011*	0.015*
	W	0.008*	0.010*	0.015*
(50,50,50)	F	0.012*	0.018	0.020
	BF	0.012*	0.017	0.020
	MBF	0.012*	0.016	0.018
	W	0.013*	0.015*	0.027
(5,10,15)	F	0.007*	0.006*	0.003
	BF	0.011*	0.008*	0.007*
	MBF	0.007*	0.006*	0.004
	W	0.013*	0.012*	0.012*
(25,30,35)	F	0.009*	0.008*	0.014*
	BF	0.013*	0.012*	0.020
	MBF	0.013*	0.012*	0.012*
	W	0.007*	0.009*	0.020
(45,50,55)	F	0.012*	0.013*	0.013*
	BF	0.014*	0.015*	0.017
	MBF	0.014*	0.015*	0.014*
	W	0.010*	0.012*	0.015*

* Test statistic is shown the probability of type I error between (0.005,0.015).

From table 10, the results appear following :

For $\phi=1$, W can control the probability of type I error in all cases. The F BF and MBF can control the probability of type I error in almost all cases except sample sizes as (10,10,10).

For $\phi=1.9$, W can control the probability of type I error in all cases. The F BF and MBF can control the probability of type I error in almost all cases except sample sizes as (10,10,10), and (50,50,50).

For $\phi=5.6$, MBF can control the probability of type I error in almost all cases except sample sizes as (50,50,50), and (5,10,15). The F can control the probability of type I error in almost all cases when sample sizes as (10,10,10), (25,30,35), and (45,50,55). The W can control the probability of type I error in almost all cases when sample sizes as (30,30,30), (5,10,15), and (45,50,55). The BF can control the probability of type I error in almost all cases especially sample sizes are (10,10,10), and (5,10,15).

Table 11-12 show the probability of type I error from normal and gamma distribution at the significance level at 0.05.

Table 11 The probability of type I error of test statistic from normal distribution at significance level 0.05.

Sample sizes (n_1,n_2,n_3)	Test statistics	Non-Centrality Parameters (ϕ)		
		1	1.9	5.6
(10,10,10)	F	0.051*	0.054*	0.061*
	BF	0.050*	0.046*	0.052*
	MBF	0.048*	0.042*	0.048*
	W	0.050*	0.052*	0.055*
(30,30,30)	F	0.058*	0.055*	0.053*
	BF	0.058*	0.054*	0.050*
	MBF	0.058*	0.054*	0.050*
	W	0.057*	0.049*	0.048*
(50,50,50)	F	0.046*	0.064*	0.067*
	BF	0.045*	0.064*	0.067*
	MBF	0.045*	0.064*	0.067*
	W	0.043*	0.059*	0.047*
(5,10,15)	F	0.021	0.026*	0.018
	BF	0.046*	0.050*	0.051*
	MBF	0.039*	0.045*	0.038*
	W	0.045*	0.046*	0.040*
(25,30,35)	F	0.047*	0.037*	0.049*
	BF	0.059*	0.051*	0.060*
	MBF	0.059*	0.048*	0.060*
	W	0.061*	0.043*	0.056*
(45,50,55)	F	0.046*	0.060*	0.056*
	BF	0.056*	0.075*	0.071*
	MBF	0.056*	0.075*	0.071*
	W	0.050*	0.058*	0.056*

* Test statistic is shown the probability of type I error between (0.025,0.075).

From table 11, the results appear following :

For $\phi = 1$ and 5.6, BF MBF and W can control the probability of type I error in all cases. The F can control the probability of type I error in almost all cases except sample sizes as (5,10,15).

For $\phi = 1.9$, the results appear that all test statistics can control the probability of type I error in all cases.

Table 12 The probability of type I error of test statistic from gamma distribution at significance level 0.05.

Sample sizes (n_1,n_2,n_3)	Test statistics	Non-Centrality Parameters (ϕ)		
		1	1.9	5.6
(10,10,10)	F	0.067*	0.058*	0.056*
	BF	0.063*	0.052*	0.045*
	MBF	0.061*	0.050*	0.041*
	W	0.060*	0.054*	0.072*
(30,30,30)	F	0.060*	0.056*	0.064*
	BF	0.059*	0.054*	0.063*
	MBF	0.059*	0.053*	0.056*
	W	0.052*	0.041*	0.067*
(50,50,50)	F	0.055*	0.063*	0.062*
	BF	0.052*	0.063*	0.060*
	MBF	0.052*	0.062*	0.053*
	W	0.059*	0.056*	0.067*
(5,10,15)	F	0.030*	0.026*	0.022
	BF	0.055*	0.045*	0.048*
	MBF	0.050*	0.038*	0.036*
	W	0.060*	0.048*	0.059*
(25,30,35)	F	0.045*	0.043*	0.054*
	BF	0.058*	0.055*	0.062*
	MBF	0.058*	0.054*	0.055*
	W	0.058*	0.052*	0.058*
(45,50,55)	F	0.047*	0.039*	0.054*
	BF	0.057*	0.044*	0.065*
	MBF	0.057*	0.044*	0.062*
	W	0.050*	0.050*	0.059*

* Test statistic is shown the probability of type I error between (0.025,0.075).

From table 12, the results appear following :

For $\phi=1$ and 1.9, the results appear that all test statistics can control the probability of type I error in all cases.

For $\phi=$ 5.6, BF MBF and W can control the probability of type I error in all cases. The F can control the probability of type I error in almost all cases except sample sizes as (5,10,15).

3.2.2 Power of The Test

Table 13-14 show the power of a test from normal and gamma distribution at the significance level 0.01.

Table 13 The highest power of a test of test statistic at significance level 0.01.

Distribution	ϕ	Sample sizes (n_1,n_2,n_3)					
		(10,10,10)	(30,30,30)	(50,50,50)	(5,10,15)	(25,30,35)	(45,50,55)
Normal	1	F,BF, MBF,W	W	F,BF, MBF,W	BF,MBF, W	BF,MBF, W	F,BF, MBF,W
	1.9	BF,MBF, W	W	W	MBF,W	BF,MBF, W	F,W
	5.6	W	BF,MBF, W	W	BF,MBF, W	F	F,W
Gamma	1	W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W
	1.9	W	F,BF, MBF,W	W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W
	5.6	F,BF, MBF	MBF,W	-	BF, W	BF, W	F,MBF, W

From table 13, the results from normal distribution appear following :

- W shows the highest power of a test in almost all cases except on ϕ = 5.6 and sample sizes as (25,30,35).
- MBF shows the highest power of a test in almost all cases except on ϕ = 1 and sample sizes as (30,30,30), on ϕ = 1.9 and sample sizes as (30,30,30), (50,50,50), and (45,50,55), on ϕ = 5.6 and sample sizes as (10,10,10), (50,50,50), (25,30,35), and (45,50,55).
- BF shows the highest power of a test in almost all cases except on ϕ = 1 and sample sizes as (30,30,30), on ϕ = 1.9 and sample sizes as (30,30,30), (50,50,50), (5,10,15), and (45,50,55), on ϕ = 5.6 and sample sizes as (10,10,10), (50,50,50), (25,30,35), and (45,50,55).
- F shows the highest power of a test in almost all cases especially on ϕ = 1 and sample sizes as (10,10,10), (50,50,50), and (45,50,55), on ϕ = 1.9 and sample sizes as (45,50,55), on ϕ = 5.6 and sample sizes as (25,30,45), and (45,50,55).

The results from gamma distribution appear following :

- W shows the highest power of a test in almost all cases except on ϕ=5.6 and sample sizes as (10,10,10), and (50,50,50).
- BF shows the highest power of a test in almost all cases except on ϕ = 1 and sample sizes as (10,10,10), on ϕ = 1.9 and sample sizes as (10,10,10), and (50,50,50), on ϕ=5.6 and sample sizes as (30,30,30), (50,50,50), and (45,50,55).
- MBF shows the highest power of a test in almost all cases except on ϕ = 1 and sample sizes as (10,10,10), on ϕ = 1.9 and sample sizes as (10,10,10), and (50,50,50), on ϕ = 5.6 and sample sizes as (50,50,50), (5,10,15), and (25,30,35).
- F shows the highest power of a test in almost all cases except on ϕ = 1 and sample sizes as (10,10,10), on ϕ = 1.9 and sample sizes as (10,10,10), and (50,50,50), on ϕ = 5.6 and sample sizes as (30,30,30), (50,50,50), (5,10,15), and (25,30,35).

Table 14 The highest power of a test of test statistic at significance level 0.05.

Distribution	ϕ	Sample sizes (n_1,n_2,n_3)					
		(10,10,10)	(30,30,30)	(50,50,50)	(5,10,15)	(25,30,35)	(45,50,55)
Normal	1	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	BF,MBF, W	F,BF, MBF,W	F,BF, MBF,W
	1.9	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W
	5.6	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	BF,MBF, W	F,BF, MBF,W	F,BF, MBF,W
Gamma	1	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	BF,MBF, W	F,BF, MBF,W	F,BF, MBF,W
	1.9	W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W
	5.6	F	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W	F,BF, MBF,W

From table 14, the results from normal distribution appear following :

- BF, MBF, and W show the highest power of a test in all cases.
- F shows the highest power of a test in almost all cases except on $\phi = 1, 5.6$ and sample sizes as (5,10,15).

The results from gamma distribution appear following :

- W shows the highest power of a test in almost all cases except on $\phi = 5.6$ and sample sizes as (10,10,10).
- F shows the highest power of a test in almost all cases except on $\phi = 1$, and sample sizes as (5,10,15), on $\phi = 1.9$ and sample sizes as (10,10,10).
- BF and MBF show the highest power of a test in almost all cases except on $\phi = 1.9, 5.6$ and sample sizes as (10,10,10).

4. Conclusions and Recommendations

In the general research, we recommend that the researchers should to consider the test statistic for testing means of three populations. In case of homogeneity of variance, we have suggested that the analysis of variance using F-test (F), Brown–Forsythe's test (BF), and modified Brown–Forsythe's test (MBF) approach in order to test the mean of three populations when the data is normal distribution even for small sample sizes at significance levels 0.01. On the other hand the analysis of variance using F-test (F) is a good performance at significance levels 0.05. For gamma distribution and significance levels 0.01, the analysis of variance using F-test (F) and Brown–Forsythe's test (BF) outperform modified Brown–Forsythe's test (MBF) and Welch's test (W) especially the small sample size, but analysis of variance using F-test (F) and Welch's test (W) perform better than Brown–Forsythe's test (BF) and modified Brown–Forsythe's test (MBF) at significance levels 0.05. In case of heterogeneity of variance, for normal and gamma distribution and significance levels 0.01, Welch's test (W) is a good test statistic better than the analysis of variance using F-test (F), Brown–Forsythe's test (BF), modified Brown–Forsythe's test (MBF) at equal sample sizes. For gamma distribution, analysis of variance using F-test (F), Brown–Forsythe's test (BF), modified Brown–Forsythe's test (MBF) are efficiency as good as Welch's test (W) significance levels 0.05.

5. References

[1] Montgomery, D.C. Design and Analysis of Experiments. New York: John Wiley & Sons ; 2013.

[2] Brown, M.B. and A.B. Forsythe. The Small Sample Behavior of Some Statistical which Test the Equality of Several Means. Technometrics. 1974; 16(1):129-132.

[3] Mehrotra, D. Improving the brown-forsythe solution to the generalized behrens fisher problem. Communications in Statistics Simulation and Computation. 1997; 26(3):1139–1145.

[4] Welch, B.L. On the comparison of several mean values: an alternative approach. Biometrika. 1951; 38(1):330-336.

[5] Games, P.A., Winkler, H.B., and Probert, D.A. Robust Tests for Homogeneity of Variance. Educational and Psychological Measurement. 1972, 32(1):887-909.

[6] Bradley, J. V. Robustness?. The British Journal of Mathematical and Statistical Psychology. 1978; 31(2):144–152.