Revista Colombiana de Estadística
January 2016, Volume 39, Issue 1, pp. 17 to 32
DOI: http://dx.doi.org/10.15446/rce.v39n1.55135

Robust Brown-Forsythe and Robust Modified Brown-Forsythe ANOVA Tests Under Heteroscedasticity for Contaminated Weibull Distribution

ANOVAS robustos heterocedásticos Brown-Forsythe y Brown-Forsythe modificado para la distribución Weibull contaminada

Derya Karagöz[1,a], Tülay Saraçbaşi[1,b]

[1]Department of Statistics, Faculty of Science, Hacettepe University, Ankara, Turkey

Abstract

In this study, robust Brown-Forsythe and robust Modified Brown-Forsythe ANOVA tests are proposed to take into consideration heteroscedastic and non-normality data sets with outliers. The non-normal data is assumed to be a two parameters Weibull distribution. Robust proposed tests are obtained by using robust mean and variance estimators based on $median/MAD$ and $median/Q_n$ methods instead of maximum likelihood. The behaviors of the robust proposed and classical ANOVA tests are examined by simulation study. The results shows that the proposed robust tests have good performance especially in the presence of heteroscedasticity and contamination.

***Key words*:** Brown-Forsythe, Modified Brown-Forsythe, ANOVA, Weibull Distribution.

Resumen

En este estudio se proponen tests Brown-Forsythe y robustos Brown-Forsythe ANOVA para tener en cuenta la no-normalidad en datos debida a la presencia de datos atípicos. Se asume que los datos no-normales tienen una distribución Weibull de dos parámetros. Estos tests se construyen en base a estimadores robustos de media y varianza obtenidos con métodos basados en la mediana en vez de métodos de máxima verosimilitud. Se examina en comportamiento de estos tests con datos simulados. Los resultados muestran que éstos tienen un buen desempeño, especialmente en presencia de atípicos y datos contaminados.

***Palabras clave*:** ANOVA, Brown-Forsythe, Brown-Forsythe modificado, distribución Weibull.

[a]PhD. E-mail: deryacal@hacettepe.edu.tr
[b]Professor. E-mail: toker@hacettepe.edu.tr

1. Introduction

The procedures to test equal means in the conventional analysis of variance (ANOVA) are based on assumptions of normality, independence, and equality of the error variances.

If the assumptions of normality and homogeneity of variances are invalid and also outliers are present, classical ANOVA does not give accurate results. The one-way ANOVA under the violation of assumptions has been studied extensively. To deal with non-normal data and/or heteroscedastic variances across groups, many alternatives such as Q, Welch, Brown-Forsythe and Modified Brown-Forsythe tests have been developed instead of classical ANOVA. The statistic Q has been extensively studied by many authors under a variety of assumptions. James (1951) and Welch (1951) derived improved approximations for the distribution of Q under the null hypothesis; these approximations are more accurate for small sample sizes of groups. A number of authors have discussed extensions of the Welch methods based on the use of robust estimators for the population location and scale parameters. Notable among these are the efforts of Wilcox (1995), Wilcox (1997), and the references contained in these papers. Kulinskaya & Dollinger (2007) consider three common robust estimators: Huber's proposed two estimators of location and scale, Hampel's M-estimator of location with scale estimated by the median absolute deviation (MAD), and the trimmed mean with scale estimated by the Winsorized standard deviation. In the presence of heteroscedasticity, the actual size of the ANOVA (which uses the F-test) can exceed the nominal level. The challenge of how to accurately compare independent heteroscedastic normal means of three or more groups is referred to as the generalized Behrens-Fisher problem. Various solutions to this problem have been proposed. In practice, some approximate procedures such as the weighted F, Welch, Brown Forsythe (BF) and modified Brown Forsythe (MBF) tests are widely used (cf. Gamage & Weerahandi 1998, Welch 1951, Brown & Forsythe 1974). Weerahandi (1995) introduced a test using the notion of generalized p-value for comparing the means of k populations with unequal error variances. Senoglu (2005) and Senoglu (2007) obtained robust and efficient estimators of the parameters in the different factorial designs by using the methodology known as modified maximum likelihood (MML) and proposed new test statistics based on MML estimators to test the main effects and the interactions when the distribution of error terms is generalized logistic. In this paper we consider the problem of comparing the means of k populations not only with heteroscedastic variances and non-normality, but also with the outliers.

Normality is the one of the important assumptions of ANOVA. However, in the application normality assumption does not work for the real life data modeled by the exponential, Weibull or lognormal distributions; especially in the field of reliability, engineering and life sciences. The characteristics of these distributions can be explained by the Weibull distribution, which is also known as Extreme Value Type III minimum distribution. This has made it extremely popular in reliability engineering, biology and medicine. This distribution is the most commonly used distribution for modeling reliability data, because it represents a wide range of asymmetric distributions. Moreover, ANOVA cannot handle censored or interval data because of non-normality. The simplest possible lifetime distribution is exponential distribution. However, its constant hazard rate is improper and unrealistic in many cases. Gamma distribution is another candidate distribution for lifetimes. Nevertheless, distribution function or survival function of gamma distribution cannot be expressed in a closed form if the shape parameter is not an integer. It is in terms of an incomplete gamma function, that one needs to obtain the distribution function, survival function or the hazard rate by numerical integration. This makes gamma distribution a little unpopular compared to the Weibull distribution, which has a nice distribution function, survival function and hazard function Gupta & Kundu (2001). It has been used in many different fields such as material science, engineering, physics, chemistry, meteorology,medicine, pharmacy, quality control, biology, geology, geography, economics and business.

The purpose of this study is to develop test statistics by using robust methods for the non-normal distribution with an outlier for the one-way ANOVA. For this reason the Weibull distribution with the parameters λ and β is discussed the most frequently encountered in the literature as a non-normal distribution. Robust estimators of Weibull distribution parameters are based on the median/median absolute deviation ($median/MAD$) and $median/Q_n$ methods. To test the hypothesis of equal means under the assumption of heteroscedastic variances the standard BF and MBF tests were proposed. To obtain these standard tests, the maximum likelihood (ML) mean and variance estimates are used. However, incase of the outliers these tests can exceed the nominal level. This is why we recommend replacing the ML mean and variance estimates with the robust estimates in the formulation of the tests. The equations from the robust tests are obtained by using the robust estimators of mean and variance based on $median/MAD$ and $median/Q_n$ methods and tests are called Robust Brown-Forsythe (RBF) and robust Modified Brown-Forsythe ($RMBF$). The behavior of the developed robust tests is examined by simulation study.

The robust mean and variance estimations of Weibull distribution is one of the important subjects of this paper, since proposed robust tests are obtained based on these estimators. The mean and variance of a Weibull random variable are the functions of the parameters. Therefore, we consider robust estimators of Weibull distribution parameters that are the $median/MAD$ proposed by Olive (2006) and the $median/Q_n$ proposed by Boudt, Caliskan & Croux (2011). These are two robust estimator alternatives to the maximum likelihood estimator. In this paper they use the log-Weibull density as a location-scale family with location $\mu = \log \lambda$ and scale $\sigma = 1/\beta$ (see Boudt et al. 2011). An initial or auxiliary estimate of scale is frequently required in robust estimation. This paper usually uses the median absolute deviation $MAD_n = 1.4826 med_i |x_i - med_j x_j|$, because it has a simple explicit formula, needs little computation time, and is very robust as witnessed by its bounded influence function and its 50% breakdown point. However MAD_n is aimed at symmetric distributions and it has a low (37%) Gaussian efficiency. Rousseeuw & Croux (1993) set out to construct explicit and 50% breakdown scale estimators that are more efficient. Rousseeuw & Croux (1993) considered the estimator Q_n given by the 0.25 quantile of the distances $|x_i - x_j|; i < j$. The advantages of the Q_n are that it does not need any location estimate and the Gaussian efficiency of Q_n is 82%, Rousseeuw & Croux (1993).

The remainder of our paper is organized as follows. Section 2 introduces robust and explicit estimators of the mean and variance of Weibull distribution. Section 3 explains the robust proposed test statistics. A numerical example is given in Section 4. To show the performance of proposed test statistics, a simulation study and the results are presented in Section 5. Finally, the last section summarizes the study's.

2. Robust Estimators

The density and cumulative distribution functions of Weibull distribution are given by

$$f_{\lambda,\beta}(x) = \frac{\beta}{\lambda}(x/\lambda)^{\beta-1}\exp[-(x/\lambda)^{\beta}], \tag{1}$$
$$F_{\lambda,\beta}(x) = 1-\exp[-(x/\lambda)^{\beta}] \tag{2}$$

$x, \lambda, \beta > 0$. Since $f_{\lambda,\beta}(x) = \frac{1}{\lambda} f_{1/\lambda,\beta}(x)$, the parameter λ is called the scale parameter. The parameter β is the shape parameter. The mean and variance of a Weibull random variable can be expressed as

$$E(X) = \lambda\Gamma(1+1/\beta) \tag{3}$$
$$Var(X) = \lambda^2[\Gamma(1+2/\beta)-\Gamma^2(1+1/\beta)]. \tag{4}$$

As shown in equation (3) and equation (4), the mean and variance of a Weibull random variable are the functions of shape β and scale λ parameters of Weibull distribution. After obtaining robust parameters estimators, the robust mean and variance estimators of this distribution can be obtained. We consider robust estimators of Weibull distribution parameters that are $median/MAD$, proposed by Olive (2006) and $median/Q_n$ proposed by Boudt et al. (2011). In the next Section 2.1. these robust location-scale estimators with be explained.

2.1. Location-Scale Estimators

The proposed robust estimators are based on the log-Weibull density which is a location-scale family with location $\mu = \log\lambda$ and scale $\sigma = 1/\beta$. Estimation of Weibull parameters can thus be seen as an estimation problem of the location and scale of $\log x_1, \ldots, \log x_n$. Olive (2006) presents the correction factors making these estimators consistent for the log-Weibull distribution. Olive (2006) proposed the standard location and scale estimators with a 50% breakdown point, which were the median and median absolute deviation

$$\hat{\sigma}_{\text{MAD}} = 1.3037\text{med}_j|\log x_j - \text{med}_i \log x_i| \tag{5}$$
$$\hat{\mu}_{\text{med/MAD}} = \text{med}_i \log x_i - \hat{\sigma}_{\text{MAD}} \log\log 2, \tag{6}$$

where $x_i \sim W(\lambda, \beta)$ and $y_i = logx_i$ have log-Weibull distribution. By using the location and scale estimators based on med/MAD given by (equation 5) and (equation 6), respectively, Boudt et al. (2011) proposed the med/MAD estimators of scale and shape parameters of Weibull distribution:

$$\hat{\lambda}_{\text{med/MAD}} = \exp(\hat{\mu}_{\text{med/MAD}}) \tag{7}$$

$$\hat{\beta}_{\text{med/MAD}} = 1/\hat{\sigma}_{\text{MAD}}. \tag{8}$$

By considering robust med/MAD estimators given by equation (7) and equation (8), the med/MAD estimators of the mean and the variance of Weibull distribution are obtained as per the following

$$\hat{\mu}_{\text{Wmed/MAD}} = \hat{\lambda}_{\text{med/MAD}}\Gamma(1 + 1/\hat{\beta}_{\text{med/MAD}}) \tag{9}$$

$$\hat{\sigma}_{\text{Wmed/MAD}} = \hat{\lambda}^2_{\text{med/MAD}}[\Gamma(1 + 2/\hat{\beta}_{\text{med/MAD}}) - \Gamma^2(1 + 1/\hat{\beta}_{\text{med/MAD}})]. \tag{10}$$

Q_n scale-estimator is as a more efficient alternative and the the same breakdown point of 50%. Boudt et al. (2011) recommended estimating σ using the Q_n scale-estimator proposed by Rousseeuw & Croux (1993). It was given by Boudt et al. (2011)

$$\hat{\sigma}_{\text{Qn}} = 1.9577\{|\log x_i - \log x_j|; i < j\}_{(l)} \tag{11}$$

$$\hat{\mu}_{\text{med/Qn}} = \text{med}_i \log x_i - \hat{\sigma}_{\text{Qn}} \log\log 2, \tag{12}$$

where the last part is the lth ordered value among this set of $\binom{n}{2}$ values, where $l = \binom{h}{2} \approx \binom{n}{2}/4$ with $h = \lfloor n/2 \rfloor + 1$. The correction factor 1.9577 ensures Fisher consistency. It equals the inverse of the 1/4 quantile of the distribution of the absolute difference between two log-Weibull random variables. These estimators are called med/Q_n estimators for scale and shape parameters. By using the med/Q_n location-scale estimators in equation (11) and equation (12), Boudt et al. (2011) proposed the robust scale and shape estimators of Weibull distribution based on med/Q_n

$$\hat{\lambda}_{\text{med/Qn}} = \exp(\hat{\mu}_{\text{med}}) \tag{13}$$

$$\hat{\beta}_{\text{med/Qn}} = 1/\hat{\sigma}_{\text{Qn}}. \tag{14}$$

By considering robust estimators in equation (13) and equation (14), the med/Q_n mean and variance estimators of Weibull distribution are obtained as per the following

$$\hat{\mu}_{\text{Wmed/Qn}} = \hat{\lambda}_{\text{med/Qn}}\Gamma(1 + 1/\hat{\beta}_{\text{med/Qn}}) \tag{15}$$

$$\hat{\sigma}_{\text{Wmed/Qn}} = \hat{\lambda}^2_{\text{med/Qn}}[\Gamma(1 + 2/\hat{\beta}_{\text{med/Qn}}) - \Gamma^2(1 + 1/\hat{\beta}_{\text{med/Qn}})]. \tag{16}$$

In this study, to define the breakdown point of estimations, previous studies are taken into account, especially Boudt et al. (2011). The breakdown point of the mean estimator and the variance estimator for Weibull distribution based on the proposed robust estimator are obtained a 50% as the breakdown of mean and variance depends on the the breakdown point of shape and scale estimators of the Weibull distribution.

3. Robust BF and MBF ANOVA Tests

The ANOVA for non-normal data with heteroscedastic variance has been studied extensively. In the case of disruption of assumptions, instead of classical ANOVA, many tests have been developed such as BF and MBF. Brown & Forsythe (1974) proposed the BF test that can be shown by

$$BF = \frac{\sum_{i=1}^{k} n_i(\hat{\mu}_{i.} - \hat{\mu}_{..})^2}{\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_i^2} \tag{17}$$

where $\hat{\mu}_{i.}$, $\hat{\sigma}_i$ and n_i denoted the ML estimator of the sample mean and variance and size for the ith k group, respectively, $\hat{\mu}_{..}$ is the ML estimator of overall mean, and $N = \sum_{i=1}^{k} n_i$. The BF test has a F_{k-1,v_1} distribution with $k-1$ and v_1 degrees of freedom. v_1 is defined as

$$v_1 = \frac{[\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_i^2]^2}{\sum_{i=1}^{k}(1 - n_i/N)^2\hat{\sigma}_i^4/(n_i - 1)}. \tag{18}$$

Mehrotra (1997) proposed the MBF test given by

$$MBF = \frac{\sum_{i=1}^{k} n_i(\hat{\mu}_{i.} - \hat{\mu}_{..})^2}{\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_i^2}, \tag{19}$$

which is the same as the equation (17). But the MBF test has F_{v_2,v_1} distribution with v_2 and v_1 degrees of freedom. By using Mehrotra's (1997) approach the numerator degrees of freedom v_2 was defined as

$$v_2 = \frac{[\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_i^2]}{\sum_{i=1}^{k}\hat{\sigma}_i^4 + [\frac{\sum_{i=1}^{k} n_i\hat{\sigma}_i^2}{N}]^2 - 2\frac{\sum_{i=1}^{k} n_i\hat{\sigma}_i^4}{N}}. \tag{20}$$

In this study, two robust tests are proposed to test the equality of population means under the Weibull distribution. The proposed tests correspond to the standard BF and MBF test statistics in which the maximum likelihood mean and variance estimators are replaced with robust estimators. We propose using the RBF and $RMBF$ tests based on the med/MAD and med/Q_n mean and variance estimators of Weibull distribution. The RBF test is given by

$$RBF = \frac{\sum_{i=1}^{k} n_i(\hat{\mu}_{ri.} - \hat{\mu}_{r..})^2}{\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_{ri}^2}, \tag{21}$$

where $\hat{\mu}_{ri.}$, $\hat{\sigma}_{ri}$ and n_i denoted the robust estimators of mean and variance and size for the ith k group, $\hat{\mu}_{r..}$ is the robust estimator of overall mean, and $N = \sum_{i=1}^{k} n_i$. The RBF test has a F_{k-1,v_r} distribution with $k-1$ and v_r degrees of freedom. v_r is defined as

$$v_r = \frac{[\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_{ri}^2]^2}{\sum_{i=1}^{k}(1 - n_i/N)^2\hat{\sigma}_{ri}^4/(n_i - 1)}. \tag{22}$$

The $RMBF$ tests is given by

$$RMBF = \frac{\sum_{i=1}^{k} n_i(\hat{\mu}_{ri.} - \hat{\mu}_{r..})^2}{\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_{ri}^2}, \tag{23}$$

which is the same as the equation (21). But the $RMBF$ test has a F_{v_{r2},v_r} distribution with v_{r2} and v_r degrees of freedom. By using Mehrotra's (1997) approach the numerator degrees of freedom v_{r2} is defined as

$$v_{r2} = \frac{[\sum_{i=1}^{k}(1 - n_i/N)\hat{\sigma}_{ri}^2]}{\sum_{i=1}^{k} \hat{\sigma}_{ri}^4 + [\frac{\sum_{i=1}^{k} n_i \hat{\sigma}_{ri}^2}{N}]^2 - 2\frac{\sum_{i=1}^{k} n_i \hat{\sigma}_{ri}^4}{N}}. \tag{24}$$

In the above equations equation (21), equation (22), equation (23) and equation (24), $\hat{\mu}_{ri.}$, $\hat{\mu}_{r..}$, $\hat{\sigma}_{ri}$ robust mean and variance estimators are based on med/MAD and med/Q_n methods.

4. A Numerical Example

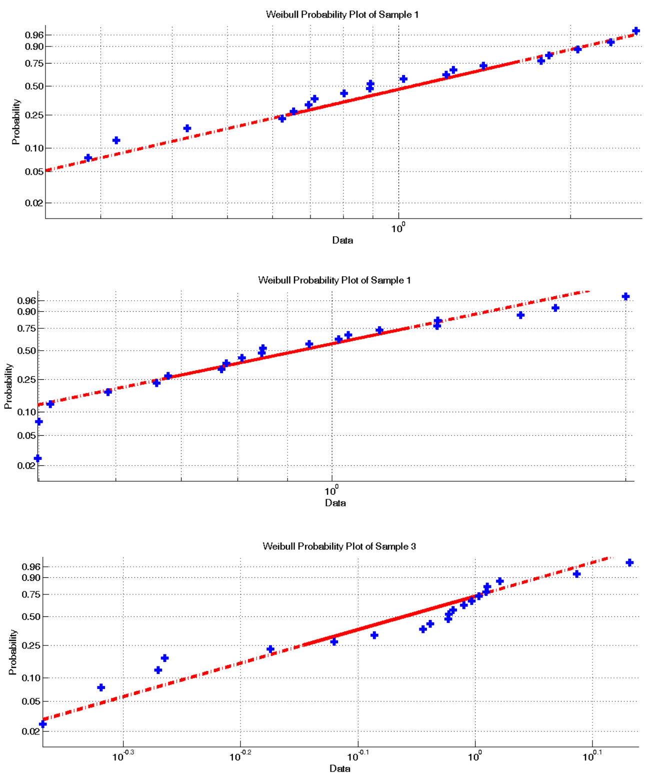
As an example, three Weibull populations, each with a size of 5,000, are generated with shape the following and scale parameters: The shape parameter of the first population is 1.5 and the scale parameter is 1. With the second population, the shape parameter is 2.5 and the scale parameter is 1. The shape parameter of the third population is 3.5 and the scale parameter is 1. Supposing that three random samples with a size of 20 are taken from each Weibull population, the sample data are shown in Table 1.

TABLE 1: Three random samples with a size of 20 taken from each of three Weibull populations.

Sample 1	Sample 2	Sample 3
0.7110	0.6601	0.9159
0.4247	0.9464	0.5436
0.8905	0.7786	0.9487
1.4064	0.7703	0.9932
0.8012	1.2813	0.9775
2.0574	0.5002	1.0214
0.6533	1.6926	0.9030
1.7747	1.0385	1.3542
0.3197	0.6782	0.9577
1.8344	0.8490	0.9496
0.6231	1.2822	0.6689
0.6943	1.1174	1.2202
1.2458	0.5139	1.0077
1.0185	0.8083	0.4797
0.2401	0.8469	1.0241
2.6065	1.0146	0.8206
0.8898	0.5889	0.5370
0.2850	2.0002	0.7586
1.2093	1.5597	1.0501
2.3521	0.4992	0.4278

The Weibull P-P plot for each sample of this data is presented in Figure 1. The results show that each sample of data is Weibull.

Figure 1: Weibull Probability Plot of Three Samples



The Anderson-Darling test Anderson & Darling (1952) is used to test if a sample of data comes from a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than the K-S test. The K-S test is distribution free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling (A-D) test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution Trujillo-Ortiz, Hernandez-Walls, Barba-Rojo, Castro-Perez & Lavaniegos-Espejo (2007). The sampled populations have a Weibull distribution. The results of the Anderson-Darling tests for three samples are shown in Table 2. Thus, these samples have been drawn from a Weibull population with estimated parameters that are given in Table 2.

Weibull data are used to test the equality of the population means using ANOVA. We consider not only clean samples but also contaminated samples. To generate contaminated samples, 4 observations from each sample are randomly replaced with 100. The classical ANOVA, RBF and RGBF tests are obtained for

Table 2: Results of Anderson-Darling statistic.

	Sample 1	Sample 2	Sample 3
A-D statistic	0.2440	0.4757	0.6327
A-D adjusted statistic	0.2549	0.4970	0.6610
Prob. A-D statistic	0.6577	0.2210	0.0872
Significance	0.050	0.050	0.050
Weibull population $W(\hat{\lambda}, \hat{\beta})$	$W(1.2409, 1.7039)$	$W(1.0974, 2.5538)$	$W(0.9661, 4.2887)$

clean and contaminated samples. The results are shown in Table 3. F_{obt}, RBF_{obt} and $RMBF_{obt}$ indicate the calculated tests. F_t, RBF_t and $RMBF_t$ indicate the F table value at the significant of 5% ($\alpha = 0.05$). RBF_{obt} and $RMBF_{obt}$ are based on med/MAD and med/Q_n robust methods. In the contamination case classical ANOVA is badly deteriorated. For clean and contaminated data, since calculated tests < F table value, there are not significant differences between the three population means. The results shows that proposed robust tests give accurate results.

Table 3: Three random samples with a size of 20 taken from each of three Weibull populations.

Tests	Clean Sample		Contaminated Sample	
	med/MAD	med/Q_n	med/MAD	med/Q_n
F_{obt}	0.6905	1.0558	0.0003	0.0009
F_t	3.1588	3.1588	3.1588	3.1588
RBF_{obt}	0.6219	0.4548	1.7068	0.4410
RBF_t	3.2287	3.2715	3.3308	3.2954
$RMBF_{obt}$	0.6219	0.4528	1.7068	0.4410
$RMBF_t$	5.1432	3.3778	4.7062	3.5294

5. Simulation Study

In the simulation study the reference distribution is $W(1, \beta)$. The value of the shape parameters are selected, as in Table 4, with respect to the different experimental designs we want to create. In this table for equal means, homogeneous variances EA, EB are used for a balanced sample size, EC, ED are used for an unbalanced sample size. For unequal means, heterogenous variances UA, UB are used for balanced sample size and UC, UD are used for an unbalanced sample size. As we mentioned, previously to generate data with equal means + homogeneous variances and unequal means + heterogenous variances, we just change the shape parameter of Weibull distribution. Since the mean and variance of the distribution have been changed with respect to the shape and scale of the parameters of the Weibull distribution, the creation of different experiment designs depends on the parameters of the distribution.

In this table

- Design EA: balanced sample size ($n = 5$) with equal means and homogenous variances,
- Design UA: balanced sample size ($n = 5$) with unequal means and heterogenous variances,

- Design EB: balanced sample size ($n = 10$) with equal means and homogenous variances,
- Design UB: balanced sample size ($n = 10$) with unequal means and heterogenous variances,
- Design EC: unbalanced sample size ($n = 5, 10, 15$) with equal means and homogenous variances,
- Design UC: unbalanced sample size ($n = 5, 10, 15$) with unequal means and heterogenous variances,
- Design ED: unbalanced sample size ($n = 10, 20, 30$) with equal means and homogenous variances,
- Design UD: unbalanced sample size ($n = 10, 20, 30$) with unequal means and heterogenous variances.

Table 4: Experimental designs for $k = 3$, $k = 6$ and $k = 9$.

		$K=3$			$K=6$						$K=9$								
EA	n_i	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	β	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
UA	n_i	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	β	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5
EB	n_i	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	β	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
UB	n_i	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	β	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5
EC	n_i	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15
	β	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
UC	n_i	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15
	β	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5
ED	n_i	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	β	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
UD	n_i	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	β	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5

The simulation experiment stork with $k = 3$ groups. The balanced and unbalanced sample sizes with homogenous and heterogeneous variances are considered. To see the effect of the number of groups, k, on the level of the tests, we replicated the experiment for $k = 3$ groups two and three times to provide the simulation experiment for $k = 6$ and $k = 9$ groups, respectively. By using the proposed estimators med/MAD and med/Q_n the Type I errors of the robust test statistics and classical ANOVA are obtained with 10,000 repetitions. Four different models are discussed below to test the behaviors of the tests, when the model is deteriorated and in the presence of outliers. In models the response variable is assumed to have a Weibull distribution. The models for simulation study are:

1. Clean sample: reference distribution $W(1, \beta)$,
2. Dixon model: $n-1$ observations from $W(1, \beta)$, 1 observation from $W(2, \beta)$,
3. Mixture model: $0.80W(1, \beta) + 0.20\, W(2, \beta/2)$,
4. Contaminated model: $0.80\ W(1, \beta) + 0.20\, (100\ Uniform(0, 1))$.

We obtain $\tau = \frac{\sum_{i=1}^{M} F_{H_i} > F_{T_i}}{M} * 100$ value with respect to the classical F, RBF and $RMBF$ tests with 10.000 repetitions. In this equation F_{H_i} indicates the calculated tests and F_{T_i} indicates the F table value at the significant of 5% for the ith simulation, so the desirable value of τ is close to $\tau \cong 5$.

The results of $\boldsymbol{\tau}$ values for experimental design EA with equal means and homogeneous variances are given in Table 5. While the classical ANOVA does not deteriorate for the clean model (model 1), it badly deteriorates for the contaminated model. When $RMBF$ based on med/MAD has good performs well for a mixture model, RBF based on med/MAD gives desirable results for contaminated model. As can be seen in this table, the RBF test based on med/Q_n only works for the clean and Dixon models for $k = 3$. Robust tests based on med/Q_n do not work for $k = 6, 9$. The results of $\boldsymbol{\tau}$ for experimental design UA are given in Table 5. For the balanced small sample size of each group the classical ANOVA test does not deteriorate, and in terms of type I error it gives the optimal result. Because the observation number is small so it does not effect the results. The classical ANOVA is only deteriorated in the case of contamination. The contaminated model $RMBF$ test based on med/MAD gives the optimal results, which are close to 5. The RBF test based on med/MAD can be used as an alternative to $RMBF$. For the Dixon and mixture models RBF and $RMBF$ based on med/Q_n perform well for $k = 3$. However, for $k = 9$ only the performance of $RMBF$ test based on med/Q_n is good for the clean and Dixon models. The mixture model MBF test based on med/Q_n can be used for $k = 3, 6$.

The results of $\boldsymbol{\tau}$ values for experimental design EB with equal means and homogeneous variances are given in Table 6. The classical ANOVA badly deteriorates for contaminated model especially for $k = 9$. When $RMBF$ based on med/MAD performs well for mixture and contaminated models, RBF based on med/Q_n gives desirable results for the Dixon model. The results of $\boldsymbol{\tau}$ for experimental design UB are given in Table 6. It can be seen that classical ANOVA is deteriorated especially in the contamination case. The $RMBF$ test based on med/MAD can be used for mixture and contaminated models for $k = 3, 6, 9$. The $RMBF$ test statistic based on med/Q_n is desirable for clean and Dixon models for $k = 3, 6$. The optimal test statistic for the mixture model is *RBF*. RBF is based on med/Q_n for $k = 3, 6, 9$.

Table 5: The results of $\boldsymbol{\tau}$ values for experimental designs EA and UA.

Experimental design EA										
RE	Model	$k=3$			$k=6$			$k=9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	4.70	13.02	10.0	4.71	12.68	87.3	4.69	13.05	8.50
	2	3.55	11.34	8.61	3.81	11.42	7.71	4.47	18.86	11.36
	3	3.32	10.31	7.67	3.37	8.40	5.66	3.09	7.34	4.50
	4	0.10	7.06	5.24	0.19	6.30	3.90	0.10	5.45	3.23
med/Q_n	1	4.33	5.43	3.87	4.61	3.90	2.36	4.96	2.79	1.42
	2	3.94	5.07	3.35	3.74	3.65	2.25	3.59	2.20	1.12
	3	3.11	3.71	2.66	3.08	1.58	0.90	3.15	1.03	0.61
	4	0.20	1.79	1.22	0.12	0.75	0.45	0.08	0.45	0.24
Experimental design UA										
RE	Model	$k=3$			$k=6$			$k=9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	5.78	15.68	12.28	5.63	16.60	11.88	5.80	18.14	12.27
	2	4.97	13.38	10.19	4.55	14.48	10.32	4.33	15.75	10.54
	3	3.78	11.94	9.16	3.46	11.58	7.88	3.40	10.97	7.06
	4	0.13	7.49	5.45	0.12	7.37	4.57	0.15	6.86	4.12
med/Q_n	1	5.90	7.23	5.51	5.29	5.28	3.20	5.97	4.63	2.80
	2	4.75	6.95	4.96	4.60	5.11	3.32	4.01	3.49	1.92
	3	3.90	4.99	3.66	3.22	2.93	1.80	3.27	1.97	1.04
	4	0.17	1.93	1.36	0.18	0.99	0.56	0.08	0.59	0.22

Table 6: The results of $\boldsymbol{\tau}$ values for experimental designs EB and UB.

Experimental design EB										
RE	Model	$k=3$			$k=6$			$k=9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	4.93	11.58	7.54	4.98	14.73	9.35	5.15	16.54	9.90
	2	4.40	11.32	6.90	4.19	14.16	8.38	4.17	15.84	9.14
	3	3.87	9.14	5.44	3.14	9.31	5.16	3.02	9.26	4.51
	4	0.04	9.06	4.72	1.14	10.82	5.40	22.37	12.16	5.29
med/Q_n	1	4.91	6.27	3.64	5.09	6.43	3.26	4.76	6.62	3.17
	2	4.43	5.79	3.01	4.43	6.32	3.28	4.06	5.99	2.79
	3	3.61	3.69	1.84	3.03	2.92	1.32	3.10	2.33	0.91
	4	0.05	2.16	0.58	0.84	1.88	0.45	22.12	1.69	0.34
Experimental design UB										
RE	Model	$k=3$			$k=6$			$k=9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	5.36	13.35	9.47	5.91	16.76	10.88	5.92	19.62	12.33
	2	5.32	13.47	9.17	5.01	15.86	9.95	4.97	17.96	11.22
	3	3.92	10.57	6.64	4.03	12.18	7.54	3.35	12.69	7.51
	4	0.08	9.97	5.17	0.91	12.72	6.31	22.86	14.62	6.58
med/Q_n	1	5.49	7.29	4.65	5.83	8.22	4.83	5.71	8.41	4.51
	2	5.19	7.55	4.62	5.15	7.14	4.03	4.81	7.87	3.88
	3	3.68	4.74	2.72	4.05	4.43	2.19	3.49	4.10	1.92
	4	0.04	2.25	0.72	1.00	2.52	0.72	9.35	2.62	0.57

The results of $\boldsymbol{\tau}$ values for experimental design EC with equal means and homogeneous variances are given in Table 7. The classical ANOVA badly deteriorates especially for contaminated model. When $RMBF$ based on med/MAD performs well for mixture and contaminated models, RBF based on med/Q_n gives desirable results for the mixture model. The results of $\boldsymbol{\tau}$ for experimental design UC are given in Table 7. In this table the classical ANOVA is deteriorated for all methods. When $RMBF$ based on med/Q_n perform well for the Dixon model for $k = 6, 9$, $RMBF$ based on based on med/MAD can be used in contaminated models. When the number of groups is big, the type I error of the $RMBF$ test statistic based on the med/Q_n method is not at a desirable level in the contaminated model, doe to a large amount of deterioration in the sample.

Table 7: The results of $\boldsymbol{\tau}$ values for experimental designs EC and UC.

Experimental design EC										
RE	Model	$k = 3$			$k = 6$			$k = 9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	6.60	17.28	11.10	5.62	17.76	10.99	5.33	19.97	12.24
	2	5.62	16.60	10.59	4.86	17.64	10.65	4.79	19.54	11.59
	3	4.84	12.84	7.63	4.06	12.04	6.58	3.88	11.56	6.37
	4	30.82	9.70	6.02	60.69	9.96	5.65	67.13	10.10	5.24
med/Q_n	1	6.75	9.10	4.86	5.97	8.46	4.48	5.82	8.04	4.32
	2	5.41	9.10	4.83	4.34	7.98	4.04	4.44	7.10	3.60
	3	4.95	6.36	3.01	4.37	4.51	2.19	4.06	3.67	1.70
	4	30.71	2.82	1.36	60.69	1.83	0.70	68.21	1.54	0.46
Experimental design UC										
RE	Model	$k = 3$			$k = 6$			$k = 9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	12.26	20.10	14.99	12.86	20.44	15.38	14.03	23.04	16.47
	2	11.53	20.96	15.5	11.03	20.98	15.53	12.10	21.20	15.23
	3	9.48	17.13	12.49	8.86	15.19	10.81	8.85	14.26	9.94
	4	31.97	10.81	7.42	62.67	11.47	7.83	68.74	11.89	7.42
med/Q_n	1	12.37	11.84	8.71	6.08	8.28	4.62	13.65	8.96	6.03
	2	11.10	12.06	8.62	10.93	9.92	6.84	11.16	8.02	5.45
	3	9.24	9.40	6.37	8.46	5.46	3.56	8.77	4.76	2.85
	4	32.29	3.77	2.25	63.23	2.55	1.39	68.58	1.74	0.79

Table 8: The results of $\boldsymbol{\tau}$ values for experimental designs ED and UD.

Experimental design ED										
RE	Model	$k = 3$			$k = 6$			$k = 9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	6.78	17.21	10.75	5.79	20.60	13.55	5.60	24.33	16.37
	2	5.74	16.86	10.31	5.34	20.12	13.24	5.05	23.78	15.97
	3	5.03	13.15	7.58	4.51	14.07	8.47	4.18	15.89	9.25
	4	82.61	13.29	7.58	86.17	16.72	9.17	87.84	20.04	10.69
med/Q_n	1	6.45	11.08	7.04	5.69	12.92	8.01	5.79	14.37	9.52
	2	6.23	11.77	6.46	4.98	12.16	7.46	5.24	13.33	8.55
	3	4.79	7.71	3.98	4.47	7.12	3.92	4.14	7.28	3.94
	4	82.94	4.48	2.07	85.90	4.88	1.68	87.28	5.90	2.11
Experimental design UD										
RE	Model	$k = 3$			$k = 6$			$k = 9$		
		F	RBF	RMBF	F	RBF	RMBF	F	RBF	RMBF
med/MAD	1	12.31	17.55	13.23	13.10	20.43	14.85	14.47	22.53	16.49
	2	11.44	18.26	13.81	11.93	21.21	15.76	11.86	22.17	15.62
	3	9.76	14.76	10.67	9.30	14.85	10.68	9.81	16.16	11.03
	4	83.30	14.63	9.04	86.10	17.28	9.80	88.04	19.88	10.95
med/Q_n	1	12.47	11.36	8.06	13.61	12.11	7.96	14.16	13.27	8.62
	2	12.20	11.78	8.58	11.57	11.36	8.03	12.31	12.03	7.69
	3	10.40	8.92	6.17	9.17	7.43	4.53	10.09	7.39	4.47
	4	83.06	4.99	2.31	86.79	5.79	1.94	87.96	6.43	1.88

The results of $\boldsymbol{\tau}$ values for experimental design ED with equal means and homogeneous variances are given in Table 8. The classical ANOVA badly deteriorates especially for the contaminated model. When $RMBF$ based on med/Q_n

works well for the mixture model, RBF based on med/Q_n gives desirable results for contaminated models. Robust tests based on med/MAD do not work in the case of big size unbalanced samples. The results of $\boldsymbol{\tau}$ for experimental design UD are given in Table 8. In this experimental design, while proposed test statistics based on med/MAD do not work, the RBF test based on med/Q_n gives desirable results in contaminated models. The $RMBF$ test based on med/Q_n can be used in mixture model.

6. Conclusion

ANOVA is one of the most commonly used models in many fields such as medicine, engineering, psychology, sociology, etc. In general, the main interest of this paper was testing the homogeneity of group means using the classical ANOVA, which uses the F-test statistic. One-way ANOVA is based on assuming the normality of the observations and the homogeneity of group variances. If the assumptions of normality and homogeneity of variances are invalid and there are also outliers are present, classical ANOVA does not give accurate results. Therefore, test statistics based on robust methods should be used instead of the classical ANOVA.

The aim of this paper was to test the equality of population means of groups under heteroscedasticity for contaminated Weibull distribution in the one-way ANOVA. To deal with non-normal contaminated data and heteroscedastic variances across groups, RBF and $RMBF$ tests were developed by utilizing robust mean and variance estimators, which are also proposed. Robust proposed estimators and tests are obtained by using robust med/MAD and med/Q_n estimators instead of maximum likelihood estimators. The behaviors of the robust tests that were developed are examined in terms of the type I-errors via a Monte-Carlo simulation study. In the simulation study, various experimental designs are considered such as balanced and unbalanced sample sizes for $k = 3, 6, 9$ groups with homogenous and heterogeneous variances. The type I errors of the improved robust tests and classical ANOVA under the Weibull distribution were obtained. In terms of homogenous experimental designs: While the classical ANOVA did not deteriorate for clean models (model 1), it badly deteriorates for contaminated models. When $RMBF$ based on med/MAD performs well for mixture models, RBF based on med/MAD gives desirable results for contaminated models. While robust tests work well for balanced samples, especially for contaminated models, robust tests based on med/MAD do not work in the case of large-size unbalanced samples. In terms of heterogeneous experimental designs: The performance of the $RMBF$ based on med/Q_n is the best for mixed models. Also, RBF based on med/Q_n is an alternative to this test statistic. In unbalanced experimental designs, classical ANOVA badly deteriorates, especially for contaminated models.

[Received: October 2014 — Accepted: February 2015]

References

Anderson, T. W. & Darling, D. A. (1952), 'Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes', *Annals of Mathematical Statistics* **23**, 193–212.

Boudt, K., Caliskan, D. & Croux, C. (2011), 'Robust explicit estimators of weibull parameters', *Metrika* **73**, 187–209.

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

Gamage, J. & Weerahandi, S. (1998), 'Size performance of some tests in one-way anova', *Communications in Statistics - Simulation and Computation* **27**, 625–640.

Gupta, R. & Kundu, D. (2001), 'Exponentiated exponential family: An alternative to gamma and weibull distributions', *Biometrical Journal* **43**(1), 117–130.

James, G., S. (1951), 'The comparison of several groups of observations when the ratios of the population variances are unkown', *Biometrika* **38**, 324–329.

Kulinskaya, E. & Dollinger, M. (2007), 'Robust weighted One-Way ANOVA: Improved approximation and efficiency', *Journal of Statistical Planning and Inference* **137**, 462–472.

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

Olive, D. (2006), 'Robust estimators for transformed location scale families', `https://www.researchgate.net/profile/David_Olive2/publication/228390054_Robust_estimators_for_transformed_location_scale_families/links/02bfe51015be6ca4fa000000.pdf`. Accessed: 2015-12-24.

Rousseeuw, P. & Croux, C. (1993), 'Alternatives to the median absolute deviation', *Journal of the American Statistical Association* **88**, 1273–1283.

Senoglu, B. (2005), 'Robust 2^k factorial design with Weibull error distributions', *Journal of Applied Statistics* **32**, 1051–1066.

Senoglu, B. (2007), 'Robust estimation and hypothesis testing in 2^k factorial design', *Communications de la Faculte des Sciences de l'Universite. Series A1* **56**, 39–50.

Trujillo-Ortiz, A., Hernandez-Walls, R., Barba-Rojo, K., Castro-Perez, A. & Lavaniegos-Espejo, B. (2007), 'Andarwtest:anderson-darling test for assessing weibull distribution of a sample data', http://www.mathworks.com/matlabcentral/fileexchange/15745-anderwtest.

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

Welch, B. (1951), 'On the comparison of several mean values', *Biometrika* **38**, 330–336.

Wilcox, R. (1995), 'The practical importance of heteroscedastic methods, using trimmed means versus means, and designing simulation studies', *British Journal of Mathematical and Statistical Psychology* **48**, 99–114.

Wilcox, R. (1997), *Introduction to Robust Estimation and Hypothesis Testing*, Academic Press, New York.