Comparison of Several Means under Heterogeneity: Over-mean-rank Function Approach

Elsayed A. H. Elamir[1]

Department of Statistics and Mathematics, Benha University, Egypt & Management & Marketing Department, College of Business, University of Bahrain, P.O. Box 32038, Kingdom of Bahrain

Abstract

In practice it is often of interest to compare the means of several populations that are not assumed to have equal variances. A method is proposed that depends on the over-mean-rank function that is defined as the percentage of the ranks over the global mean rank in each group. Chi-square distribution is found to give a very good fit for this function until for small sample sizes. The main advantages for the proposed method are: does not require normality and equal variance assumptions; stables in terms of Type I error; and can be shown graphically. Comparison with Kruskal-Wallis, Welch and ANOVA methods are given for unbalanced designs and not equal variances from normal and non-normal populations in terms of Type I error. The simulation results are shown that the proposed method improves the Type I error and its performance exhibits superior robustness over the studied methods.

Keywords**:** ANOVA; F-distribution; Hypothesis testing; Kruskal-Wallis test; Welch-test.

1 Introduction

The seminal work by [1] provided us a robust rank-based test for comparison of several means, complementing the parametric approaches. [2], [3] and [4] among others (see [5] and [6]) have presented approximate test statistics for testing for mean equality when there are more than two groups and when population variances are not presumed to be equal. Unfortunately, these procedures have not proven to be uniformly successful in controlling test size when data are heterogeneous as well as non-normal, particularly in unbalanced designs. Although there are parametric solutions have been presented by [7] and [8], it will be only focused on the nonparametric approach especially rank-approach.

To compare for means under heterogeneity using nonparametric approach a method is derived based on over-mean-rank function that is defined as the percentage of the ranks more than the global mean rank in each group. Chi-square distribution is found to give a very good fit for this function until for small sample sizes. This method does not require normality and equal variance assumptions, stables in terms of Type I error and is shown graphically. Comparison with Kruskal-Wallis, Welch and ANOVA methods are given for unbalanced designs and not equal variances from symmetric and asymmetric populations in terms of Type I error. The simulation results are shown that the proposed method improves the Type I error and its performance exhibits superior robustness over the studied methods.

Over-mean-rank function and over-mean-rank plot are introduced in Section 2. The simulation results are presented in Section 3 . An application is given in Section 4. Section 5 is devoted for conclusion.

[1]Email: shabib40@gmail.com

2 Comparison of several means or medians

2.1 Over-mean-rank function approach

Suppose independent random observations $Y_{gi}(g = 1, \ldots, G, i = 1, \ldots, n_g,$ and $n_1 + \cdots + n_g = n$) are obtained from a continuous population with mean μ_g and variance σ_g^2. G is the number of groups or treatments and n_g is the sample size in each group. Then, let $\bar{Y}_g = \sum_{i=1}^{n_g} Y_{gi}/n_g$ and $\bar{Y} = \sum_{g=1}^{G} \sum_{i=1}^{n_g} Y_{gi}/n$ are the group and overall means. The model is

$$Y_{gi} = \mu + \alpha_g + \epsilon_{gi}$$

μ is the global mean of the data, $\alpha_g$the difference to the mean of the g-th group and ϵ_{gi} is the residual error. Thus the null hypothesis can be expressed as

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_G = \mu$$

versus at least two means are not equal.

The rank function can be defined as

$$R = R_{gi} = \text{rank}(Y_{gi}), \quad g = 1, \ldots, G, \quad i = 1,2, \ldots, n_g$$

and the ranks in each group is

$$R_g = R_{gi}, \quad \text{for each } g = 1, \ldots, G$$

The R_{gi} has a discrete uniform distribution with probability mass function

$$f(R_{gi} = r) = \frac{1}{n}, \qquad r = 1,2, \ldots\ldots, n$$

with

$$E(R) = (n+1)/2$$

If all means are equal, R will have average equals to the average for each group. On the other hand, if the means are not equal the averages of at least two groups are not equals.

Therefore, under H_0 the average of ranks in each group equal to the overall average as

$$E(R_g) = E(R) = 0.5(n+1), \qquad g = 1, \ldots, G$$

The over-mean-rank functions can be defined as

$$\pi = p(R > E(R))$$

and

$$\pi_g = p(R_{gi} > E(R)), \qquad i = 1,2, \ldots, n_g$$

It is clearly that under H_0

$$\pi_1 = \pi_2 = \cdots = \pi_G = \pi = 0.5$$

Note that

$$\pi = \begin{cases} \dfrac{\#(R > 0.5(n+1))}{n} = \dfrac{1}{2}, & \text{if } n \text{ is even} \\ \dfrac{\#(R > 0.5(n+1))}{n-1} = \dfrac{1}{2}, & \text{if } n \text{ is odd} \end{cases}$$

Therefore the null hypothesis

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_G = \mu$$

is equivalent to

$$H_0: \pi_1 = \pi_2 = \cdots = \pi_G = \pi = 0.5$$

Consequently, the proposed test for equal means is

$$\mathrm{E}^2 = \sum_{g=1}^{G} \left(\frac{\hat{\pi}_g - 0.5}{\sqrt{0.25/n_g}} \right)^2$$

Since $\hat{\pi}_g$ is a sample mean, if n_g is large, the central limit theorem allows to approximate

$$\mathrm{E} = \frac{\hat{\pi}_g - 0.5}{\sqrt{0.25/n_g}} \approx N(0,1)$$

Consequently, E^2 can be approximated by chi-square distribution with $G-1$ degrees of freedom, therefore,

$$\mathrm{E}^2 \approx \chi^2(G-1)$$

The approximate size α rejection is $\mathrm{E}^2 \geq \chi^2_{\alpha,G-1}$.
To prove this let

$$H = \sum_{g=1}^{G} \left(\frac{\hat{\pi}_g - \pi}{\sigma_{\hat{\pi}_g}} \right)^2 = \sum_{g=1}^{G} \left(\frac{\hat{\pi}_g - \overline{\hat{\pi}_g} + \overline{\hat{\pi}_g} - \pi}{\sigma_{\hat{\pi}_g}} \right)^2 \approx \chi^2(G)$$

Since

$$\sum_{g=1}^{G} (\hat{\pi}_g - \overline{\hat{\pi}_g}) = 0$$

Then

$$\sum_{g=1}^{G} \left(\frac{\hat{\pi}_g - \pi}{\sigma_{\hat{\pi}_g}} \right)^2 = \sum_{g=1}^{G} \left(\frac{\hat{\pi}_g - \overline{\hat{\pi}_g}}{\sigma_{\hat{\pi}_g}} \right)^2 + G \left(\frac{\overline{\hat{\pi}_g} - \pi}{\sigma_{\hat{\pi}_g}} \right)^2$$

Hence,

$$\sum_{g=1}^{G} \left(\frac{\hat{\pi}_g - \overline{\hat{\pi}_g}}{\sigma_{\hat{\pi}_g}} \right)^2 = \chi^2(G) - \chi^2(1) = \chi^2(G-1)$$

Under H_0

$$\sum_{g=1}^{G} \left(\frac{\hat{\pi}_g - 0.5}{0.25/\sqrt{n_g}} \right)^2 \approx \chi^2(G-1)$$

The estimated over-mean-rank function for each group can be obtained as

$$\hat{\pi}_g = \frac{\#\big(R_{gi} > 0.5(n+1)\big)}{n_g}, \qquad i = 1,2,\ldots,n_g$$

For $g = 1,2,\ldots,G$. Note that, when n is even, it can use $n_g - 1$ instead of n_g in a group that contains rank $(n+1)/2$. Also if there are ties only equal $0.5(n+1)$ it can make half of them less than $0.5(n+1)$ and other half more than $0.5(n+1)$.

Table 1 gives the empirical four moments of E^2 using data simulated from normal and exponential distributions with different sample sizes and variances along with the first theoretical four moments of chi square distribution. Actually, the chi square distribution gives a very good fit to E^2.

Table 1 empirical mean, variance, skewness and kurtosis fo E^2 using data simulated from normal and exponential distributions with different sample sizes and variances.

				$G = 4$							
	Parameters			normal			|Exponential				
n_g	n	$variances$	$mean$	$var.$	$sk.$	$ku.$	$mean$	$var.$	$sk.$	$ku.$
(5,7,9,10)	31	(25,49,64,144)	3.21	6.02	1.26	5.10	3.20	6.05	1.27	5.08
(10,15,17,20)	62	(25,49,64,144)	3.11	6.17	1.59	6.82	3.17	6.55	1.50	6.17
(20,26,29,30)	105	(25,49,64,144)	3.09	6.08	1.60	6.85	3.16	6.50	1.58	6.53
(5,7,9,10)	31	(144,64,49,25)	3.19	5.97	1.38	5.36	3.22	6.11	1.36	5.50
(10,15,17,20)	62	(144,64,49,25)	3.10	6.13	1.62	7.33	3.20	6.60	1.51	6.15
(20,26,29,30)	105	(144,64,49,25)	3.07	6.10	1.59	6.82	3.19	6.55	1.57	6.34
				First four χ_3^2 moments						
			3	6	1.63	7				

2.2 Graphical display (Over-mean-rank plot)

This is a graph for each group and consists of:

1. X-axis represents the index for group size.
2. Y-axis represents the ranks in each group.
3. The middle line at $0.5(n + 1)$

This graph should reflect the following information

1. $\hat{\pi}_g$ the percentage of ranks above the middle line in each group. If this value is more than have 0.5 that is indication of the shifting up in mean of this group and vice versa.
2. The χ^2 value that gives the contribution of each group in the test.
3. Patterns among the groups.

Figure 1 shows over-mean-rank plot for simulated data from normal distribution with four groups, same means (0,0,0,0) and different variances:

1. Values of $\hat{\pi}_g$ are near from each other and 0.5.
2. The χ^2 values are very small and group 2 has the most contribution in the test 0.06.
3. The four groups have almost the same patterns. It might conclude there are no significance differences among groups.

While Figure 2 shows over-mean-rank plot for simulated data from normal distribution with four groups, means $(0,0,5,0)$ and different variances:

1. The highest value of $\hat{\pi}_g$ is 0.8 for group 3 and the lowest value is 0.35 for group 4.
2. The χ^2 value for group 3 has the highest contribution in the test followed by group 4.
3. Groups 1 and 2 similar in patterns while groups 3 and 4 are in reverse pattern.

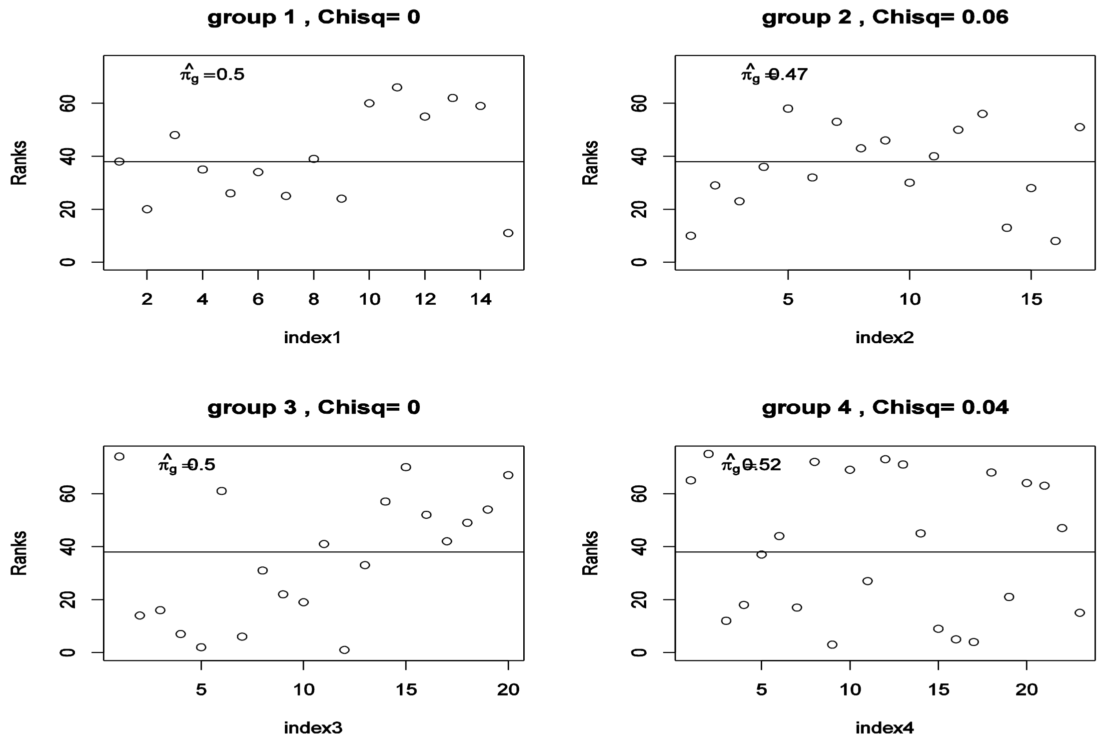

Figure 1 over-mean-rank plot with the over-mean line using simulated data from normal distribution with mean (0,0,0,0) and variances (9,49,64,100) and the sample sizes are (15,17,20,23).

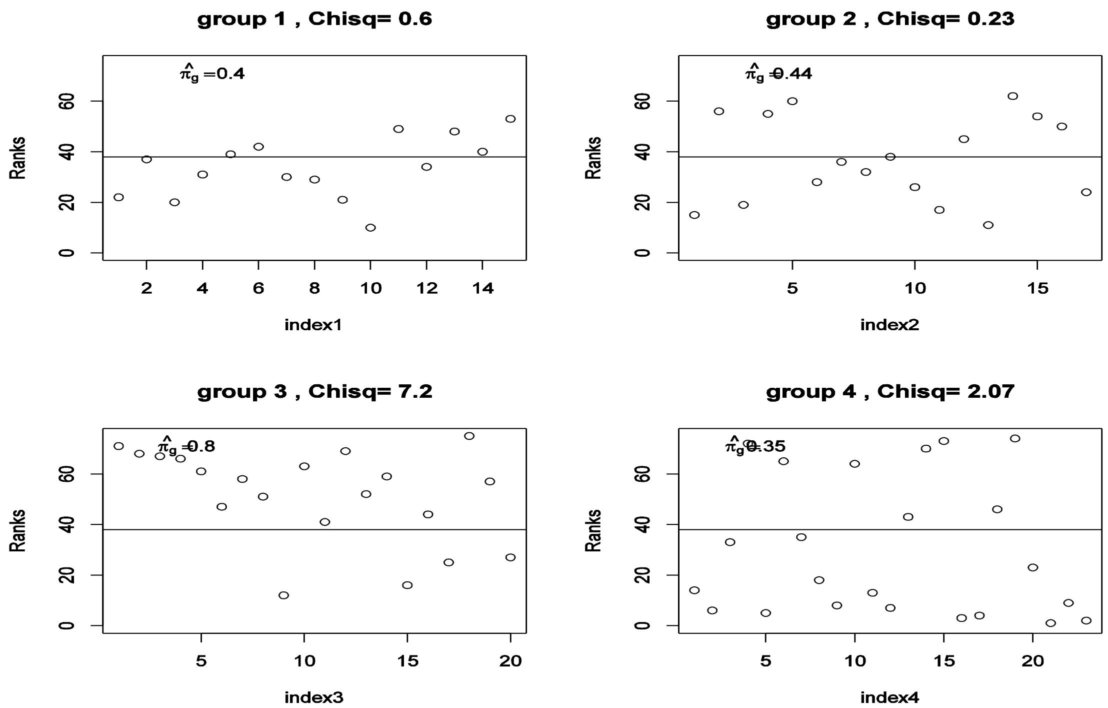

Figure 2 over-mean-rank plot using simulated data from normal distribution with mean (0,0,5,0) and variances (9,49,64,100) and the sample sizes are (15,17,20,23).

3 Simulation results

The Welch-statistic can be defined as

$$F_W = \frac{\sum_{g=1}^{G} w_g\left(\bar{X}_g - \bar{X}\right)^2/(g-1)}{1 + \frac{2(g-2)}{(g^2-1)}\sum_{g=1}^{G}\frac{\left(1-w_j/W\right)^2}{n_g-1}}$$

where $w_j = n_g/s_g^2$, $\bar{X} = \sum_{g=1}^{G} w_g\bar{X}_g/W$, $W = \sum_{g=1}^{G} w_g$.

The test statistic is approximately distributed as an F variate and is referred to the critical value $F[(1-\alpha);(g-1),\nu]$

$$\nu = \frac{(g^2-1)}{3\sum_{g=1}^{G}\frac{\left(1-w_j/W\right)^2}{n_g-1}}$$

See; for example, [9] and [10].

It is well known that the test statistic for one-way fixed effect ANOVA is

$$F = \frac{SSB/(G-1)}{SSW/(n-G)} = \frac{MSB}{MSW}$$

Where SSB is sum of squares between groups, SSW is the sum of squares within treatments, $(G-1)$ and $(n-G)$ are the degrees of freedom, between and within treatments, respectively; see, [11].

The Kruskal-Wallis test is defined as

$$KW = \frac{(n-1)\sum_{g=1}^{G} n_g\left(\bar{r}_g - \bar{r}\right)^2}{\sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(r_{gi} - \bar{r}\right)^2} = \frac{12}{n(n+1)}\sum_{g=1}^{G} n_g\left(\bar{r}_g - \frac{n+1}{2}\right)^2$$

where $\bar{r}_g = \frac{\sum_{g=1}^{n_g} r_{gi}}{n_g}$ and $\bar{r} = 0.5(n+1)$.

This is distributed as $\chi^2(G-1)$; See, [12] and [1].

The random variable Y is said to have Variance-Gamma (VG) with parameters $c, \theta \in R$, $\nu, \sigma > 0$, if it has probability density function given by

$$f(y; c, \sigma, \theta, \nu) = \frac{2e^{\frac{\theta(y-c)}{\sigma^2}}}{\sigma\sqrt{2\pi}\nu^{\frac{1}{\nu}}\Gamma\left(\frac{1}{\nu}\right)}\left[\frac{|y-c|}{\sqrt{\frac{2\sigma^2}{\nu}+\theta^2}}\right]^{\frac{1}{\nu}-1} K_{\frac{1}{\nu}-\frac{1}{2}}\left[\frac{|y-c|\sqrt{\frac{2\sigma^2}{\nu}+\theta^2}}{\sigma^2}\right], \qquad y \in R$$

Where $K_\nu(x)$ is a modified Bessel function of the third kind; see, for example, [13] and [14].

Note that there are other versions of this distribution available but this version is chosen because there is a software package in R called *gamma-variance* based on this version that be used to obtain all the simulations. The moments of this distribution are

$$E(Y) = c + \theta,$$
$$V(Y) = \sigma^2 + \nu\theta^2,$$
$$sk = \frac{2\theta^3\nu^2 + 3\sigma^2\theta\nu}{\sqrt{(\theta^2\nu + \sigma^2)^3}},$$

and

$$ku = 3 + \frac{3\sigma^4\nu + 12\sigma^2\theta^2\nu^2 + 6\theta^4\nu^3}{(\theta^2\nu + \sigma^2)^2}$$

This distribution is defined over the real line and has many distributions as special cases or limiting distributions such as Gamma distribution in the limit $\sigma \downarrow 0$ and $c = 0$, Laplace distribution as $\theta = 0$ and $\upsilon = 2$ and normal distribution as $\theta = 0, \nu = 1/r$ and $r \to \infty$.

All the simulation results use the variance gamma distribution with different choices of parameters where it can control the skewness and kurtosis by different choices of its parameters. Table 2 gives the parameters of the variance-gamma distribution used in this study.

Table 2 variance-gamma distribution parameters used in the study

		Parameters			
c	θ	σ^2	ν	skewness	Kurtosis
0	0	variances	0.001	0	3
0	0	Variances	4	0	15
0	0	variances	6	0	21
0	3	Variances	0.90	1	6
0	2	Variances	12	6	12

Four variables were manipulated in the study: (a) number of groups (4 and 5), (b) sample size (small-medium- large), (c) population distribution (variance-gamma distribution), and (d) degree/pattern of variance heterogeneity (moderate and large/all (mostly) unequal). Variances and group sizes were both positively and negatively paired. For each design size, three sample size cases were investigated. In our unbalanced designs, the smaller of the three cases investigated for each design has an average group size of less than 10, the middle has an average group size less than 20 while the larger case in each design had an average group size less than 30. With respect to the effects of distributional shape on Type I error, we chose to investigate conditions in which the statistics were likely to be prone to an excessive number of Type I errors as well as a normally distributed case. For positive (negative) pairings, the group having the smallest number of observations was associated with the population having the smallest (largest) variance, while the group having the greatest number of observations was associated with the population having the greatest (smallest) variance. These conditions were chosen since they typically produce distrorted Type I error rates; see, [9].

Table 3 empirical rates of type I error ($G = 4$), EE for E^2 test, KW for Kurskal-Wallis test, W for Welch test and ANOVA for analysis of variance test.

		$G = 4$					
Sample sizes	n	Group variances	$(Sk., Ku.)$	EE	KW	Welch	ANOVA
			Symmetric				
(5,7,10,13)	35	(100,324,400,625)	(0,3)	0.052	**0.030**	0.046	**0.030**
(5,7,10,13)	35	(625,400,324,100)	(0,3)	0.054	**0.077**	0.057	**0.118**
(15,17,20,23)	75	(100,324,400,625)	(0,3)	0.053	0.042	0.047	0.041
(15,17,20,23)	75	(625,400,324,100)	(0,3)	0.054	0.065	0.051	**0.078**
(25,27,31,34)	117	(100,324,400,625)	(0,3)	0.053	0.055	0.052	0.051
(25,27,31,34)	117	(625,400,324,100)	(0,3)	0.057	0.059	0.049	0.056
(5,7,10,13)	35	(100,324,400,625)	(0,15)	0.051	0.035	**0.019**	**0.021**
(5,7,10,13)	35	(625,400,324,100)	(0,15)	0.049	0.055	**0.017**	**0.090**
(15,17,20,23)	75	(100,324,400,625)	(0,15)	0.053	0.043	**0.030**	0.036
(15,17,20,23)	75	(625,400,324,100)	(0,15)	0.047	0.056	**0.028**	**0.069**
(25,27,31,34)	117	(100,324,400,625)	(0,15)	0.055	0.046	0.037	0.041
(25,27,31,34)	117	(625,400,324,100)	(0,15)	0.050	0.052	0.036	**0.064**
			Asymmetric				
(5,7,10,13)	35	(150,75,60,50)	(1,6)	0.054	0.064	0.043	**0.091**
(5,7,10,13)	35	(50,60,75,150)	(1,6)	0.051	0.035	0.035	**0.030**
(15,17,20,23)	75	(150,75,60,50)	(1,6)	0.048	0.054	0.043	**0.072**
(15,17,20,23)	75	(50,60,75,150)	(1,6)	0.052	0.045	0.042	0.041
(25,27,31,34)	117	(150,75,60,50)	(1,6)	0.055	0.054	0.045	**0.066**
(25,27,31,34)	117	(50,60,75,150)	(1,6)	0.052	0.047	0.045	0.045
(5,7,10,13)	35	(150,75,60,50)	(6,60)	0.049	0.058	**0.008**	0.057
(5,7,10,13)	35	(50,60,75,150)	(6,60)	0.057	0.041	**0.009**	**0.025**
(15,17,20,23)	75	(150,75,60,50)	(6,60)	0.055	0.065	**0.025**	0.049
(15,17,20,23)	75	(50,60,75,150)	(6,60)	0.057	0.056	**0.021**	0.035
(25,27,31,34)	117	(150,75,60,50)	(6,60)	0.058	**0.069**	0.044	0.055
(25,27,31,34)	117	(50,60,75,150)	(6,60)	0.056	0.064	0.039	0.042

To evaluate the particular conditions under which a test was insensitive to assumption violations, the idea of Bradley's (1978) of robustness criterion was employed. According to this criterion, in order for a test to be considered robust, its empirical rate of Type I error α must be contained in the interval $\alpha \pm \varepsilon$. The choice of Bradley is $\varepsilon = 0.25$ and this makes the interval is liberal. Therefore, in this study the choice of $\varepsilon = 0.15$ a something in the middle between nothing and 0.25. Therefore, for the five percent level of significance used in this study, a test was considered robust in a particular condition if its empirical rate of Type I error fell within the interval $0.35 \leq \hat{\alpha} \leq 0.65$. Correspondingly, a test was considered to be nonrobust if, for a particular condition, its Type I error rate was not contained in this interval. Nonetheless, there is no one universal standard by which tests are judged to be robust, so different interpretations of the results are possible. In the tables, boldfaced entries are used to denote these latter values.

Table 4 empirical rates of type I error ($G = 5$), EE for E^2 test, KW for Kurskal-Wallis test, W for Welch test and ANOVA for analysis of variance test

Sample sizes	n	Group variances	$G = 5$ $(S., K.)$	EE	KW	W	ANOVA
			Symmetric				
(5,7,10,13,14)	49	(150,75,60,51,50)	(0,15)	0.044	0.048	**0.018**	**0.086**
(5,7,10,13,14)	49	(50,51,60,75,150)	(0,15)	0.048	0.038	**0.018**	**0.029**
(15,17,20,23,24)	99	(150,7 5,60,51,50)	(0,15)	0.052	0.057	**0.030**	**0.068**
(15,17,20,23,24)	99	(50,51,60,75,150)	(0,15)	0.052	0.044	**0.027**	0.040
(25,27,31,33,34)	150	(150,75,60,51,50)	(0,15)	0.049	0.052	0.042	0.056
(25,27,31,33,34)	117	(50,51,60,75,150)	(0,15)	0.049	0.048	0.036	0.049
			Asymmetric				
(5,7,10,13,14)	49	(150,75,60,51,50)	(6,60)	0.053	0.065	**0.005**	0.063
(5,7,10,13,14)	49	(50,51,60,75,150)	(6,60)	0.054	0.045	**0.007**	**0.029**
(15,17,20,23,24)	99	(150,75,60,51,50)	(6,60)	0.058	0.064	**0.029**	0.055
(15,17,20,23,24)	99	(50,51,60,75,150)	(6,60)	0.057	0.059	**0.024**	0.038
(25,27,31,33,34)	150	(150,75,60,51,50)	(6,60)	0.058	**0.072**	0.046	0.056
(25,27,31,33,34)	117	(50,51,60,75,150)	(6,60)	0.059	0.067	0.038	0.043
		Symmetric with very small sizes					
(5,6,7,8,9)	35	(225,144,25,9,4)	(0,3)	**0.066**	**0.101**	0.056	**0.121**
(5,6,7,8,9)	35	(4,9,25,144,225)	(0,3)	**0.075**	0.044	0.051	0.051
(5,6,7,8,9)	35	(1,2,3,4,5)	(0,3)	0.052	**0.033**	0.052	0.037
(5,6,7,8,9)	35	(5,4,3,2,1)	(0,3)	0.049	0.063	0.056	**0.091**
(5,6,7,8,9)	35	(225,144,81,49,36)	(0,3)	0.052	**0.071**	0.060	**0.121**
(5,6,7,8,9)	35	(225,144,81,49,36)	(0,21)	0.046	0.045	**0.011**	**0.060**
(5,6,7,8,9)	35	(5,4,3,2,1)	(0,21)	0.048	0.044	**0.009**	0.052
(5,5,5,5,5)	25	(10,10,10,10,10)	(0,3)	0.055	0.038	0.047	0.052
(5,5,5,5,5)	25	(10,10,10,10,10)	(0,21)	0.056	0.037	**0.008**	**0.025**

Tables 3 and 4 contain empirical rates of Type I error for a design containing four and five groups, respectively. The tabled data indicates that

1. When the observations were obtained from normal distributions (Table 3, sk.=0 and ku.=3), rates of Type I error were controlled by EE, KW and W methods while were not controlled by ANOVA where the variances were not equals.
2. When the observations were obtained from symmetric distributions with kurtosis more than 3 (Table 3 and Table 4, sk.=0 and ku.=15), rates of Type I error were controlled by EE and KW methods while were not controlled by W and ANOVA methods.
3. When the observations were obtained from non-normal distributions (Table 3 and Table 4, sk.=1 and 6 and ku.=6 and 60), rates of Type I error were controlled by EE, KW while were not controlled by W and ANOVA methods.
4. When the observations were obtained from normal with equal variances and sample sizes are equal the rates of Type I error was controlled by ANOVA method (Table 4).
5. In small sample sizes, the EE method rates of Type I error were controlled by EE better than KW method, Table 4.
6. Rates of Type I error were controlled by EE for all cases studied except two cases out of 45 cases, while rates were controlled by KW for all cases studied except seven cases out of 45.

4 Application

The RS company provides several services. It currently operates in four regions (M1, M2, M3 and M4) . recently, RW manager questioned whether the mean billing amount for the services differed by region. Simple random samples of employees served in these regions have been selected. The data are given in Table 5

Table 5 billing amount for the services in four regions with ranks in brackets

	M1	M2	M3	M4
	102.3 (36)	95.5 (21)	103.5 (39)	70.6 (2)
	101.5 (34)	99.3 (25)	103.1 (37)	69.7 (1)
	100.7 (29)	101.5 (34)	117.6 (47)	83.8 (9)
	98.1 (23)	100.3 (27)	87.9 (10)	91.9 (15)
	101.4 (32)	101.5 (34)	100.4 (28)	109.8 (45)
	100.9 (30)	93.1 (18)	104.7 (42)	88.6 (11)
	92.9 (17)	92.7 (16)	83.4 (8)	98.6 (24)
	101.3 (31)	94.4 (19)	91.7 (14)	74.4 (3)
	100.2 (26)	109.9 (46)	88.9 (12)	94.6 (20)
	104.7 (41)	96.6 (22)	103.2 (38)	75.9 (4)
		104.3 (40)	108.3 (44)	83.1 (7)
		105.5 (43)		81.3 (6)
				89.2 (13)
				80.1 (5)
sizes	10	12	11	14
Mea.	100.4	99.5	99.3	85.1
Var.	9.6	28.5	104.2	126.9

Table 5 gives the data and its ranks. Figure 3 shows the over-mean-rank plot for the data and in can conclude the following:

1. The lowest value of $\hat{\pi}_g$ is 0.08 for group 4 that showing shifting down in this group and the heist $\hat{\pi}_g$ is 0.80 for group 1 that showing up in this group.
2. Group 4 has the highest contribution in the test where its χ^2 value 9.17, followed by group 1 that has 3.6.
3. Group 4 is different in patterns with groups 1, 2 and 3. Groups 2 and 3 are nearest in patterns.

To test for equal mean the null hypothesis is

$$\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.5$$

Where $\mathrm{E}^2 = 13.94 > \chi^2(0.95,3) = 7.81,$ therefore, H_0 is rejectd.

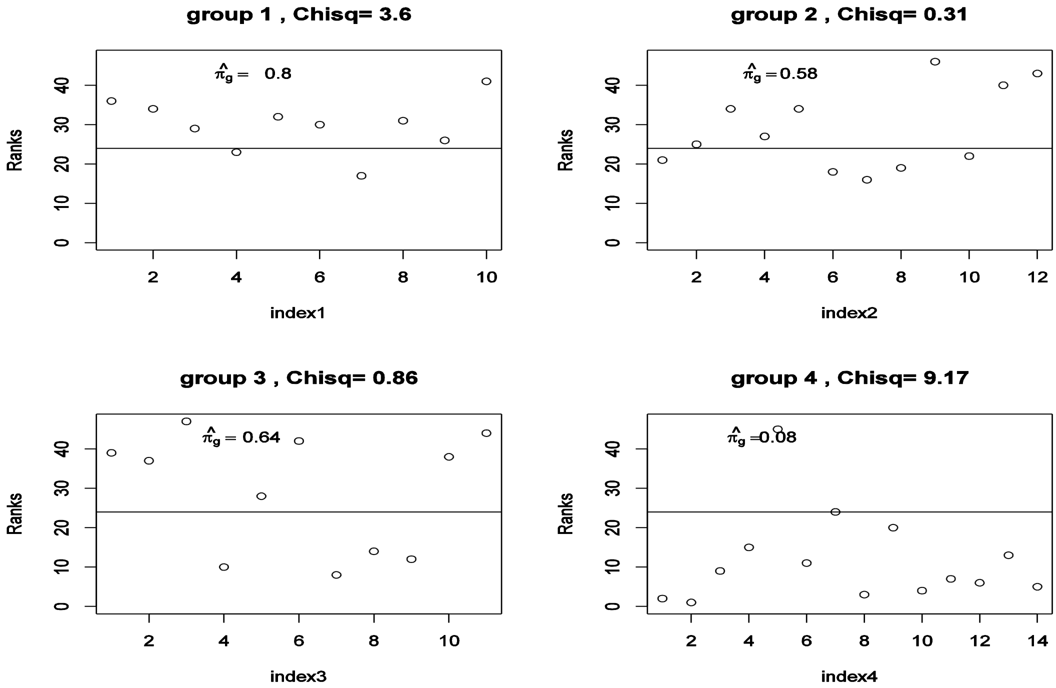

Figure 3 over-mean-rank plot for RW company data

5 Conclusion

Comparison of several means under heterogeneity is studied using over-mean-rank approach. The sampling distribution for this function was obtained and found that the chi square distribution had given a very good fit for this function until for small sample sizes.

Comparison with Kruskal-Wallis, Welch and ANOVA methods had been given for unbalanced designs and not equal variances from normal and non-normal populations in terms of Type I error and the simulation results were shown that the proposed method improved the Type I error and its performance had better robustness than the studied methods.

This approach might be extended to a multiple comparisons procedure. For example, the pair comparisons of averages can be obtained as

$$H_0: \mu_g = \mu_k, \quad g \neq k$$

or equivalently

$$H_0: \pi_g = \pi_k$$

The first approach is Behrens-Fisher approach. Following [15] procedures, the paired comparisons might be done using

$$W_r = \hat{\pi}_g - \hat{\pi}_k = q_{1-\alpha}(r, df_1)\sqrt{\frac{\hat{\pi}_g(1-\hat{\pi}_g)}{2n_g} + \frac{\hat{\pi}_k(1-\hat{\pi}_k)}{2n_k}}$$

where

$$df_1 = \frac{\left[\frac{\hat{\pi}_g(1-\hat{\pi}_g)}{n_g} + \frac{\hat{\pi}_k(1-\hat{\pi}_k)}{n_k}\right]^2}{\frac{\left(\hat{\pi}_g(1-\hat{\pi}_g)/n_g\right)^2}{n_g-1} + \frac{(\hat{\pi}_k(1-\hat{\pi}_k)/n_k)^2}{n_k-1}}$$

where r comparison and $q_{1-\alpha}(r, df_1)$ is studentized range statistics; see, [16].
Another approximation is a family wise error rate as

$$z_{gk} = \frac{\hat{\pi}_g - \hat{\pi}_k}{\sqrt{\frac{\hat{\pi}_g(1-\hat{\pi}_g)}{n_g} + \frac{\hat{\pi}_k(1-\hat{\pi}_k)}{n_k}}}$$

and comparing it to $\ddot{z} = z_{[\alpha/k(k-1)]}$, the $[\alpha/k(k-1)]$upper standard normal quantile. The quantity $[\alpha/k(k-1)]$ called the experiment wise error rate or the overall significant level, which is the probability of at least one erroneous rejection among the $k(k-1)/2$ pairwise comparisons; see, [17], [19] and [20].
These approaches need more study and comparisons with other methods and it will be left to another research.

References

[1] W.H. Kruskal, and W.A. Wallis, "Use of ranks in one-criterion variance analysis" Journal of American Statistical Association, 47, 1952,583–621.

[2] B.L.Welch,"On the comparison of several mean values: an alternative approach" Biometrika, 38, 1951 933-943.

[3] G.S. James, "The comparison of several groups of observations when the ratios of the population variances are unknown. Biometrika, 38, 1951, 324-329.

[4] G.S. James, "Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown" Biometrika, 41, 1954, 19-43.

[5] J. Gamage, and S. Weerahandi, "Size performance of some tests in one-way ANOVA" Communications in Statistics-Simulation and Computation, XX, 1998, 625-639.

[6] L.M. Lix, and H.J. Keselman, "Approximate degrees of freedom tests: A unified perspective on testing for mean equality," Psychological Bulletin, 117, 1995,547-560.

[7] S. Weerahandi, "ANOVA under unequal error variances," Biometrics, 51, 1995, 589-599.

[8] S., Chen and H.J. Chen, H.J., "Single-stage analysis of variance under heteroscedasticity," Communications in Statistics-Simulation and Computation, XX, 1998, 641-666.

[9] L.M. Lix, and H.J. Keselman, "To trim or not to trim: Tests of mean equality under heteroscedasticity and nonnormality" Educational and Psychological Measurement, 58, 1998,409-429 (Errata:58, 853).

[10] R. Wilcox, H. Keselman, and R. Kowalchuk "Can tests for treatment group equality be improved?: The bootstrap and trimmed means conjecture," British Journal of Mathematical and Statistical Psychology, 51, 1998, 123-134.

[11] J. Neter, H. Kutner, C. Nachtsheim and W. Wasserman, "Applied linear statistical models" 4th ed., McGraw-Hill, 1996.

[12] W.H. Kruskal, "A nonparametric test for the several sample problem," Annals of Mathematical Statistics, 23, 1952,525–540.

[13] E. Seneta,"Fitting the variance-gamma model to financial data" Journal of Applied Probability. 41A, 2004,177-187.

[14] S. Kotz, T.J. Kozubowski, and K. Podgórski, "The Laplace Distribution and Generalizations" *Birkhauser*, Boston, 2001.

[15] P.A. Games, and J.F. Howell, "Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carlo study" Journal of Educational Statistics,1, 1976, 113-125.

[16] E.J. Dudewicz,Y. Ma,, E. Mai and H.Su, "Exact solutions to the Behrens–Fisher Problem: Asymptotically optimal and finite sample efficient choice among" Journal of Statistical Planning and Inference. 137, 2007, 1584 – 1605.

[17] O.J. Dunn, "Multiple comparisons using rank sums" Technometrics, 6, 1964, 241-252.

[18] R. Rosenthal, and R.L. Rosnow, "Contrast analysis: focused comparisons" Boston: Cambridge University Press; 1985.

[19] H. Abdi, "The Bonferonni and Šidák corrections for multiple comparisons" In: Neil Salkind (Ed.), Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage, 2007.