



Article

Empirical Likelihood-Based ANOVA for Trimmed Means

Mara Velina ¹, Janis Valeinis ^{1,*}, Luca Greco ² and George Luta ³

¹ Department of Mathematics, Faculty of Physics and Mathematics, University of Latvia, Riga LV-1002, Latvia; mara.velina@lu.lv

² Department of Law, Economics, Management and Quantitative Methods, University of Sannio, Benevento 82100, Italy; luca.greco@unisannio.it

³ Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC 20057, USA; George.Luta@georgetown.edu

* Correspondence: janis.valeinis@lu.lv; Tel.: +371-67033711

Academic Editor: Igor Burstyn

Received: 17 May 2016; Accepted: 20 September 2016; Published: 27 September 2016

Abstract: In this paper, we introduce an alternative to Yuen's test for the comparison of several population trimmed means. This nonparametric ANOVA type test is based on the empirical likelihood (EL) approach and extends the results for one population trimmed mean from Qin and Tsao (2002). The results of our simulation study indicate that for skewed distributions, with and without variance heterogeneity, Yuen's test performs better than the new EL ANOVA test for trimmed means with respect to control over the probability of a type I error. This finding is in contrast with our simulation results for the comparison of means, where the EL ANOVA test for means performs better than Welch's heteroscedastic F test. The analysis of a real data example illustrates the use of Yuen's test and the new EL ANOVA test for trimmed means for different trimming levels. Based on the results of our study, we recommend the use of Yuen's test for situations involving the comparison of population trimmed means between groups of interest.

Keywords: ANOVA; empirical likelihood; trimmed means; robust statistics; hypothesis testing

1. Introduction

The comparison of the means of several populations is frequently encountered in the statistical analysis of data from environmental research and public health studies. Typically, ANOVA is used to compare these means of interest, for example, for the comparison of means of blood lead levels between groups of children receiving different interventions. Practical situations may involve complications such as unbalanced designs (i.e., unequal sample sizes for the groups), variance heterogeneity, and departures from normality. It may be the case, for instance, that the distributions underlying the data from each group are truly heavy tailed or skewed, but it is also possible that such departures from normality are due to few observations located away from the bulk of the data in the tails of the distribution. It is well-known that the classical ANOVA F test cannot handle such violations of its assumptions, and, as a consequence, it has problems controlling the probability of the type I error at the specified nominal level. Heteroscedasticity and/or outliers can completely break down the results of the ANOVA F test when not properly taken into account (see, for example, [1]). Given this limitation of the ANOVA F test, there is a need for ANOVA type tests that are robust to both heteroscedasticity and outliers.

A statistical test that satisfies these requirements is the test developed by Yuen [2], who proposed a modified version of Welch's heteroscedastic F test [3]. The latter test is designed to deal with heteroscedasticity for normally distributed data, and it is using the sample means and sample variances

to estimate their population counterparts. Since the sample mean and the sample variance are not robust to outliers, Yuen [2] proposed to replace them with a pair of robust estimators consisting of the trimmed mean and the Winsorized variance. Such an approach provides a better control of the probability of the type I error for one-way ANOVA situations involving unbalanced designs and skewed distributions (see [4]). There are two important comments to be made. The first comment is that the construction of Yuen's test has a somewhat ad hoc nature, by replacing the least squares estimators with robust versions. The second comment is that Yuen's test is no longer a test for the comparison of populations means, but, rather, it is a test to compare population trimmed means. It may be preferable to make inferences regarding the population trimmed means rather than the population means when the underlying distributions for the groups are skewed, since the trimmed means are more representative for the bulk of the data in those situations [5].

In this paper, we present an alternative to Yuen's test, a new nonparametric test that can be used to compare several trimmed means based on the empirical likelihood (EL) approach to statistical inference [6–8]. The EL method (see [9] for a detailed overview) is a popular nonparametric approach that does not require normality (or other distributional assumptions) and can be regarded as a data adaptive method. We develop an EL-based ANOVA test for the comparison of trimmed means that takes advantage of the nonparametric nature of the EL approach, by extending the results of Qin and Tsao [8] who introduced the EL method for a trimmed mean (see also the results from [10]). All technical details regarding the tests considered in this paper (including the asymptotic results for the new EL-based ANOVA for trimmed means) are provided in the Appendices A–D.

The paper is organized as follows. In Section 2, we present and interpret the results of a simulation study that compares the performance of the EL-based ANOVA for trimmed means and means with alternative methods under several scenarios involving skewed distributions. In Section 3, we analyze a real data set using different types of tests for the comparison of population trimmed means and population means. We end the paper by presenting conclusions in Section 4.

2. Simulation Study

For simplicity, we present only situations where we are interested in the comparison of three population trimmed means or three population means ($k = 3$), while having samples of equal sizes. We consider scenarios involving skewed distributions, with and without variance heterogeneity. For the EL ANOVA for trimmed means, we consider only *symmetric* trimming, where all samples are trimmed symmetrically. We note that, although we are primarily interested in the performance of the tests for the comparison of trimmed means, EL ANOVA for trimmed means (panel *ELT*) and Yuen's test (panel *Yuen*); for completeness purposes, we are also including the results for the tests for the comparison of means, specifically the classical ANOVA *F* test (panel *F test*), Welch's heteroscedastic *F* test for means (panel *Welch*), and the EL ANOVA for means (panel *EL*). For Welch's test and Yuen's test, we have used the R function `t1way` (see Wilcox [11]). The R functions that provide the implementation of the EL ANOVA methods for trimmed means and means are available from the corresponding author upon request.

For the simulation study, we investigate the potential effect of the shape of the distributions on the estimated probability of type I errors. We consider several skewed distributions with and without variance heterogeneity. We use a simulation design similar to that from [5], where (trimmed) means of only two independent skewed populations are compared. For the scenario with homogeneous variances (scenario 1), we simulate data from three independent skewed distributions. We consider the χ_3^2 distribution, the lognormal distribution with normal mean $\mu = 0$ and normal scale $\sigma = 1$, the gamma distribution with shape parameter $\alpha = 2$ and scale parameter $\sigma = 1$, and the skew-normal distribution with location parameter $\zeta = 0$, scale parameter $\omega = 1$, and slant parameter $\alpha = 1$ (see [12]). For the scenario with heterogeneous variances (scenario 2), we further transform the data simulated from the three independent skewed distributions as to have the ratios between variances to be either 1:4:9 or 1:1:36. To ensure that the relevant H_0^T of equal trimmed means or H_0 of equal means are

true, before altering the variances, we center the data using the theoretically determined trimmed means (when using tests for the comparison of trimmed means) or means (when using tests for the comparison of means). We use 10,000 Monte Carlo simulations to calculate the empirical probability of type I errors for the tests performed at the nominal 0.05 significance level.

Table 1 presents the empirical probability of type I errors for the different tests for the situation involving skewed distributions with homogeneous variances (scenario 1). Regarding the comparison of trimmed means, the results for Yuen’s test are closer to the nominal significance level than those for the EL ANOVA test for trimmed means. By contrast, among the tests that compare means, the results of the EL ANOVA test for means are closest to the nominal significance level. Tables 2 and 3 present the corresponding results for the same tests for situations involving skewed distributions with heterogeneous variances (scenario 2). We note that it is more difficult to control the probability of a type I error when the ratios between variances are 1:1:36 than when they are 1:4:9. Similar to the homogeneous variances scenario, the results for the heterogeneous variances scenario suggest that Yuen’s test performs best among the tests for the comparison of trimmed means, while the EL ANOVA test performs best among the tests for the comparison of means.

Table 1. Empirical probability of type I error for various tests for the equality of means and trimmed means of three independent skewed distributions with homogeneous variances. For methods involving trimmed means, symmetric trimming at level $\alpha_i = \beta_i = c, i = 1, 2, 3$ is used.

χ^2_3									
Trimming level									
$c = 5\% \quad c = 10\% \quad c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.047	0.076	0.052	0.050	0.079	0.050	0.080	0.049	0.090
30	0.048	0.070	0.054	0.054	0.055	0.053	0.075	0.054	0.079
40	0.044	0.062	0.052	0.049	0.063	0.048	0.063	0.050	0.069
50	0.048	0.058	0.049	0.046	0.047	0.047	0.060	0.049	0.067
100	0.049	0.055	0.051	0.050	0.056	0.050	0.056	0.049	0.056
200	0.051	0.055	0.053	0.050	0.053	0.051	0.053	0.051	0.056
500	0.051	0.049	0.049	0.048	0.050	0.050	0.051	0.051	0.052
Lognormal ($\mu = 0, \sigma = 1$)									
Trimming level									
$c = 5\% \quad c = 10\% \quad c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.044	0.073	0.047	0.040	0.069	0.040	0.070	0.040	0.081
30	0.045	0.069	0.050	0.048	0.049	0.046	0.068	0.046	0.072
40	0.044	0.063	0.049	0.046	0.062	0.046	0.062	0.045	0.066
50	0.045	0.065	0.054	0.049	0.048	0.047	0.059	0.046	0.062
100	0.049	0.059	0.055	0.049	0.055	0.049	0.057	0.049	0.057
200	0.050	0.053	0.050	0.045	0.048	0.046	0.049	0.048	0.052
500	0.051	0.053	0.052	0.051	0.052	0.050	0.050	0.052	0.054
Gamma ($\alpha = 2, \sigma = 1$)									
Trimming level									
$c = 5\% \quad c = 10\% \quad c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.052	0.078	0.050	0.050	0.077	0.052	0.079	0.052	0.096
30	0.049	0.069	0.053	0.052	0.053	0.050	0.070	0.052	0.080
40	0.050	0.062	0.052	0.052	0.064	0.052	0.067	0.053	0.074
50	0.050	0.060	0.051	0.048	0.048	0.050	0.062	0.052	0.069
100	0.052	0.057	0.052	0.051	0.057	0.050	0.056	0.048	0.056
200	0.052	0.056	0.053	0.052	0.055	0.053	0.055	0.052	0.055
500	0.049	0.052	0.051	0.050	0.051	0.049	0.051	0.051	0.052

Table 1. Cont.

Skew-normal ($\zeta = 0, \omega = 1, \alpha = 1$)									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.055	0.077	0.049	0.049	0.077	0.051	0.083	0.054	0.099
30	0.048	0.065	0.050	0.051	0.051	0.050	0.068	0.051	0.078
40	0.049	0.061	0.049	0.049	0.062	0.051	0.065	0.051	0.073
50	0.051	0.058	0.049	0.049	0.050	0.050	0.061	0.052	0.071
100	0.055	0.055	0.052	0.051	0.056	0.050	0.056	0.052	0.060
200	0.052	0.051	0.049	0.050	0.052	0.051	0.054	0.052	0.056
500	0.046	0.048	0.047	0.047	0.048	0.048	0.049	0.048	0.049

Table 2. Empirical probability of type I error for various tests for the equality of means and trimmed means of three independent skewed distributions with the ratios between variances being 1:4:9. For methods involving trimmed means, symmetric trimming at level $\alpha_i = \beta_i = c, i = 1, 2, 3$ is used.

χ^2_3									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.086	0.101	0.071	0.064	0.096	0.061	0.094	0.062	0.109
30	0.083	0.088	0.071	0.060	0.062	0.064	0.084	0.064	0.091
40	0.080	0.075	0.062	0.060	0.072	0.055	0.072	0.054	0.078
50	0.079	0.067	0.057	0.050	0.051	0.051	0.066	0.052	0.071
100	0.079	0.063	0.058	0.055	0.060	0.055	0.060	0.054	0.062
200	0.085	0.060	0.057	0.057	0.059	0.054	0.058	0.054	0.058
500	0.073	0.052	0.051	0.053	0.054	0.052	0.054	0.051	0.053

Lognormal ($\mu = 0, \sigma = 1$)									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.110	0.146	0.113	0.078	0.106	0.070	0.106	0.066	0.114
30	0.109	0.131	0.111	0.063	0.064	0.065	0.088	0.063	0.090
40	0.100	0.115	0.100	0.066	0.081	0.062	0.083	0.059	0.084
50	0.098	0.110	0.099	0.060	0.060	0.060	0.074	0.059	0.077
100	0.097	0.090	0.083	0.058	0.063	0.055	0.062	0.055	0.063
200	0.082	0.071	0.069	0.051	0.055	0.052	0.055	0.054	0.058
500	0.077	0.061	0.060	0.051	0.053	0.052	0.053	0.051	0.054

Gamma ($\alpha = 2, \sigma = 1$)									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.086	0.099	0.070	0.061	0.093	0.062	0.097	0.065	0.111
30	0.083	0.080	0.063	0.058	0.062	0.056	0.079	0.060	0.090
40	0.079	0.074	0.061	0.058	0.073	0.056	0.073	0.059	0.083
50	0.079	0.068	0.057	0.049	0.050	0.053	0.066	0.058	0.075
100	0.078	0.059	0.055	0.053	0.060	0.053	0.060	0.053	0.061
200	0.078	0.057	0.054	0.053	0.055	0.053	0.057	0.053	0.057
500	0.079	0.052	0.050	0.049	0.051	0.050	0.050	0.052	0.053

Table 2. Cont.

Skew-normal ($\xi = 0, \omega = 1, \alpha = 1$)									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.080	0.079	0.048	0.048	0.081	0.052	0.085	0.054	0.106
30	0.075	0.069	0.050	0.051	0.054	0.053	0.073	0.057	0.085
40	0.080	0.060	0.049	0.050	0.064	0.050	0.069	0.053	0.075
50	0.076	0.060	0.050	0.050	0.051	0.051	0.064	0.053	0.073
100	0.079	0.053	0.049	0.048	0.054	0.048	0.055	0.051	0.059
200	0.078	0.054	0.051	0.051	0.054	0.051	0.054	0.052	0.056
500	0.075	0.049	0.047	0.048	0.049	0.048	0.049	0.048	0.049

Table 3. Empirical probability of type I error for various tests for the equality of means and trimmed means of three independent skewed distributions with the ratios between variances being 1:1:36. For methods involving trimmed means, symmetric trimming at level $\alpha_i = \beta_i = c, i = 1, 2, 3$ is used.

χ^2_3									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.124	0.090	0.067	0.059	0.087	0.056	0.088	0.056	0.102
30	0.119	0.080	0.064	0.056	0.058	0.058	0.078	0.058	0.086
40	0.116	0.070	0.058	0.055	0.067	0.053	0.067	0.052	0.072
50	0.112	0.063	0.052	0.048	0.049	0.051	0.063	0.052	0.069
100	0.111	0.061	0.055	0.053	0.059	0.051	0.058	0.050	0.058
200	0.113	0.059	0.056	0.054	0.057	0.053	0.056	0.052	0.056
500	0.102	0.053	0.053	0.053	0.054	0.050	0.052	0.050	0.052

Lognormal ($\mu = 0, \sigma = 1$)									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.168	0.126	0.101	0.071	0.095	0.062	0.095	0.058	0.103
30	0.166	0.118	0.098	0.055	0.056	0.060	0.081	0.057	0.084
40	0.153	0.104	0.089	0.060	0.076	0.061	0.077	0.057	0.080
50	0.148	0.095	0.086	0.052	0.053	0.056	0.068	0.055	0.072
100	0.136	0.080	0.075	0.054	0.061	0.053	0.062	0.054	0.061
200	0.119	0.064	0.062	0.050	0.054	0.049	0.052	0.051	0.055
500	0.112	0.056	0.055	0.052	0.053	0.050	0.051	0.052	0.054

Gamma ($\alpha = 2, \sigma = 1$)									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.123	0.089	0.064	0.058	0.089	0.059	0.093	0.060	0.107
30	0.122	0.079	0.061	0.057	0.059	0.055	0.077	0.058	0.086
40	0.116	0.071	0.057	0.056	0.069	0.054	0.070	0.055	0.077
50	0.113	0.066	0.055	0.051	0.052	0.051	0.064	0.053	0.070
100	0.110	0.059	0.053	0.053	0.058	0.052	0.057	0.051	0.060
200	0.109	0.054	0.051	0.051	0.054	0.050	0.054	0.051	0.055
500	0.108	0.052	0.051	0.049	0.050	0.050	0.051	0.050	0.051

Skew-normal ($\xi = 0, \omega = 1, \alpha = 1$)									
Trimming level									
$c = 5\%$ $c = 10\%$ $c = 20\%$									
n	F test	Welch	EL	Yuen	ELT	Yuen	ELT	Yuen	ELT
20	0.113	0.077	0.050	0.047	0.080	0.049	0.083	0.054	0.103
30	0.107	0.067	0.051	0.050	0.052	0.052	0.074	0.054	0.083
40	0.112	0.060	0.049	0.049	0.063	0.051	0.067	0.054	0.074
50	0.107	0.058	0.048	0.049	0.050	0.051	0.064	0.055	0.074
100	0.107	0.054	0.048	0.048	0.054	0.049	0.056	0.052	0.060
200	0.106	0.053	0.051	0.052	0.054	0.051	0.054	0.053	0.057
500	0.106	0.048	0.048	0.050	0.051	0.050	0.051	0.049	0.051

3. Real Data Example

To illustrate the use of the EL ANOVA for trimmed means and means, we use the Oslo Transect data set [13]. This real data set includes 360 observations corresponding to different plants collected along a 120 km transect running through the city of Oslo, Norway. The concentrations of 25 chemical elements found in these plants were recorded together with factors that may influence the mineral concentration. Except for not including two chemical elements, *Au* and *Na*, this data is available within *R* package *rrcov* [14] as *OsloTransect* dataset. We analyze this dataset, and, thus, only 23 chemical elements are included in Table 4. To preserve the skewness of the data, we have also used the raw data, as opposed to the log transformed data (as done in [13]). After removing the observations with missing values, we are left with 332 observations. We consider the 23 concentrations of chemical elements as the response variables, and the lithology as a group variable with four levels.

As for the simulation study, even though our main interest is in tests that compare population trimmed means, for completeness purposes, we also provide the results from the tests that compare population means. We consider three symmetric trimming strategies similar to those used in the simulation study. The entries from Table 4 provide the *p*-values from the tests for the comparison of population means and population trimmed means. We note that, for each trimming strategy, the *p*-values from the EL ANOVA for trimmed means (panel *ELT*) and Yuen's test (panel *Yuen*) are very similar. In addition, the *p*-values from the EL ANOVA for means (panel *EL*) and Welch's heteroscedastic *F* test (panel *Welch*) are also very similar.

Table 4. *p*-values from tests of equality of means and trimmed means of 23 chemical element concentrations in plants collected along the Oslo Transect ([13]). Symmetric trimming, $\alpha_i = \beta_i = c$, $i = 1, \dots, 4$.

Element	Trimming Level									
	F Test	Welch	EL	c = 5%					c = 20%	
				Yuen	ELT	Yuen	ELT	Yuen	ELT	
Ag	0.26	0.10	0.09	0.22	0.23	0.42	0.41	0.74	0.73	
B	0.08	0.09	0.07	0.10	0.09	0.12	0.11	0.18	0.16	
Ba	0.01	0.01	0.01	0.03	0.03	0.02	0.02	<0.01	<0.01	
Ca	0.15	0.19	0.18	0.22	0.22	0.31	0.31	0.42	0.41	
Cd	0.08	0.05	0.04	0.09	0.09	0.05	0.05	0.03	0.02	
Co	<0.01	0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	
Cr	0.17	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	
Cu	0.44	0.26	0.24	0.66	0.67	0.77	0.76	0.76	0.75	
Fe	0.03	0.02	0.01	0.04	0.04	0.02	0.02	0.04	0.03	
Hg	0.31	0.29	0.27	0.35	0.37	0.19	0.18	0.40	0.38	
K	0.47	0.28	0.26	0.50	0.52	0.53	0.52	0.58	0.57	
La	0.28	<0.01	<0.01	0.01	0.01	0.13	0.10	0.01	0.01	
Mg	0.24	0.23	0.21	0.28	0.28	0.38	0.37	0.57	0.56	
Mn	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	
Mo	0.02	<0.01	<0.01	0.02	0.02	0.04	0.03	0.17	0.15	
Ni	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.01	
P	0.28	0.25	0.24	0.39	0.40	0.43	0.43	0.58	0.57	
Pb	0.52	<0.01	<0.01	0.01	0.01	0.01	0.01	<0.01	<0.01	
S	0.58	0.55	0.54	0.70	0.72	0.78	0.78	0.81	0.81	
Sb	0.16	0.01	<0.01	0.21	0.22	0.19	0.19	0.25	0.20	
Sr	0.14	0.07	0.06	0.18	0.19	0.22	0.22	0.10	0.09	
Ti	0.01	0.01	<0.01	0.06	0.06	0.09	0.08	0.08	0.07	
Zn	0.88	0.80	0.79	0.97	0.97	0.97	0.97	0.97	0.96	

4. Conclusions

In this paper, we introduce a new nonparametric ANOVA type test for the comparison of population trimmed means. Although the new method is derived from the general principles of the empirical likelihood approach, versus the somewhat ad hoc nature of the derivation of Yuen's test from Welch's heteroscedastic *F* test, the results of our simulation study in situations involving skewed distributions indicate that, unless the sample sizes per group are very large, the new EL ANOVA method for trimmed means performs worse than Yuen's test with respect to control over the

probability of a type I error. This is in contrast with our simulation results for the comparison of means, where the EL ANOVA for means performs better than Welch's heteroscedastic F test. The analysis of the real data example provides similar p -values for the new EL ANOVA method for trimmed means and the Yuen's test for different trimming levels, and also similar p -values for the EL ANOVA and Welch's heteroscedastic F test.

Based on these results, we recommend the use of Yuen's test for situations, where the research question involves the comparison of population trimmed means between groups of interest. The choice of the specific trimming strategy is an important and complex issue, since different trimming strategies imply different null hypotheses being tested. As such, the selection of the trimming strategy should be based on subject matter reasons that take into account what is known by the experts about the data under investigation. Alternatively, in the absence of expert knowledge information, different trimming strategies could be used to evaluate the sensitivity of the results to the choice of the trimming strategy.

Acknowledgments: The authors would like to thank the academic editor and the reviewers for their thoughtful and constructive suggestions.

Author Contributions: Janis Valeinis and George Luta proposed the new EL ANOVA for trimmed means; Mara Velina proved the asymptotic results; Mara Velina and Luca Greco performed the simulation study; and Mara Velina performed the data analysis. All authors wrote and edited the paper and all authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Statistical Tests Not Based on EL

Let $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i}), i = 1, 2, \dots, k$, be independent random samples from k different distributions with population means μ_i . We are interested in testing the null hypothesis of equal population means

$$H_0 : \mu_1 = \dots = \mu_k = \mu. \quad (A1)$$

Under the assumption of equal variances (homoscedasticity) and normally distributed data in each group, i.e., $Y_{ij} \sim N(\mu_i, \sigma)$, one can use the classical ANOVA F test

$$F = \frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2 / (k-1)}{\sum_{i=1}^k (n_i - 1) s_i^2 / (N - k)},$$

where

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \text{ and } s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

are the sample mean and sample variance of the i -th group, respectively, and

$$\bar{Y}_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij} / N$$

is the pooled sample mean. The null hypothesis in (A1) is rejected at level c , if $F > F_{c,k-1,N-k}$, where $F_{c,k-1,N-k}$ is the critical value based on the F distribution with $k-1$ and $N-k$ degrees of freedom.

Let us suppose now that $Y_{ij} \sim N(\mu_i, \sigma_i)$ for $i = 1, \dots, k$. Welch's heteroscedastic F test [3] is designed to be robust to the violation of the assumption of equal group variances. The main difference with the classical ANOVA F test is that the following weights are used:

$$w_i = \frac{n_i}{s_i^2}.$$

The Welch’s heteroscedastic F test statistics is defined by

$$F_W = A \frac{\sum_{i=1}^k w_i (\bar{Y}_i - \bar{Y}')^2}{k - 1},$$

where

$$\bar{Y}' = \frac{\sum_{i=1}^k w_i \bar{Y}_i}{\sum_{i=1}^k w_i},$$

$$A = \left(1 + \frac{2(k-2)}{(k^2-1)} \sum_{i=1}^k \left(\frac{1}{n_i-1} \right) \left(1 - \frac{w_i}{\sum_{i=1}^k w_i} \right)^2 \right)^{-1}.$$

The null hypothesis (A1) is rejected at level c if $F_W \geq F_{c,k-1,\nu_W}$, where

$$\nu_W = \left[\frac{3}{k^2-1} \sum_{i=1}^k \left(\frac{1}{n_i-1} \right) \left(1 - \frac{w_i}{\sum_{i=1}^k w_i} \right)^2 \right]^{-1}.$$

Yuen’s test, i.e., the robust modification of Welch’s heteroscedastic F test, is designed to be robust to departures from normality. The test is obtained by using the sample trimmed means and Winsorized variances instead of the sample means and variances. Let $Y_{i(1)}, Y_{i(2)}, \dots, Y_{i(n_i)}$ denote the order statistics for the i th sample. Let $q_i = [n_i\alpha_i] + 1$ and $r_i = n_i - [n_i\beta_i]$, where $0 < \alpha_i < 1/2$ and $0 < \beta_i < 1/2$ represent the proportion of observations trimmed from the left and from the right tail of the distribution, respectively, and $[x]$ denotes the largest integer less than or equal to x . Then, $m_i = n_i - [n_i\alpha_i] - [n_i\beta_i]$ represents the effective sample size after trimming and the sample trimmed mean of the i th group is defined as

$$\bar{Y}_{\alpha\beta i} = \frac{1}{m_i} \sum_{j=q_i}^{r_i} Y_{i(j)}.$$

Let W_{ij} represent the new observations after replacing the trimmed observations in the lower and upper tails with the lowest and highest untrimmed values of the sample, i.e.,

$$W_{ij} = \begin{cases} Y_{i(q_i)}, & Y_{ij} \leq Y_{i(q_i)}, \\ Y_{ij}, & Y_{i(q_i)} < Y_{ij} < Y_{i(r_i)}, \\ Y_{i(r_i)}, & Y_{ij} \geq Y_{i(r_i)}. \end{cases}$$

The sample Winsorized variance for the i -th group is computed as

$$s_{wi}^2 = \frac{\sum_{j=1}^{n_i} (W_{ij} - \bar{Y}_{wi})^2}{n_i - 1},$$

where

$$\bar{Y}_{wi} = \frac{1}{n} \sum_{j=1}^{n_i} W_{ij}.$$

Yuen’s test statistics is given by

$$F_{YT} = A_T \frac{\sum_{i=1}^k w_{iT} (\bar{Y}_{\alpha\beta i} - \bar{Y})^2}{k - 1},$$

where

$$w_{iT} = \frac{m_i(m_i - 1)}{(n_i - 1)s_{wi}^2},$$

$$\tilde{Y} = \frac{\sum_{i=1}^k w_{iT} \tilde{Y}_{\alpha\beta i}}{\sum_{i=1}^k w_{iT}},$$

and

$$A_T = \left(1 + \frac{2(k-2)}{(k^2-1)} \sum_{i=1}^k \left(\frac{1}{m_i-1} \right) \left(1 - \frac{w_{iT}}{\sum_{i=1}^k w_{iT}} \right)^2 \right)^{-1}.$$

The null hypothesis of equal (trimmed) means is rejected at level c if $F_{YT} \geq F_{c,k-1,\nu_{YT}}$, where

$$\nu_{YT} = \left[\frac{3}{k^2-1} \sum_{i=1}^k \left(\frac{1}{m_i-1} \right) \left(1 - \frac{w_{iT}}{\sum_{i=1}^k w_{iT}} \right)^2 \right]^{-1}.$$

Note that this test reduces to Welch’s heteroscedastic F test when there is no trimming.

Appendix B. EL-Based ANOVA for Means

Let F_i denote a candidate for the true unknown distribution F_{i0} and $v_{ij} = F_i\{Y_{ij}\}$ denote the jump of F_i at $\{Y_{ij}\}$. The EL for the i -th sample is $L(F_i) = \prod_{j=1}^{n_i} v_{ij}$ and corresponds to a multinomial distribution defined on the i -th sample by attaching a weight v_{ij} to each Y_{ij} . The weights $v_{ij} = v_{ij}(\mu_i)$ satisfy the conditions

1. $v_{ij} \geq 0$;
2. $\sum_{j=1}^{n_i} v_{ij} = 1$;
3. $\sum_{j=1}^{n_i} v_{ij} Y_{ij} = \mu_i$.

The function $L(F_i)$ attains its maximum value when $v_{ij} = n_i^{-1}$. Similar to the classical approach based on the parametric likelihood, the profile EL ratio function is defined as

$$R(\mu_i) = \sup_{v_{ij}} \left\{ \prod_{j=1}^{n_i} n_i v_{ij}, \sum_{j=1}^{n_i} v_{ij} = 1, \sum_{j=1}^{n_i} v_{ij} (Y_{ij} - \mu_i) = 0 \right\}.$$

For an ANOVA model, Owen ([15]) defined the k -sample EL as the product of k group specific empirical likelihoods. Therefore, given the k samples, the profile EL ratio function can be defined as follows:

$$R(\mu) = \prod_{i=1}^k R_i(\mu) = \sup_{v_{ij}} \left\{ \prod_{i=1}^k \prod_{j=1}^{n_i} n_i v_{ij} : \sum_{j=1}^{n_i} v_{ij} = 1, \sum_{j=1}^{n_i} v_{ij} (Y_{ij} - \mu) = 0 \right\}, \tag{B1}$$

where $v_{ij} = v_{ij}(\mu)$. Under the null hypothesis (A1), the $k - 1$ contrasts between means are constrained to be zero, and if $\mu = \mu_0 + O(n_0^{-1/2})$, where μ_0 is the true unknown common mean and $n_0 = \min_{1 \leq i \leq k} n_i$, then

$$-2 \log \max_{\mu} R(\mu) = \sum_{i=1}^k w_i (\bar{Y}_i - \bar{Y})^2 + O_p(n_0^{-1/2}) \xrightarrow{d} \chi_{(k-1)}^2 \tag{B2}$$

as $n_0 \rightarrow \infty$, where \bar{Y}_i is the sample mean for the i -th sample, \bar{Y} is the common mean estimator

$$\bar{Y} = \frac{\sum_{i=1}^k \bar{Y}_i w_i}{\sum_{i=1}^k w_i},$$

and the weights w_i are inverse proportional with the sample variances S_i^2 , i.e.,

$$w_i = \frac{n_i}{S_i^2}.$$

It is important to note that EL-based ANOVA is robust to heteroscedasticity (see [16]).

Appendix C. EL for the Trimmed Mean

In a one sample situation, let Y_1, Y_2, \dots, Y_n be independent identically distributed with $Y_1 \sim F_0$ and let $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ be the respective order statistics. Let $q = [n\alpha] + 1$ and $r = n - [n\beta]$, where $0 < \alpha < 1/2$ and $0 < \beta < 1/2$, represent the proportion of observations trimmed from the left and right tails, respectively, and $[x]$ denotes the largest integer less than or equal to x . Then, $m = n - [n\alpha] - [n\beta]$ is the effective sample size after trimming. Let weights $p_j = 0$ for $j < q$ and $j > r$, $p_j \geq 0$ for $q \leq j \leq r$, and $\sum_{j=q}^r p_j = 1$. Then, the profile EL ratio function for the trimmed sample is defined as

$$R(\mu_{\alpha\beta}) = \sup_{p_j} \left\{ \prod_{j=q}^r m p_j : p_j \geq 0, \sum_{j=q}^r p_j = 1, \sum_{j=q}^r p_j Y_{(j)} = \mu_{\alpha\beta} \right\}.$$

Theorem C1. (Qin and Tsao, [8]) Let $\mu_{\alpha\beta}^0$ be the true value of $\mu_{\alpha\beta}$. Assume F_0 is continuous, $F_0'(\xi_\alpha) > 0$ and $F_0'(\xi_\beta) > 0$. Then,

$$-2a \log R(\mu_{\alpha\beta}) \xrightarrow{d} \chi_1^2,$$

where

$$a = \sigma_{\alpha\beta}^2 / ((1 - \alpha - \beta)\tau_{\alpha\beta}^2),$$

$$\sigma_{\alpha\beta}^2 = \frac{1}{(1 - \alpha - \beta)} \int_{\xi_\alpha}^{\xi_{1-\beta}} y^2 dF_0(y) - \mu_{\alpha\beta}^2,$$

$$\tau_{\alpha\beta}^2 = \frac{1}{(1 - \alpha - \beta)^2} ((1 - \alpha - \beta)\sigma_{\alpha\beta}^2 + \beta(1 - \beta)(\xi_{1-\beta} - \mu_{\alpha\beta})^2 - 2\alpha\beta(\xi_\alpha - \mu_{\alpha\beta})(\xi_{1-\beta} - \mu_{\alpha\beta}) + \alpha(1 - \alpha)(\xi_\alpha - \mu_{\alpha\beta})^2).$$

The unknown scaling constant a can be estimated consistently via $\hat{a} = \hat{\sigma}_{\alpha\beta}^2 / ((1 - \alpha - \beta)\hat{\tau}_{\alpha\beta}^2)$, where

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{1}{(1 - \alpha - \beta)} \int_{\hat{\xi}_\alpha}^{\hat{\xi}_{1-\beta}} y^2 dF_n(y) - \bar{Y}_{\alpha\beta}^2,$$

$$\hat{\tau}_{\alpha\beta}^2 = \frac{1}{(1 - \alpha - \beta)^2} ((1 - \alpha - \beta)\hat{\sigma}_{\alpha\beta}^2 + \beta(1 - \beta)(\hat{\xi}_{1-\beta} - \bar{Y}_{\alpha\beta})^2 - 2\alpha\beta(\hat{\xi}_\alpha - \bar{Y}_{\alpha\beta})(\hat{\xi}_{1-\beta} - \bar{Y}_{\alpha\beta}) + \alpha(1 - \alpha)(\hat{\xi}_\alpha - \bar{Y}_{\alpha\beta})^2),$$

where $\hat{\xi}_p = \inf\{y : F_n(y) \geq p\}$ for any $0 < p < 1$, and $F_n(y)$ is the empirical distribution function.

Appendix D. EL-Based ANOVA for Trimmed Means

The ELRT (B2) is still sensitive to outliers. Following the approach from [8], one could develop a version of (B2) with trimmed means. The main result stated in [8] is given in Appendix C.

Let $Y_{i(1)}, Y_{i(2)}, \dots, Y_{i(n_i)}$ denote the ordered i -th sample, set $q_i = [n_i\alpha_i] + 1$ and $r_i = n_i - [n_i\beta_i]$, where $0 < \alpha_i < 1/2$ and $0 < \beta_i < 1/2$ represent the proportion of observations trimmed from the left and the right tails, respectively, and $[x]$ denotes the largest integer less than or equal to x . Then, $m_i = n_i - [n_i\alpha_i] - [n_i\beta_i]$ is the effective sample size after trimming in each group. The group specific trimmed means and variances are

$$\begin{aligned} \bar{Y}_{\alpha\beta i} &= \frac{1}{m_i} \sum_{j=q_i}^{r_i} Y_{i(j)}, \\ S_{\alpha\beta i}^2 &= m_i^{-1} \sum_{j=q_i}^{r_i} (Y_{i(j)} - \bar{Y}_{\alpha\beta i})^2. \end{aligned}$$

Similar to (B1), define the profile EL ratio function over the trimmed samples, that is, as if the m_i observations in each sample are independent, i.e.,

$$R(\mu_{\alpha\beta}) = \prod_{i=1}^k R_i(\mu_{\alpha\beta i}) = \sup_{v_{ij}} \left\{ \prod_{i=1}^k \prod_{j=q_i}^{r_i} m_i v_{ij}, \sum_{j=q_i}^{r_i} v_{ij} = 1, \sum_{j=q_1}^{r_1} v_{ij} (Y_{i(j)} - \mu_{\alpha\beta}) = 0 \right\}. \tag{D1}$$

It is important to note that the weights are no longer a function of the common population mean but of the common population trimmed mean, that is $v_{ij} = v_{ij}(\mu_{\alpha\beta})$. As a consequence, we will obtain an ELRT for a different null hypothesis claiming the equality of population trimmed means (see [8,10]), that is

$$H_0^T : \mu_{\alpha\beta 1} = \mu_{\alpha\beta 2} = \dots = \mu_{\alpha\beta k} = \mu_{\alpha\beta}. \tag{D2}$$

When the underlying distribution of the data in each group is symmetric, the two hypotheses (A1) and (D2) are equivalent if symmetric trimming is performed. This equivalence does not hold for skewed distributions, for which it may be preferable to compare trimmed means rather than means [5]. The following result holds.

Theorem D1. Let $\mu_{\alpha\beta 0}$ be the common population trimmed mean. Assume that F_{i0} is continuous, $F'_{i0}(\xi_\alpha) > 0$ and $F'_{i0}(\xi_\beta) > 0$ for each $i = 1, \dots, k$. If $\mu_{\alpha\beta i} = \mu_{\alpha\beta 0} + O_p(m_0^{-1/2})$, $i = 1, 2, \dots, k$, where $m_0 = \min_{1 \leq i \leq k} m_i$, then

$$-2 \sum_{i=1}^k a_i \max_{\mu} \log R_i(\mu_{\alpha\beta}) = \sum_{i=1}^k a_i w_{\alpha\beta i} (\bar{Y}_{\alpha\beta i} - \bar{Y}_{\alpha\beta})^2 + O_p(m_0^{-1/2}) \xrightarrow{d} \chi^2_{(k-1)}, \tag{D3}$$

where

$$\begin{aligned} \bar{Y}_{\alpha\beta} &= \frac{\sum_{i=1}^k \bar{Y}_{\alpha\beta i} w_{\alpha\beta i}}{\sum_{i=1}^k w_{\alpha\beta i}} + o_p(m_0^{-1/2}), \\ w_{\alpha\beta i} &= \frac{m_i}{S_{\alpha\beta i}^2}, \end{aligned}$$

and the scaling factors are given by

$$a_i = \sigma_{\alpha\beta i}^2 / ((1 - \alpha_i - \beta_i) \tau_{\alpha\beta i}^2). \tag{D4}$$

The quantities $\sigma_{\alpha\beta i}^2$ and $\tau_{\alpha\beta i}^2$ for the i th trimmed sample are defined in Appendix C.

Proof. By a Lagrange multiplier argument, it can be shown (see, for example, [9]), that the $v_{ij}, i = 1, 2, \dots, k$ that maximize $R(\mu_{\alpha\beta})$ are given by

$$v_{ij} = \frac{1}{m_i (1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta}))}, j = q_i, \dots, r_i, \tag{D5}$$

where the Lagrange multiplier λ_i is the solution to

$$\sum_{j=q_i}^{r_i} \frac{Y_{i(j)} - \mu_{\alpha\beta}}{1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta})} = 0,$$

and

$$\lambda_i = \frac{\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta}}{\frac{1}{m_i} \sum_{j=q_i}^{r_i} (Y_{i(j)} - \mu_{\alpha\beta})^2} + o_p(m_i^{-1/2}).$$

Then, by substituting v_{ij} from (D5) in the expression (D1), we obtain

$$-2 \log R(\mu_{\alpha\beta}) = 2 \sum_{i=1}^k \sum_{j=q_i}^{r_i} \log(1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta})).$$

The maximum empirical likelihood estimator $\bar{Y}_{\alpha\beta}$ is the solution to $\partial R(\mu_{\alpha\beta}) / \partial \mu_{\alpha\beta} = 0$:

$$\frac{\partial R(\mu_{\alpha\beta})}{\partial \mu_{\alpha\beta}} = - \sum_{i=1}^k \sum_{j=q_i}^{r_i} \left(\frac{\partial \lambda_i}{\partial \mu_{\alpha\beta}} (Y_{i(j)} - \mu_{\alpha\beta}) - \lambda_i \right) \left(1 + \lambda_i (Y_{i(j)} - \mu_{\alpha\beta}) \right)^{-1} = \sum_{i=1}^k m_i \lambda_i.$$

It follows that $\bar{Y}_{\alpha\beta}$ satisfies

$$\sum_{i=1}^k \frac{m_i (\bar{Y}_{\alpha\beta i} - \mu_{\alpha\beta})}{m_i^{-1} \sum_{j=q_i}^{r_i} (Y_{ij} - \mu_{\alpha\beta})^2} = o_p(\sqrt{m_0}),$$

and expression (D3) follows. Since, according to Theorem C1, for each of the trimmed samples $i = 1, \dots, k$ and true value $\mu_{\alpha\beta}^0$,

$$-2a_i \log R_i(\mu_{\alpha\beta}^0) \xrightarrow{d} \chi_1^2,$$

then, summing over the groups, we prove the result stated in (D3), by using the same arguments leading to the result stated in (B2). \square

References

1. Wilcox, R. Comparing the means of two independent groups. *Biom. J.* **1990**, *32*, 771–780.
2. Yuen, K. The two-sample trimmed t for unequal population variances. *Biometrika* **1974**, *61*, 165–170.
3. Welch, B. On the comparison of several mean values: An alternative approach. *Biometrika* **1951**, *38*, 330–336.
4. Lix, L.; Keselman, H. To trim or not to trim: Tests of location equality under heteroscedasticity and nonnormality. *Educ. Psychol. Meas.* **1998**, *58*, 409–429.
5. Keselman, H.; Wilcox, R.; Kowalchuk, R.; Olejnik, S. Comparing trimmed or least squares means of two independent skewed populations. *Biom. J.* **2002**, *44*, 478–489.
6. Owen, A. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **1988**, *75*, 237–249.
7. Owen, A. Empirical likelihood ratio confidence regions. *Ann. Stat.* **1990**, *18*, 90–120.
8. Qin, G.; Tsao, M. Empirical likelihood ratio confidence interval for the trimmed mean. *Commun. Stat. Theory Methods* **2002**, *31*, 2197–2208.
9. Owen, A. *Empirical Likelihood*; Chapman & Hall/CRC: New York, NY, USA, 2001.
10. Tsao, M.; Zhou, J. A nonparametric confidence interval for the trimmed mean. *J. Nonparametr. Stat.* **2002**, *14*, 665–673.
11. Wilcox, R. *Introduction to Robust Estimation and Hypothesis Testing*, 2nd ed.; Elsevier Academic Press: Burlington, MA, USA, 2005.
12. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
13. Reimann, C.; Arnoldussen, A.; Boyd, R.; Finne, T.; Koller, F.; Nordulgen, O. Element contents in leaves of four plant species (birch, mountain ash, fern and spruce) along anthropogenic and geogenic concentration gradients. *Sci. Total Environ.* **2007**, *377*, 416–433.
14. Todorov, V.; Filzmoser, P. An object-oriented framework for robust multivariate analysis. *J. Stat. Softw.* **2009**, *32*, 1–47.

15. Owen, A. Empirical likelihood for linear models. *Ann. Stat.* **1991**, *19*, 1725–1747.
16. Tsao, W.; Wu, C. Empirical likelihood inference for a common mean in presence of heteroscedasticity. *Can. J. Stat.* **2006**, *34*, 45–59.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).