BÖLGESEL İHRACAT VERİLERİNİN MODİFİYE EDİLMİŞ GENELLEŞTİRİLMİŞ F-TESTİ İLE ANALİZİ

Mustafa CAVUS[1] *Berna YAZICI*[2] *Ahmet SEZER*[3]

Öz

Normal dağılmış anakütlelerden geldiği bilinen homojen varyanslı ikiden fazla grubun ortalamasının eşitliğinin test edilmesi için Klasik F-testi kullanılır. Klasik F-testi, grupların birbirinden bağımsız, homojen varyanslı ve normal dağıldığı varsayımları altında parametrik istatistiksel yöntemler arasında en güçlü testtir. Gerçek hayatta bahsedilen varsayımların sağlandığı durumlarla çok nadirdir. Bu nedenle araştırmacılar varsayımların sağlanmadığı durumlar için yöntemler geliştirmeye yönelmişlerdir. Welch, Genelleştirilmiş F, Parametrik Bootstrap testleri varyans homojenliğinin sağlanmadığı durumlarda normal dağılmış grupların ortalamalarının eşitliğinin test edilmesi için geliştirilmiştir. Yalnızca varyans homojenliği sağlanmadığı durumlarda doğru sonuçlar veren bu yöntemler normal dağılım varsayımının bozulması durumunda performanslarını kaybettiklerinden birçok çalışmada bahseilmiştir. Bu çalışmada aykırı değerden kaynaklı normal dağılmama ve homojen olmayan varyanslılık durumunda grup ortalamalarının karşılaştırılabilmesi için kullanılabilen Modifiye Edilmiş Genelleştirilmiş F-testi ele alınmıştır. Bahsedilen koşullar altında bu yöntemin etkinliğinin ortaya konulabilmesi için homojen varyansa sahip olmayan ve aykırı değerden kaynaklı normal dağılmayan Türkiye'deki coğrafi bölgelerin ortalama ihracat tutarları karşılaştırılmıştır. Sonuç olarak bölgeler arasındaki istatistiksel olarak anlamlı farkların Modifiye edilmiş Genelleştirilmiş F-testi ile tespit edilebileceği ortaya konulmuştur.

Anahtar Kelimeler: *Heterojen varyans, Normal dağılmama, Aykırı değer, ANOVA*
JEL Sınıflandırması: *C90, C120, C150*

ANALYSING REGIONAL EXPORT DATA BY THE MODIFIED GENERALIZED F-TEST

Abstract

Classical F-test is used for testing equality of more than two group means under normality and variance homogeneity. Classical F test is most powerful parametric method among the parametric statistical methods in case of the assumptions are hold. However, the assumptions are not always satisfied in real life. Thus researchers study on improving methods to solve this problem. Welch, Generalized F, Parametric Bootsrap tests are proposed for testing equality of group means under variance heterogeneity. These methods just give better results under variance heterogeneity but they are not same in case of violation of normality assumption due to researches. In this article, modified generalized F-test is considered which is proposed for variance heterogeneity and non-normality caused by outlier. To show the efficiency of this method, testing equality of annual export amounts of geographical regions under variance heterogeneity and non-normality caused by outlier. As a result, it is stated that significant differences between regions are detected only by modified generalized F-test.

Keywords: *Variance heterogeneity, Non-normality, Outlier, ANOVA*
JEL Classification: *C90, C120, C150*

[1] *Anadolu University, Department of Statistics, mustafacavus@anadolu.edu.tr*
[2] *Anadolu University, Department of Statistics, bbaloglu@anadolu.edu.tr*
[3] *Anadolu University, Department of Statistics, a.sezer@anadolu.edu.tr*

DOI: 10.18092/ulikidince.348070
Makalenin Geliş Tarihi (Recieved Date): 31-10-2017
Yayına Kabul Tarihi (Acceptance Date): 16-01-2018

1. Introduction

Testing the equality of group means is one of the most common statistical problem in many disciplines such as econometrics, industrial issues, engineering, biostatistics, pharmacology, agriculture and etc. Researchers want to obtain if there is a significant difference between groups in these disciplines. It is called the analysis of variance (ANOVA). CF test is used when three assumptions are hold (Fisher, 1925). These assumptions are independently and normally distributed groups have homogeneity variance. When one or more of these assumptions are violated, CF test gives wrong results so researchers do not detect significant difference between groups or determine insignificant difference between groups (Gamage and Weerahandi, 1998). In this case, it must be obtained which asssumption is violated hence solution method can be decided.

The rest of this article is organized as follows. Section 2 provides a literature review about the testing equality of group means. Moreover, the alternative ways in case of assumption violation are described. Section 3 describes the methods are used to test the equality of group means and the proposed method. An illustrative example is given in Section 4, and Section 5 gives some concluding remarks.

2. Literature Review

CF is the most commonly used procedure in testing equality of group means when the assumptions are hold. When one or more of these assumptions are violated, CF test can give wrong results. Some modifications to test statistics based on weighting are used solving this problem in case of variance homogeneity violation. For example, Cochran (1937) and Welch (1951) proposed test procedures based on weightining for variance heterogeneity. Box (1954) and Brown and Forsythe (1974) proposed some adjusment to degrees of freedom of CF test for providing better results under variance heterogeneity.

More powerful methods are began to improved with the development of Monte-Carlo simulation method. Firstly Weerahandi (1995) introduced the Generalized F-test based on generalized p-value approach and Krishnamoorty et al. (2007) proposed the Parametric Bootstrap test and then Alvandi (2012) proposed a new test procedure based on generalized p-value approach which depend on Monte-Carlo simulation method.

There are numerous studies about the performance comparison of these methods in terms of type 1 error rates and power of the test. Gamage and Weerahandi (1998), Hartung et al. (2002) and Alvandi et al. (2012) are the most well known among them. The results of these studies are similar, GF and PB tests are more powerful than CF and the adjusted versions of CF in most cases under variance heterogeneity.

Besides the violation of variance homogeneity assumption, the non-normality is another common violation. In recent studies the researchers focused on testing equality of group means under assumption violations. It is obviously that type 1 error rates of the methods are inflated because of non-normality so Tan and Tabatabai (1985) modified the Brown-Forsythe test with Huber's M-estimators, Wilcox (1995) proposed modification to CF test using trimmed mean towards the outlier effect cause non-normality. Karagoz (2015) tried to test the equality of non-normal group means with modified Welch F-test with robust estimators. Karagoz and Saracbasi (2016) proposed a modification to Brown-Forsythe test for same purposes.

3. Methodology

Consider the problem of testing equality of group means of k populations. Assume $X_{1n_1}, X_{2n_2}, \dots, X_{in_i}$, $i = 1, 2, \dots, k$ are observations of k independent populations from normal distributions . The maximum likelihood estimators of sample mean and sample variance of the k independent groups are given in the following equations respectively

$$\bar{X}_i = \frac{\sum_{i=1}^{n_i} X_{ij}}{n_i} \tag{1}$$

$$S_i^2 = \frac{\sum_{i=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2}{n_i - 1} \tag{2}$$

the observed values of sample mean in Equation (1) and sample variance in Equation (2) are $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)$ and $s^2 = (s_1^2, s_2^2, \dots, s_k^2)$ respectively. The hypotheses of the problem are as follows

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k \text{ vs } H_1: \mu_i \neq \mu_j \text{ for } i \neq j$$

To test hypotheses given above, CF test is used under some assumptions. These assumptions are as follows:

- Each group is from normally distributed populations[1]
- All populations have equal variance[2]
- All groups are independenly distributed of each other[3]

and they will be mentioned abbreviately as [1]normality, [2]variance homogeneity and [3]independence in this article. Alternative methods are developed in violation of the variance homogeneity assumption. Generalized F-test is one of these developed methods and it is powerful than others in many cases. In violation of variance homogeneity and normality caused by outlier, Cavus et al. (2007) proposed the modified generalized F-test and showed that the power of MGF against the alternatives.

3.1. Classical F-Test

Population variances are equal $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$, CF is powerful method to test the equality of group means.

$$CF = \frac{\left.\sum_{i=1}^{k} n_i \bar{x}_i^2 - n\bar{x}^2 \middle/ (k-1)\right.}{\left.\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{k} n_i \bar{x}_i^2 \middle/ (n-k)\right.} \tag{3}$$

where $n = \sum_{i=1}^{k} n_i$ is the total number of observations and $\bar{x} = \sum_{i=1}^{k} \bar{x}_i$ is the grand mean average of observations. CF test statistic in Equation (3) has an F distribution with $k-1$ and $n-k$ degrees of freedom.

3.2. Generalized F-Test

Generalized F-test is proposed by Weerahandi (1995) under variance heterogeneity and the p-value of the test statistic is computed by the Monte-Carlo simulation method. Consider the following standardized sum of squares between groups

$$\tilde{s}s_G = \sum_{i=1}^{k} \frac{n_i \bar{x}_i^2}{s_i^2} - \frac{\left[\sum_{i=1}^{k} n_i \bar{x}_i / s_i^2\right]^2}{\sum_{i=1}^{k} n_i / s_i^2} \tag{4}$$

Let the nuisance parameter s_i^2 replaced by random chi-squared random variables $\chi_{n_i-1}^2$ with $n_i - 1$ degrees of freedom.

$$GF = E\left[\tilde{s}s_G\left(\frac{n_1 s_1^2}{U_1}, \frac{n_2 s_2^2}{U_2}, \dots, \frac{n_k s_k^2}{U_k}\right)\right] \tag{5}$$

where GF is distributed chi-squared with $k-1$ degrees of freedom and the expectation is taken with respect to the independent U_i random variables. The computation of p-value of GF test can be given in the following algorithm.

Algorithm 1

1. Compute the maximum likelihood estimators of sample mean and sample variance for k groups
2. Compute the standardized sum of squares between groups $\tilde{ss}_G$
3. Repeat the loop for r times
    - Generate $U_i \sim \chi^2_{n_i-1}$ independent random variables
    - Compute GF test statistic in Equation (5) with respect to U_i
    - If $GF > \tilde{ss}_G$ in Equation (4), set counter variable $Q_i = 1$
    - End loop
4. Monte-Carlo estimate of p-value is $\sum_{i=1}^{r} Q_i / r$

3.3. Modified Generalized F-Test

In GF test, $\bar{x}$ and s^2 denoted as the maximum likelihood estimators of sample mean and sample variance respectively. The proposed modified generalized F-test in which the maximum likelihood estimators of sample mean and sample variance are replaced with Huber's M-estimators. Consider the following standardized sum of squares between groups with Huber's M-estimators:

$$\tilde{ss}_G{}^{*} = \sum_{i=1}^{k} \frac{n_i \bar{x}_i^{2^*}}{s_i^{2^*}} - \frac{\left[\sum_{i=1}^{k} n_i \bar{x}_i{}^{*} / s_i^{2^*}\right]^2}{\sum_{i=1}^{k} n_i / s_i^{2^*}} \tag{6}$$

where $\bar{x}_i^{2^*}$ and $s_i^{2^*}$ are Huber's M-estimators of sample mean and sample variance repectively. Let the nuisance parameter s_i^2 replaced by random chi-squared random variables $\chi^2_{n_i-1}$ with $n_i - 1$ degrees of freedom.

$$GF^{*} = E\left[\tilde{ss}_G{}^{*}\left(\frac{n_1 s_1^{2^*}}{U_1}, \frac{n_2 s_2^{2^*}}{U_2}, \dots, \frac{n_k s_k^{2^*}}{U_k}\right)\right] \tag{7}$$

p-value of the modified generalized F-test can be calculated easily with Monte-Carlo simulation method using following algorithm.

Algorithm 2

1. Compute Huber's M-estimators of sample mean and sample variance for k groups
2. Compute the proposed standardized sum of squares between groups $\tilde{ss}_G{}^{*}$
3. Repeat the loop for r times
    - Generate $U_i \sim \chi^2_{n_i-1}$ independent random variables
    - Compute GF^{*} test statistic in Equation (7) with respect to U_i
    - If $GF^{*} > \tilde{ss}_G{}^{*}$ in Equation (6), set counter variable $Q_i = 1$
    - End loop
4. Monte-Carlo estimate of p-value is $\sum_{i=1}^{r} Q_i / r$

4. Illustrative Example

In this part of the study, an illustrative example is examined to show the efficieny of the proposed method. Classical F, Generalized F and Modified Generalized F tests are used for testing equality of the mean export amounts of the regions. Data which is used in example are taken from Turkish Statistical Instutie Database. It consists of the 2015 total export amounts of 81 cities in 7 geographical regions as currency Euro (**€**). The mean and the total amounts of the annual exports of the regions are showed in Table 1. Marmara has extremely higher export amounts and the lowest total export amount of the regions is Eastern Anatolia.

Table 1: **Total Export Amounts of Cities in Turkey in 2015 (€)**

REGION	Mediterranean	Aegean	Marmara	Black Sea
	Adana[1]	Afyon[3]	Balıkesir[10]	Amasya[5]
	39242	11045	12838	3944
	Antalya[7]	Aydın[9]	Bilecik[11]	Artvin[8]
	35508	13753	2653	1980
	Burdur[15]	Denizli[20]	Bursa[16]	Bolu[14]
	3052	13830	42996	3412
	Hatay[31]	İzmir[35]	Çanakkale[17]	Çorum[19]
	31040	55553	5339	6622
	Isparta[32]	Kütahya[43]	Edirne[22]	Giresun[28]
	5184	6896	4234	4588
	Mersin[33]	Manisa[45]	İstanbul[34]	Gümüşhane[29]
	28291	19362	241121	1821
	Kahramanmaraş[46]	Muğla[48]	Kırklareli[39]	Kastamonu[37]
	21818	11318	3695	4020
	Osmaniye[80]	Uşak[64]	Kocaeli	Ordu[52]
	9257	4522	30749	8924
			Sakarya[54]	Rize[53]
			14153	4194
			Tekirdağ[59]	Samsun[55]
			14077	17127
			Yalova[77]	Sinop[57]
			2991	2249
				Tokat[60]
				7266
				Trabzon[61]
				10409
				Zonguldak[67]
				6726
				Bayburt[69]
				1225
				Bartın[74]
				2119
				Karabük[78]
				2499
				Düzce[81]
				5098
AVERAGE	802.357.040	1.596.739.792	7.975.928.736	190.457.691
TOTAL	6.418.856.323	12.773.918.337	87.735.216.097	3.428.238.430

It is known that CF test gives correct results in testing the equality of group means when the necessary assumptions are hold. Results of Shapiro-Wilk (SW) normality test are given in Table 2 and the bounds of interquantile range (IQR) for detecting outliers in the data are given in Table 1. (*) and (**) shows the non-normality of the regions in 95% and 99% confidence level respectively. According to the p-values of SW test, regions are not distributed normal except Mediterrenean. Black sea and Southeastern Anatolia are not distributed normal because of outliers when the results of SW test are compared in case of without outliers. It is obviously that when the non-normality problem caused by outlier occurs, the alternative methods of CF test under variance heterogeneity can not be used.

Table 1(Continue): **Total Export Amounts of Cities in Turkey in 2015 (€)**

REGION	Central Anatolia	Eastern Anatolia	Southeastern Anatolia
	Ankara[6]	Ağrı[4]	Adıyaman[2]
	76944	15750	13338
	Çankırı[18]	Bingöl[12]	Diyarbakır[21]
	2151	5438	43321
	Eskişehir[26]	Bitlis[13]	Gaziantep[27]
	9923	8972	48227
	Kayseri[38]	Elazığ[23]	Mardin[47]
	22879	8916	21318
	Kırşehir[40]	Erzincan[24]	Siirt[56]
	2894	3081	8507
	Konya[42]	Erzurum[25]	Şanlıurfa[63]
	36080	15173	62056
	Nevşehir[50]	Hakkari[30]	Batman[72]
	4140	5767	14451
	Niğde[51]	Kars[36]	Şırnak[73]
	5662	6029	14431
	Sivas[58]	Malatya[44]	Kilis[79]
	9139	11962	2890
	Yozgat[66]	Muş4[9]	
	6029	11105	
	Aksaray[68]	Tunceli[62]	
	6664	1073	
	Karaman[70]	Van[65]	
	3823	29567	
	Kırıkkale[71]	Ardahan[75]	
	3327	1527	
		Iğdır[76]	
		4489	
AVERAGE	806.956.226	56.996.153	884.141.294
TOTAL	10.490.430.934	797.946.143	7.957.271.645

Table 2: **Results of the Important Statistics About Data (€)**

Region	Mean Export Amount (€)	Outlier (€)	Lower bound of IQR (€)	Upper bound of IQR (€)
Mediterranean	802.357.040	-	- 1.554.688.636	3.052.855.482
Aegean	1.596.739.792	İzmir[35] 7.481.020.537	- 2.002.783.396	4.019.387.232
Marmara	7.975.928.736	İstanbul[34] 69.373.002.022	- 5.851.955.700	10.214.694.665
Black Sea	190.457.691	Trabzon[61] 1.394.191.986	- 217.998.556	436.527.875
Central Anatolia	806.956.226	Ankara[6] 6.333.049.126	- 1.023.120.597	1.836.602.563
Eastern Anatolia	56.996.153	-	- 155.567.449	267.734.356
Southeastern Anatolia	884.141.294	Gaziantep[27] 5.695.911.335	- 555.776.807	1.115.033.457

Table 3: **Normality Test Results**

Region	p-value of SW Test (with outliers)	p-value of SW Test (without outliers)
Mediterranean	0.2186	0.2186
Aegean	0.0005**	0.0223*
Marmara	0.0000**	0.0002**
Black Sea	0.0000**	0.0531
Central Anatolia	0.0000**	0.0007**
Eastern Anatolia	0.0015**	0.0015**
Southeastern Anatolia	0.0000**	0.3139

According to the exports of the regions, p-value of the Levene Variance Homogeneity test is 0.001^{**}. It is seen that export amount of regions data do not hold for both variance homogeneity and normality assumptions. In this case, equality of some region means combinations are tested by CF, GF and MGF. Hypotheses of the combinatinations are as follows:

Case 1: $H_0: \mu_{Med.} = \mu_{Aeg.} = \mu_{Mar.}$ vs $H_1: At\ least\ one\ of\ \mu_i \neq \mu_j,\ \ i \neq j$

Case 2: $H_0: \mu_{Med.} = \mu_{Aeg.} = \mu_{Mar.} = \mu_{Bla.}$ vs $H_1: At\ least\ one\ of\ \mu_i \neq \mu_j,\ \ i \neq j$

Case 3: $H_0: \mu_{Aeg.} = \mu_{Mar.} = \mu_{Bla.}$ vs $H_1: At\ least\ one\ of\ \mu_i \neq \mu_j,\ \ i \neq j$

Case 4: $H_0: \mu_{Cen.} = \mu_{Eas.} = \mu_{Sou.}$ vs $H_1: At\ least\ one\ of\ \mu_i \neq \mu_j,\ \ i \neq j$

The p-values of the CF, GF and MGF are given in Table 4. (*) and (**) shows the significant difference between group means in the combination in 95% and 99% confidence level respectively.

Table 4: **p-values of the CF, GF and MGF tests**

Case	Test Combination	CF	GF	MGF
1	Mediterranean, Aegean, Marmara	0.4326	0.3380	0.0466*
2	Mediterranean, Aegean, Marmara, Black Sea	0.2388	0.0220*	0.0036**
3	Aegean, Marmara, Black Sea	0.1912	0.1304	0.0209*
4	Central Anatolia, Eastern Anatolia, Southeastern Anatolia	0.0995	0.0739	0.0225*

Conclusions and interpretations about the results in Table 4 are given as follows:

Case 1: CF and GF tests give the same result about the equality of Mediterranean, Aegean and Marmara region annual export means, there is no difference between regions. Unlike CF and GF test, the conclusion about the equality of region means are changed with MGF. p-value of MGF test is lower than significance level $\alpha = 0.05$ so the annual export means of the regions are not same. It is obvious that the difference between the region means can be detected by MGF.

Case 2: By adding of Black Sea region into the combinations, the difference between regions can be detected by GF. However, the significant difference can be detected in 95% confidence level by GF, it could be detected by MGF in 99% significance level. In this case, the significant difference can be detected in more confidently by MGF.

Case 3: It is clear in Table 2 that Black Sea region is not normally distributed because of the existence of an outlier. CF and GF tests give the same result about the equality of region annual export means, there is no difference between regions. Unlike CF and GF test, the difference between the region means can be detected by MGF.

Case 4: It is clear in Table 2 that Southeastern Anatolia region is not normal distributed because of the existence of outlier. CF and GF tests give the same result about the equality of region annual export means, there is no difference between regions. Unlike CF and GF test, the difference between Central Anatolia, Eastern Anatolia and Southeastern Anatolia region means can be detected by MGF.

5.Results and Discussion

CF is the most powerful method to test the equality of group means when the assumptions are hold. GF test is proposed as an alternative to CF in case of the violation of variance homogeneity. Despite GF is more powerful then CF under heteroskedasticity, it is not the same under non-normality. To achive more powerful result under both variance heterogeneity and non-normality, Cavus et al. (2017) proposed MGF test. This method is a modification of GF test with replacing maximum likelihood estimator of sample mean and sample variance with Huber's M-estimators. This article focused to show the efficiency of the proposed MGF test in testing equality of group means under heteroscedasticity and non-normality caused by outliers. As seen in the results in Table 4, significant differences between region means can be detected by MGF test but not by CF and GF. Also, detections of differences between region means are made more confidently by MGF compared to CF and GF. As a result, MGF test should be used under variance heterogeneity and non-normality caused by outlier to make the right decision in testing equality of group means.

References

Alvandi, S. M., Jafari, A. A. and Mardani, F. A. (2012). One-Way ANOVA with Unequal Variances. *Communications in Statistics*, 41, 4200-4221.

Box, G. E. P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, *Annals of Mathematical Statistics*, 25, 290-302.

Brown, M.B. and Forsythe, A.B. (1974). The Small Sample Behavior of Some Statistics which Test the Equality of Several Means, *Technometrics*, 16, 129-132.

Çavuş, M., Yazıcı, B. and Sezer, A. (2017). Modified Tests for Comparison of Group Means Under Heteroskedasticity and Non-Normality Caused by Outlier(s). *Hacettepe Journal of Mathematics and Statistics*, 46(3), 492-510.

Çavuş, M., Yazıcı, B. and Sezer, A. (2017). Comparison of the Average Amounts of Exports for Geographical Regions in Turkey with the Modified Generalized F-test. *18.International SYmposium on Econometrics Operations Research and Statistics*, Trabzon.

Fisher, R. A. (1925). Statistical Methods and Scientific Inference. Oxford, England: Oliver and Boyd.

Gamage, J. and Weerahandi, S. (1998). Size Performance of Some Tests in One-Way ANOVA. *Communications in Statistics*, 27(3), 625-640.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 69, 73-101.

Karagoz, D. (2015). Modified Welch test statistic for ANOVA under Weibull distribution. *Hacettepe Journal of Mathematics and Statistics*, 27(3), 625-640.

Karagoz, D. and Saraçbaşı, T. (2009). Robust Brown-Forsythe and Robust Modified Brown-Forsythe ANOVA Tests Under Heteroscedasticity for Contaminated Weibull Distribution. *Revista Colombiana de Estadistica*, 39, 17-32.

Krishnamoorthy, K., Lu, F. and Mathew, T. (2007). A Parametric Bootstrap Approach for ANOVA with Unequal Variances. *Computational Statistics and Data Analysis*, 51, 5731-5742.

Li, X. (2009). A Generalized p-value Approach for Comparing the Means of several Log-Normal Populations. *Statistics and Probability Letters*, 79, 1404-1408.

Tan, W. Y. and Tabatabai, M. A. (1985). Some Robust ANOVA Procedures under Heteroscedasticity and Non-Normality. *Communications in Statistics*, 14(4), 1007-1026.

Tsui, K. and Weerahandi, S. (1989). Generalized p-values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*, 84, 602-607.

Welch, B. L. (1951). On the Comparison of Several Group Mean Values. *Biometrika*, 38, 330-336.

Weerahandi, S. (1995). ANOVA Under Unequal Error Variances. *Biometrics,* 51, 589-599.

Wilcox, R. R. (1995). Simulation Results on Solutions to the Multivariate Behrens-Fisher Problem via Trimmed Means. *The Statistician*, 44(2), 213-225.

ANALYSING REGIONAL EXPORT DATA BY THE MODIFIED GENERALIZED F-TEST

Extended Abstract

Aim: Classical F-test is used for testing equality of more than two group means under normality and variance homogeneity. Classical F test is most powerful parametric method among the parametric statistical methods in case of the assumptions are hold and when the groups are independent. However, the assumptions are not always satisfied in real life. Welch, Generalized F, Parametric Bootsrap tests are proposed for testing equality of group means under violation of variance homogeneity assumption. These methods just give better results under variance heterogeneity but the results are not same in case of violation of normality assumption due to researches. In this article, modified generalized F-test is considered which is proposed for variance heterogeneity and non-normality caused by outlier. To show the efficiency of this method, testing equality of annual export amounts of geographical regions under both variance heterogeneity and non-normality caused by outlier.

Method: An illustrative example is examined to show the efficieny of the proposed method. Classical F, Generalized F and Modified Generalized F tests are used for testing equality of the mean export amounts of the regions. Data which is used in example are taken from Turkish Statistical Instutie Database. It consists of the 2015 total export amounts of 81 cities in 7 geographical regions as currency Euro (€). Marmara has extremely higher export amounts and the lowest total export amount of the regions is Eastern Anatolia.

Findings: Export amount of regions data do not hold for both variance homogeneity and normality assumptions. In this case, equality of some region means combinations are tested by CF, GF and MGF. CF and GF tests give the same result about the equality of Mediterranean, Aegean and Marmara region annual export means, there is no difference between regions. CF and GF tests give the same result about the equality of Mediterranean, Aegean and Marmara region annual export means, there is no difference between regions. Unlike CF and GF test, the conclusion about the equality of region means are changed with MGF. p-value of MGF test is lower than significance level $\alpha = 0.05$ so the annual export means of the regions are not same. It is obvious that the difference between the region means can be detected by MGF. By adding of Black Sea region into the combinations, the difference between regions can be detected by GF. However, the significant difference can be detected in 95% confidence level by GF, it could be detected by MGF in 99% significance level. In this case, the significant difference can be detected in more confidently by MGF. It is clear in Table 2 that Black Sea region is not normally distributed because of the existence of an outlier. CF and GF tests give the same result about the equality of region annual export means, there is no difference between regions. Unlike CF and GF test, the difference between the region means can be detected by MGF. It is clear in Table 2 that Southeastern Anatolia region is not normal distributed because of the existence of outlier. CF and GF tests give the same result about the equality of region annual export means, there is no difference between regions. Unlike CF and GF test, the difference between Central Anatolia, Eastern Anatolia and Southeastern Anatolia region means can be detected by MGF.

Conclusion: CF is the most powerful method to test the equality of group means when the assumptions are hold. GF test is proposed as an alternative to CF in case of the violation of variance homogeneity. Despite GF is more powerful then CF under heteroskedasticity, it is not the same under non-normality. To achive more powerful result under both variance heterogeneity and non-normality, Cavus et al. (2017) proposed MGF test. This method is a modification of GF test with replacing maximum likelihood estimator of sample mean and sample variance with Huber's M-estimators. This article focused to show the efficiency of the proposed MGF test in testing equality of group means under heteroscedasticity and non-normality caused by outliers. As seen in the results in Table 4, significant differences between region means can be detected by MGF test but not by CF and GF. Also, detections of differences between region means are made more confidently

by MGF compared to CF and GF. As a result, MGF test should be used under variance heterogeneity and non-normality caused by outlier to make the right decision in testing equality of group means.