Comparing the Means of Normal Populations with Unequal Variances

Abbas GERAMI (*),

Ministry of Agriculture, Iran (a.gerami @ mailcity.com)

Ali Reza ZAHEDIAN

Statistical centre of Iran

1. Abstract

The t- and F-test to test the equality of normal populations means heavily based on the assumption of the equality of populations variances. Investigations have shown that these tests are not roboust under violation of equal variances. In this paper several solutions for two populations have been compared by monte carlo method and best solutions have been identified. The solutions have been extended to more than two populations and best ones been recommended. Finally some point have have been mentioned about determination of sample size.

2. Introduction

In applied statistics the experimenter often wants to compare two or more treatments measured on independent samples. For example it may be of interest to compare two different fertilizers, or two teaching methods. In many cases, it has been assumed that populations have normal distributions with the same (common)variances, known or unknown. Suppose we are concerned with normal data, then if the common variance is known, an exact normal test procedure is satisfactory and if the common variance is unknown an exact student test approach could be employed.

The assumption of common variance is a great convenience but it is not the general case. In general, if variances are known an exact standardized normal test can be used subject to normality. If the variances were not known, then the student's t-test is no longer valid and this is known as "Behrenes-Fisher" problem.

Given k independent samples of independent observations X_{ij} from $N(\mu_i, \sigma_i^2);\ i = 1,2,\ldots,k,\ j = 1,2,\ldots,n_i\,(k > 2)$, we are concerned with the equality of means.

The usual and common method to test the equality of means when the variances are equal but unknown, is called "Analysis of variance" or ANOVA. The lack of robustness of the on-way ANOVA to the heterogeneity of the variances is well known.

For this problem several parametric and non parametric methods has been suggested which we only consider those which are parametric. Since Buning (1995) shows that in the case of heterogeniety of variances Welch's test as a parametric solution is better than non parametric approaches.

3. Different solutions

In this section we briefly review some solutions. We need to introduce some notations.

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}\right)^2$$

$$SSB = \sum_{i=1}^{k} n_i\left(\overline{X}_i - \overline{X}\right)^2$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2$$

$$\overline{X} = \frac{1}{N}\sum_{j}^{i=1}\sum_{j}^{j=1} X_{ij}$$

$$N = \sum n_i$$

$$\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$$

3.1 Welch' s method

Welch (1951) gives the following test statistics

$$F_w = \frac{\dfrac{1}{k-1}\sum_{i=1}^{k} g_i\left(\overline{X}_i - \tilde{X}\right)^2}{1 + \dfrac{2(k-2)}{k^2-1}\sum_{i=1}^{k}\dfrac{(1-g_i/g)^2}{n_i - 1}}$$

$$g_i = n_i / s_i^2 \qquad s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2$$

$$\tilde{X} = \sum_{i=1}^{k} g_i\,\overline{X}_i / g$$

$$g = \sum_{i=1}^{k} g_i$$

If Ho is true, then the approximate distribution of is

$F\left(k-1,\ \hat{f}_w\right)$; where

$$\hat{f}_w = \left(\frac{3}{k^2 - 1} \sum_{i=1}^{k} \frac{\left(1 - g_i/g\right)^2}{n_i - 1} \right)^{-1}$$

and Ho is rejected of a leuel if $F_w > F\left(k-1,\ \hat{f}_w, a\right)$

3.2 Scott and Smith (1971) method

In this method the test statics is

$$F_s = \sum_{i=1}^{k} \frac{n_i \left(\overline{X}_i - \overline{X}\right)^2}{S_i^{*2}} \quad ; \quad S_i^{*2} = \frac{n_i - 1}{n_i - 3} S_i^2$$

If Ho is true then asymptotically F_s has $x^2(k)$ and Ho is rejected if at a Level $F_s > x^2(k, a)$.

3.3 Weerahandi (1995).

In this methode following notation are used:

$$\tilde{SS}T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{1}{s_i^2} \left(X_{ij} - \tilde{\overline{X}}\right)^2$$

$$\tilde{SS}B = \sum_{i=1}^{k} \frac{n_i}{s_i^2} \left(\overline{X}_i - \tilde{\overline{X}}\right)^2$$

$$\tilde{SS}E = \sum_{i} \sum_{j} \frac{1}{s_i^2} \left(\overline{X}_{ij} - \overline{X}_i\right)^2 = \sum_{i} \frac{n_i s_i^2}{s_i^2}$$

$$\tilde{\overline{X}} = \left(\sum n_i \overline{X}_i / s_i^2\right) \Big/ \left(\sum n_i / s_i^2\right)$$

Then

$$B_j = \frac{\sum_{i=1}^{j} \frac{n_i s_i^2}{s_i^2}}{\sum_{i=1}^{j+1} \frac{n_i s_i^2}{s_i^2}} \sim Beta\left(\frac{1}{2} \sum_{i=1}^{j} \left(n_i - 1\right), \frac{1}{2}\left(n_{j+1} - 1\right) \right)$$

$\tilde{SS}B$ is defined based on weights n_i / σ_i^2. Now if we denote this by $\tilde{SS}B(\sigma_1^2, \sigma_2^2, ..., \sigma_k^2,)$, then for weights $\frac{n_i \sigma_i^2 s_i^2}{s_i^2}$, then SS is denoted by $\tilde{ss}b\left(\frac{s_1^2 \sigma_1^2}{s_1^2}, \frac{s_2^2 \sigma_2^2}{s_2^2}, ..., \frac{s_k^2 \sigma_k^2}{s_k^2}\right)$ whre s_i^2 is the sample variane based on observation of population.

Weerahandi (1995) defined generalized p-value for testing hypothesis H_o against H_1 as following.

$$\text{p-value} = pr\left(\frac{\tilde{SS}B/(k-1)}{\tilde{SSE}/(n-1)} \geq \frac{N-k}{k-1}\tilde{ss}b\left(\frac{n_1 s_1^2}{B_1 B_2 ..B_{k-1}}, ..., \frac{n_k s_k^2}{1-B_k}\right)\right) = 1 - H(k-1, N-k)A$$

where

$$A = \frac{N-k}{k-1}\tilde{SS}B\left(\frac{n_1 s_1^2}{B_1 B_2 ..B_{k-1}}, \frac{n_2 s_2^2}{(1-B_1)B_2 ..B_{k-1}}, ..., \frac{n_k s_k^2}{1-B_{k-1}}\right)$$

and $H(k-1, N-k)(o)$ is cumulative distribution function of F distribution. Therefore

p-value = 1 – E (H (k-1 , N – k) (A))

In order to compute p-value we can generate random variables with large sample size (let say 1000) from Beta distribution with different degrees of freedom for $B'_j s$.

Then if $p-value \leq \alpha$ Ho will be rejected at α level.

3.4 Chen' s method

Chen (1998) difines $\bar{X}_i^* = \frac{1}{n_i - 1}\sum_{j=1}^{n_i - 1} X_{ij}$, $S_i^{*2} = \frac{1}{n_i - 2}\sum_{j=1}^{n_i - 1}(X_{ij} - \bar{X}_i^*)^2$.

Then let $U_i = \frac{1}{n_i} + \frac{1}{n_i}\sqrt{\frac{1}{n_i - 1}\left(\frac{S_{max}^2}{S_i^2} - 1\right)}$

$$V_i = \frac{1}{n_i} - \frac{1}{n_i}\sqrt{\frac{1}{n_i - 1}\left(\frac{S_{max}^2}{S_i^2} - 1\right)}$$

Where $S_{max}^2 = \max\left(s_1^{*2}, s_2^{*2}, ... s_k^{*2}\right)$. Then define

$$\tilde{X}_i = \sum_{j=1}^{n_i} W_{ij} X_{ij} \; ; \; W_{ij} = \begin{cases} U_i & if \quad 1 \leq j \leq n_i - 1 \\ V_i & if \quad j = n_i \end{cases}$$

Then if we define $\tilde{F}$ as following

$$\tilde{F} = \sum_{i=1}^{k} \left(\frac{\tilde{X}_i - \overline{\tilde{X}}}{\sqrt{S^2_{max} \ / n_i}} \right)^2$$

when $n_1 = n_2 = ... = n_k = n$, then Bishop et.al (1978) show the approximate distribution of as following

$$\tilde{F} \sim \frac{n-1}{n-3} X^2 (k-1)$$

By Simulation studies Chen (1998) Shows that when sample sizes are unequal we can use

$n = \sum_{i=1}^{k} n_i / k = N / k . H_o$ is rejected if $\tilde{F} > \frac{n-1}{n-3} X^2 (k-1, \alpha)$

4. Comparieson between methods

Simulation study is used to compare methods. For this purpose sample sizes (5,5,5),(10,10,10),…(5,10,15,20) from normal distributions with means, (0,0,0),(1,2,3),…(1,3,5,7) and variances (1,1,1),(1,2,4),…(30,20,10,1) were used and for each 1000 replications is used for the values $\alpha = 5\%$ and 1% Results are shown in tables A–1 and A–2.

For simplicity we used the following abbreviations

"F" for usual F

"We" // Weerahandi

"Ch" // Chen

"S" // Scott and Smith

"W" // Welch

As it can be seen from Tables A.1. and A.2 in Appendix the order of goodness of different dests are changed due to change in components, sample size, value of variances. Without regard to F-test Ch has least power, then W, We and S such that S has maximum power.

For equal sample sizes

- F has significant level above 5%, while We and W are similar and has significant level less than the nominal value comparing with two other tests. Althoug when sample sizes are greater than 10 Ch behaves like We and W.

In all cases S have significant leval more than nominal value and Ch almost in all cases and We and W in some cases have signincant level bigger than 5%.

For unequal sample sizes for cases which bigger sample sizes in assigned to population with bigger variances F-test shows smaller significant level and power. Amongst the other We has least significant level while Ch, W and S are in order to next places.

For cases when bigger sample are aissigned to population with smaller variances, then F-test has significant level more than the nominal value, but it has bigger power than the other. We has least significant level W,Ch and S are in order.

In addition to above findings, it is observed that size and power of F-test has drastic changes and this shows that F-test is not robust against changes in variances.

5. Conclusion

Based on the simulation studies, F-test is not an appropriate test for heterescedasticity of variances. The same is true for S test.

Ch is fairly appropriate test when sample sizes are proportional to variances, but We is always better than Ch.

In spite of the fact that when sample sizes are equal W test seems reasonable but it is close to We from the size of the test aspect but has less power relative to We.

In conclusion We seems abetter method, but this test is not good in size of test for the cases in which variances and sample sizes are in inverse order. In addition difficulties in computation in this method is a matter fact and there is a need to seek for a better test.

References:

[1] Bishop, T.A. Dudewicz, J.M. and Stephan, M. A. (1978): Percentage Points of a Quaratic Form in Student t Variates.Biometrika,65,435-439.

[2] Buning,H.(1995). Robust ANOVA the Case of Two- Sided Alternatives. Proceedings of Interational Conference on Statistical Computing for Quality and Productivity Improvement, August 17-19, Seoul, Korea.

[3] Chen, S.Y. and Chen, H.J.(1998). Single-Stage Analysis of Variance Under Heteroscedasticity. Communications in Statistics, Simulation and Computation, 27(3), 641-666.

[4] Scott, A.J. and Smith, T. M. F (1971). Interval Estimates for Linear Combinations of Means. Applied Statistics, Vol. 20, No.3, 276-285.

[5] Weerahandi, S. (1995). Exact Statistical Methods for Data Analysis. New York: Springer-Verlag.

[6] Welch, B.L. (1947). The Generalization of Studen't Problem when Several Different Population Variances are Involved. Biometrika, 34,28-35.

[7] Welch, B. L. (1951). On the Comparison of Several Means Values: An Alternative Approach. Biometrika, 38,330-336.

Appendix

Table A.1. Size and power of tests for $a = 5$ percent

n_i	m_i	s_i^2	F	We	Ch	W	S
5,5,5	**0,0,0**	1,10,20	60	50	92	48	222
		1,2,4	68	40	68	42	82
		1,1,1	50	38	90	40	62
		4,2,1	68	40	68	42	82
		20,10,1	60	50	92	48	222
	1,2,3	1,10,20	90	72	78	62	406
		1,2,4	166	134	128	132	400
		1,1,1	708	624	462	620	646
		4,2,1	166	134	128	132	400
		20,10,1	90	72	78	62	406
	1,3,5	1,10,20	98	94	94	88	744
		1,2,4	418	517	310	502	906
		1,1,1	998	998	974	994	996
		4,2,1	418	517	310	502	906
		20,10,1	98	94	94	88	744
10,10,10	**0,0,0**	1,10,20	82	64	60	62	154
		1,2,4	64	52	54	54	92
		1,1,1	50	50	56	52	42
		4,2,1	64	52	54	54	92
		20,10,1	82	64	60	62	154
	1,2,3	1,10,20	74	50	48	48	776
		1,2,4	268	310	164	310	656
		1,1,1	982	958	920	947	676
		4,2,1	268	310	164	310	656
		20,10,1	74	50	48	48	776
	1,3,5	1,10,20	104	102	64	94	834
		1,2,4	784	904	518	892	986
		1,1,1	1000	1000	1000	1000	1000
		4,2,1	784	904	518	892	986
		20,10,1	104	102	64	94	834
5,10,15	**0,0,0**	1,10,20	24	30	41	46	168
		1,2,4	16	32	38	44	90
		1,1,1	50	48	54	54	46
		4,2,1	194	54	94	58	78
		20,10,1	278	62	92	66	270
	1,2,3	1,10,20	22	68	54	64	568
		1,2,4	98	296	276	288	660
		1,1,1	962	918	916	912	1000
		4,2,1	458	278	260	254	504
		20,10,1	272	68	65	54	548
	1,3,5	1,10,20	34	120	72	116	942
		1,2,4	542	902	730	908	1000
		1,1,1	1000	1000	1000	1000	1000
		4,2,1	878	824	764	772	998
		20,10,1	246	86	126	84	928

Table A.1. Continiued…

n_i	m_i	s_i^2	F	We	Ch	W	S
5,5,5,5	**0,0,0,0**	1,10,20,30	76	66	90	66	206
		1,2,4,6	56	50	90	68	76
		1,1,1,1	50	50	78	48	54

		4,2,1,1	56	50	90	48	76
		30,20,10,1	76	66	90	66	206
	1,2,3,4	1,10,20,30	164	214	132	190	558
		1,2,4,6	526	516	342	486	670
		1,1,1,1	974	948	836	934	934
		4,2,1,1	526	516	342	486	670
		30,20,10,1	164	214	132	190	558
	1,3,5,7	1,10,20,30	476	686	290	592	952
		1,2,4,6	990	998	834	988	998
		1,1,1,1	1000	1000	998	1000	1000
		6,4,2,1	990	998	834	988	998
		30,20,10,1	476	686	290	592	952
10,10,10,10	**0,0,0,0**	1,10,20,30	68	58	54	56	260
		1,2,4,6	64	42	44	42	62
		1,1,1,1	50	50	54	50	60
		4,2,1,1	64	42	44	42	62
		30,20,10,1	68	58	54	56	260
	1,2,3,4	1,10,20,30	262	408	142	366	840
		1,2,4,6	906	922	626	918	980
		1,1,1,1	1000	1000	998	1000	1000
		6,4,2,1	906	922	626	918	980
		30,20,10,1	262	408	142	366	840
	1,3,5,7	1,10,20,30	1000	1000	944	1000	1000
		1,2,4,6	1000	1000	990	1000	1000
		1,1,1,1	1000	1000	1000	1000	1000
		6,4,2,1	1000	1000	990	1000	1000
		30,20,10,1	1000	1000	944	1000	1000
5,10,15,20	**0,0,0,0**	1,10,20,30	14	38	45	50	164
		1,2,4,6	18	36	40	46	74
		1,1,1,1	50	51	54	56	54
		6,4,2,1	174	56	68	58	82
		30,20,10,1	224	68	100	74	320
	1,2,3,4	1,10,20,30	118	498	434	510	770
		1,2,4,6	804	934	936	938	976
		1,1,1,1	1000	1000	1000	1000	1000
		6,4,2,1	952	930	901	989	962
		30,20,10,1	474	386	308	296	818
	1,3,5,7	1,10,20,30	696	996	894	992	998
		1,2,4,6	1000	1000	1000	1000	1000
		1,1,1,1	1000	1000	1000	1000	1000
		6,4,2,1	1000	1000	998	1000	1000
		30,20,10,1	934	944	848	908	1000

Table A.2. size and power of tets for a =L percent

n_i	m_i	s_i^2	F	We	Ch	W	S
5,5,5	**0,0,0**	1,10,20	18	12	32	14	108
		1,2,4	12	4	42	4	28
		1,1,1	10	10	42	8	18
		4,2,1	12	4	42	4	28
		20,10,1	18	12	32	14	108
	1,2,3	1,10,20	92	66	46	54	690
		1,2,4	174	156	138	144	422
		1,1,1	718	626	550	606	628
		4,2,1	174	156	138	144	422
		20,10,1	92	66	46	54	690
	1,3,5	1,10,20	86	82	70	78	666
		1,2,4	416	498	302	488	804
		1,1,1	1000	994	974	988	964
		4,2,1	416	498	302	488	804
		20,10,1	86	82	70	78	666
10,10,10	**0,0,0**	1,10,20	22	11	14	12	174
		1,2,4	11	8	17	9	34
		1,1,1	10	8	15	8	20
		4,2,1	11	8	17	9	24
		20,10,1	22	11	14	12	74
	1,2,3	1,10,20	94	54	46	50	738
		1,2,4	304	350	172	342	620
		1,1,1	972	952	894	948	960
		4,2,1	304	350	172	342	640
		20,10,1	94	54	46	50	738
	1,3,5	1,10,20	128	96	68	82	804
		1,2,4	802	908	528	890	936
		1,1,1	1000	1000	998	1000	1000
		4,2,1	802	908	528	890	976
		20,10,1	128	96	68	82	878
5,10,15	**0,0,0**	1,10,20	4	6	6	10	78
		1,2,4	6	6	8	10	40
		1,1,1	10	6	10	10	20
		4,2,1	36	4	34	16	24
		20,10,1	66	16	28	18	90
	1,2,3	1,10,20	18	90	78	82	750
		1,2,4	220	302	280	298	662
		1,1,1	816	696	651	682	970
		4,2,1	502	288	254	266	650
		20,10,1	178	50	44	46	779
	1,3,5	1,10,20	150	568	436	560	882
		1,2,4	946	966	910	943	998
		1,1,1	1000	1000	1000	1000	1000
		4,2,1	994	952	898	924	994
		20,10,1	538	286	198	204	890

Table A.2. Continued …

n_i	m_i	s_i^2	F	We	Ch	W	S
5,5,5,5	**0,0,0,0**	1,10,20,30	8	16	36	18	108
		1,2,4,6	24	14	48	14	48
		1,1,1,1	10	6	56	6	22
		4,2,1,1	24	14	48	14	48
		30,20,10,1	8	16	36	18	108
	1,2,3,4	1,10,20,30	24	178	150	166	412
		1,2,4,6	526	816	796	802	890
		1,1,1,1	1000	1000	990	994	996
		4,2,1,1	526	816	796	802	890
		30,20,10,1	24	178	150	166	412
	1,3,5,7	1,10,20,30	408	938	768	910	946
		1,2,4,6	1000	1000	1000	1000	1000
		1,1,1,1	1000	1000	1000	1000	1000
		6,4,2,1	1000	1000	1000	1000	1000
		30,20,10,1	408	938	768	910	956
10,10,10 ,10	**0,0,0,0**	1,10,20,30	26	14	14	14	98
		1,2,4,6	12	6	8	8	36
		1,1,1,1	10	5	10	7	16
		4,2,1,1	12	6	8	8	36
		30,20,10,1	26	14	14	14	98
	1,2,3,4	1,10,20,30	48	78	58	66	422
		1,2,4,6	256	242	186	224	514
		1,1,1,1	854	762	650	702	820
		6,4,2,1	256	242	186	224	514
		30,20,10,1	48	78	58	66	422
	1,3,5,7	1,10,20,30	580	854	302	796	998
		1,2,4,6	1000	1000	972	1000	1000
		1,1,1,1	1000	1000	1000	1000	1000
		6,4,2,1	1000	1000	972	1000	1000
		30,20,10,1	580	854	302	796	998
5,10,15, 20	**0,0,0,0**	1,10,20,30	4	12	13	14	68
		1,2,4,6	6	8	10	12	28
		1,1,1,1	10	12	8	14	24
		6,4,2,1	82	20	30	26	30
		30,20,10,1	96	24	36	28	160
	1,2,3,4	1,10,20,30	40	220	190	214	632
		1,2,4,6	512	788	760	778	870
		1,1,1,1	1000	996	971	989	1000
		6,4,2,1	870	746	558	656	910
		30,20,10,1	306	182	114	128	744
	1,3,5,7	1,10,20,30	420	934	764	918	992
		1,2,4,6	1000	1000	1000	1000	1000
		1,1,1,1	1000	1000	1000	1000	1000
		6,4,2,1	1000	1000	996	1000	1000
		30,20,10,1	824	756	594	614	998